


# Observed changes in temperature extremes over Asia and their attribution

Siyan Dong<sup>1</sup> · Ying Sun<sup>1,2</sup>  · Enric Aguilar<sup>3</sup> · Xuebin Zhang<sup>4</sup> · Thomas C. Peterson<sup>5</sup> · Lianchun Song<sup>1</sup> · Yingxian Zhang<sup>1</sup>

Received: 22 March 2017 / Accepted: 21 September 2017 / Published online: 2 November 2017  
© The Author(s) 2017. This article is an open access publication

**Abstract** This study presents trends in a newly compiled temperature extreme indices dataset for Asia covering the period of 1958–2012. Daily data were homogenized prior to the calculation of the indices. A clear warming trend was observed in all indices, which is consistent with the global warming. In most of the indices, larger warming was observed at high latitudes than at low latitudes. We also compared observations with simulations from the Coupled Model Inter-comparison Project Phase 5 for some indices using an optimal fingerprinting method. These indices include the number of days with daily maximum or minimum temperatures greater than their 90th percentiles or smaller than their 10th percentiles, the annual highest daily maximum and minimum temperatures, and the annual lowest daily maximum and minimum temperatures. We determined that the warming trend was inconsistent with the natural variability of the climate system but agreed with climate responses to external forcing as simulated by the models. The anthropogenic and natural signals could be detected and separated from each other in the region for almost all

indices, indicating the robustness of the warming signal as well as the attribution of warming to external causes.

**Keywords** Extreme temperatures · Observations in Asia · Attribution · Anthropogenic forcing

## 1 Introduction

Changes in climate extremes have received an increased amount of attention over the last several decades because the impacts of climate change on human society and natural systems are caused by changes in extremes (IPCC 2012). Warming trends and increasing temperature extremes have been observed across most of Asia. This observed warming is projected to continue. It has also been projected that more frequent and intense heat waves in this region will increase mortality and morbidity in vulnerable groups and that increases in heavy rain and temperature will increase the risk of diarrheal diseases, dengue fever, and malaria (Hijioka et al. 2014).

Changes in temperature extremes have been studied extensively over the past several decades. Significant changes in extreme temperatures have been observed on global and regional scales and within different datasets (Alexander et al. 2006; Donat et al. 2013a; Dong et al. 2015; Bindoff et al. 2013). Donat et al. (2013b) updated the global analysis of Alexander et al. (2006) and showed widespread significant changes in temperature extremes that are consistent with warming. These changes are more pronounced in indices related to daily minimum temperature than they are in indices related to daily maximum temperature. At regional scales, Zhang et al. (2005) found increasing trends in the frequency of warm extremes and significant decreasing trends in the frequency of cold extremes in the Middle East over the

✉ Ying Sun  
sunying@cma.gov.cn

<sup>1</sup> National Climate Center, Laboratory for Climate Studies, China Meteorological Administration, Beijing, China  
<sup>2</sup> Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing University of Information Science and Technology, Nanjing, China  
<sup>3</sup> Centre for Climate Change, C3, University Rovira i Virgili, Vilaseca, Tarragona, Spain  
<sup>4</sup> Climate Research Division, Environment and Climate Change Canada, Toronto, Ontario, Canada  
<sup>5</sup> Commission for Climatology, World Meteorological Organization, Geneva, Switzerland

period of 1950–2003. Caesar et al. (2011) demonstrated that widespread significant warming in temperature extremes, especially in the number of warm nights in the Indo-Pacific region during 1971–2005. Donat et al. (2014) found increasing number of warm days and nights, higher extreme temperature values, shorter cold spell durations and fewer cold days and nights in the Arab regions since the mid-1950s. Increases in warm extremes and decreases in cold extremes have also been observed in Chinese temperature data (You et al. 2008, 2009; Yin et al. 2016; Lu et al. 2016).

It is important to understand the causes of long-term trends in the observed climate, especially the possible influence of external forcings on the climate system. This is typically accomplished using detection and attribution analysis, in which observations are compared with the simulated responses of climate models to natural and external forcings. Evidence of anthropogenic influence through the emission of greenhouse gases and aerosols, and sometimes evidence of influence from natural external forcings, can be clearly detected in the magnitudes of extreme temperatures, such as the highest annual values of daily maximum and minimum temperatures (TXx and TNx, respectively) and annual lowest daily maximum and minimum temperatures (TXn and TNn, respectively), on both global and regional scales (e.g., Christidis et al. 2005; Zwiers et al. 2011; Wen et al. 2013; Yin et al. 2016; Kim et al. 2016). In general, climate models can accurately simulate the observed changes in the warmest night-time temperatures, but they underestimate the observed changes in the coldest temperatures (e.g., Kim et al. 2016). Anthropogenic influence has also been detected in the frequency of moderate temperature extremes, such as the number of days with daily maximum or minimum temperatures greater than the 90th percentile (TX90p and TN90p, respectively) or smaller than the 10th percentile (TX10p and TN10p, respectively). Morak et al. (2011, 2013) found a detectable signal of external forcing in the frequencies of warm days and nights in most regions of global land. Lu et al. (2016) found similar results in Chinese temperature data. In addition to daily temperature extremes, Sun et al. (2014) showed anthropogenic influences on the highest seasonal mean temperatures and projected a rapid increase in the frequency of such events in the future due to the anthropogenic emissions of greenhouse gases.

Although global studies have included Asia and some regional studies have covered some countries in Asia, changes in extreme temperatures and their causes in Asia have not been comprehensively studied. This is partly due to the lack of a comprehensive dataset for this region. To address this issue and to facilitate data exchange in this

region, the World Meteorological Organization<sup>1</sup> (WMO) Expert Team on Climate Change Detection and Indices (ETCCDI) organized a regional workshop in Nanjing, China, in March 2013, with support from the China Meteorological Administration and the WMO. As with the other workshops organized by ETCCDI (e.g., Peterson and Manton 2008; Zhang et al. 2005), this workshop sought to enhance the regional capacity to monitor changes in climate extremes as well as to make precipitation and temperature extreme indices from this region available to the international research community. This workshop gathered scientists and data from different parts of Asia, including Saudi Arabia, Bahrain, Bangladesh, Bhutan, China, Japan, Indonesia, North Korea, Laos, Malaysia, the Maldives, Mongolia, Russia, Thailand and Vietnam. All of these countries provided long-term daily weather observing station time series data and computed the indices that were included in this analysis. The extreme indices provided by different countries were preliminarily checked and tested for homogeneity using the RHtestV3 (Wang and Feng 2010); they then underwent a second round of quality control, which involved manually verifying outliers to avoid the inclusion of erroneous observations in this analysis. The workshop yielded a unique dataset that provides detailed information about temperature extremes in the region; this dataset features more stations and better spatial and temporal coverage than the global HadEX2 dataset (Donat et al. 2013b). The outcome of this Nanjing workshop has thus provided a unique opportunity to examine past changes and to attribute the causes of these changes in temperature extremes in Asia.

Here, we conducted two types of “detection” analysis. The first one is a broader detection analysis, in which we examine if there exists a statistically significant change in the observations that is represented by a linear trend. The second one is a narrower detection analysis, in which we examine if an expected change in response to external forcings, as simulated by climate models, is detectable in the observations. In this paper, to avoid confusion, we refer to the first type as trend analysis and the second type as detection and attribution analysis. We conducted trend analysis for all of the ETCCDI extreme indices derived from daily temperature observations. Our detection and attribution analysis is performed on eight indices, including the annual maxima and minima of daily maximum and daily minimum temperatures, as well as the frequencies of moderate extreme temperatures, which are defined as the daily maximum or minimum temperatures above their 90th or below their 10th percentiles. These analyses were performed to provide a basic understanding of how external forcing may have

<sup>1</sup> Specifically, it is a joint team sponsored by the WMO Commission for Climatology, the World Climate Research Programme, and the Joint Commission for Oceanography and Marine Meteorology.

influenced temperature extremes in Asia. The structure of the paper is as follows. Section 2 describes the observational and model datasets, as well as the methods that were used for trend analysis and detection and attribution analysis. The main results are presented in Sect. 3. Section 4 presents the discussion and conclusions.

## 2 Data and methodology

### 2.1 Observational data

The observed daily station data used in this analysis were obtained from three sources. One source represents daily station data that were contributed by workshop participants from 15 countries. This is an important data source, as data from many of these stations were not available elsewhere; as such, they filled a significant spatial gap. This data source was then augmented using station data available from the Global Historical Climate Network Daily (GHCND) dataset (Menne et al. 2012). The GHCND dataset was used to supplement additional stations or to expand the temporal completeness of the station data provided by the workshop participants. In the latter case, station coordinates, names and ID numbers were manually searched and compared to ensure that station data from two different sources belonged to the same station. Station data were blended only if the data from the two sources agreed during overlapping periods. The third data source was the homogenized daily temperature dataset from the China National Meteorological Information Center (NMIC). Although data from a limited number of Chinese stations were brought to the workshop and GHCND has over 100 Chinese stations, the NMIC dataset is the best available Chinese dataset because it has the highest spatial coverage; it contains over 2400 stations, and its data have undergone strict quality control and careful homogenization (Cao et al. 2016). For this reason, we use the NMIC dataset for China instead of the data collected at the workshop or those that are available in the GHCND dataset. Although both daily temperature and precipitation data were available from the workshop, for the purpose of this paper, we focused only on daily maximum and minimum temperatures.

The homogeneity of these data was carefully examined and adjusted. The Chinese data were homogenized using RHtests (Wang and Feng 2010) in the NMIC (Cao et al. 2016). This dataset has been extensively analyzed in several recent studies (Sun et al. 2014; Yin et al. 2016; Cao et al. 2016). The quality and homogeneity of the data from the non-Chinese stations were checked preliminarily at the workshop. These procedures included the application of RCLimindex-extraqc (see [http://www.c3.urv.cat/data/manual/Manual\\_rclimdex\\_extraQC.r.pdf](http://www.c3.urv.cat/data/manual/Manual_rclimdex_extraQC.r.pdf)) to check the data quality and RHtests to test the homogeneity of the data. The

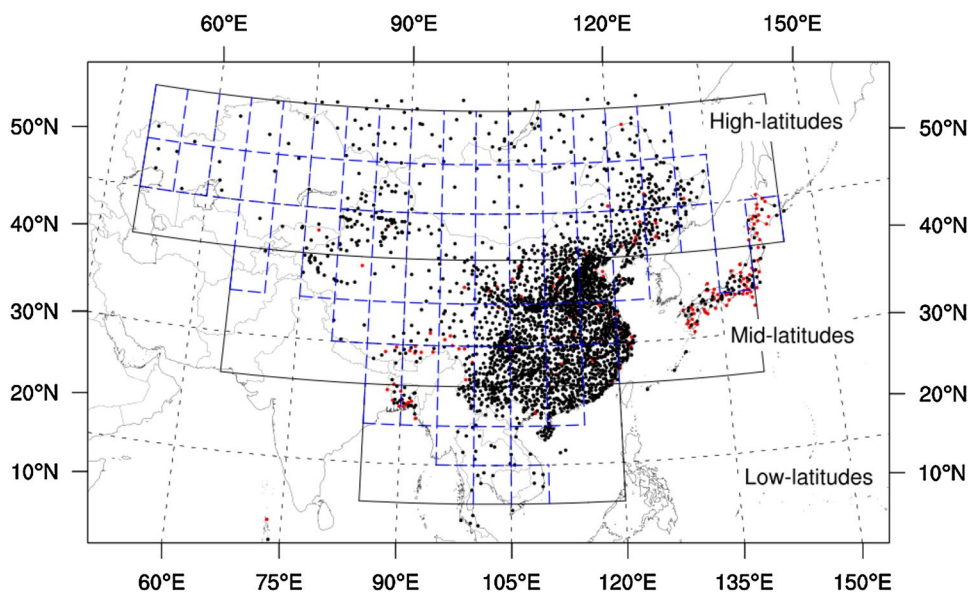
quality-controlled data were then subjected to the homogeneity adjustment of their monthly mean values using HOMER (Mestre et al. 2013), and their daily factors were interpolated using the approach described by Vincent et al. (2002).

The 29 indices defined by the ETCCDI were computed using quality-controlled and homogenized daily temperature and precipitation data. The indices were computed using the RCLimDex/FCLimdex software package (Zhang et al. 2011). The climatological base period of 1961–1990 was used for the indices that are based on their comparisons with base period values. The calculations of the ETCCDI indices require nearly complete daily values to compute the indices for a particular year. We also excluded stations for which there were fewer than 15 complete years of data during the base period. As a result, the indices computed for approximately 2300–2400 stations were retained for subsequent analysis. Among these stations, approximately 2000–2100 stations are Chinese stations. The number of stations varies slightly for different extreme indices due to their different data availability requirements. To provide a general picture of the spatial distribution of these stations, Fig. 1 displays the locations of stations that yielded a sufficient amount of TXx data.

The spatial distribution of these stations is uneven, and most of these stations are located in the eastern part of China. For some countries that sent representatives to the workshop, the daily data that were provided were too poor or too limited to be included in this analysis. Additionally, in some countries, the international exchange of data was too limited for them to be included in the GHCN-daily dataset. As a result, there are noticeably fewer stations in southern and western Asia. Thus, the region of Asia that can be studied is limited to the region with data coverage. To make the best use of the available information and to avoid overweighting areas with a dense distribution of observing stations, we produced gridded values on  $5^\circ \times 5^\circ$  grids by averaging station anomalies relative to the 1961–1990 mean for each grid containing at least one station. The values for grid boxes without any station data are marked as missing. The grid boxes with data are marked by blue dashed lines in Fig. 1. We analyzed the trends of these gridded values. More analyses were conducted on regional averages over three latitudinal bands with an equal latitudinal range. These three bands are referred to as low-, mid- and high-latitude regions, and they cover the areas from  $10^\circ\text{N}$  to  $25^\circ\text{N}$ ,  $25^\circ\text{N}$  to  $40^\circ\text{N}$ , and  $40^\circ\text{N}$  to  $55^\circ\text{N}$ , respectively. The regional values were obtained by averaging all available grid box values within each region. Because a much fewer number of grids with data are available prior to 1958, only data from 1958 to 2012 are included in these analyses.

For the trend analysis, we considered all of the temperature indices defined in Table 1. For the detection and

**Fig. 1** Illustration of station locations used for TXx. The grid boxes with sufficient data are marked by blue dashed lines, and regional boundaries are marked by black lines that separate the three study regions: low-latitude region: (10–25°N, 85–120°E); mid-latitude region: (25–40°N, 65–140°E); and high-latitude region: (40–55°N, 50–145°E). Red dots represent excluded stations and black dots represent those used in the calculation of TXx. The distributions of stations for the other indices are similar to that shown here



**Table 1** Description and units of the temperature indices used in this study

Index	Indicator name	Definition	Unit
TXx	Max Tmax	The annual maxima of daily maximum temperature	°C
TNx	Max Tmin	The annual maxima of daily minimum temperature	°C
TXn	Min Tmax	The annual minima of daily maximum temperature	°C
TNn	Min Tmin	The annual minima of daily minimum temperature	°C
TX90p	Warm days	Percentage of days with daily maximum temperature greater than the 90th percentile of the base period	% of days
TN90p	Warm nights	Percentage of days with daily minimum temperature greater than the 90th percentile of the base period	% of days
TX10p	Cold days	Percentage of days with daily maximum temperature less than the 10th percentile of the base period	% of days
TN10p	Cold nights	Percentage of days with daily minimum temperature less than the 10th percentile of the base period	% of days
DTR	Diurnal temperature range	Mean difference between maximum temperature and minimum temperature	°C
GSL	Growing season Length	Annual (1st Jan to 31st Dec in NH, 1st July to 30th June in SH) length between first span of at least 6 days with TG > 5 °C and first span after July 1 (January 1 in SH) of 6 days with TG < 5 °C	Days
FD	Frost days	Number of days with minimum temperature less than 0 °C	Days
SU	Summer days	Number of days with maximum temperature greater than 25 °C	Days
TR	Tropical nights	Number of days with minimum temperature greater than 20 °C	Days
ID	Ice days	Number of days with maximum temperature less than 0 °C	Days
HWDI	Heat wave duration index	Longest days in interval of at least 5 consecutive days with daily maximum temperature of 5 °C greater than the base period mean	Days
CWDI	Cold wave duration index	Longest days in interval of at least 5 consecutive days with daily minimum temperature of 5 °C less than the base period mean	Days

attribution analysis, we focused on two groups of variables. One group represented the severity of extreme daily temperatures, including TXx, TNx, TXn and TNn, which are herein referred to as “intensity indices”. Another group represented the frequencies of moderate extreme temperatures, which are defined as the daily maximum or daily

minimum temperatures that are greater than the 90th percentile (TX90p, TN90p) and the daily maximum or daily minimum temperatures that are smaller than the 10th percentile (TX10p, TN10p). These indices are referred to as “frequency indices”. The intensity indices are expressed in units of °C, and the frequency indices are expressed in units

of percent of days. Together, they provide insights into the behavior of a variety of different aspects of the daily temperature distribution, and there is not necessarily a correspondence between these two groups.

## 2.2 Model simulations

The detection and attribution analysis requires the estimation of the expected responses of the climate system to external forcing as well as the quantification of the natural internal variability of the climate. The expected climate responses and internal variability are estimated based on the model simulations of the Coupled Model Inter-comparison Project Phase 5 (CMIP5, Taylor et al. 2012). Sillmann et al. (2013a, b) calculated the ETCCDI indices for some of the CMIP5 simulations of the twentieth century (historical simulations) and the twenty-first century under different representative concentration pathways (RCPs) emission scenarios. Data from these indices available from Sillmann et al. (2013a) are used in this study. Indices data for simulations that were not computed by Sillmann et al. (2013a), but whose daily temperature data were available at the time of analysis, were computed using the same software. Although CMIP5 historical simulations generally end in 2005, some simulations have been extended to 2012. For this reason, our detection and attribution analysis is conducted over a period of 55 years, from 1958 to 2012. Some studies have shown that the CMIP5 models are generally able to simulate climate extremes and their trend patterns (Sillmann et al. 2013a, b). For the sake of simplicity, we opted not to evaluate model performance at the regional scale in this study.

In total, the indices derived from the CMIP5 simulations include 82 runs conducted using 17 models under the combined effects of historical anthropogenic and natural external forcings (ALL) (Table 2). For the ALL simulations that were not extended to 2012, we used RCP4.5 (an emission scenario in which the radiative forcing value in the year 2100 is approximately  $4.5 \text{ W/m}^2$  higher than the preindustrial value) simulations to extend the model data to 2012. We also used single forcing simulations, including those under natural external forcing (NAT) and those under greenhouse gases forcing (GHG). For these simulations, we only used the output from the models that offered extended simulations to 2012. In total there were 26 runs conducted with six models under the NAT forcing and 23 runs conducted with five models under the GHG forcing. Pre-industrial control (CTL) simulations in which external forcing are kept constant at the pre-industrial level conducted with output from 28 models were used for estimating internal variability. For this purpose, the simulations were divided into chunks of 55 years. There were 260 chunks of the control simulations.

The regional averages from model indices were computed using a similar method as the observations. As model

**Table 2** Available simulations for the different forcing experiments used in this study

Model name	All	CTL	NAT	GHG
BCC-CSM1-1	3	8		
BCC-CSM1-1-m	3	7		
CanESM2	5	19	5	5
CCSM4		2	4*	3*
CESM1-BGC		8		
CESM1-FASTCHEM	3*			
CMCC-CESM		4		
CMCC-CM		5		
CMCC-CMS		8		
CNRM-CM5	10	15	6	6
CSIRO-Mk3-6-0	10	8	5	5
FGOALS-s2	3	8		
GFDL-CM3	5*	14	3*	3*
GFDL-ESM2G		8		
GFDL-ESM2M		8		
GISS-E2-R	3			
HadCM3	10			
HadGEM2-CC		4		
HadGEM2-ES	4	10	4	4
IPSL-CM5A-LR	6	17	3	3
IPSL-CM5A-MR	3*	5	3	3*
IPSL-CM5B-LR		5		
MIROC-ESM	3	11		
MIROC-ESM-CHEM		4		
MIROC4h	3 <sup>#</sup>	1		
MIROC5	5	11		
MPI-ESM-MR	3	17		
MPI-ESM-LR	3	17		
MPI-ESM-P		20		
MRI-CGCM3	5	8		
NorESM1-M	3	8		
SUM (models)	82 (17)	260 (28)	26 (6)	23 (5)

Numbers represent simulation ensemble sizes or the number of 55-year chunks of the CTL simulations. The models that are used only for the calculations of signal or noise are marked with hash signs and asterisks, respectively. The model simulations that do not extend to 2012 are used for noise estimation (e.g., CESM1-FASTCHEM), and those with less than 110 years of historical simulation data are not used for noise estimation (e.g., MIROC4h)

simulations were conducted at various spatial resolutions, we first converted the model indices to a  $5^\circ \times 5^\circ$  resolution, which is consistent with that of observations. These model indices were then masked to mimic the availability of the observations over both space and time (i.e., the first year of model data corresponds to the first year of observational data) by defining the model values as missing if the observational data for the grid at that time were missing. Since all observations are land-based, we only used land grids in the

model simulations. The climate response to external forcing was calculated as the arithmetic average of the ensemble mean values of individual models.

## 2.3 Methods

### 2.3.1 Trend estimation

We used trends as a measure of long-term changes in these temperature indices. These indices exhibit different probabilistic properties and may also be autocorrelated. For these reasons, we used a non-parametric method (Sen 1968) to compute trends and their confidence intervals. Additionally, an iterative procedure was used to consider autocorrelation when determining the statistical significance of a trend (Zhang et al. 2000; Wang and Swail 2001). This approach offers a robust trend estimate that excludes outliers and provides a proper test of statistical significance against the misspecification of the probability distributions of the indices and colored noise. It has been extensively used to estimate trends in similar datasets (e.g., Alexander et al. 2006). Due to data availability, trends are computed for the time period of 1958–2012.

### 2.3.2 Optimal fingerprint method

We used a standard optimal fingerprint method to ascertain the influence of external forcing. This method is based on a generalized linear regression (Allen and Tett 1999; Allen and Stott 2003; Ribes et al. 2013). It regresses the observations ( $\mathbf{Y}$ ) onto multimodel mean signal patterns ( $\mathbf{X}$ ), which is expressed as  $\mathbf{Y} = (\mathbf{X} - \mathbf{v})\boldsymbol{\beta} + \boldsymbol{\epsilon}$ . Here,  $\boldsymbol{\beta}$  is the regression coefficient or scaling factor;  $\mathbf{v}$  reflects the effects of internal variability that remain in the signal  $\mathbf{X}$  because the multimodel ensemble mean does not completely remove all vestiges of internal variability; and  $\boldsymbol{\epsilon}$  represents the regression residual, which reflects the internal variability. The internal variability is estimated using model simulations, and the scaling factor  $\boldsymbol{\beta}$  is obtained using the total least squares (TLS) method (Allen and Stott 2003; Ribes et al. 2013). A residual consistency test (Ribes et al. 2013) is conducted to determine whether the variance of the residuals of the regression is inconsistent with the variance of the unforced variability estimates in the model simulations. If the variability simulated by the model is not too much smaller than the variance of the residuals, then an inference about scaling factors can be used to make detection and attribution statements. A scaling factor whose 5th percentile is greater than zero implies that the signal can be detected at the 5% significance level. If the 90% confidence interval for a scaling factor also includes unity, then these observations are considered to be consistent with the model response to that external forcing. This can then be used to formulate attribution statements.

We conducted single- and two-signal analyses to detect the relative role of individual forcing in changes in the observational data. For the single-signal analysis, the observational data were regressed onto the multimodel averaged responses to ALL forcing to determine if the observed changes were consistent with external forcing or natural variability. We used the ALL forcing response because this represents the combined effect of all known external forcings. In the two-signal analyses, observed changes were simultaneously regressed onto ANT and NAT signals to determine whether these two signals could be detected and whether the influences of ANT could be separated from those of NAT and internal variability. The model response to ANT forcing was calculated as the difference between the ALL and NAT simulations from all available simulations, assuming that these models are interchangeable. In reality, these models may not be completely interchangeable, as the differences between ALL and NAT can contain difference due to different models being used in the estimation. Nevertheless, the detection results did not appear to be greatly affected by different ANT signal estimates when different sets of models were used.

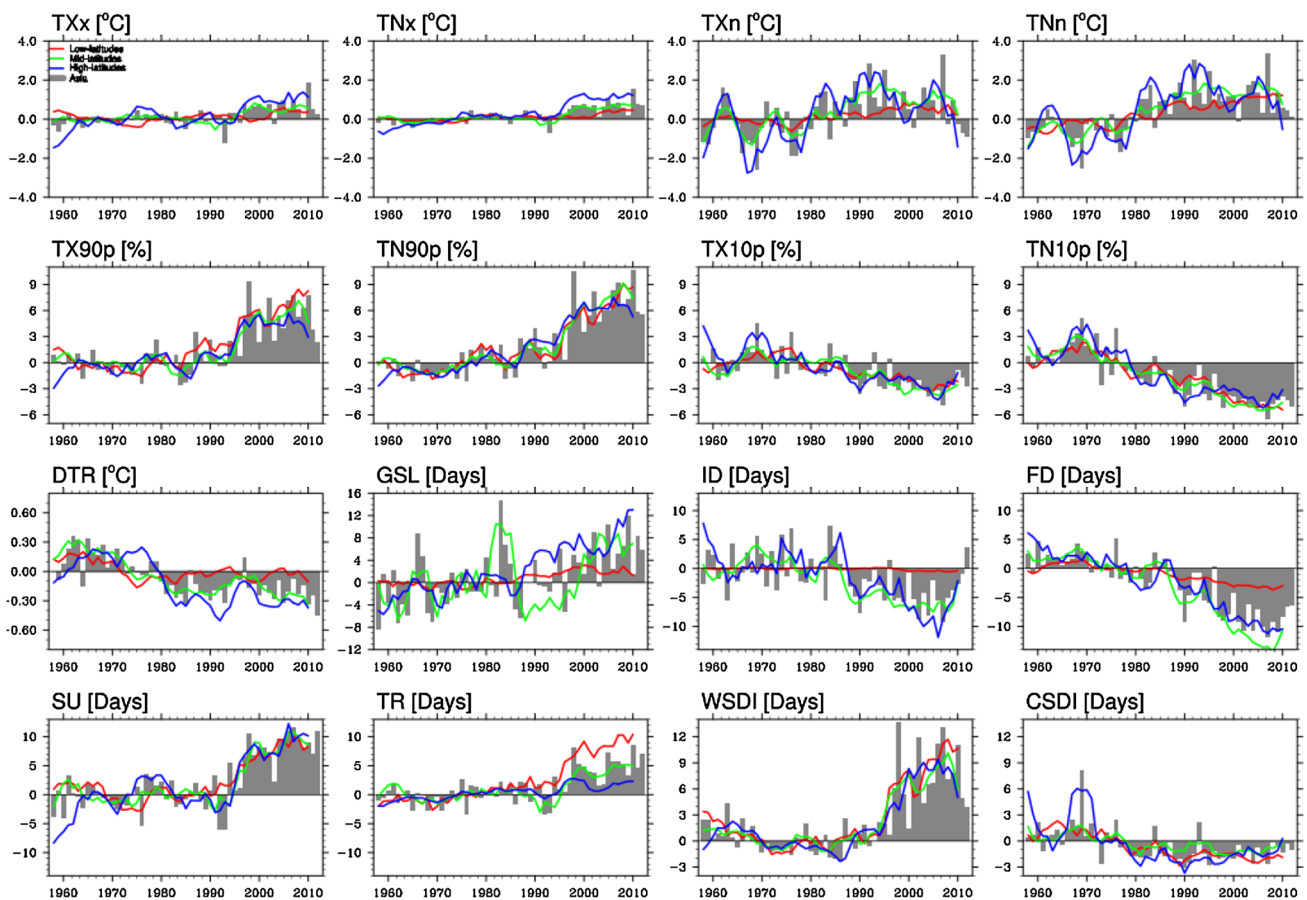
Our fingerprinting analyses were conducted on the non-overlapping 3- or 5-year mean values of regional average series. The detection results were insensitive to the use of 3- or 5-year means; thus, here, we report results only from the analyses of the 5-year mean series. These analyses were performed on the time series for the three latitudinal regions separately to determine if the climate response to external forcing was detectable in the subregions of Asia. We also conducted these analyses throughout all of Asia and assessed the space–time patterns of temperature responses by including data from the three individual regions as three spatial dimensions.

The internal variability was estimated based on model simulations. We used both within-ensemble differences (i.e., the residuals of the ensemble simulations after the removal of the ensemble mean) and preindustrial control simulations. Two independent estimates were obtained, each of which used 274 chunks of the 55-year series from the within-ensemble differences and half of the preindustrial run data. One of the estimates was used to calculate scaling factors, while the other was used to estimate the 90% confidence interval of the scaling factors as well as to test residual consistency.

## 3 Results

### 3.1 Observed changes in extremes in Asia

Figure 2 displays the time evolution of the annual series for the 16 extreme temperature indices in Asia (bar plots) and



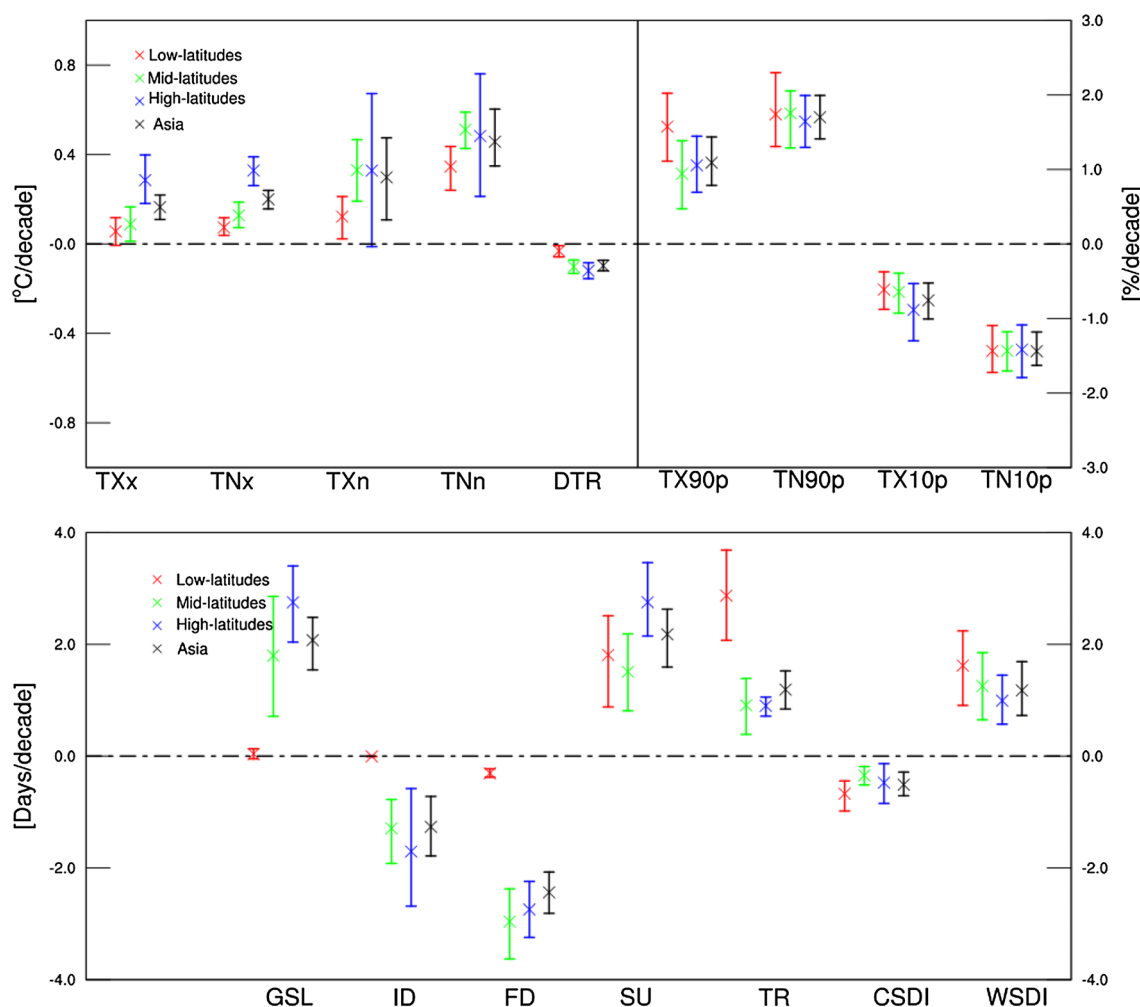
**Fig. 2** Annual anomalies (relative to the 1961–1990 mean) of area-averaged 16 temperature indices during 1958–2012. Gray bars and blue, green and red lines represent regional averages over Asia and its three subregions: the high-, mid- and low-latitude regions, respectively

its three subregions. The most salient feature is the increase in warm extremes and the decrease in cold extremes, which are consistent with warming. Changes in the indices over Asia are generally larger than those in the global averages of the indices (Donat et al. 2013b). These changes are more pronounced after the 1980s but become flatter after the late 1990s. As our data end in 2012, the very warm temperatures that have occurred since 2012 (NOAA National Centers for Environmental information 2017) are not captured. For the absolute annual warmest and coldest temperatures (TXx, TNx, TXn and TNn, top panel of Fig. 2), the magnitudes of change in the warmest extremes (TXx and TNx) are less than those in the coldest extremes (TXn and TNn) in the region and all subregions. Across the subregions, both the long-term trends and interannual variabilities are larger at higher latitudes than they are at lower latitudes.

The percentile indices also show clear warming, with increases in the exceedances of high percentiles (e.g., TX90p and TN90p) and decreases in the occurrences of lower percentiles (e.g., TX10p and TN10p). The largest positive anomalies in the TX90p and TN90p indices are greater than the absolute values of the largest anomalies

in the TX10p and TN10p indices; these data reflect the effects caused by the low exceedances being bounded by zero in a warming environment, while larger changes can occur in the high exceedances of percentiles. Although diurnal temperature range (DTR) decreases, the magnitude of its decrease is far smaller than trends of day-time and night-time extreme temperatures (i.e., TXx vs TNx and TXn vs TNn). The other indices also consistently exhibit the effects of warming, as increases are observed in the numbers of summer days (SU), tropical nights (TR), and warm spells (WSDI), and decreases are observed in the numbers of frost days (FD), ice days (ID), and cold spells (CSDI). Notably, because the fixed-threshold indices (e.g., SU, TR, FD, ID) are applicable only over some regions (Alexander et al. 2006), their average values across the research domain do not necessarily represent those of the entire region.

These trends and their 90% confidence intervals (CI) are displayed in Fig. 3. In general, trends in almost all indices in Asia and its subregions are significant at the 5% level, as the 90% CI very rarely crosses zero. The exception to this rule is related to the length of the growing season (GSL) and the



**Fig. 3** Linear trends of area-averaged extreme temperature indices and their 5–95% confidence intervals in Asia and its subregions during 1958–2012. Black, blue, green and red represent regional aver-

ages over Asia and its three subregions: the high-, mid- and low-latitude regions, respectively

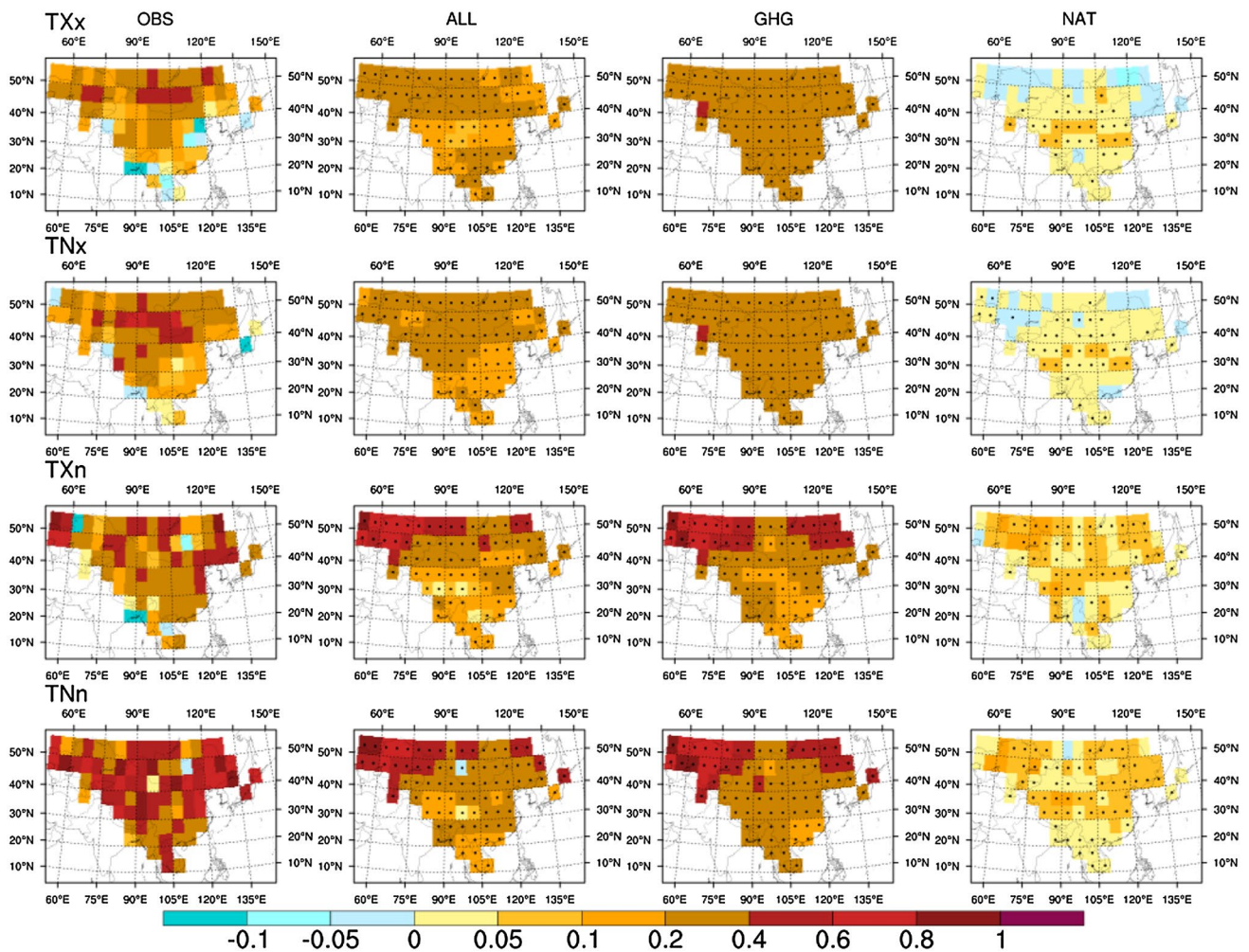
number of ice days (ID) at low latitudes, where the growing season is year-long and ice days do not occur.

### 3.2 Spatial and temporal patterns in observations and simulations

In this subsection, we focus on the spatial patterns of trends and the temporal evolution of the regional averages of the observed and simulated values of eight indices. The spatial patterns of trends in terms of their intensity and frequency indices are shown in Figs. 4 and 5, respectively. Positive trends appear in the observed intensity indices almost everywhere that sufficient data are available, thus reflecting the strengthening of warm extremes and the weakening of cold extremes. For the same index, trends tend to be larger at higher latitudes or in areas of higher elevation. The warm extremes experienced less warming than the cold extremes.

Cooling trends appear in patches in Eastern China, Japan and Southeast Asia. Some researchers have speculated that cooling in Eastern China may be related to increased air pollution (e.g., Kaiser and Qian 2002; Zhou and Ren 2014; Liao et al. 2015).

The spatial patterns of the multimodel ensemble mean intensity indices in the ALL and GHG simulations are consistent with those of observations, as stronger trends are obtained at high latitudes and in the cold extremes indices. The spatial pattern is smooth compared to that of observations, thus reflecting the effect of averaging across multiple simulations. The magnitudes of the simulated trends are also comparable to those of the observed trends, except for TNn, in which the simulated trend is weaker. The small cooling trend observed in parts of China, Japan and Southeastern Asia is not reproduced in these simulations. Multiple factors may have been involved, including natural internal



**Fig. 4** Spatial distribution of linear trends (°C/decade) of annual maxima of daily maximum and minimum temperatures (TXx, TNx) and annual minima of daily maximum and minimum temperatures (TXn, TNn) during 1958–2012 in the observations (OBS) and model responses to ALL, GHG and NAT forcings. Model trends were calculated based on the multimodel ensemble mean. The crosses on the model trend maps indicate the agreement of the trend sign in model simulations. A grid cell is marked with a dot if the trends of at least 75% of the individual simulations show the same sign

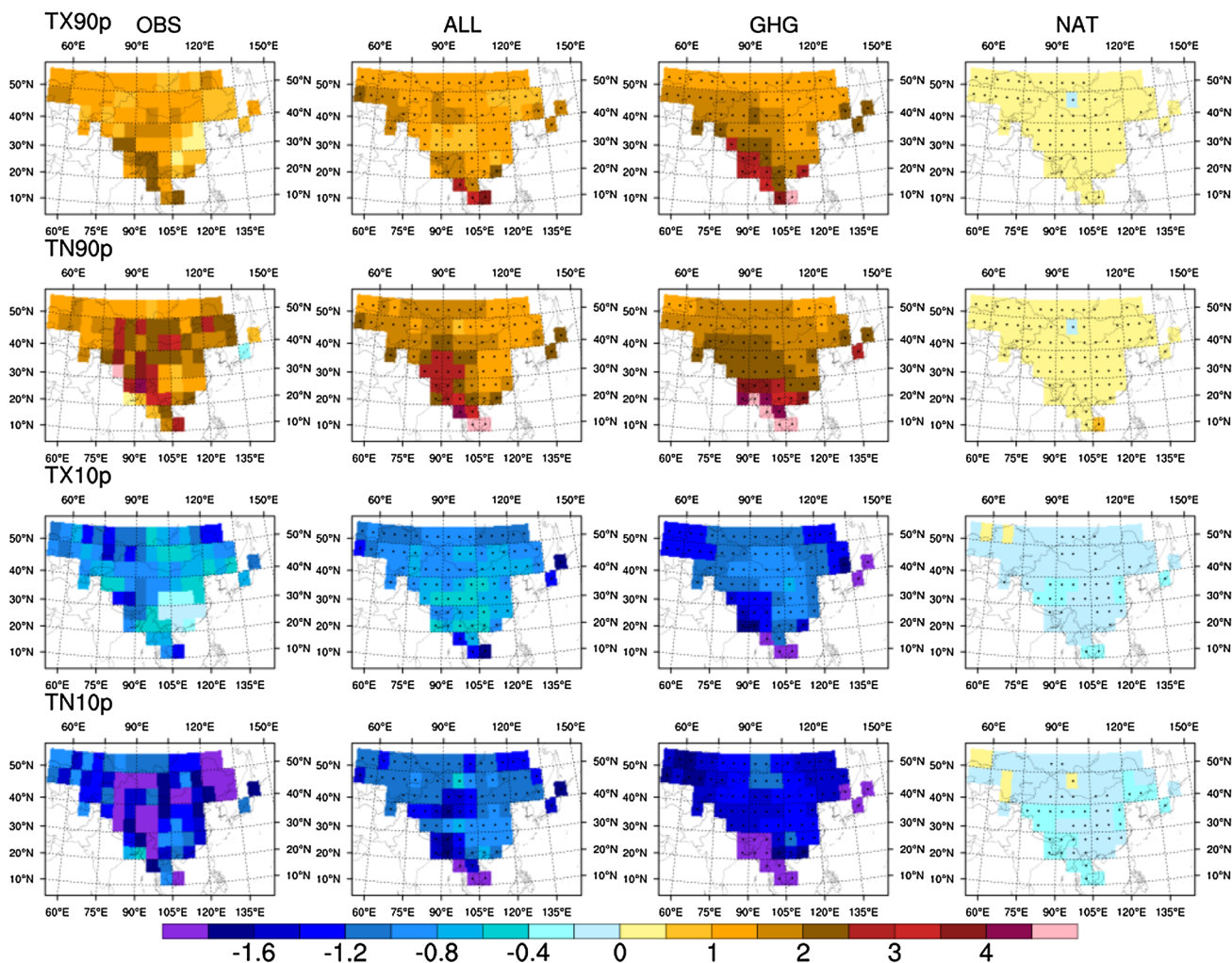
culated based on the multimodel ensemble mean. The crosses on the model trend maps indicate the agreement of the trend sign in model simulations. A grid cell is marked with a dot if the trends of at least 75% of the individual simulations show the same sign

variability and strong local aerosol forcing, that may have not been fully reflected in the CMIP5 forcing field. The response to the NAT forcing is a weak but spatially consistent warming trend. The magnitude of this trend is much smaller than that obtained in the GHG simulation and seen in the observations. This small warming trend reflects the combined effects of solar and volcanic forcing, as it is characterized by a small increase in solar forcing and the lack of major volcanic eruptions in recent years (Jones et al. 2013; Kim et al. 2016).

The observations (Fig. 5) show the increased frequency of warm extremes and the decreased frequency of cold extremes, which are consistent with warming. Large positive trends appear in the frequency of warm extremes. This should be interpreted in the proper context. The frequency of cold extremes has a lower bound of zero, thereby limiting

the magnitude of a negative trend. The trends are also stronger for the indices that are associated with night-time temperatures than those associated with day-time temperatures, i.e., trends in TN10p are stronger than those in TX10p, and trends in TN90p are also stronger than those in TX90p.

The signs of trends in the multimodel frequency indices under the ALL and GHG forcings are similar to those of observations. Some regional features in the magnitudes of trends are reproduced. For example, stronger TN90p trends in high-elevation areas and larger decreases in cold extremes in the subtropics are reproduced. However, smaller trends in observations near East Asia are not reproduced in the model simulation, thus suggesting that the weakening of the trends in the observations might be related to natural decadal variability. The signs of the trends of the NAT simulations are similar to those of observations and the ALL and



**Fig. 5** Same as Fig. 4, but for trends (%/decade) in frequency indices

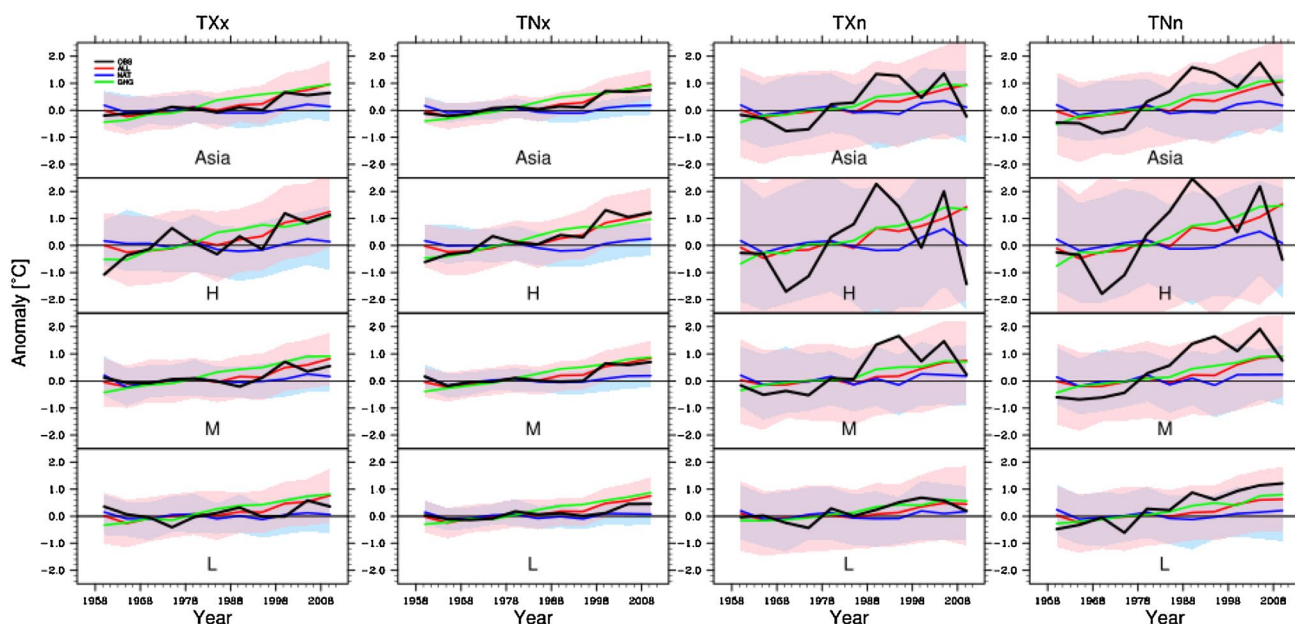
GHG forcings, but they are of a much smaller magnitude. This indicates that while NAT may have contributed to the observed changes in the frequency indices, natural forcings are unlikely to be the major cause of the observed changes.

The 5-year mean time series provides some information about the temporal evolution of these temperature indices. Figure 6 shows the observed and simulated intensity series for Asia and its latitudinal subregions. For the warmest day (TXx) and warmest night (TNx) temperatures, the observations are well within the range of the simulated responses to ALL and GHG forcing. For the coldest day (TXn) and coldest night (TNn) temperatures, the long-term changes in the simulated responses to ALL or GHG are similar to those of the observations; additionally, 90% of the range of model simulations does not completely cover the observed variability, thus indicating that the variability simulated by the model may be smaller than that seen in the observations. This is especially true at high latitudes. Figure 7 displays the observed and simulated frequency indices. In general, the

observations are well within the 90% range of the model-simulated response to ALL or GHG forcings and are consistent with the multimodel ensemble averages. However, they are not consistent with the simulated response to NAT forcing. These data suggest that GHG forcing can explain the observed changes, but NAT forcing cannot.

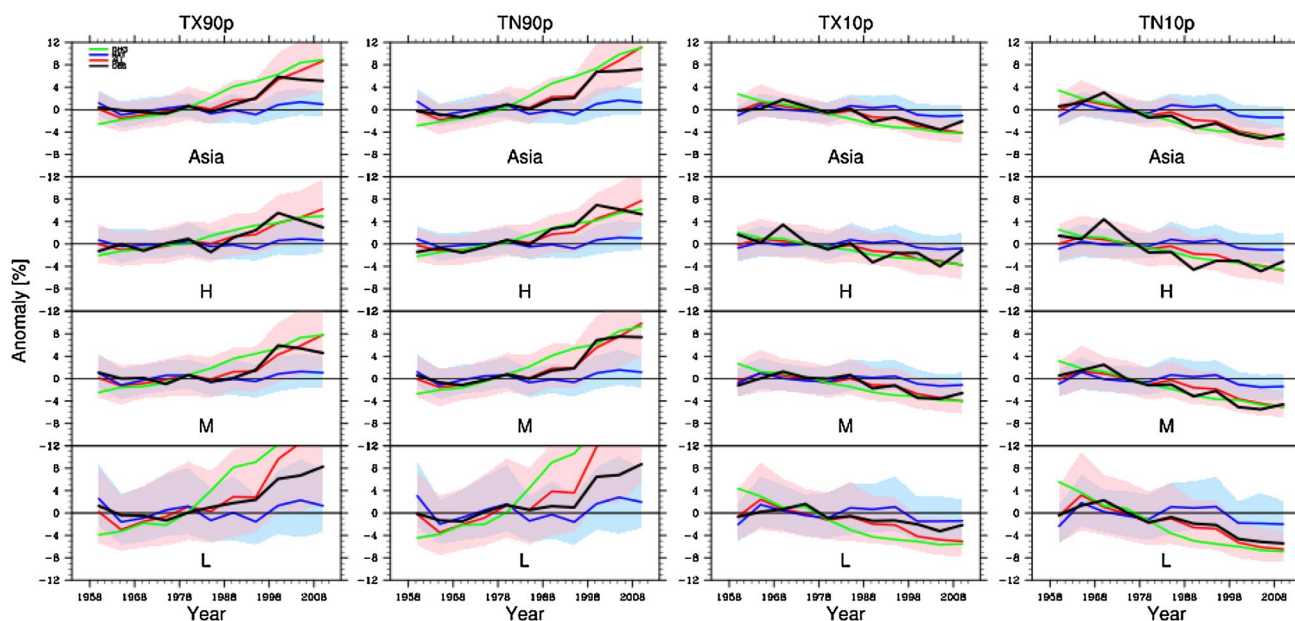
### 3.3 Detection results

Figure 8 summarizes the results of the single-signal analyses, including the best estimates of the scaling factor and their 90% confidence intervals. These results indicate that the simulated model response to ALL forcing can be detected in all intensity indices in both all of Asia and all three of its subregions, with the exceptions of TXx and TXn at lower latitudes. Residual consistency tests failed in three cases, which may affect the interpretation of detection results. Although these tests failed for TNx and TXn at low latitudes, this does not negate the detection results, as model-simulated



**Fig. 6** Time series of 5-year mean regional average anomalies (°C, relative to 1961–1990) for TXx, TNx, TXn, and TNn in the observations (OBS, black lines) and model simulations. The red, green and blue lines represent multimodel ensemble means in the ALL, GHG and NAT simulations. The light pink and light blue shadings indicate

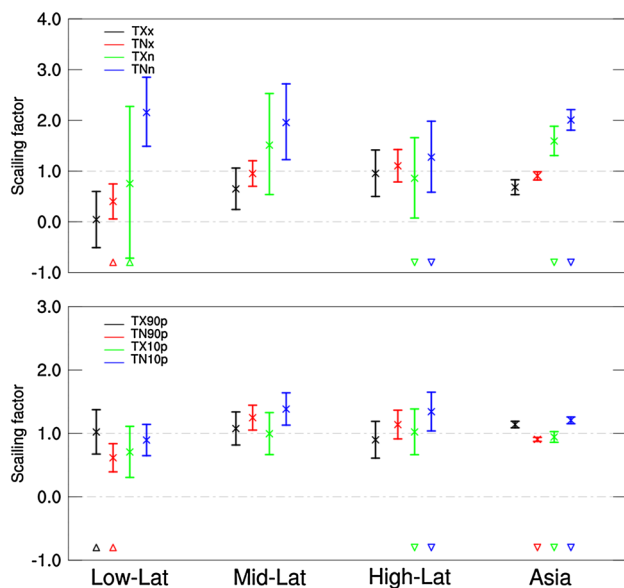
the 5–95% range of the individual model results under ALL and NAT forcings. The top (Asia), upper-middle (H), lower-middle (M) and bottom panels (L) show the regional averages for Asia and the high-, mid-, and low-latitude regions, respectively



**Fig. 7** Same as Fig. 6, but for 5-year mean anomalies (% relative to 1961–1990) of frequency indices

variability appeared to be too large in that region. This test also failed for TXn and TNn at high latitudes and in Asia; in this case, the model-simulated variability is likely too small, thus suggesting that the detection of ALL for TXn and TNn in that region may not be credible. The best estimates of

the scaling factor for the warm extremes TXx are smaller than one, but they are not inconsistent with the value of one for Asia as a whole and for the mid- and low-latitudes, thus indicating that models may overestimate the observed changes. The best estimates of the scaling factor for TNn



**Fig. 8** Best estimates of the scaling factors and their 5–95% confidence intervals from single-signal analyses, in which the observations are regressed to the model-simulated response to ALL forcing for the period of 1958–2012. Upward and downward triangles indicate that models may over- or under-simulate observed variability, respectively, according to the residual consistency tests. The upper panel shows the results for the intensity indices (e.g., TXx, TNx, TXn, TNn), while the lower panel shows the results for frequency indices. Low-Lat, Mid-Lat, High-Lat, and Asia represent the low-, mid-, and high-latitude regions and Asia, respectively

and TXn are significantly greater than one in Asia and at the mid-latitudes, thus reflecting the underestimation of the changes simulated by the model. The scaling factor is quite close to one at high latitudes. The model-simulated response for TNx appears to be consistent with the observed values across all subregions. Overall, our results support those of previous studies, which demonstrated that models underestimate the changes in TNn and TXn, there is good consistency between the model-simulated and observed values of TNx, and models overestimate the changes in TXx (Zwiers et al. 2011; Yin et al. 2016; Kim et al. 2016). Detection becomes more difficult at low latitudes, perhaps because the region containing available data is too small.

The 90% confidence intervals of the scaling factors for the frequency indices are all greater than zero, thus suggesting that the model-simulated response to ALL forcing can be detected in these indices. The variability in TX90p and TN90p simulated by models appears to be too large at low latitudes. Models also appear to simulate variability in TX10p and TN10p in Asia and at high latitudes that is too small, thus suggesting that detection may be less credible in these cases. The 90% confidence intervals are much narrower when compared with the detection results for the intensity indices. This may be because frequency indices

can more efficiently use information contained in daily data because they are based on data collected on all days, rather than just the largest or smallest values of annual temperature, which are used to create the intensity indices. The best estimates of the scaling factor are close to one in most cases, except for TN10p in mid- to high-latitude areas, where models appear to underestimate the observed changes, and for TN90p at low latitudes, where models appear to overestimate the observed changes.

Figure 9 plots the best estimates of the scaling factors and the 90% confidence intervals for the ANT and NAT experiments from the two-signal analyses, in which observational data were simultaneously regressed onto the model-simulated responses to ANT and NAT forcings. For the intensity indices, both ANT and NAT can be detected and separated from each other at mid-latitudes and in Asia as a whole. ANT is also detected at high latitudes for warm extremes. The scaling factors for NAT include negative values, thus indicating that the NAT signal is not detectable in the changes of the coldest and warmest extremes at high latitudes. Similar to the single-signal detection results, model simulations overestimate the observed changes in warm extremes but underestimate the observed changes in cold extremes. Overall, it appears that long-term changes in the observed intensity indices over Asia are influenced by both anthropogenic and natural external forcings and that this influence can even be detected in a relatively small area in the mid-latitudes, thus providing strong evidence for the influence of external forcings on extreme temperatures in this region.

For the frequency indices, the influences of ANT forcing are clearly detected in both Asia and its three subregions, and for all indices; this represents very robust evidence that anthropogenic external forcings have affected the occurrences of those moderate extremes. The 90% confidence intervals of the scaling factors are smaller than those of the intensity indices for the same region and forcing, which suggests that evidence of the influence of external forcing may be stronger in the moderate extremes because more information from the available data have been used. These results are consistent with those of previous studies that have been performed in parts of Asia (e.g., Morak et al. 2013; Lu et al. 2016).

## 4 Discussion and conclusions

In this paper, we first reported trends in extreme temperature indices in Asia calculated using observed daily data collected from over 2400 stations. The collection of station data, as well as follow-up work involving careful data homogenization, was organized by the ETCCDI to fill in the data gap over this region. Data from these indices will

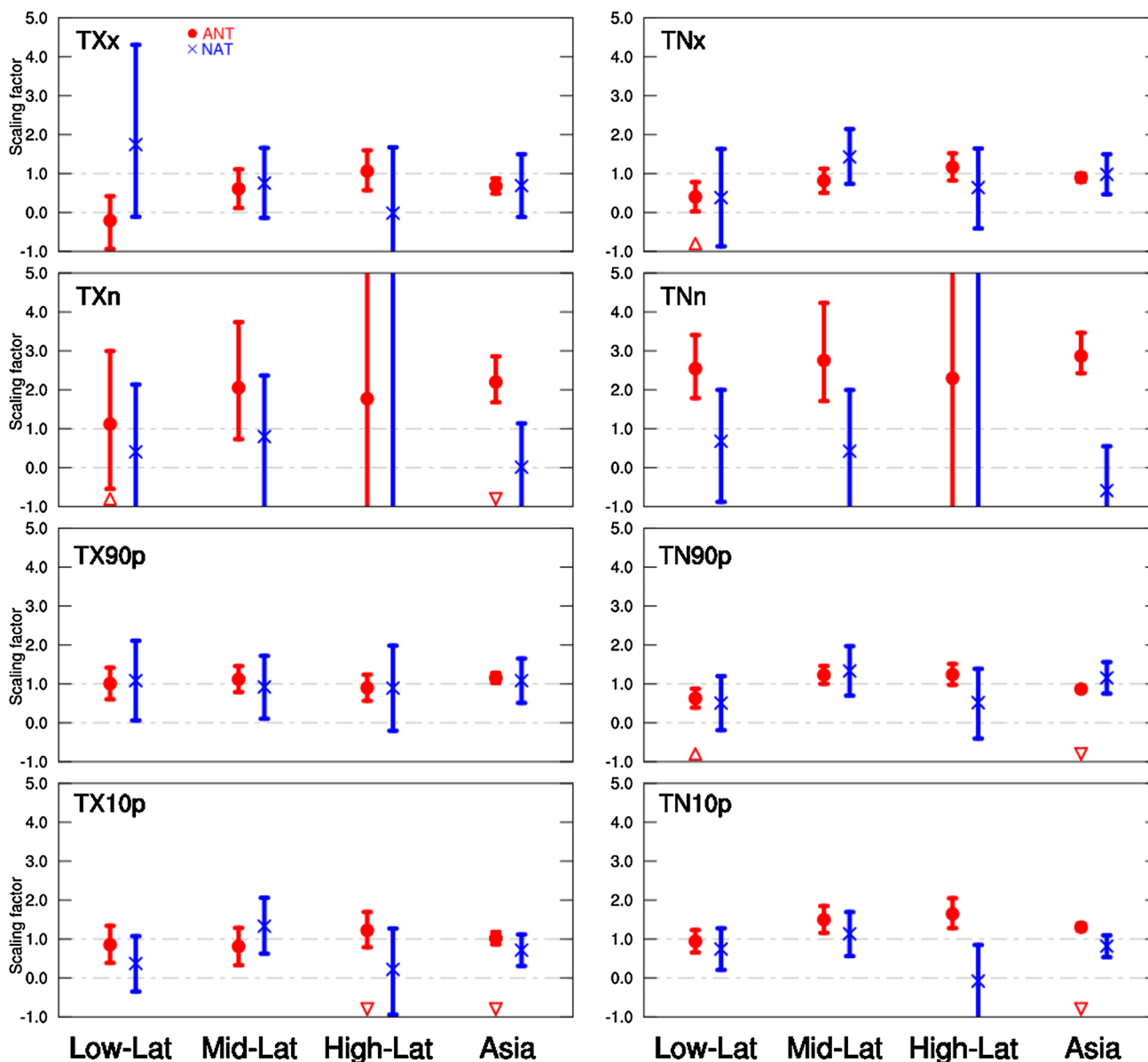


Fig. 9 Same as Fig. 8, but for scaling factors and 5–95% confidence intervals for ANT (red) and NAT (blue) from the two-signal analyses

be incorporated into the global dataset that feeds into the production of HadEX2 (Donat et al. 2013b). Compared with earlier studies performed in this region, we were able to use station observations in many countries in this region that were not previously publicly available. The indices we examined show significant trends that are consistent with warming in the low-, mid-, and high-latitudes of Asia. These include increases in the warm extremes, the length of the growing season and the number of summer days and tropical nights, as well as decreases in cold extremes and the number of frost or ice days. These trends are generally stronger at higher latitudes. These results are consistent

with those of earlier studies, thus indicating the robustness of the warming trend.

To understand the causes of these trends, we then performed detection and attribution analyses by comparing observations with the results of simulations conducted using the CMIP5 models. The indices being compared included four intensity indices that represent the annual maxima and minima of daily maximum and daily minimum temperatures and four frequency indices that characterize the occurrence of moderate extremes in daily maximum and daily minimum temperatures. The results clearly detected the influence of anthropogenic and natural

external forcings on the changes in these temperature indices in both Asia and its latitudinal subregions. We also found that models generally overestimate changes in warm extremes but underestimate changes in cold extremes.

Using data from stations that were not previously publicly available, we showed that not only the trends but also the detection and attribution results are consistent with the results of previous studies that covered all of Asia or parts of Asia (e.g., Donat et al. 2013b; Zwiers et al. 2011; Morak et al. 2013; Min et al. 2013; Kim et al. 2016; Lu et al. 2016). That these findings are consistent despite the use of slightly different methods to delineate regions, the use of different datasets, and that the influence of anthropogenic and natural external forcings can be detected and separated from each other for some indices, even in relatively small regions at low latitudes, clearly indicates that the warming trend is robust, as is the attribution of this warming trend to external forcings.

**Acknowledgements** We thank Dáithí Stone and two anonymous reviewers for their very helpful comments that led to the significant improvement of the paper. The International Workshop on Climate Data Requirements and Applications that was held in Nanjing, China, from 4–8 March 2013 was jointly funded by the World Meteorological Organization and the China Meteorological Administration. This workshop enabled the collection and dissimulation of the observation-based indices data in this paper. We acknowledge the World Climate Research Programme's Working Group on Coupled Modeling, as well as the climate modeling groups, for producing and making available their model output. S. Dong and Y. Sun are supported by National Key R&D Program of China 2017YFA0603804, China NSF 41675074, Climate Change Project CCSF201704 and GYHY201406020. We thank workshop participants from the following countries for providing daily data and computing the indices during and after the Nanjing Workshop: Bangladesh: Nurul Karim; Bhutan: Sonam Doril; Hong Kong: Tsz-cheung Lee; Indonesia: Urip Haryoko; Japan: Kenji Kamiguchi; Kazakhstan: Aliya Mamytova; Kyrgyzstan: Zoia Kretova; Laos: Surinh Kousoonsavath; Malaysia: Farah Safura Othman; the Maldives: Zahid Hameed; Mongolia: Davaanyam Enkhbaatar; Russia: Oleg Alexandrovich Anisimov and Vasily Kokorev; Thailand: Yuwadee Suwananee; Vietnam: Nguyen Mau Dang.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Alexander LV et al (2006) Global observed changes in daily climate extremes of temperature and precipitation. *J Geophys Res* 111:D05109
- Allen MR, Stott PA (2003) Estimating signal amplitudes in optimal fingerprinting, part I: theory. *Clim Dyn* 21:477–491. doi:10.1007/s00382-003-0313-9
- Allen MR, Tett SFB (1999) Checking for model consistency in optimal fingerprinting. *Clim Dyn* 15:419–434
- Bindoff NL et al (2013) Detection and attribution of climate change: from global to regional. In: Stocker TF et al (eds) *The physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change*. Cambridge University Press, Cambridge
- Caesar J, Alexander LV, Trewin B et al (2011) Changes in temperature and precipitation extremes over the Indo-Pacific region from 1971 to 2005. *Int J Climatol* 31:791–801. doi:10.1002/joc.2118
- Cao L, Zhu Y, Tang G, Yan Z (2016) Climatic warming in China according to a homogenized data set from 2419 stations. *Int J Climatol* 4392:4384–4392. doi:10.1002/joc.4639
- Christidis N, Stott PA, Brown S et al (2005) Detection of changes in temperature extremes during the second half of the 20th century. *Geophys Res Lett*. doi:10.1029/2005GL023885
- Donat MG, Alexander LV, Yang H et al (2013a) Global land-based datasets for monitoring climatic extremes. *Bull Am Meteor Soc* 94:997–1006. doi:10.1175/BAMS12-00109.1
- Donat MG et al (2013b) Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: the HadEX2 dataset. *J Geophys Res Atmos* 118:2098–2118
- Donat MG, Peterson TC, Brunet M et al (2014) Changes in extreme temperature and precipitation in the Arab region: long-term trends and variability related to ENSO and NAO. *Int J Climatol* 34(3):581–592
- Dong SY, Xu Y, Zhou BT, Shi Y (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
- Hijioka Y, Lin E, Pereira JJ et al (2014) Aisa. In: Barros VR et al (eds) *Climate change 2014 impacts, adaptation, and vulnerability part B: regional aspects. Contribution of working group II to the fifth assessment report of the intergovernmental panel on climate change*. Cambridge University Press, Cambridge
- IPCC (2012) Summary for policymakers. In: Field CB, Barros V, Stocker TF, Qin D, Dokken DJ, Ebi KL, Mastrandrea MD, Mach KJ, Plattner G-K, Allen SK, Tignor M, Midgley PM (eds) *Managing the risks of extreme events and disasters to advance climate change adaptation. A special report of working groups I and II of the intergovernmental panel on climate change*. Cambridge University Press, Cambridge, pp 1–19
- Jones GS, Stott PA, Christidis N (2013) Attribution of observed historical near-surface temperature variations to anthropogenic and natural causes using CMIP5 simulations. *J Geophys Res* 118:4001–4024. doi:10.1002/jgrd.50239
- Kaiser DP, Qian Y (2002) Decreasing trends in sunshine duration over China for 1954–1998: indication of increased haze pollution?. *Geophys Res Lett* 29(21):38-1–38-4. doi:10.1029/2002GL016057
- Kim YH, Min SK, Zhang X et al (2016) Attribution of extreme temperature changes during 1951–2010. *Clim Dyn* 46(5–6):1769–1782. doi:10.1007/s00382-015-2674-2
- Liao H, Chang WY, Yang Y (2015) Climatic effects of air pollutants over China: a review. *Adv Atmos Sci* 32(1):115–139. doi:10.1007/s00376-014-0001-1
- Lu C, Sun Y, Wan H et al (2016) Anthropogenic influence on the frequency of extreme temperatures in China. *Geophys Res Lett* 43(12):6511–6518. doi:10.1002/2016GL069296
- Menne MJ, Durre I, Vose RS et al (2012) An overview of the global historical climatology network-daily database. *J Atmos Ocean Technol* 29(7):897–910
- Mestre O, Domonkos P, Picard F et al (2013) HOMER: a homogenization software-methods and applications. *Q J Hung Meteor Serv* 117:47–67

- Min SK, Zhang X, Zwiers FW, Shiogama H, Tung YS, Wehner M (2013) Multimodel detection and attribution of extreme temperature changes. *J Clim* 26:7430–7451
- Morak S, Hegerl GC, Kenyon J (2011) Detectable regional changes in the number of warm nights. *Geophys Res Lett* 38:L17703. doi:10.1029/2011GL048531
- Morak S, Hegerl GC, Christidis N (2013) Detectable changes in the frequency of temperature extremes. *J Clim* 26:1561–1574
- NOAA National Centers for Environmental Information (2017) Climate at a glance: global time series. Published May 2017 <http://www.ncdc.noaa.gov/cag/>. Accessed 16 June 2017
- Peterson TC, Manton MJ (2008) Monitoring changes in climate extremes: a tale of international collaboration. *Bull Am Meteorol Soc* 89(9):1266–1271. doi:10.1175/2008BAMS2501.1
- Ribes A, Planton S, Terray L (2013) Application of regularized optimal fingerprinting to attribution. Part I: method, properties and idealized analysis. *Clim Dyn* 41:2817–2836
- Sen PK (1968) Estimates of the regression coefficient based on Kendall's tau. *J Am Stat Assoc* 63(324):1379–1389
- Sillmann J, Kharin VV, Zhang X, Zwiers FW, Bronaugh D (2013a) Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. *J Geophys Res* 118:1716–1733. doi:10.1002/jgrd.50203
- Sillmann J, Kharin VV, Zwiers FW, Zhang X, Bronaugh D (2013b) Climate extremes indices in the CMIP5 multimodel ensemble: part 2. Future climate projections. *J Geophys Res* 118:2473–2493. doi:10.1002/jgrd.50188
- Sun Y, Zhang X, Zwiers FW et al (2014) Rapid increase in the risk of extreme summer heat in Eastern China. *Nat Clim Change* 4:1082–1085. doi:10.1038/nclimate2410
- 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
- Vincent L, Zhang X, Bonsal B, Hogg W (2002) Homogenization of daily temperatures over Canada. *J Clim* 15:1322–1334
- Wang XL, Feng Y (2010) RHtestsV3 user manual, UserManual.doc. Climate Research Division, Science and Technology Branch, Environment Canada, Toronto. <http://cccma.seos.uvic.ca/ETC-CDMI/RHtest/RHtestsV3>. Accessed 1 May 2014
- Wang XL, Swail VR (2001) Changes of extreme wave heights in Northern Hemisphere oceans and related atmospheric circulation regimes. *J Clim* 14:2204–2220
- Wen QH, Zhang X, Xu Y et al (2013) Detecting human influence on extreme temperatures in China. *Geophys Res Lett* 40:1171–1176. doi:10.1002/grl.50285
- Yin H, Sun Y, Wan H, Lu C (2016) Detection of anthropogenic influence on the intensity of extreme temperatures in China. *Int J Climatol*. doi:10.1002/joc.4771
- You Q, Kang S, Aguilar E et al (2008) Changes in daily climate extremes in the eastern and central Tibetan Plateau during 1961–2005. *J Geophys Res* 113:D07101. doi:10.1029/2007JD009389
- You Q, Kang S, Flügel WA et al (2009) Does a weekend effect in diurnal temperature range exist in the eastern and central Tibetan Plateau? *Environ Res Lett* 4(4):045202. doi:10.1088/1748-9326/4/4/045202
- Zhang X, Vincent LA, Hogg WD, Niitsoo A (2000) Temperature and precipitation trends in Canada during the 20th century. *Atmos Ocean* 38:395–429
- Zhang X, Aguilar E, Sensoy S et al (2005) Trends in Middle East climate extreme indices from 1950 to 2003. *J Geophys Res Atmos*. doi:10.1029/2005JD006181
- Zhang X, Alexander L, Hegerl GC et al (2011) Indices for monitoring changes in extremes based on daily temperature and precipitation data. *Wiley Interdiscip Rev Clim Change* 2:851–870. doi:10.1002/wcc.147
- Zhou YQ, Ren GY (2014) Urbanization effect on long-term trends of extreme temperature events in North China. *Plateau Meteorol* 33:1589–1598
- Zwiers FW, Zhang XB, Feng Y (2011) Anthropogenic influence on long return period daily temperature extremes at regional scales. *J Clim* 24:881–892