

Detection and Attribution of Changes in Extreme Temperatures at Regional Scale

ZHUO WANG

College of Economics, Shenzhen University, Shenzhen, Guangdong, China

YUJING JIANG

Department of Statistics, University of Connecticut, Storrs, Connecticut

HUI WAN

Climate Data Analysis Section, Environment and Climate Change Canada, Toronto, Ontario, Canada

JUN YAN

Department of Statistics, University of Connecticut, Storrs, Connecticut


XUEBIN ZHANG


Climate Data Analysis Section, Environment and Climate Change Canada, Toronto, Ontario, Canada

(Manuscript received 14 November 2015, in final form 4 April 2017)

ABSTRACT

This paper improves an extreme-value-theory-based detection and attribution method and then applies it to four types of extreme temperatures, annual minimum daily minimum (TNn) and maximum (TXn) and annual maximum daily minimum (TNx) and maximum (TXx), using the HadEX2 observation and the CMIP5 multimodel simulation datasets of the period 1951–2010 at 17 subcontinent regions. The methodology is an analog of the fingerprinting method adapted to extremes using the generalized extreme value (GEV) distribution. The signals are estimated as the time-dependent location parameters of GEV distributions fitted to extremes simulated by multimodel ensembles under anthropogenic (ANT), natural (NAT), or combined anthropogenic and natural (ALL) external forcings. The observed extremes are modeled by GEV distributions whose location parameters incorporate the signals as covariates. A coordinate descent algorithm improves both computational efficiency and accuracy in comparison to the existing method, facilitating detection of multiple signals simultaneously. An overall goodness-of-fit test was performed at the regional level. The ANT signal was separated from the NAT signal in four to six regions. In these analyses, the waiting times of the 1951–55 20-yr return level in the 2006–10 climate for the temperature of the coldest night and day were found to have increased to over 20 yr; the corresponding waiting times for the warmest night and day were found to have dropped below 20 yr in a majority of the regions.

 Denotes content that is immediately available upon publication as open access.

 Supplemental information related to this paper is available at the Journals Online website: <http://dx.doi.org/10.1175/JCLI-D-15-0835.s1>.

Corresponding author: Jun Yan, jun.yan@uconn.edu

1. Introduction

Increases in extreme temperatures have been observed since the 1950s (e.g., [Brown et al. 2008](#); [Donat et al. 2013](#)), and demonstration of the influence of anthropogenic (ANT) and/or natural (NAT) external forcing on the changes in extreme temperatures has gained sharpened focus. The increase in the frequency of warm extremes and decrease in the frequency of cold

DOI: 10.1175/JCLI-D-15-0835.1

© 2017 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#) (www.ametsoc.org/PUBSReuseLicenses).

extremes are very likely due to human influence (Pachauri et al. 2014). Some detection and attribution studies have used the standard optimal fingerprinting method (e.g., Allen and Stott 2003), a widely accepted method designed for mean climatic states, directly to certain measures of temperature extremes. Christidis et al. (2005) compared the observations of the warmest days and nights with the corresponding values from climate model simulations under various forcings at the global scale. The anthropogenic influence was found in changes in extremely warm nights and in extremely cold nights and days but less robustly. Similar findings were reported in Shiogama et al. (2006) with different climate models. Morak et al. (2011) studied the global and regional changes in the number of warm nights annually and concluded that the anthropogenic influence contributed to the increasing trend in the number of warm nights. Morak et al. (2013) further investigated changes in the frequency of warm and cold extremes in warm and cold seasons. These applications of the standard optimal fingerprinting method did not account for the unique distributional properties of the extreme values.

Some studies used the extreme value theory indirectly by applying the standard optimal fingerprinting to estimated parameters of extreme value models for the climate simulations and observations. Using the threshold approach for extreme value analysis, Christidis et al. (2011) fitted a generalized Pareto (GP) distribution with location parameter varying every decade at each grid box for both simulated and observed data and then applied the optimal fingerprinting method to the estimated location parameters. They found the combined influence of anthropogenic and natural external forcings (ALL) in the increasing of warm extremes. Min et al. (2013) transformed the extreme temperatures into probability-based indices through fitted generalized extreme value (GEV) distribution and then applied the optimal fingerprinting to the standardized indices. Their single-signal and two-signal detection analyses confirmed previous findings and, for the first time, showed that ANT signals were separable from NAT signals at global and subcontinental scales. Using the same method but updated observations and multimodel simulations, Kim et al. (2016) found the results to be generally insensitive to either different observed data coverage or different model samples. While the methods in these studies were able to infer changes in some mean conditions of extremes, they did not provide easy ways to assess changes in rare, high-impact, long-return-period events.

Zwiers et al. (2011) explicitly built extreme value theory into a new framework for the detection and attribution of changes in extremes. In their approach, the expected responses to external forcing were estimated

by fitting GEV distributions, with decadal varying location parameters, to model-simulated annual maximum or minimum values of daily maximum or minimum temperatures at each grid box under each forcing and treating the estimated location parameters as signals. The GEV distributions were also fitted to the observed extreme temperatures with the signals as a covariate in the location parameters. Similar to the standard fingerprinting, it was assumed that the coefficient of the signal, also known as scaling factor, was the same across the whole region under consideration; see section 3 for more details. The scaling factor was estimated from an independence likelihood profiled over a prespecified grid of possible values. A block bootstrap method that retains the spatial and temporal dependence was used to assess the uncertainty of the estimator. The method of Zwiers et al. (2011) is an important methodological advance for detection and attribution analysis of climate extremes. Nonetheless, their profiling algorithm when combined with the required bootstrap procedure for inferences is very time-consuming. It took weeks in a small computer cluster to complete the one-signal analysis at 17 subcontinental regions presented in their paper. Considering multiple signals, a must for properly attributing observed changes to various external forcings such as greenhouse gases, anthropogenic aerosols, and natural external forcing, would be computationally prohibitive. In particular, the amount of computation for a two-signal analysis would be quadratic to that for one-signal analysis. Further, the accuracy of estimated scaling factor depends on the fineness and range of the grid of possible values.

This study provides a solution to the computational problem of the fingerprinting framework in Zwiers et al. (2011). The contribution of this article is twofold. The first is methodological—a coordinate descent algorithm to maximize the independence likelihood improves both computational efficiency and accuracy relative to the profile approach. The coordinate descent algorithm is much faster than the profile approach, facilitating multiple-signal analyses, and its accuracy does not depend on any prespecified grid. Our second contribution is more substantive. This is the first multiple-signal detection and attribution analysis for extreme temperatures that explicitly considers distributional properties of extreme values based on the concept proposed by Zwiers et al. (2011). We focus on the two-signal analyses in which both the ANT and the NAT signals are incorporated as covariates in the location parameter of the GEV distribution simultaneously. Two-signal analyses were performed on certain summary statistics of the temperature extremes (regional averages of probability-based indices) (Min et al. 2013; Kim et al. 2016), but our

TABLE 1. Summaries of the names and coordinate ranges of the 17 regions analyzed in the study. The Mediterranean basin (MED) and Greenland (GRL) in Giorgi and Francisco (2000) are renamed as SEU and CGL, respectively, to be consistent with those in Zwiers et al. (2011).

| Acronym | Region name | Latitude (°N) | Longitude (°E) | Acronym | Region name | Latitude (°N) | Longitude (°E) |
|---------|----------------------------|---------------|----------------|---------|-----------------|---------------|----------------|
| ALA | Alaska | 60/72 | -170/-103 | SAF | Southern Africa | -35/-12 | 10/52 |
| CGL | Canada and Greenland | 50/85 | -103/-10 | NAS | North Asia | 50/70 | 40/180 |
| WNA | Western North America | 30/60 | -130/-103 | CAS | Central Asia | 30/50 | 40/75 |
| CNA | Central North America | 30/50 | -103/-85 | TIB | Tibet | 30/50 | 75/100 |
| ENA | Eastern North America | 25/50 | -85/-50 | EAS | East Asia | 20/50 | 100/145 |
| CAM | Central America and Mexico | 10/30 | -116/-83 | SAS | South Asia | 5/30 | 65/100 |
| SSA | Southern South America | -56/-20 | -76/-40 | SEA | Southeast Asia | -11/20 | 95/155 |
| NEU | Northern Europe | 48/75 | -10/40 | AUS | Australia | -45/-11 | 110/155 |
| SEU | Southern Europe | 30/48 | -10/40 | | | | |

analysis directly works on the GEV distributions of the temperature extremes. We also did one-signal analyses that differ from Zwiers et al. (2011) in two aspects. First, the signals were assumed to be constant in every 5 yr instead of 10 yr, which increases the chances to capture short-lived responses to events such as volcanic eruptions (much of the volcanic signal would be averaged out if 10-yr window is used). Second, our data cover 17 subcontinent regions over 1951–2010, 10 yr longer than those used in Zwiers et al. (2011), enabling us to examine the robustness of previous results.

The rest of this paper is organized as follows. The observational and climate model simulation data for the attribution analyses are described in section 2. The methodology is presented in section 3, covering the extreme value models for both signals and observations, parameter estimation with a coordinate descent algorithm, assessment of uncertainty in inferences with a block bootstrap procedure, and a goodness-of-fit test for the models for the observed data. The results are discussed in detail in section 4. Conclusions are given in section 5.

2. Data

Four types of extreme temperatures were considered: annual minimum daily minimum (TN_n) and maximum (TX_n) and annual maximum daily minimum (TN_x) and maximum (TX_x). The observations were obtained from the HadEX2 dataset (Donat et al. 2013), with 3.75° longitude × 2.5° latitude grid, of period 1951–2010. Our

analyses were conducted on Giorgi regions, which were subjectively defined by dividing the world land into 21 regions for evaluating climate models at the regional scale (Giorgi and Francisco 2000, their Fig. 1). These regions represent different climatic regimes and physiographic settings and have been widely used in other climate studies at regional scales including climate change (e.g., Zwiers et al. 2011); see also Figs. SPM 4A and 4B of IPCC (2012). As HadEX2 has very poor data coverage for some Giorgi regions, we only include 17 Giorgi regions in this study (see Table 1). In each region, the grid boxes were selected such that each one has at least 50-yr nonmissing annual extreme values for the 60-yr period for each of the four types of temperature extremes.

Climate model simulations were obtained from phase 5 of the Coupled Model Intercomparison Project (CMIP5) multimodel datasets (Taylor et al. 2012). Most CMIP5 historical simulations end in 2005. There were, however, 22 ensembles from six climate models under the ALL forcing and 26 ensembles from six climate models under the NAT forcing that had extended simulations to 2010. These simulations were used here; see Table 2. For each of the four extreme temperatures from each ensemble, annual extreme temperatures were first regridded to match the spatial resolution of the HadEX2 data. Because of the spatial contiguity in temperature, this regridding process should not alter the spatial patterns of the signals. The regridded data were then masked by the observational coverage to match the spatial and temporal coverage of the observed data. Finally, grid boxes were retrieved for each region using

TABLE 2. List of CMIP5 models analyzed in this study. The model names and ensemble sizes are given for ALL and NAT forcing.

| ALL | | | | NAT | | | |
|---------------|------|-----------|------|--------------|------|---------------|------|
| Model | Size | Model | Size | Model | Size | Model | Size |
| BCC_CSM1.1 | 3 | MIROC5 | 5 | CanESM2 | 5 | CSIRO Mk3.6.0 | 5 |
| BCC_CSM1.1(m) | 3 | MRI-CGCM3 | 3 | IPSL-CM5A-MR | 3 | HadGEM2-ES | 4 |
| CanESM2 | 5 | NorESM1-M | 3 | CNRM-CM5 | 6 | IPSL-CM5A-LR | 3 |

the same criteria for data availability as for the observations.

3. Methods

The fingerprinting model of Zwiers et al. (2011) to be used in this study takes the distributional properties of extremes into consideration. This is in contrast to and has an important advantage over the standard fingerprinting method, which was designed for mean climate states and cannot appropriately account for the distributional properties of extremes. The standard fingerprinting model is a linear model where the space–time variation of an observed climate variable is regressed on the expected climate responses (or fingerprints) to external forcings to determine whether the signals are present in the observed data (e.g., Hasselmann 1997; Hegerl et al. 2007; Ribes et al. 2013). The signals of external forcings are estimated from climate model simulations under the relevant forcings (Allen and Stott 2003). It is typically assumed that the climate models are able to simulate climate response patterns (in both space and time) correctly while the magnitude of the responses may differ from those in the observations. Based on this assumption, the task of a detection and attribution analysis becomes an estimation and inference problem of the scaling factors, common for all space and time dimensions, which scale model-simulated patterns to provide the best match to the observations. Detection of a signal is claimed if the corresponding scaling factor

is significantly greater than zero, and attribution is claimed if the confidence interval of the scaling factor also includes one. The inferences are made with quite mature practical guidance in a weighted least squares framework with weight matrix chosen to reflect the natural climate variability over space and time (Hegerl et al. 2010; Hegerl and Zwiers 2011; Zwiers et al. 2014). To account for the extreme value distributions of climate extremes, Zwiers et al. (2011) adapted the concept of fingerprinting to extremes, by fitting a GEV distribution to observed extremes at each grid box with climate-model-simulated signals as a covariate in the location parameter. The parameters of the GEV distributions for a region are estimated jointly such that the coefficient (or scaling factor) for every grid box is the same across the entire spatial domain. The R code for the proposed method is available in the online supplemental material.

a. Fingerprinting extremes

The GEV distribution arises as the limit distribution of properly normalized maximum of a sequence of independent and identically distributed random variables [see, e.g., Dey et al. (2015) for a recent review]. The probability density function of a GEV distribution is

$$f(y; \mu, \sigma, \xi) = \frac{1}{\sigma} g(y; \mu, \sigma, \xi)^{\xi+1} e^{-g(y; \mu, \sigma, \xi)}, \quad (1)$$

with

$$g(y; \mu, \sigma, \xi) = \begin{cases} \left[1 + \left(\frac{y - \mu}{\sigma} \right) \xi \right]^{-1/\xi}, & 1 + \left(\frac{y - \mu}{\sigma} \right) \xi > 0, \quad \xi \neq 0; \\ e^{(y - \mu)/\sigma}, & \xi = 0, \end{cases}$$

where μ , σ , and ξ are known as the location, scale, and shape parameters, respectively. The shape parameter ξ determines the tail behavior; negative indicates light tail while positive indicates heavy tail. When $\xi = 0$, the distribution becomes the Gumbel distribution. The GEV distribution will be used to model the maximum extremes (TNx and TXx) of both the observed and simulated data. For minimum extremes (TNn and TXn), we take the negative values of the data before applying the GEV modeling because the GEV distribution is meant for maxima instead of minima; the interpretation is still on the original scale.

Following the framework set up by Zwiers et al. (2011), a GEV distribution is specified for the observed extremes at each grid box, with the signals from certain external forcings for this grid box used as covariates for

the location parameters of the GEV distribution. The coefficients of the covariates, shared at all grid boxes in the region, are the common scaling factors of interest in a fingerprinting method. The scale and shape parameters are allowed to vary across grid boxes. Specifically, suppose that extreme temperatures are available at m grid boxes over n years in a region. The observed annual extreme climate variable $\mathbf{Y}_{t,s}$ at grid box s in year t , $s = 1, \dots, n$, is modeled by a GEV distribution $f(\cdot; \mu_{t,s}, \sigma_s, \xi_s)$ with

$$\mu_{ts} = \alpha_s + \mathbf{X}_{t,s}^T \boldsymbol{\beta}, \quad (2)$$

where $\mathbf{X}_{t,s}$ is a $p \times 1$ vector of the signals of p external forcings of interest at grid box s in year t , $\boldsymbol{\beta}$ is the vector of scaling factors shared by all grid boxes within the

region, and α_s , σ_s , and ξ_s are gridbox-specific location, scale, and shape parameters, respectively. The signals $\mathbf{X}_{t,s}$ are to be estimated from climate model output data as described in the next subsection and treated as known inputs here. This framework could be applied to partial duration series, but in practice such application is constrained by two factors: the availability of observed daily temperature data are much more limited, and automated threshold selection is challenging. So we focus on annual maxima/minima data as other works.

The assumptions in Zwiers et al. (2011) remain unchanged, which could be grouped into two categories for ease of discussion. The first category is statistical on the observed and simulated extremes. A GEV distribution is assumed at each grid box for the annual extremes, which is widely accepted (e.g., Kharin and Zwiers 2000). Only the location parameter of the GEV distribution is modified by the signals from the external forcings, while the scale and shape are unaffected. This is reasonable because non-stationarity has been reported to be primarily associated with the location parameter at most grid boxes (Kharin et al. 2005; Brown et al. 2008). We recognize that the scale parameter may change in some areas due to, for example, land-atmosphere feedback or changes in circulation (Seneviratne et al. 2006; Screen 2014), but at present overall and over large regions changes in scale parameter are less important than changes in location parameter. This is in no conflict with the conclusion “variability is more important than averages” of Katz and Brown (1992) in that their variability and average refer to those of daily climate variable instead of annual extremes. The annual extremes can be spatially and temporally dependent as long as the dependence disappears when two observations are sufficiently far apart. The multiple signals can be correlated, but the design matrix they form at each grid box needs to be of full rank as in a multiple regression. The second category of assumptions is physical. As in other detection and attribution analysis, we assume that the climate models are representative of the real-world climate system and can correctly simulate the spatial and temporal patterns of climate responses to external forcings, although the magnitudes of the response patterns may differ from those in the observations. This assumption implies that the scaling factor β is shared by all grid boxes over all years in the same region. Note that the “signals” estimated from model simulation are still different from one location to another location within the same region.

The focus of the detection analysis is the inference about β , which is analogous to that in the standard fingerprinting method. Similar to the standard fingerprinting setting, a response to external forcing is said to be “detected” if the corresponding scaling factor is found to be significantly greater than zero; further, if the

corresponding confidence interval also covers unity, then there is no evidence that observation and model simulations are inconsistent, and an attribution is claimed. The observed change is considered to be underestimated by the climate models if the interval lies above 1 and overestimated by the climate models if the interval lies between 0 and 1.

b. Signal estimation

The signals $\mathbf{X}_{t,s}$ in the fingerprinting model in (2) are obtained from the estimated location parameters of the GEV distribution of the extremes from ensemble simulations by multiple climate models. We first estimate signals based on simulations by individual models. Signals under each forcing (ALL or NAT) are estimated from the multiple members of the ensemble simulation under the corresponding forcing. Let $Z_{t,s,u}$ be the extreme temperature model output at grid box s in year t from ensemble u for $s = 1, \dots, m$, $t = 1, \dots, n$, and $u = 1, \dots, l$. A GEV distribution $f(\cdot; \mu_{t,s,u}, \sigma_{t,s,u}, \xi_{t,s,u})$ is assumed for $Z_{t,s,u}$ with parameters

$$\mu_{t,s,u} = \mu_{b(t),s}, \sigma_{t,s,u} = \sigma_s, \xi_{t,s,u} = \xi_s, \quad (3)$$

where $b(t) = \text{ceiling}(t/5)$, $b = 1, \dots, B = n/5$. That is, the grid boxes are assumed to be independent of each other, and at grid box s , the GEV distributions have the same scale and shape parameters, while the location parameters vary every 5 yr. The parameters can be easily estimated through a maximum likelihood method at each grid box s , with ensembles treated as replicates. The stepwise signals $\{\mu_{1,s}, \dots, \mu_{B,s}\}$ at each grid box s are then smoothed with a locally weighted scatterplot smoother to form a smoothed signal for the climate model under consideration.

The signal estimation procedure is similar to that in Zwiers et al. (2011) except for two aspects. First, the GEV location parameters change every 5 instead of 10 yr. The estimation of the signals is affected by the bias-variance trade-off (e.g., Sippel et al. 2015). The 5-yr-piecewise signal estimates would follow the data more closely but with higher variation; the 10-yr-piecewise signal estimates would have lower variation but may be more biased locally. The former have a better chance to capture short-lived signals in response to events such as volcanic eruptions, which could be smoothed out much more if a longer time window such as 10 yr is used. Second, the smoothed estimates obtained from locally weighted scatterplot smoothing on the raw signal estimates were used to form the final signal. This may further reduce the variation in estimating the signal and has the potential to increase the power in detection.

For each of the four extreme temperatures, the estimation process is repeated for each climate model under each external forcing. Once these signals are estimated they are then averaged with equal weight to all climate models to form pooled estimate of the multimodel signals for the subsequent detection and attribution analysis. As there are only a few climate models providing ANT simulations, to use as many simulations as possible, the ANT signal is obtained by subtracting the NAT signal from the ALL signal based on the additive assumption of the two signals (e.g., Meehl et al. 2004). Not all models provide daily temperatures under both the ALL and the NAT forcings for the time period we considered. In fact, only the CanESM2 model produces extended simulations for both ALL and NAT forcings with the resulting daily temperature data available at the time of analysis. As a result, the difference between multiple model ensemble means for ALL and NAT is also influenced by this modeling uncertainty in addition to that of ANT forcing. This caveat needs to be considered when interpreting our results.

c. Detection analysis

As in Zwiers et al. (2011), we estimate β by maximizing the independence log-likelihood across all grid boxes. Let $\zeta_s = (\alpha_s, \sigma_s, \xi_s)$. For each extreme temperature in a region, the log-likelihood at grid box s is

$$l_s(\beta, \zeta_s) = \sum_{t=1}^n \log f(\mathbf{Y}_{t,s}; \alpha_s + \mathbf{X}_{t,s}^T \beta, \sigma_s, \xi_s),$$

where $f(y, \mu, \sigma, \xi)$ is the density of the GEV distribution with location μ , scale σ , and shape ξ in (1). The independence log-likelihood function is then

$$L(\beta, \zeta_1, \dots, \zeta_m) = \sum_{s=1}^m l_s(\beta, \zeta_s). \quad (4)$$

This is a form of pseudo-log-likelihood that discards the spatiotemporal dependence in the data. It is only used for point estimation; the uncertainty assessment of the estimator will need to account for the spatiotemporal dependence (see section 3d).

There are $3m + p$ unknown parameters in (4): p in the scaling factor β and 3 in ζ_s for each s , $s \in \{1, \dots, m\}$. Jointly estimating these $3m + p$ unknown parameters is challenging. Zwiers et al. (2011) used a profile approach. For each value of β in a grid of possible values, estimation of ζ_s can be easily done separately at each grid box s by maximizing $l_s(\beta, \zeta_s)$ with respect to ζ_s . The objective function L is then profiled with respect to β on the grid of possible values, and the value that gives the highest evaluation of L is the estimate of β . This method

is computationally very intensive, especially when bootstrapping is necessary for inferences.

We use a coordinate descent approach, which updates a small set of parameters at a time with others being held fixed until convergence (e.g., Tseng 2001; Tseng and Yun 2009). In particular, this approach is a two-step iterative process: 1) given the current estimate $\hat{\beta}$ of β , the remaining m unknown sets ζ_s , $s = 1, \dots, m$, are completely separable and the estimation can be done in parallel [the estimate $\hat{\zeta}_s$ of ζ_s is obtained separately at each grid box s by maximizing $l_s(\hat{\beta}, \zeta_s)$ with respect to ζ_s], and 2) given the current estimate $\hat{\zeta}_s$, $s \in \{1, \dots, m\}$, estimate β by maximizing $l_s(\hat{\beta}, \zeta_s)$ with respect to β . These two steps may start from, for example, $\beta = 1$ and iterate until $\hat{\beta}$ converges.

Compared to the profile approach, the coordinate descent approach has advantages in computation efficiency and accuracy. The profile approach requires that a predetermined grid of possible values of β be wide enough to cover the optimal value. The range of the predetermined grid needs to be set from exploratory analysis, such as the smallest and largest values of the scaling factors estimated from individual grid boxes (Zwiers et al. 2011). The estimate can only take values on the grid points since the likelihood is only evaluated at such points, and the estimate is the one with the highest likelihood evaluation. The accuracy of the estimate depends on the fineness of the grid; a finer grid means increased computing burden. The coordinate descent algorithm does not need any preset grid, and the estimate ends up wherever it needs to be to maximize the likelihood. Using the northern Europe (NEU) region for illustration, with a single signal, for each of the four extreme temperatures, the profile approach took about 46–53 s to search β among 301 possible values on a laptop computer with an Intel Core 2.50-GHz CPU, while the coordinate descent approach took only 2–4 s. When $p = 2$ in the model in (2), this approach only took 4–5 s; in contrast, the profile approach would need a two-dimensional grid and the computing task would increase quadratically. This is the main reason why two-signal analysis was not practical for the method of Zwiers et al. (2011).

The efficient estimation method makes it possible to perform detection analyses with both ANT and NAT signals in the model. Such analyses examine whether the responses to different forcings can be separately detected in the presence of the other one. Including multiple signals simultaneously allows one to properly attribute observed changes to different external forcings. Specifically, in two-signal detection, $\mu_{t,s}$ in (2) becomes

$$\mu_{t,s} = \alpha_s + \mathbf{X}_{A,t,s}^T \beta_A + \mathbf{X}_{N,t,s}^T \beta_N, \quad (5)$$

where \mathbf{X}_A and \mathbf{X}_N are the relative signals of ANT and NAT forcings, respectively, and $\boldsymbol{\beta}_A$ and $\boldsymbol{\beta}_N$ are the corresponding scaling factors. Note that \mathbf{X}_A and \mathbf{X}_N are allowed to be correlated, as the ANT and NAT signals could be, as long as they are not perfectly dependent.

d. Uncertainty assessment

To assess the uncertainty of $\hat{\boldsymbol{\beta}}$, we follow the block bootstrap in Zwiers et al. (2011). Because the signals $\mathbf{X}_{t,s}$ are estimated quantities, uncertainties in their estimation, which also affect the uncertainty in $\hat{\boldsymbol{\beta}}$, need to be accounted for in the bootstrap procedure appropriately. The bootstrap process is a two-level, 32×32 block bootstrap that preserves both temporal and spatial dependence in both the simulated and observed data. At the first level, the ensembles from each climate model are bootstrapped to obtain 32 bootstrap samples of the signals. At the second level, for each of the 32 signals, the observed extremes are bootstrapped 32 times, each of which provides a bootstrap sample of $\boldsymbol{\beta}$. The first-level and the second-level bootstrap account for the variation from the signal estimation and the fingerprinting, respectively. The choice of 32×32 (instead of 1024×1) is a compromise between computing time and accuracy because estimating the signals at each grid box (first level) is much more time-consuming than fitting the fingerprinting model with given signal estimates (second level). This choice of 32 for the first level is big enough in practice since the second-level variation dominates the first-level variation: the fingerprinting model uses observed data with only one replicate while the signal estimation uses climate model simulation data with multiple replicates.

Our bootstrap procedure for the signal estimates is adapted from that in Zwiers et al. (2011) with the following steps:

- 1) For each ensemble, transform the simulation data into Gumbel residuals by the fitted GEV parameters in the model in (3) at each grid box (Kharin and Zwiers 2005), and divide the Gumbel residuals into B nonoverlapping 5-yr blocks.
- 2) With each 5-yr block of m grid boxes treated as an entity, randomly sample B 5-yr blocks of Gumbel residuals with replacement.
- 3) Transform the sampled Gumbel residuals back to the GEV scale by the fitted GEV parameters in the model in (3).
- 4) Estimate the signals from the transformed data.
- 5) Repeat steps two to four 32 times.

For each bootstrap sample of the signal estimates, the bootstrap sampling procedure of the observational data

accounting for the natural internal variability in the climate system is given below:

- (i) Subtract the scaled signal $\mathbf{X}_{t,s}^T \hat{\boldsymbol{\beta}}$ from $\mathbf{Y}_{t,s}$ to obtain the residuals, and divide the residuals into B nonoverlapping 5-yr blocks.
- (ii) With each 5-yr block of m grid boxes treated as an entity, randomly reorder the B 5-yr blocks of residuals.
- (iii) Add the scaled signal $\mathbf{X}_{t,s}^T \hat{\boldsymbol{\beta}}$ back to the reordered residuals, and denote it as.
- (iv) Estimate the scaling factor from $\tilde{\mathbf{Y}}_{t,s}$ with the given signal.
- (v) Repeat steps (ii) to (iv) 32 times for each of the 32 bootstrap samples of signal.

The two-level bootstrap procedure leads to 1024 bootstrap samples of the $\boldsymbol{\beta}$ that can be used for inferences. Both the effects of the signal uncertainty (level 1) and the natural internal variability (level 2) are considered, and the spatial–temporal dependence has been retained. The 1024 samples give a sampling distribution of $\boldsymbol{\beta}$. The 5% quantile and 95% quantile lead to an approximate 90% confidence interval for each component in $\boldsymbol{\beta}$. We claim that an external forcing is detected if its confidence interval is above zero.

e. Goodness-of-fit test

Goodness-of-fit test for the nonstationary GEV distribution in the model in (2) can be done for each region separately. At each region, there are three challenges in constructing a goodness-of-fit test: 1) the uncertainty in estimating the scaling factor $\boldsymbol{\beta}$ shared by all grid boxes, 2) the spatiotemporal dependence among the grid boxes over time, and 3) the assessment of the overall fit at the regional level as opposed to the fit at individual grid boxes. We propose a goodness-of-fit test procedure for each of the extreme indices at the regional level.

We first transform the observed extremes into Gumbel residuals by the fitted GEV models in (2) at each grid box. The Anderson–Darling (AD) statistic for testing the goodness of fit of the standard Gumbel distribution for the Gumbel residuals at each grid box was obtained as in Zwiers et al. (2011) except that they used the Kolmogorov–Smirnov (KS) test; we use the AD test because it is more sensitive to extremes than the KS test (Heo et al. 2013). A bootstrap procedure that accounts for the uncertainty in estimating the shared scale factor could be carried out to assess the p value of the AD statistic at each grid box, but this leads to multiple testing issues for each extreme temperature in the region. The approach of “field significance” uses the number of rejected individual tests as a testing statistic (Livezey and Chen 1983), but the sampling distribution

of the statistic depends on the spatial dependence across grid boxes, which is left unspecified and hard to simulate from. Further, it ignores the fact that some individual test has been rejected with confidence, and the discrete nature of the test statistic can potentially compromise the statistical power (Wilks 2006). In contrast, we construct a regional-level goodness-of-fit test statistic by taking the maximum of the AD statistics from the individual goodness-of-fit tests at all the grid boxes in the region. As all grid boxes have the same length of data period, the AD statistics on the Gumbel scale are comparable. The process is essentially equivalent to one that combines the p values of all p values from the individual goodness-of-fit tests at all grid boxes by taking their minimum, one of the classical ways to combine p values dating back to Tippett (1931). The significance of the regional-level goodness-of-fit test can be assessed from the same bootstrap procedure used to assess the significance of the individual goodness-of-fit tests—for each bootstrap replicate, we just need to take the maximum of the individual AD bootstrap realizations to form an approximate draw from the null distribution of the regional maximum of the AD statistics.

To preserve the spatial dependence in the observed data, we use the same block bootstrap procedure for the observational data in the uncertainty assessment in section 3d. Specific steps to generate one realization from the null sampling distribution of the test statistic are the following:

- 1) Follow steps (i)–(iii) in section 3d to form a bootstrap sample of the observed data $\{\tilde{\mathbf{Y}}_{t,s}\}$.
- 2) Fit the model in (2) to the bootstrap sample, and denote the point estimates of $\boldsymbol{\beta}$, α_s , σ_s and ξ_s by $\hat{\boldsymbol{\beta}}$, $\hat{\alpha}_s$, $\hat{\sigma}_s$ and $\hat{\xi}_s$, respectively.
- 3) Transform the bootstrap sample $\{\tilde{\mathbf{Y}}_{t,s}\}$ to Gumbel residuals by the estimated parameters.
- 4) Calculate the AD test statistic for the standard Gumbel distribution using the Gumbel residuals at each grid box, and take the maximum AD statistics over the whole region as a realization of the regional testing statistic.

The steps are repeated 1000 times, resulting in a bootstrap sample of size 1000 for the regional testing statistic at each region. The 90% quantile of 1000 bootstrap samples of test statistics at each region is used as the critical value for a size 10% goodness-of-fit test.

f. Attribution to external forcing

As in Zwiers et al. (2011), possible changes in the extreme temperatures attributable to external forcing are studied by estimating the waiting time of the 20-yr return value from the furthest past in the most recent

climate. The waiting time for an extreme of a given size is defined as the inverse of the probability of occurrence of an annual extreme at least as extreme as the one that is specified. The waiting time of 20-yr return value should remain as 20 yr in a stationary climate, and the difference from 20 yr reflects the influence of the external forcing in the changes in the extreme temperatures (Stone and Allen 2005). Changes in the waiting time or the odds of a particular extreme event (here an event that occurs once in 20 yr on average) provides useful information for the assessment of impacts of climate change on extremes.

Specifically, we calculate the waiting time in the 2006–10 climate for the 20-yr event in the 1951–55 climate based on the fitted GEV parameters of the model in (2). The 1951–55 climate 20-yr return value is the temperature that is exceeded once in every 20 yr on average based on the fitted GEV parameters in the model in (2) for 1951–55. The waiting time in the 2006–10 climate of the 20-yr event in the 1951–55 climate is the inverse of the probability of the 1951–55 event to be exceeded based on the fitted GEV parameters in the model in (2) for the period of 2006–10. In the analysis for each region, we first calculate the exceeding probability at each grid box by its own signal, scale, and shape parameters, then take the average from all grid boxes in the region, and use its inverse as the waiting time for this region. The average here is necessary because parameters of the GEV distributions can be quite different across grid boxes within the same region although the scaling factor is the same. If the extreme temperature has increased over time, then we expect the waiting time for the 20-yr return level in the 1951–55 climate to be longer than 20 yr for TNn and TXn and shorter than 20 yr for TNx and TXx in the 2006–10 climate.

4. Results

The estimates of the scaling factors and their 90% bootstrap confidence intervals for single-signal detection of ALL forcing and ANT forcing are summarized in Fig. 1. The ANT and ALL signals are detected in TNx in almost all regions except southern South America (SSA), southern Africa (SAF), and South Asia (SAS). The results are in good agreement with those reported by Min et al. (2013), who did not detect the ANT and ALL signals in TNx in region SAF and SAS either. Both ANT and ALL signals are detected in TNn in most regions, except northern Europe (NEU), SAF, and SAS, and ANT in eastern North America (ENA). For TXn and TXx, the influences of ANT and ALL forcings are detected in more than 10 regions. For the number of detected regions of each of the four extreme

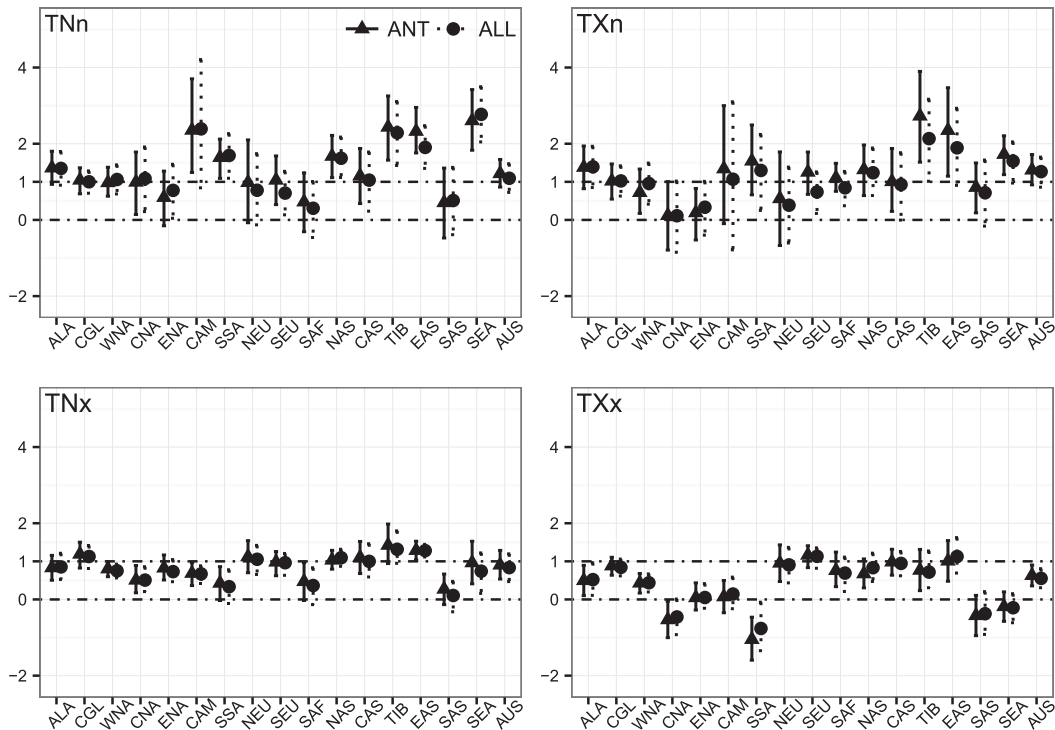


FIG. 1. Point estimates of the scaling factors—TNn, TXn, TNx, and TXx—and the corresponding 90% confidence intervals for single-signal detection of ANT forcing (\blacktriangle symbols) and ALL forcing (\bullet symbols), respectively, at each region.

temperatures, TNx is associated with most regions while TXx with least regions, which shows that, among the four extreme temperatures, the influence of external forcings on the annual maximum of daily minimum temperature is the most detectable while that on the annual maximum daily maximum temperature is the least detectable. This is consistent with the findings in Zwiers et al. (2011) and Min et al. (2013). Compared to the analyses using 10-yr-mean probability-based index (PI) in Min et al. (2013), we detect both ANT and ALL forcing in TXx in Australia (AUS) region, while they did not detect either. Kim et al. (2016) used 5-yr-mean PI in their analyses and our results are in large agreement with theirs.

The results for the simultaneous two-signal (ANT and NAT) detection for the 17 regions are summarized in Fig. 2. The ANT signal is detected in TNx in all 17 regions. Compared to the single-signal analyses, the ANT signal is detected in two more regions for TNn, one more for TXn, and one less for TXx. The results suggest that a more proper statistical model that includes all known external factors may improve the power of detecting the responses than the model that includes only some of the external factors. The influence of ANT forcing is found to be separable from that of NAT forcing in four to six

regions for each of the four extremes. In both north Asia (NAS) and Tibet (TIB), the ANT signal is separable from NAT signal in three temperature indices—TNn, TNx, and TXx. The detection results of the two-signal analyses of TNx are in large agreement with the two-signal detection results of TNx presented in Min et al. (2013): among our six detected regions and their five detected regions, four are in common—central North America (CNA), southern Europe (SEU), NAS, and TIB. We detect more regions/temperature indices than Min et al. (2013): they did not detect ANT forcing in TNn, TNx, and TXx in AUS, TNn and TNx in South America (consists of CAM and SSA), and TNn in Europe (consists of NEU and SEU), while we detect ANT forcing in these cases. Nor did Kim et al. (2016) detect ANT forcing in TNn in Europe.

The detection and attribution results of the single-signal and two-signal analyses as well as the total number of grid boxes for each type of extreme temperature at each region are summarized in Table 3. Most confidence intervals that lie above zero also include unity, indicating consistency between model-simulated responses and the observed changes. The confidence intervals of β in the analyses of TNn and TXn for some regions lie above unity, indicating that the climate

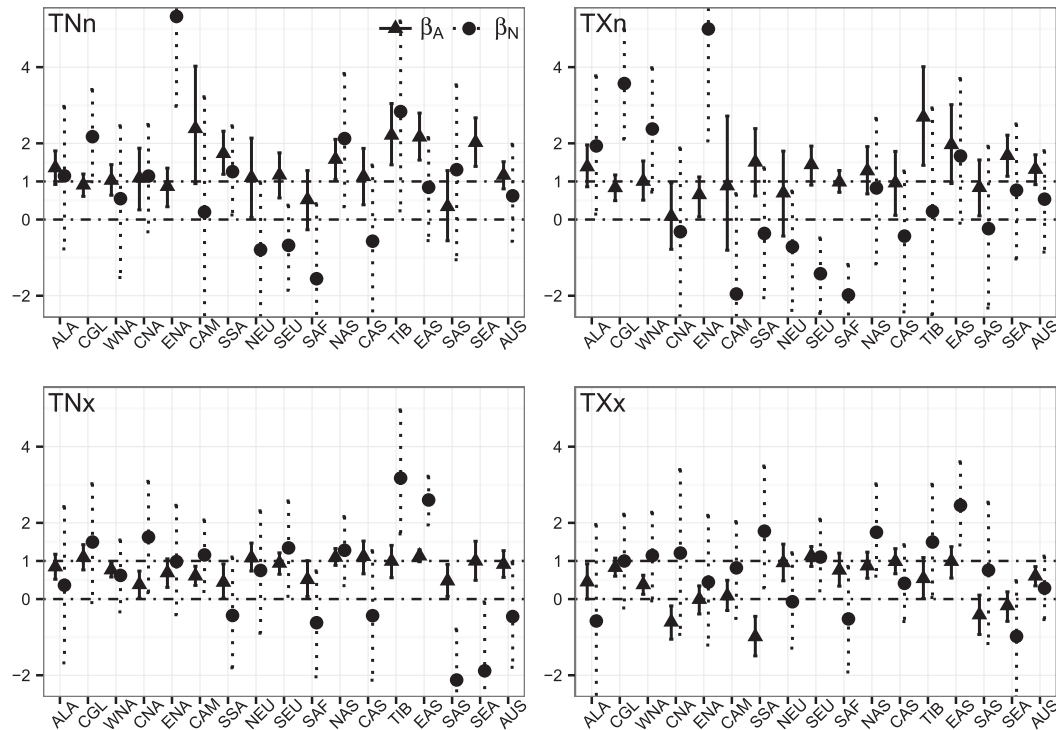


FIG. 2. Point estimates of the scaling factors and the corresponding 90% confidence intervals for two-signal (ANT and NAT) detection at each region. Parameters β_A and β_N are the scaling factors corresponding to ANT forcing and NAT forcing in Fig. 1, respectively. [Note: the 90% confidence interval of β_N for TNn in SEA region is (7.23, 14.90), which is too large to be plotted in the figure.]

models underestimate the observed changes in annual minima of daily maximum and daily minimum temperatures in these regions. The confidence intervals of β from the analyses of TNx and TXx lie between zero and unity for some regions, implying that the climate models overestimate changes in the annual maxima of daily minimum and maximum temperatures in these regions. The regional-level goodness-of-fit tests for single-signal analyses and two-signal analyses indicate that the GEV distribution with the location parameters specified in the model in (2) is appropriate in general with only four exceptions (Table 3). These findings echo those in Zwiers et al. (2011), Min et al. (2013), and Kim et al. (2016).

Figure 3 shows the waiting time in the 2006–10 climate for the 20-yr events in the 1951–55 climate due to external forcing. Most of the 90% confidence intervals of the waiting time for the 1951–55 20-yr events in the 2006–10 climate lie above 20 yr for TNn and TXn and below 20 yr for TNx and TXx, indicating a significant increase in the waiting time for cold extremes and a significant decrease in the waiting time for warm extremes due to external forcing. Among the 17 regions, the median point estimates of the waiting times for TNn,

TXn, TNx, and TXx are 107 yr in SSA, 45 yr in CAS, 4 yr in WNA, and 8 yr in AUS, respectively; the corresponding 90% confidence intervals are (68, 160), (20, 90), (3, 5), and (6, 12), respectively. These indicate that the changes in the extreme temperatures in terms of waiting time are much larger for the nighttime temperatures than those for daytime temperatures. These findings are also in good agreement with Zwiers et al. (2011). The estimated changes in waiting time from the 1951–60 climate to 1991–2000 reported in their study are noticeably smaller than reported here because of differences in the base periods and their lengths.

5. Conclusions

Our detection and attribution analyses for changes in four extreme temperature indices from the HadEX2 use signals estimated from CMIP5 models during 1951–2010 over 17 subcontinent regions. The extreme value theory is applied in the fingerprinting; a GEV distribution is fitted to the observed data with the location parameter proportional to the model-simulated responses to ANT or ALL forcing in single-signal analysis and to ANT and NAT forcings jointly in two-signal analysis. A

TABLE 3. Summaries of detection and attribution analysis results with single-signal ANT forcing and ALL forcing, respectively, and two-signal detection (ANT and NAT) at each region. Parameters β_A and β_N are the scaling factors corresponding to ANT forcing and NAT forcing in the two-signal detection, respectively. Letter “D” means the corresponding 90% confidence interval lies above zero but does not contain unity, and “A” means the corresponding 90% confidence interval lies above zero and contains unity; blanks mean the corresponding 90% confidence interval does not lie above zero. Entries in bold font mean that the corresponding goodness-of-fit tests were rejected at the 10% significance level.

| Region | Total grids | | | | TNn | | | | TXn | | | | TNx | | | | TXx | | | |
|--------|-------------|-----|-----|-----|-----|-----|-----------|-----------|-----|-----|-----------|-----------|-----|-----|-----------|-----------|-----|-----|-----------|-----------|
| | TNn | TXn | TNx | TXx | ALL | ANT | β_A | β_N | ALL | ANT | β_A | β_N | ALL | ANT | β_A | β_N | ALL | ANT | β_A | β_N |
| ALA | 73 | 73 | 73 | 73 | A | A | A | | A | A | A | A | A | A | A | | D | D | D | |
| CGL | 167 | 167 | 95 | 114 | A | A | A | A | A | A | A | D | A | A | A | | A | A | A | |
| WNA | 75 | 75 | 75 | 75 | A | A | A | | A | A | A | A | D | D | D | | D | D | D | |
| CNA | 43 | 43 | 43 | 43 | A | A | A | | | | | | D | D | D | A | | | | |
| ENA | 31 | 31 | 31 | 31 | A | | A | D | | | A | D | A | A | A | | | | | |
| CAM | 29 | 29 | 25 | 28 | A | D | A | | | | | | D | D | D | A | | | | |
| SSA | 43 | 43 | 26 | 32 | D | D | D | A | A | A | A | | | | D | | | | | A |
| NEU | 67 | 67 | 67 | 67 | | | A | | | | | | A | A | A | | A | A | A | |
| SEU | 64 | 64 | 46 | 51 | A | A | A | | A | A | A | | A | A | A | A | A | A | A | A |
| SAF | 28 | 27 | 18 | 20 | | | | | A | A | A | | | | A | | A | A | A | |
| NAS | 277 | 277 | 265 | 246 | D | D | D | A | A | A | A | | A | A | A | A | A | A | A | A |
| CAS | 85 | 85 | 62 | 71 | A | A | A | | | A | A | | A | A | A | | A | A | A | |
| TIB | 63 | 63 | 56 | 61 | D | D | D | A | D | D | D | | A | A | A | D | A | A | A | A |
| EAS | 94 | 95 | 89 | 91 | D | D | D | | A | D | A | | D | D | A | D | A | A | A | D |
| SAS | 48 | 48 | 24 | 33 | | | | | | A | A | | | | D | | | | | |
| SEA | 19 | 19 | 15 | 17 | D | D | D | D | A | D | D | | A | A | A | | | | | |
| AUS | 64 | 61 | 31 | 41 | A | A | A | | A | A | A | | A | A | A | | D | D | | |

coordinate descent algorithm is used in the estimation of the scaling factors from maximizing an independence likelihood. The method surpasses that in Zwiers et al. (2011) in both computational efficiency and estimation accuracy. In particular, the improved computational efficiency makes the Zwiers et al. (2011) method tractable for multiple-signal detection and attribution analysis, which is crucial for properly attributing observed changes in extremes to different external forcings. The results show that the ANT signals are clearly detectable in many subcontinent regions, with more frequent detection for TNx. These findings are in agreement with earlier studies (Zwiers et al. 2011; Min et al. 2013). As reported in Zwiers et al. (2011), influence of external forcing and anthropogenic forcing in particular has resulted in significantly longer waiting time for cold extremes and significantly shorter waiting time for warm extremes in most of the regions we analyzed.

The two-signal detection analyses based on the extreme value theory are a highlighted important contribution of this study. Other studies (Min et al. 2013; Kim et al. 2016) also performed two-signal detection analyses on these temperature indices, but their methods were based on region-average probability transferred data and the standard fingerprinting method. In contrast, we explicitly consider the distributional properties of extreme values as well as spatial and temporal evolution of extreme temperatures. Our analyses make it easier to

link anthropogenic influences to changes in the risks of extreme temperatures. As we used a very different method than theirs, our analysis is also complementary to these earlier studies. The fact that consistent conclusions are obtained based on different methods does enhance our confidence on detection results.

Detection and attribution analyses have been conducted on different aspects of extreme temperatures, including the magnitude of extremes such as annual maximum or minimum of daily temperatures (Christidis et al. 2011; Zwiers et al. 2011; Min et al. 2013; Kim et al. 2016) and frequency of moderate extreme events (Morak et al. 2011, 2013). The existing analyses and ours show very clear anthropogenic influence on extreme temperatures at global and regional scales. Anthropogenic influence on nighttime extreme temperatures is more readily detectable than that in daytime extreme temperatures. Additionally, it appears that climate models tend to underestimate the observed changes in cold extremes but overestimate the observed in warm extremes in certain regions. These systematic differences between model-simulated responses and observations are poorly understood and should be investigated in the future.

A few directions of methodological development are worth exploring. Wang et al. (2014) have recently taken into account the spatial dependence to increase the estimation efficiency with a combined-score equations

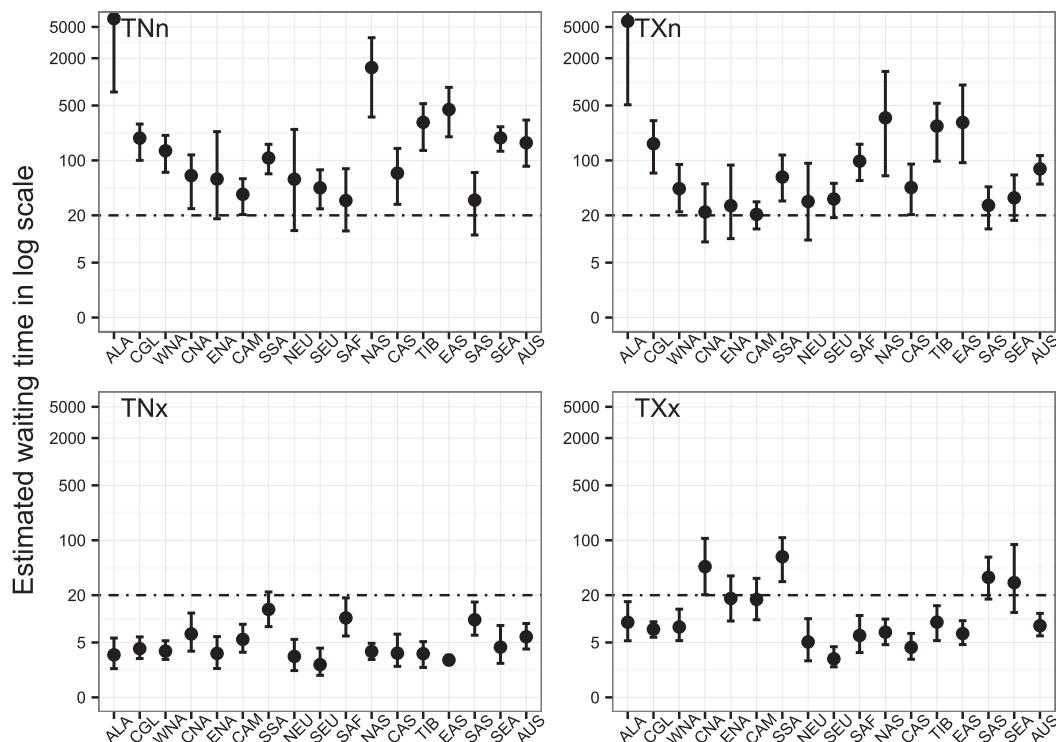


FIG. 3. Point estimates and the 90% confidence intervals of the 2006–10 waiting time of the 1951–55 climate 20-yr return value for TNn, TXn, TNx, and TXx in log scale for two-signal (ANT and NAT) detection at each region.

method, which only assumes marginal GEV distributions at each grid box, without specification of spatial dependence. The scores of the marginal GEV distribution at each grid box are combined through some working correlation structure to improve the estimation efficiency and, hence, the detection power. Developing a closer analog to standard optimal fingerprinting that accounts for spatial dependence merits further investigation along this direction. The framework we proposed here assumes that the scale (and shape) parameters of the GEV distribution at each grid box do not change over time. This would fail to capture the reported changes in variability in regions strongly affected by land–atmosphere feedback or changes in atmospheric circulation (Seneviratne et al. 2006; Screen 2014). It is possible in principle to include such changes in our framework by constructing a second or a third set of signals from the GEV scale or shape parameters for each external forcing and incorporate them as covariates in the scale or shape parameters of the observed data. It is unclear, however, if such a more complete model would lead to better insight of anthropogenic influence on temperature extremes as additional parameters into the model in (2) would unavoidably introduce larger uncertainty in the model fitting and changes in

temperature extremes in most regions are associated with a shift in the mean condition.

Acknowledgments. This research was partially supported by an NSF grant (DMS 1521730), a University of Connecticut Research Excellence Program grant, a Shenzhen University grant (17QNFC61), and Environment and Climate Change Canada.

REFERENCES

- Allen, M. R., and P. A. Stott, 2003: Estimating signal amplitudes in optimal fingerprinting, part I: Theory. *Climate Dyn.*, **21**, 477–491, doi:10.1007/s00382-003-0313-9.
- Brown, S. J., J. Caesar, and C. A. T. Ferro, 2008: Global changes in extreme daily temperature since 1950. *J. Geophys. Res.*, **113**, D05115, doi:10.1029/2006JD008091.
- Christidis, N., P. A. Stott, S. J. Brown, G. C. Hegerl, and J. Caesar, 2005: Detection of changes in temperature extremes during the second half of the 20th century. *Geophys. Res. Lett.*, **32**, L20716, doi:10.1029/2005GL023885.
- , —, and —, 2011: The role of human activity in the recent warming of extremely warm daytime temperatures. *J. Climate*, **24**, 1922–1930, doi:10.1175/2011JCLI4150.1.
- Dey, D. K., D. Roy, and J. Yan, 2015: Univariate extreme value analysis. *Extreme Value Modeling and Risk Analysis: Methods and Applications*, D. K. Dey and J. Yan, Eds., CRC Press, 1–22.

- 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. Atmos.*, **118**, 2098–2118, doi:10.1002/jgrd.50150.
- Giorgi, F., and R. Francisco, 2000: Uncertainties in regional climate change prediction: A regional analysis of ensemble simulations with the HADCM2 coupled AOGCM. *Climate Dyn.*, **16**, 169–182, doi:10.1007/PL00013733.
- Hasselmann, K., 1997: Multi-pattern fingerprint method for detection and attribution of climate change. *Climate Dyn.*, **13**, 601–611, doi:10.1007/s003820050185.
- Hegerl, G. C., and F. Zwiers, 2011: Use of models in detection and attribution of climate change. *Wiley Interdiscip. Rev.: Climate Change*, **2**, 570–591, doi:10.1002/wcc.121.
- , and Coauthors, 2007: Understanding and attributing climate change. *Climate Change 2007: The Physical Science Basis*, S. Solomon et al., Eds., Cambridge University Press, 663–745.
- , O. Hoegh-Guldberg, G. Casassa, M. P. Hoerling, R. S. Kovats, C. Parmesan, D. W. Pierce, and P. A. Stott, 2010: Good practice guidance paper on detection and attribution related to anthropogenic climate change. *IPCC Expert Meeting on Detection and Attribution of Anthropogenic Climate Change*, Geneva, Switzerland, WMO, 1–8.
- Heo, J.-H., H. Shin, W. Nam, J. Om, and C. Jeong, 2013: Approximation of modified Anderson–Darling test statistics for extreme value distributions with unknown shape parameter. *J. Hydrol.*, **499**, 41–49, doi:10.1016/j.jhydrol.2013.06.008.
- IPCC, 2012: Summary for policymakers. *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*, C. B. Field et al., Eds., Cambridge University Press, 1–19.
- Katz, R. W., and B. G. Brown, 1992: Extreme events in a changing climate: Variability is more important than averages. *Climatic Change*, **21**, 289–302, doi:10.1007/BF00139728.
- Kharin, V. V., and F. W. Zwiers, 2000: Changes in the extremes in an ensemble of transient climate simulations with a coupled atmosphere–ocean GCM. *J. Climate*, **13**, 3760–3788, doi:10.1175/1520-0442(2000)013<3760:CITEIA>2.0.CO;2.
- , and —, 2005: Estimating extremes in transient climate change simulations. *J. Climate*, **18**, 1156–1173, doi:10.1175/JCLI3320.1.
- , —, and X. Zhang, 2005: Intercomparison of near-surface temperature and precipitation extremes in AMIP-2 simulations, reanalyses, and observations. *J. Climate*, **18**, 5201–5223, doi:10.1175/JCLI3597.1.
- Kim, Y.-H., S.-K. Min, X. Zhang, F. Zwiers, L. V. Alexander, M. G. Donat, and Y.-S. Tung, 2016: Attribution of extreme temperature changes during 1951–2010. *Climate Dyn.*, **46**, 1769–1782, doi:10.1007/s00382-015-2674-2.
- Livezey, R. E., and W. Y. Chen, 1983: Statistical field significance and its determination by Monte Carlo techniques. *Mon. Wea. Rev.*, **111**, 46–59, doi:10.1175/1520-0493(1983)111<0046:SFSRID>2.0.CO;2.
- Meehl, G. A., W. M. Washington, C. M. Ammann, J. M. Arblaster, T. M. L. Wigley, and C. Tebaldi, 2004: Combinations of natural and anthropogenic forcings in twentieth-century climate. *J. Climate*, **17**, 3721–3727, doi:10.1175/1520-0442(2004)017<3721:CONAAF>2.0.CO;2.
- Min, S.-K., X. Zhang, F. W. Zwiers, H. Shiogama, Y.-S. Tung, and M. Wehner, 2013: Multimodel detection and attribution of extreme temperature changes. *J. Climate*, **26**, 7430–7451, doi:10.1175/JCLI-D-12-00551.1.
- Morak, S., G. C. Hegerl, and J. Kenyon, 2011: Detectable regional changes in the number of warm nights. *Geophys. Res. Lett.*, **38**, L17703, doi:10.1029/2011GL048531.
- , —, and N. Christidis, 2013: Detectable changes in the frequency of temperature extremes. *J. Climate*, **26**, 1561–1574, doi:10.1175/JCLI-D-11-00678.1.
- Pachauri, R. K., and Coauthors, 2014: *Climate Change 2014: Synthesis Report*. Cambridge University Press, 151 pp.
- Ribes, A., S. Planton, and L. Terray, 2013: Application of regularised optimal fingerprinting to attribution. Part I: Method, properties and idealised analysis. *Climate Dyn.*, **41**, 2817–2836, doi:10.1007/s00382-013-1735-7.
- Screen, J. A., 2014: Arctic amplification decreases temperature variance in northern mid- to high-latitudes. *Nat. Climate Change*, **4**, 577–582, doi:10.1038/nclimate2268.
- Seneviratne, S. I., D. Lüthi, M. Litschi, and C. Schär, 2006: Land–atmosphere coupling and climate change in Europe. *Nature*, **443**, 205–209, doi:10.1038/nature05095.
- Shiogama, H., N. Christidis, J. Caesar, T. Yokohata, T. Nozawa, and S. Emori, 2006: Detection of greenhouse gas and aerosol influences in changes in temperature extremes. *SOLA*, **2**, 152–155, doi:10.2151/sola.2006-039.
- Sippel, S., D. Mitchell, M. T. Black, A. J. Dittus, L. Harrington, N. Schaller, and F. E. L. Otto, 2015: Combining large model ensembles with extreme value statistics to improve attribution statements of rare events. *Wea. Climate Extremes*, **9**, 25–35, doi:10.1016/j.wace.2015.06.004.
- Stone, D. A., and M. R. Allen, 2005: The end-to-end attribution problem: From emissions to impacts. *Climatic Change*, **71**, 303–318, doi:10.1007/s10584-005-6778-2.
- 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.
- Tippett, L. H. C., 1931: *The Methods of Statistics*. Williams and Norgate, 222 pp.
- Tseng, P., 2001: Convergence of a block coordinate descent method for nondifferentiable minimization. *J. Optim. Theory Appl.*, **109**, 475–494, doi:10.1023/A:1017501703105.
- , and S. Yun, 2009: A coordinate gradient descent method for nonsmooth separable minimization. *Math. Program.*, **117**, 387–423, doi:10.1007/s10107-007-0170-0.
- Wang, Z., J. Yan, and X. Zhang, 2014: Toward optimal fingerprinting in detection of changes in climate extremes with combined score equations. University of Connecticut Dept. of Statistics Tech. Rep. 10, 41 pp.
- Wilks, D. S., 2006: On “field significance” and the false discovery rate. *J. Appl. Meteor. Climatol.*, **45**, 1181–1189, doi:10.1175/JAM2404.1.
- Zwiers, F. W., X. Zhang, and Y. Feng, 2011: Anthropogenic influence on long return period daily temperature extremes at regional scales. *J. Climate*, **24**, 881–892, doi:10.1175/2010JCLI3908.1.
- , G. C. Hegerl, X. Zhang, and Q. Wen, 2014: Quantifying the human and natural contributions to observed climate change. *Statistics in Action*, J. F. Lawless, Ed., Chapman and Hall/CRC Press, 349–370.