Toward Optimal Fingerprinting in Detection and Attribution of Changes in Climate Extremes

Zhuo Wang, Yujing Jiang, Hui Wan, Jun Yan & Xuebin Zhang

View supplementary material

Published online: 30 Mar 2020.

Submit your article to this journal

Article views: 718

View related articles

View Crossmark data

Taylor & Francis
Taylor & Francis Group

Toward Optimal Fingerprinting in Detection and Attribution of Changes in Climate Extremes

Zhuo Wang[a], Yujing Jiang[b], Hui Wan[c], Jun Yan[d], and Xuebin Zhang[c]

[a]College of Economics, Shenzhen University, Shenzhen, Guangdong, China; [b]Department of Statistics, Colorado State University, Fort Collins, CO; [c]Climate Data Analysis Section, Environment and Climate Change Canada, Toronto, ON, Canada; [d]Department of Statistics, University of Connecticut, Storrs, CT

ABSTRACT

Detection and attribution of climate change plays a central role in establishing the causal relationship between the observed changes in the climate and their possible causes. Optimal fingerprinting has been widely used as a standard method for detection and attribution analysis for mean climate conditions, but there has been no satisfactory analog for climate extremes. Here, we turn an intuitive concept, which incorporates the expected climate responses to external forcings into the location parameters of the marginal generalized extreme value (GEV) distributions of the observed extremes, to a practical and better-understood method. Marginal approaches based on a weighted sum of marginal GEV score equations are promising for no need to specify the dependence structure. The computational efficiency makes them feasible in handling multiple forcings simultaneously. The method under working independence is recommended because it produces robust results where there are errors-in-variables. Our analyses show human influences on temperature extremes at the subcontinental scale. Compared with previous studies, we detected human influences in a slightly smaller number of regions. This is possibly due to the under-coverage of the confidence intervals in existing works, suggesting the need for careful examinations of the properties of the statistical methods in practice. Supplementary materials for this article, including a standardized description of the materials available for reproducing the work, are available as an online supplement.

1. Introduction

The United Nations Framework Convention on Climate Change reached the landmark Paris Agreement in December 2015 that, for the first time, commits nearly every country to lowering the emission of greenhouse gases to reduce the impact of human-induced climate change. Of fundamental importance to support such a global policy is the scientific underpinning of a causal relationship between the human emission of greenhouse gas and the warming of planet Earth, in which detection and attribution of climate changes has played a central role. In climate science, detection refers to the process of determining whether there is a discernible change in some aspect of the climate over time; attribution refers to the process of attributing causes of the detected change (e.g., Hegerl et al. 2010). The classical problem is to determine whether some combination of external forcings to the climate system have caused a detected change in climate and, subsequently, quantify the size of the influence. Examples of external forcings are increase in greenhouse gas concentrations caused by the use of fossil fuels and eruption of volcanoes, which affect atmospheric optical properties and thereby the amount of solar radiation entering the atmosphere. Typically, physical arguments are made to identify the most important external forcings that could plausibly be responsible

for the observed change. Then, physics-based numerical climate models are used to estimate the expected climate responses to the forcings, termed signals, or fingerprints; and statistical inference is used to determine whether those fingerprints are present in observations and if so, whether they have the expected amplitude.

For mean climate conditions, optimal fingerprinting (e.g., Hasselmann 1997; Allen and Stott 2003) has been used as a standard method to quantify influences from external forcings. This method regresses observed measures onto the expected responses to, or signals of, specific forcings of interest to determine whether the signals are present in the observations and, if so, with what amplitude. The regression errors contain information about the natural internal (chaotic) variability of the climate system. Decisions about the presence or absence of the signals in the observations, termed fingerprinting, are based on the statistical inferences about their coefficients, termed scaling factors, which adjust the amplitudes of the signals to best match the observations. Estimation of the scaling factors faces two complications: both the optimal weight matrix and the signals are not observed but estimated. The optimal weight matrix is the inverse of the covariance matrix of the regression errors, which optimizes the efficiency of the inferences; that is, it min-

imizes the uncertainty of the estimator of the scaling factors as measured in mean squared error. Estimation of the covariance matrix is challenging because the number of climate model simulations is often too small relative to the matrix dimension to produce consistent estimators. This has traditionally been addressed by restricting the number of orthogonal components used in the analysis but recent works have also emphasized alternative regularization approaches (e.g., Ribes, Planton, and Terray 2013). Estimation of the signals brings an errors-in-variables (EIV) issue, which is addressed by total least squares (Allen and Stott 2003) and likelihood-based EIV regression (e.g., Huntingford et al. 2006; Hannart, Ribes, and Naveau 2014). Optimal fingerprinting has been used in many studies that have detected human influence on different aspects of the climate system. The resulting literature has been assessed in successive reports of the Intergovernmental Panel on Climate Change (IPCC) (e.g., Mitchell and Karoly 2001; Hegerl et al. 2007; Bindoff et al. 2013). Those assessments have led to the IPCC statement: "It is extremely likely [probability greater than 0.95] that human influence has been the dominant cause of the observed warming since the mid-20th century" (IPCC 2013b, p. 17).

Much of the impact of climate is caused by extremes. For example, extremely hot weather can cause devastating impacts on human health even in the United States and Europe (Karl and Knight 1997; Schär and Jendritzky 2004). Understanding the causes of observed changes is important for confidently projecting future changes in climate extremes and thereby for developing sound adaptation strategies. An analog of the optimal fingerprinting method has, however, not been well established for extremes. Several approaches have been proposed, most of which avoid directly addressing the probabilistic properties of extremes by applying the standard fingerprinting method to, for example, transformed extreme values (Min et al. 2011), fitted parameters of extreme value distributions (Christidis, Stott, and Brown 2011), or averages of extremes over large regions (Wen et al. 2013; Zhang et al. 2013). These studies did not exploit the unique properties of extremes and cannot assess changes in rare, high impact events.

Zwiers, Zhang, and Feng (2011) conducted the first detection and attribution analysis of changes in climate extremes based on extreme value analysis. Annual temperature extremes at each grid box in a study region are assumed to follow a generalized extreme value (GEV) distribution. Over the study period, the scale and shape parameters of the GEV distribution at each grid box are assumed to remain unchanged, while the location parameters incorporate the signals of external forcings as covariates. These signals are assumed to be in the form of time-evolving location parameters and are estimated based on climate model simulations. Similarly to standard fingerprinting, the coefficients of the signals or the scaling factors are shared by all grid boxes, see Equation (1), which adjusts the amplitude of the expected time-evolution to best fit the observed extremes. This framework points to an exciting possibility for an analog of the standard optimal fingerprinting for extreme values. The method of Zwiers, Zhang, and Feng (2011) has limitations. The use of a profile independence likelihood (IL) method is computationally too expensive to be feasible when multiple signals are involved, which are often required in detection and attribution analyses. More importantly, there is little understanding on the statistical

properties of the framework, making it difficult to interpret the results based on the method. In particular, the two challenges shared with the standard fingerprinting, the efficiency issue and the EIV issue, and their implications in practical applications are not addressed or discussed.

It is tempting to model the spatial dependence under the framework of Zwiers, Zhang, and Feng (2011). Spatial dependence can be introduced to extremes by a latent Gaussian process, conditional on which the observed data follow independent GEV distributions (Casson and Coles 1999; Cooley, Nychka, and Naveau 2007; Sang and Gelfand 2009; Fuentes, Henry, and Reich 2013). The resulting marginal distributions are, however, no longer GEV distributions. Alternatively, spatial extremes can be directly modeled through max-stable processes; see Davison, Padoan, and Ribatet (2012) and Ribatet, Dombry, and Oesting (2016) for recent reviews. Max-stable processes are typically used when the aim is the prediction of joint/conditional events or realistic simulations. Since the joint density is intractable for dimension greater than two or three, inferences are mostly based on pairwise likelihood (PL) (Smith and Stephenson 2009; Padoan, Ribatet, and Sisson 2010; Davison and Gholamrezaee 2012). Such a fully parametric model for spatial dependence, when correctly specified, may lead to fairly efficient estimation of the regression coefficients in marginal GEV models, but misspecification of bivariate distributions may lead to inconsistent estimators for the marginal parameters of primary interest. In reality, it is likely that a max-stable process will fail goodness-of-fit tests when the study region is of practical size (Kojadinovic, Shang, and Yan 2015). Additionally, the computational cost of the PL method increases drastically as the number of sites increases.

The goal of this article is to advance toward an optimal fingerprinting in the extreme value modeling framework. As in standard fingerprinting, efficiency and EIV are two major methodological concerns. We propose three methods to approach these issues. The first is a combined score equation (CSE) method based on weighted sum of marginal GEV scores where, for improved efficiency, the weights are estimated to account for spatial dependence. The idea is similar to the generalized estimating equations (GEE) approach for clustered data (Liang and Zeger 1986). The second method is a variant of the CSE method with an identity weight matrix, that we call CSE-I. It does not use spatial dependence in point estimation but, like the CSE method, uses bootstrap to adjust for spatial dependence in constructing confidence intervals. CSE-I is conceptually equivalent to the profile IL method of Zwiers, Zhang, and Feng (2011) but computationally much faster. The third method is to fully specify the joint distribution of spatial extremes with a max-stable process and estimate the parameters by maximizing the PL. This method may lead to bias in estimation if the PL is misspecified; further, it is computationally much more expensive than the marginal methods. Numerical studies where the signals are observed suggest that the CSE estimator is the most preferred in terms of mean squared error. When the signals are estimated, however, the CSE estimator is more sensitive to the EIV issue resulting from the signal estimation than the CSE-I estimator due to the spatial dependence in EIV. The bias of the CSE estimator makes the coverage rate of the confidence intervals unreliable.

The methods were applied to regional detection and attribution analyses of four temperature extremes including annual maximum of the daily maxima (TXx), annual maximum of the daily minima (TNx), annual minimum of the daily maxima (TXn), and annual minimum of the daily minima (TNn). Important practical issues were investigated in numerical studies mimicking real detection and attribution analyses as well as in perfect model detection using simulated extremes from the second-generation Canadian Earth System Model (CanESM2) (Fyfe et al. 2017). The bootstrap confidence intervals were found to have coverage rates lower than the nominal levels even for the less biased CSE-I estimator, which has important implications on the conclusions based on confidence intervals in existing studies. The two-signal perfect model detection study over 13 subcontinental regions showed no evidence against the additivity between the anthropogenic (ANT) and natural (NAT) signals in GEV location parameter, at least for the CanESM2 model, which is reassuring for existing studies relying on the assumption (e.g., Wang et al. 2017). Analyses with the CSE-I method reported human influences on temperature extremes in a slightly smaller number of regions than previous studies, which may indicate under-coverage of the confidence intervals in earlier works.

2. Fingerprinting With Climate Extremes

Following Zwiers, Zhang, and Feng (2011), we need to specify GEV distributions for both the observed extremes and the climate model extremes under external forcings.

2.1. Observation Extremes

Suppose that observed extremes are available at m grid boxes over n years. Let $F(\cdot \mid \mu, \sigma, \xi)$ be a GEV distribution function with location μ, scale σ, and shape ξ. The observed annual extreme Y_{ts} at grid box s in year t, $s = 1, \ldots, m$, $t = 1, \ldots, n$, is characterized by

$$Y_{ts} \sim F(\cdot \mid \mu_{ts}, \sigma_s, \xi_s), \qquad \mu_{ts} = \alpha_s + X_{ts}^\top \beta, \qquad (1)$$

where $X_{ts} = (X_{ts1}, \ldots, X_{tsp})^\top$ is a $p \times 1$ vector of the signals (fingerprints) of certain external forcings at grid box s in year t obtained from climate model simulations under the forcings (see details below), β is the vector of scaling factors that adjusts the amplitude of the signals to best match the observed data, and α_s, σ_s, and ξ_s are grid-box-specific location, scale, and shape parameters, respectively. Although covariates could be incorporated into the scale and shape parameters (e.g., Risser and Wehner 2017), we do not pursue this direction here. Equation (1) only specifies marginal GEV distributions; no spatial/temporal dependence specification is needed for this application. The GEV distribution for modeling annual climate extremes has been widely accepted in climate research (e.g., Kharin, Zwiers, and Zhang 2005).

The constant scaling factor β in (1) across all grid boxes is an important assumption in the standard fingerprinting framework (Allen and Stott 2003; Zwiers, Zhang, and Feng 2011; Zwiers et al. 2014). This is based on the assumption that climate models properly reproduced the patterns of climate responses to external forcing even though the magnitude of the responses may differ because of differences in climate sensitivity.

This assumption is reasonable in that while the magnitude of response is uncertain, the spatial-temporal pattern of the response is constrained by physical understanding. For example, the increase in the greenhouse gases must result in an increase in temperature, but northern high-latitude should warm more than lower latitude because of the reduction of snow and ice as a result of warming. The land should warm more than water body because of difference in heat capacity. Of course, this assumption would fail where physical processes are not properly represented in a region, such as snow albedo feedback or soil moisture feedback. Projections of future change mostly bear this out. Projected temperature change patterns under different transient forcing scenarios are remarkably similar, in both means and "extremes" (IPCC 2013a) to the extent that most differences can be accounted for by simple pattern scaling. The task of detection and attribution analysis is to examine whether the modeled patterns (fingerprints) are present in the observed data and whether their amplitudes are consistent with the observed data, through inferences about the constant scaling factors that produce the best match between the fingerprints and the observations.

Among the $3m + p$ parameters in Equation (1), detection and attribution analyses focus on inference about the scaling factors β shared by all grid boxes, which is to be distinguished from the scale parameters σ_s's in terminology. Based on physical judgement in detection and attribution analyses, it is assumed that no important external forcing factors are left out. If the coefficient in β corresponding to an external forcing is significantly greater than zero, then that forcing's fingerprint is said to be "detected" in the observed data. If, in addition, the hypothesis that this coefficient equals one is not rejected, then this is statistical evidence supporting that the changes in the observed data can be "attributed" to the corresponding external forcing; the actual attribution requires, in addition, careful physical reasoning to ensure that the responsible physical processes are understood and elimination of the possibility that forcing agents not considered could have been responsible (Mitchell and Karoly 2001). If the scaling factor has a confidence interval in between zero and one, this may be an indication of over estimation in the climate model response. If the scaling factor's confidence interval is above one, this may be an indication of under estimation in the climate model response.

Although not specified in Equation (1), spatial and temporal dependence affects the efficiency in inferences about β. Annual extreme temperatures at one year have some degree of dependency over the space, while at any single location, the temporal dependency is relatively small. A working assumption is that, conditional on the signals, the annual extremes are spatially dependent in one year and temporally independent at one location (e.g., Padoan, Ribatet, and Sisson 2010; Blanchet and Davison 2011; Davison and Gholamrezaee 2012; Davison, Padoan, and Ribatet 2012). The justification of temporal independence is that daily measures only exhibit short-range dependence and that the conditions for independence among well-separated extremes in time are likely to be satisfied (Leadbetter, Lindgren, and Rootzén 1983, sec. 3.2). Empirically, known low-frequency phenomena such as the El Niño/Southern Oscillation (ENSO) may affect the distribution of climate extremes and introduce some temporal and spatial dependence (Zhang et al. 2010). Such dependence at multi-year scale, usually not too strong, is

not an issue for the CSE and CSE-I methods proposed in the next section because the block bootstrap approach of Zwiers, Zhang, and Feng (2011) have taken both spatial and temporal dependence into consideration.

2.2. Climate Model Extremes

In detection and attribution analyses, each component of the signal vector X_{ts} in Equation (1) is unknown and needs to be estimated. The signals at each s are the fingerprints of the external forcings of interest, which are assumed to be in the form of a time-evolving GEV location parameter as in Zwiers, Zhang, and Feng (2011). We assume that the runs in the ensemble under each external forcing come from a single climate model or multi-models that are interchangeable. Accounting for uncertainty from heterogeneity climate models (Huntingford et al. 2006) is an interest of future work.

Consider estimating X_{tsi}, $i = 1, \ldots, p$, the signal of the ith external forcing. Suppose that an ensemble of size l_i is available from a climate model under the ith forcing. The ensemble members were obtained under the same forcing scenario but from different initial conditions, essentially forming a random sample of size l_i of the same stochastic process of climate dynamics. Therefore, at grid box s, this leads to l_i independent time series of annual extremes, that is, one time series for each ensemble member.

Let $Z_{tsi}^{(r)}$ be the annual extreme in ensemble member r of the climate model simulations under the ith external forcing at grid box s in year t, $s = 1, \ldots, m$, $t = 1, \ldots, n$, and $r = 1, \ldots, l_i$. A GEV distribution $F(\cdot \mid X_{tsi}, \sigma'_{si}, \xi'_{si})$ is assumed for $Z_{tsi}^{(r)}$, with scale σ'_{si}, shape ξ'_{si}, and location parameter X_{ts} varying smoothly over time:

$$Z_{tsi}^{(r)} \sim F(\cdot \mid X_{tsi}, \sigma'_{si}, \xi'_{si}), \quad X_{tsi} = \mu(t; \gamma_{si}), \quad r = 1, \ldots, l_i, \tag{2}$$

where $\mu(t; \gamma_{si})$ is a function of t with parameters γ_{si}. Note that X_{tsi} is the fingerprint of forcing i in Equation (1). A simple yet flexible characterization for μ is

$$\mu(t; \gamma_{si}) = \sum_{k=1}^{K} \gamma_{si,k} B_k(t),$$

where $B_k(t)$'s, $k = 1, \ldots, K$, are a set of B-splines basis with K degrees of freedom, and $\gamma_{si} = (\gamma_{si,1}, \ldots, \gamma_{si,K})$ is the coefficient vector of the basis. Higher degrees of freedom allows more flexibility. The parameters $(\sigma'_{si}, \xi'_{si}, \gamma_{si})$ can be estimated separately with ease at each grid box s under working temporal independence, with the l_i series of simulated annual extremes treated as independent replicates. Let $\hat{\gamma}_{si}$ be an estimate of γ_{si}, $s = 1, \ldots, m$. The signal X_{tsi} under forcing i is then estimated by $\hat{X}_{tsi} = \mu(t; \hat{\gamma}_{si})$. Using $\hat{X}_{tsi}$ in place of X_{tsi} in Equation (1) is the source of the EIV issue.

3. Estimating the Scaling Factor

We present the estimation methods for the case where X_{ts}'s are observed and then discuss practical issues including EIV. In addition to scaling factor β, there are grid-box-specific parameters $\zeta_s = (\alpha_s, \sigma_s, \xi_s)^\top$, $s = 1, \ldots, m$.

3.1. CSE Method

The CSE method combines marginal GEV score equations over all s with certain weight matrix. Let $f(\cdot; \mu_{ts}, \sigma_s, \xi_s)$ be the density function of the GEV distribution with location μ_{ts}, scale σ_s, and shape ξ_s. For each $t \in \{1, \ldots, n\}$, $s \in \{1, \ldots, m\}$, let $S_{ts} = d \log f(Y_{ts}; \mu_{ts}, \sigma_s, \xi_s)/d\mu_{ts}$. Correct marginal specification implies $E(S_{ts}) = 0$. Under working temporal independence, the score equation for β with data at site s is

$$\sum_{t=1}^{n} X_{ts}^\top S_{ts} = 0, \quad s = 1, \ldots, m. \tag{3}$$

In a slightly more general setting, a known link function could be used to connect μ_{ts} to the linear predictor; we use the identity link here. The CSE is

$$\sum_{t=1}^{n} X_t W_t^{-1} S_t = 0, \tag{4}$$

where $X_t^\top = (X_{t1}, \ldots, X_{tm})$, W_t^{-1} is a $m \times m$ weight matrix, and $S_t = (S_{t1}, \ldots, S_{tm})^\top$. When W_t is the identity matrix, the CSE is score equation of the independence likelihood under working independence. A more efficient estimator is possible with an appropriate choice of W_t that accounts for spatial dependence to some extent.

The optimal W_t has the form $W_t = \Omega_t \Delta_t^{-1}$ (Nikoloulopoulos, Joe, and Chaganty 2011) where $\Omega_t = \text{cov}(S_t)$ and $\Delta_t = -\text{diag}(\Delta_{t1}, \ldots, \Delta_{tm})$, with $\Delta_{ts} = E\{d^2 \log f(Y_{ts}; \mu_{ts}, \sigma_s, \xi_s)/d\mu_{ts}^2\}$, $s = 1, \ldots, m$. Under spatial independence, W_t reduces to the identity matrix as $\Omega = \Delta$, and the estimating equation (4) reduces to the score equations of the IL method. Since Δ_t is known for the marginal GEV distributions (Prescott and Walden 1980), we just need to approximate the covariance matrix Ω_t of the score functions S_t.

With years treated as clusters, we propose to use a simple working spatial correlation structure for S_t. Since the univariate marginal GEV distributions are assumed to be correctly specified, we have $\text{var}(S_{ts}) = \Delta_{ts}$ as long as $\xi_s > -0.5$ (Smith 1985; Bücher and Segers 2017). As in a GEE setting, suppose all clusters share the same correlation matrix, R, of the score functions. Then Ω_t can be written as $\Omega_t = \Delta_t^{1/2} R \Delta_t^{1/2}$. We need to find a reasonably flexible correlation structure to approximate the spatial dependence among the location-wise scores within a year. As the spatial dependence between the observations from two sites decays as their distance increases in general, we suggest using simple one-parameter working spatial correlation functions such as exponential or spherical. In four scenarios reported in the supplementary materials, Section 1, the exponential working correlation fits well the true correlation functions of the scores, which were approximated by Monte Carlo. In practice, the score functions can be evaluated at a parameter estimate under working independence and their empirical correlation coefficients can be used to fit the working correlation structure through, for instance, nonlinear least squares.

An estimate of the parameters in Equation (1), $\Theta = (\beta^\top, \zeta_1^\top, \ldots, \zeta_m^\top)$, can be obtained with a block coordinate descent algorithm cycling through all its components (e.g., Tseng 2001; Tseng and Yun 2009). The CSE (4) is viewed as the gradient of some objective function with respect to β that

does not have a closed-form. We solve for the root of the CSE by alternating between two steps until convergence:

1. Given an estimate $\hat{\beta}_n$ of β, obtain likelihood estimates $\hat{\zeta}_{s,n}$ of ζ_s separately at each grid box $s \in \{1, \ldots, m\}$.
2. Given estimates $\hat{\zeta}_{s,n}$ of ζ_s, $s \in \{1, \ldots, m\}$, obtain the CSE estimate $\hat{\beta}_n$ of β by solving Equation (4) with an appropriately chosen working correlation structure.

The initial value of $\hat{\beta}_n$ can be obtained by estimating a β at each s separately and taking the average of the estimated β's. The working correlation parameters do not need to be updated if they are estimated based on a good initial value $\hat{\beta}_n$ of β_n. To ensure $\xi_s > -0.5$ so that the marginal Fisher information is finite at each s, we impose a penalty that restricts the range of each ξ_s to be within $(-0.5, 0.5)$ using a Beta$(9, 6)$ prior distribution. The idea, suggested by Martins and Stedinger (2000), has been used in many applications with GEV modeling (e.g., Cooley and Sain 2010). The estimating equations in each step can be solved with the R package nleqslv (Hasselman 2018).

Inference about β can be based on a block bootstrap that preserves spatial and temporal dependence (Zwiers, Zhang, and Feng 2011). Let $\hat{\Theta}$ be the CSE estimate of Θ. Define the residual series $\hat{\epsilon}_{ts} = Y_{ts} - \hat{\alpha}_s - \hat{\beta}X_{ts}$ at each grid box $s \in \{1, \ldots, m\}$ for $t \in \{1, \ldots, n\}$. Divide the residual series into 5-year non-overlapping blocks and randomly reorder the sequence of 5-year blocks over time. The choice of 5-year blocks is such that the temporal dependence introduced by low frequency phenomena such as ENSO is retained approximately. Let $\tilde{\epsilon}_{ts}$ be the reordered residual series. A bootstrap sample of the data is obtained by $\hat{\alpha}_s + \hat{\beta}X_{ts} + \tilde{\epsilon}_{ts}$ at each s. The CSE method is then applied to this sample to estimate β. Confidence intervals for β can be constructed using a bootstrap sample of $\hat{\beta}$.

As a general method, CSE can be applied for potential efficiency gains in settings where the marginal parameters are of primary interest, with no distributional assumption beyond univariate marginal distributions. Another example in spatial extremes modeling is regional frequency analysis. Covariates can also be incorporated into the GEV scale and shape parameters (e.g., Wang, Yan, and Zhang 2014), in which case, the CSE method can be applied to one set of parameters at a time given others in an iterative estimation procedure.

3.2. CSE-I Method

When W_t is the identity matrix (i.e., under working independence), the CSE method reduces to the CSE-I method. It is equivalent to the IL method of Zwiers, Zhang, and Feng (2011), but has a clear advantage in computational efficiency and accuracy in inferences about β. The profile method used in Zwiers, Zhang, and Feng (2011) requires that a predetermined search grid covers the solution; the finer the grid, the closer approximation to the solution. In our single-signal analyses of East Asia in Section 6, the profile method on a grid of 301 points took about 35–39 sec for each of the four temperature extremes, while the CSE-I method took only 1–3 sec. When multiple signals are of interest, which is often the case in practice, the CSE-I method is even more advantageous combined with bootstrapping. In our two-signal analyses, the CSE-I method only took 1–4 sec

for parameter estimation, which makes the bootstrap much more manageable. The computing time of the profile method is expected to increase about 300 times if a grid of 301 points is selected for the scaling factor of the second signal. Another advantage of the CSE-I method, as will be seen later, is its robustness to EIV, in which case, the lower variance of the CSE estimator is completely offset by its more severe bias than the CSE-I estimator

3.3. PL Method

The fingerprinting formulation (1) does not specify any multi-dimensional distribution beyond the marginal GEV specifications. One may assume a max-stable process model for spatial dependence (e.g., Davison, Padoan, and Ribatet 2012) and estimate Θ along with additional dependence parameters using the PL method. Nonetheless, using a max-stable process for the dependence structure needs much stronger justification as the convergence of the dependence structure of annual extremes to that of a limiting multivariate extreme value distribution may occur at a very different rate than that of the marginal distributions (Ledford and Tawn 1996, 1997; Resnick 2002). Further, high-order dependence structures are difficult to diagnose and test for goodness of fit (Kojadinovic, Shang, and Yan 2015). Our numerical studies (not reported) revealed that the PL estimator of β is not necessarily more efficient than the CSE estimator even when the dependence structure of the PL method is correctly specified. This may seem surprising at first, but the PL estimator is not, after all, the true maximum likelihood estimator. Its advantage in estimation under correct specification occurs for the shape parameter, not for location parameters such as β.

3.4. Practical Issues

We focus on the CSE method and the CSE-I method in the sequel because, compared to the PL method, they spare specification of spatial dependence in the extremes which is difficult to test. In addition, the computation of the PL method is much more intensive than the two marginal methods.

3.4.1. EIV From Estimated Signals

When X_{ts} is replaced with $\hat{X}_{ts}$ in Equation (1), the uncertainty in $\hat{X}_{ts}$ has an effect of EIV. When X_{ts}'s are known, both CSE and CSE-I estimators are unbiased, and the CSE estimator gives a notable reduction in the variance compared to the CSE-I estimator, as reported in the supplementary materials, Section 3. In presence of EIV, however, both estimators are biased but of different extent, which complicates the comparison. The EIV issue in this setting has not been studied in the literature. There are several unique challenges. The EIV are spatially and temporally dependent. The EIV affect the estimating equations for both β and ζ_s's. Sometimes, the errors in one signal might be correlated with the errors in another signal. Specifically, in order to maximize the effectiveness of computing power in climate model simulation, it is desirable to design simulation experiments focusing on the ALL (combination of anthropogenic and natural) forcing and some individual forcings with weaker signals. For example, Ribes, Gillett, and Zwiers (2015) recom-

mended to simulate responses to the ALL and NAT forcings, and derive the ANT influence from these two sets of simulations. If the ANT signal is inferred as the difference between the ALL and NAT responses, the errors in ANT could be correlated with errors in NAT, which is the case in Section 6.

The bias of the CSE and CSE-I estimators in presence of EIV comes from the fact that the marginal GEV score equations based on $\hat{X}_{ts}$'s in place of X_{ts} are no longer unbiased. For the CSE, an additional layer of bias comes from the spatial dependence of the EIV, which contaminates the Type I condition (Lai and Small 2007) of the marginal GEV score functions. The Type I condition requires the expectation of the marginal estimating function at each s conditional on $\hat{X}_{ts}$ to be the same as that conditional on $(\hat{X}_{t1}, \ldots, \hat{X}_{tm})$. When it does not hold, CSE which uses a nonidentity weight matrix is more biased than CSE-I which uses the identity weight matrix. The larger bias of the CSE method in comparison to the CSE-I method, as shown in Section 4, could lead to unacceptable coverage rates of confidence intervals.

Correcting the bias caused by the EIV issue is challenging and merits an independent study. Functional modeling techniques are preferred to structural modeling techniques in this application because of no requirement on the distribution of the true signals (e.g., Huang 2014). None of the existing correction approaches, however, can be applied to amend the estimating Equation (4) due to its special structure. Expanding the estimating function with a second-order Taylor approximation did not give satisfactory results in our investigation. For lack of a satisfying correction, we may still estimate the scaling factors discarding the EIV issue, but then natural questions arise such as: whether the 50-run ensembles used in the data analyses in Section 6 are sufficiently large to give ignorable bias; whether one method is more robust than the other in response to the EIV issue. These issues are investigated in a numerical study in Section 4.

3.4.2. Bootstrap Confidence Interval

Confidence intervals for the scaling factors need to be constructed carefully to account for the spatial dependence and the EIV issue. We propose to use the two-level bootstrap procedure of Zwiers, Zhang, and Feng (2011) to construct confidence intervals. The first level accounts for the uncertainty in signal estimation. A block bootstrap procedure with non-overlapping 5-year blocks is applied to the runs in the ensemble under each forcing to form a bootstrap sample of the signal estimates. At the second level, for each set of the bootstrapped signal estimates, a block bootstrap procedure is applied to the observed data to form a bootstrap sample of scaling factor estimates. For instance, if we get 32 bootstrap replicates of signal estimates and then for each of them get 32 bootstrap replicates of scaling factor estimates, then we will obtain a bootstrap sample of size 1024 for the scaling factors, which can then be used to construct confidence intervals.

Little attention has been paid to the coverage property of the confidence intervals. For the standard optimal fingerprinting with linear regression that has been used for over 20 years, it is only recently reported that the confidence intervals widely used in practice have smaller coverage rate than their nominal

intervals (DelSole et al. 2019). Neither Wang et al. (2017) nor Zwiers, Zhang, and Feng (2011) studied the coverage rate of the two-level bootstrap confidence intervals in the extreme value setting. If the bias resulting from EIV is severe, the intervals may have lower coverage rate, leading to over statement of detection and incorrect statement of attribution. This issue is investigated in our numerical study in Section 4.

3.4.3. Additive ANT and NAT Signals

Similar to standard fingerprinting, Equation (1) depends on the assumption that signals are additive. That is, climate response to the combined effect of different forcings is statistically indistinguishable from the sum of the responses to individual forcings. This seems to be the case at least for mean temperature at large scale (Marvel et al. 2015). With 50-run ensembles under the ALL and NAT forcing from the CanESM2 model, it is possible to study whether the additivity holds in the GEV location parameters of extreme temperatures in a two-signal perfect model detection study, which has not been done before. Specifically, one run under the ALL forcing is used as the observed data; the remaining runs under the ALL forcing and all the runs under the NAT forcing are used as climate model simulations. If the additivity of signals holds, the coefficients of the ANT signal and NAT signals will both be surrounding one, which can be used as a diagnosis for the additivity assumption. This issue is investigated in Section 5.

Since the scaling factor of the ANT signal is of primary interest, one may include directly the ALL and NAT signals as covariates, and then appeal to the additivity of the two signals to make inferences about the scaling factors of the ANT and NAT signals. Alternatively, one may estimate the ANT signal as the difference between the estimated ALL and NAT signals, and then use the estimated ANT and NAT signals in fingerprinting. Both approaches have advantages. For the latter, the two signals are much closer to being orthogonal, and thus the design matrix would be better conditioned. On the other hand, the ANT signal estimate that is obtained by subtracting the NAT signal from the ALL signal has contributions to uncertainty from both the ALL and NAT signal estimates. Moreover, this also produces dependence between the EIV in the ANT signal estimate and that in the NAT signal estimate. We used the latter approach in the numerical studies and regional analyses in the next sections.

4. Numerical Study

A numerical study was designed to further investigate the performance of the CSE and CSE-I methods in terms of their point estimates and confidence intervals. The PL method was not included because of its need to specify the dependence structure without obvious gain and its high computational cost. The study mimics a real detection and attribution analysis in East Asia (EAS) over the period of 1951–2010. The EAS region is bounded by latitude 20N–50N and longitude 100E–145E as defined in Giorgi and Francisco (2000). We chose to mimic the analyses of TXx and TNn as TNn is known to have heavier tails than TXx. The grid boxes are of resolution $5° \times 5°$ (latitude × longitude). To match the analyses in Section 6, the TXx data had 44 grid boxes and the TNn data had 48 grid boxes. The difference in the

Table 1. Summaries of the estimated scaling factors of ANT and NAT signals from 400 replicates in the numerical study.

Type	Ensemble size	Dependence	External forcing	CSE-I				CSE			
				Mean	SD	RMSE	CP	Mean	SD	RMSE	CP
TNn	25	Moderate	ANT	0.88	0.20	0.24	85	0.65	0.14	0.38	27
			NAT	0.88	0.35	0.37	90	0.65	0.23	0.42	65
		Strong	ANT	0.89	0.25	0.28	86	0.66	0.14	0.36	32
			NAT	0.87	0.51	0.52	89	0.66	0.21	0.40	62
	50	Moderate	ANT	0.96	0.21	0.21	89	0.78	0.16	0.27	67
			NAT	0.96	0.43	0.43	90	0.81	0.28	0.33	84
		Strong	ANT	0.91	0.27	0.29	86	0.79	0.15	0.26	69
			NAT	0.94	0.54	0.54	90	0.79	0.26	0.33	84
TXx	25	Moderate	ANT	0.95	0.16	0.17	86	0.78	0.12	0.25	50
			NAT	0.95	0.36	0.36	89	0.80	0.22	0.30	83
		Strong	ANT	0.94	0.20	0.21	89	0.76	0.12	0.27	46
			NAT	0.94	0.47	0.47	90	0.78	0.23	0.32	80
	50	Moderate	ANT	0.97	0.16	0.16	88	0.88	0.12	0.17	80
			NAT	0.99	0.37	0.37	92	0.90	0.23	0.26	92
		Strong	ANT	0.96	0.19	0.19	90	0.87	0.12	0.18	81
			NAT	0.93	0.52	0.52	90	0.89	0.24	0.26	92

NOTE: CP is the coverage percentage of 95% confidence intervals.

number of grid boxes was due to different missing patterns in the observed data. In each replicate for each outcome (TXx or TNn), we generated 50-run ensembles under the ALL and NAT forcings, respectively, from Equation (2); we then generated one additional run under the ALL forcing and used it as the observed data. This setup assumes that the ALL signal is the sum of the ANT and NAT signals. When performing fingerprinting analysis, we include two signals, ANT and NAT, where the ANT signal was the difference between the ALL and NAT signals. Our setting ensures that the true scaling factors are both one.

Specifically, each run under each forcing was generated under temporal independence. For a given forcing ALL or NAT, annual extremes were generated from a spatial model whose marginal GEV distributions were specified in Equation (2). At each site s, the marginal GEV distribution was set to be the fitted distributions based on the 50-run ensembles under the corresponding forcing from CanESM2. The location parameters were characterized by a quadratic spline with knots at every 5 years. The spatial dependence was set to be a mixture of a geometric Gaussian max-stable (GGMS) process and a Gaussian copula (GC) model with mixing rate of half-half, with the region treated as a planar surface for simulating GGMS processes; see the supplementary materials for some background on the GGMS process. Two dependence levels were considered, strong and moderate. For each dependence level, the parameters of the GGMS model were determined first and then the parameter of the GC model was tuned such that the empirical correlation of the scores from a large sample matches that from the GGMS model closely. Following the rationale of Shang, Yan, and Zhang (2015), we fixed one parameter of the GGMS model such that a close-to-full range of dependence for pairwise extremal coefficients (Schlather and Tawn 2003) is offered with upper bound 1.96. The correlation function of the underlying Gaussian process of the GGMS model was exponential $\rho(h) = \exp(-h/\phi)$, where h is the distance between two locations and the range parameter ϕ controls the dependence. The strong and moderate dependence setting had $\phi = 4200$ and 2100 miles, respectively; the estimated ϕ by fitting the extremal coefficient function of GGMS model to the

observed TXx in EAS through nonlinear least squares was 3150 miles, which is midway between the two settings for ϕ. The matching GC model had an exponential correlation function with range $\tau = 1120$ and 630 miles corresponding to the strong and moderate dependence level, respectively.

Table 1 summarizes the mean, standard deviation (SD), and root mean squared error (RMSE) of the point estimates of the two scaling factors, as well as the coverage percentage (CP) of the 95% confidence intervals based on 400 replicates. Both estimators show downward bias due to the EIV issue from the signal estimation. The bias is larger in the analyses of TNn than in the analyses of TXx, as TNn has heavier tails and larger variability than TXx. More runs in the ensembles used in signal estimation helps to reduce the bias. Stronger dependence led to higher variation in the point estimation of both scaling factors due to less information from the data. As expected, the estimated scaling factor of the ANT signal has much smaller variation than that of the NAT signal possibility due to weaker signal to noise ratio for NAT.

In all cases the CSE estimates appear to be more severely biased than the CSE-I estimates, although they have notably reduced standard deviations, especially for NAT. In contrast to the scenarios where there are no EIV (see the supplementary materials), CSE is no longer a clear winner in terms of RMSE. The CSE estimates still have smaller RMSE than the CSE-I estimates for the NAT coefficient, but the CSE-I estimates have an edge over or are comparable with the CSE estimates for the ANT coefficient. The empirical coverage percentage of the 95% confidence intervals from the CSE method are far off from the nominal level due to the severe bias; even with 50-run ensembles, it is only 67% for ANT in the analyses of TNn under moderate dependence. The CSE-I method, on the other hand, yielded much more reliable coverage percentages of 86% or higher when the ensemble size is 50. Given that the scaling factor of the ANT signal is of major concern in fingerprinting, and that the decisions are based on confidence intervals, the CSE-I method is preferred. The results of the coverage percentage also suggest that, in an "ad hoc" way, the 95% confidence interval may be viewed as calibrated 90% confidence intervals by increasing the critical values in construction (Loh 1987).

5. Perfect Model Detection

5.1. Single-Signal Perfect Model Detection

Perfect model detection refers to the setting where each run of an ensemble of climate model simulations is used as the observed data, while the remaining runs are used to estimate the fingerprints. This detection setting has a flavor of leave-one-out cross-validation. It is "perfect" in the sense that the observed data and the climate model simulations carry the same fingerprints and natural variability. Such studies serve to establish the detectability of the signals under a perfect situation, where detection is not hampered by possible mismatch in natural variability or forcing response between observations and climate model simulations, and to understand the power of detection of different methods. Perfect model detection has been done for extreme temperature and precipitation by adapting the standard fingerprinting framework (Hegerl et al. 2004; Min et al. 2009), but no works exist within the extreme value analysis setting.

From the 50-run ensembles of the CanESM2 model under the ALL forcing, we extracted the four annual temperature extremes over the period of 1951–2010 over 5° × 5° (latitude × longitude) grid boxes. For a given region, a perfect model detection for each of the four temperature extremes led to a set of 50 analyses. If the GEV distributions fit well and the estimation is unbiased, the estimated scaling factors will be clustered around 1. The perfect model detection was performed on all four temperature extremes for the 13 subcontinental regions analyzed in Section 6. Known as Giorgi regions (Giorgi and Francisco 2000), these regions are widely used in regional climate analyses. The grid boxes for each temperature extreme were selected to match those determined by the available observed data analyzed in Section 6 in order shed light on the real data analyses. The 13 regions all had 20 or more grid boxes for all four temperature extremes. Table 2 summarizes the boundaries and number of grid boxes for the 13 regions.

Figure 1 shows the boxplots of the 50 estimated scaling factors of the ALL signal using the CSE and the CSE-I method. The centers of the estimates from both methods appear to be lower than one, which is expected since the EIV issue was not accounted for. The downward patterns are much more severe in the CSE estimates than in the CSE-I estimates, especially for the two minima temperatures which have higher variation than the two maxima. On the other hand, the CSE estimates have clearly a smaller variation than the CSE-I estimates in most of the results. For both estimates, the variations are higher in the analyses of the two minima than in the analyses of the two maxima, which is

Table 2. Summaries of ranges and numbers of grid boxes of the 13 regions analyzed in the study.

Acronym	Region name	Latitude (°N)	Longitude (°E)	Number of grid boxes			
				TNn	TXn	TNx	TXx
ALA	Alaska	60/70	−165/−105	37	37	37	37
CGL	Canada and Greenland	50/80	−100/−10	87	87	60	68
WNA	Western North America	30/60	−130/−105	34	34	34	34
CNA	Central North America	30/50	−100/−85	20	20	20	20
ENA	Eastern North America	25/50	−85/−55	23	23	22	23
NEU	Northern Europe	50/70	−10/40	40	40	40	40
SEU	Southern Europe	30/45	−10/40	37	37	29	33
NAS	North Asia	50/70	40/175	123	123	121	112
CAS	Central Asia	30/50	40/75	40	40	36	37
TIB	Tibet	30/50	75/100	30	30	30	30
EAS	East Asia	20/50	100/145	48	48	44	44
SAS	South Asia	5/30	65/100	29	29	20	27
AUS	Australia	−45/−15	115/150	38	38	21	26

NOTE: The Greenland (GRL) in Giorgi and Francisco (2000) is renamed as CGL to be consistent with that in Zwiers, Zhang, and Feng (2011).
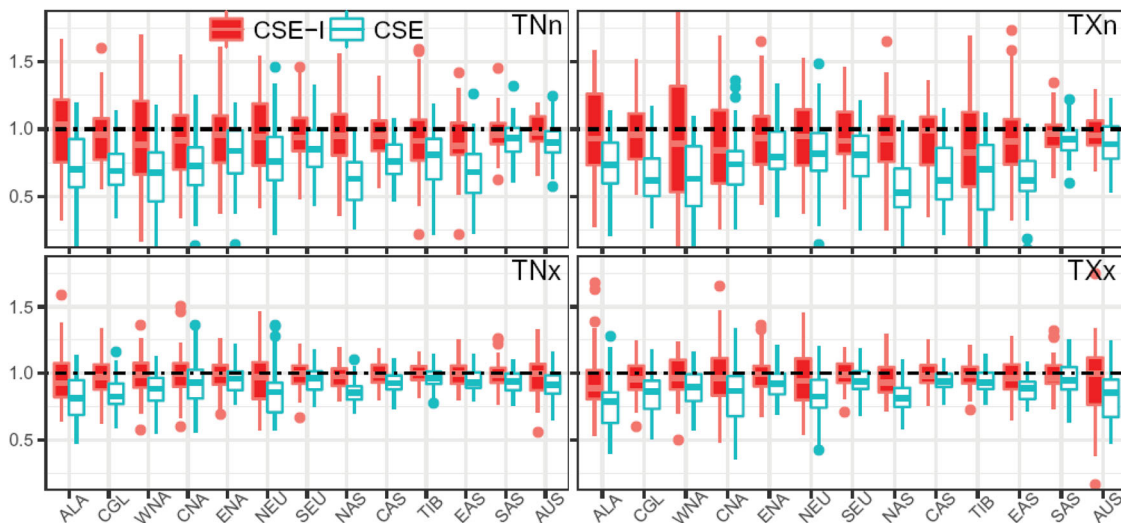

Figure 1. Boxplots of estimated scaling factors from perfect model detection with the 50-run ensembles under the ALL forcing from climate model CanESM2.
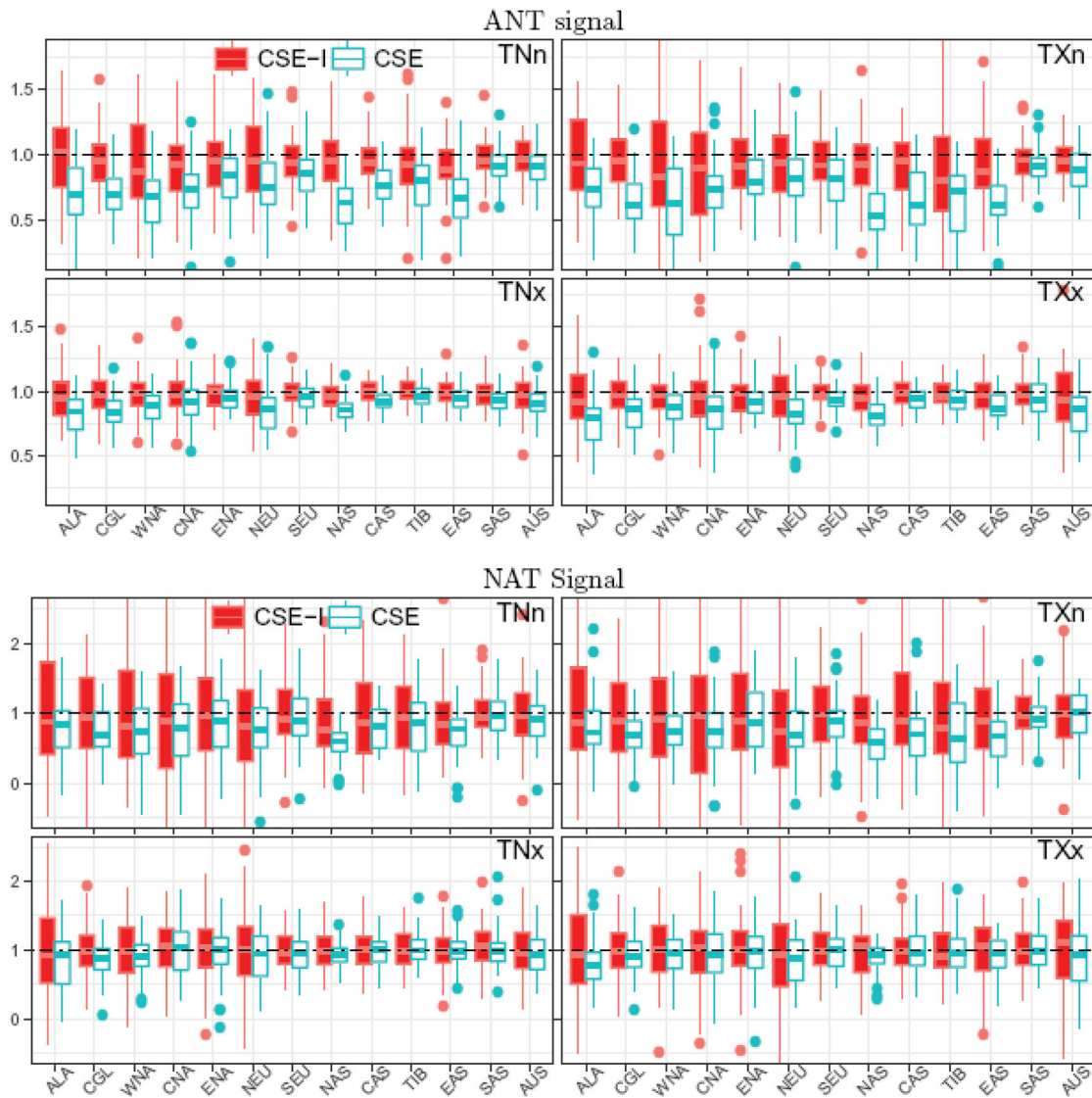
Figure 2. Boxplots of estimated scaling factors from perfect model detection with two signals, ANT and NAT.

expected as the minima have heavier tails and larger variability than the maxima such that the uncertainty in signal estimation is higher for the minima than for the maxima. These results echo those in Section 4.

5.2. Two-Signal Perfect Model Detection

Perfect model detection can be extended to have multiple signals where one run from the combined forcings is used as the observed data while the ensembles of individual forcings are used to estimate the signals. The CanESM2 model provides a 50-run ensemble under the NAT forcing in addition to the 50-run ensemble under the ALL forcing. Each run under the ALL forcing was used as the observed data; the other 49-runs under the ALL forcing and the 50-runs under the NAT forcing were used to estimate the signals under the corresponding forcing. If the ANT and NAT signals are additive, and if the ANT signal estimate is taken as the difference between the ALL signal estimate and the NAT signal estimate, then we expect that the estimated scaling factors of both the ANT and NAT forcings from a reasonably good method will be surrounding 1.

The boxplots of the 50 estimates of the scaling factors of the ANT and the NAT signals, respectively, from the two-signal perfect model detection using CSE-I method and the CSE method are presented in Figure 2. Similar observations to those from Figure 1 can be made on the estimated scaling factors for both signals. The downward pattern toward zero and the reduction in variation of the CSE estimates are most obvious in the estimates of the scaling factor for the NAT signal in the analyses of the two minima. When conclusions are drawn based on confidence intervals, the coverage of confidence intervals from the CSE methods may be far off, as seen in Section 4, which may lead to over statement of the detection results. Focusing on the CSE-I estimates, all the coefficients appear to be centered at 1, although with some noticeable deviation for the two minima in some regions (e.g., EAS). The results suggest no obvious evidence against the additive assumption of the two signals with the 50-run ensembles from CanESM2.

6. Regional Detection and Attribution

Detection and attribution analyses were performed on four temperature extremes of 13 Giorge regions summarized in Table 2.
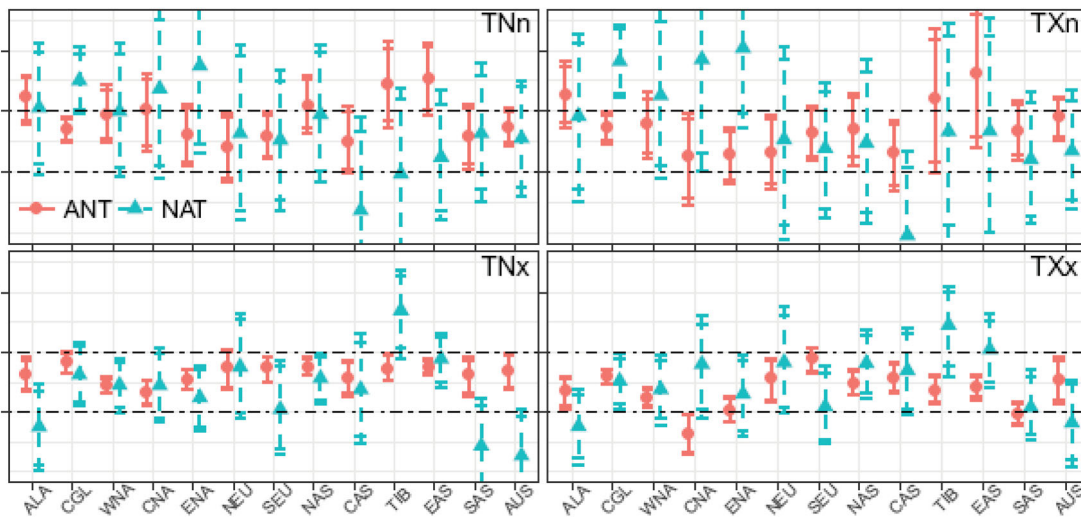
Figure 3. Summaries of estimated scaling factors of the ANT and NAT signals with 90% and 95% confidence intervals for four temperature extremes in 13 regional analyses using the CSE-I method.

The observed annual temperature extremes over the 60-year period of 1951–2010 were obtained from the HadEX2 data (Donat et al. 2013) for grid boxes of resolution 5° × 5°. In each region, grid boxes with more than 10 years of missing data in the 60-year period were discarded, resulting in 13 regions which had 20 or more grid boxes for all four temperature extremes. To avoid accounting for the differences among multiple climate models in signal estimation, we used a single climate model, the CanESM2 model, with 50-run ensembles under the ALL and NAT forcings; most other climate models have much smaller ensemble sizes. Annual extremes were extracted from the CanESM2 outputs on grid boxes matching the availability of the observed data. The signals of primary interest, ANT, were estimated as the difference between the ALL signal and the NAT signal, assuming that the ANT and NAT fingerprints are additive when forming the ALL signal in the GEV location parameter. The signals under the ALL or the NAT forcing at each grid box were estimated by fitting (2) to the 50 runs with quadratic splines for μ, where the knots were placed evenly every 5 years. Knot selection through certain information criterion, although intriguing, is beyond the scope of this article.

Figure 3 presents the estimated scaling factors of the ANT signal and the NAT signal for each of the four extremes in each of the 13 regions, using the CSE-I method. The choice of the CSE-I method is based on its robustness in presence of the EIV issue and its reliability in confidence intervals, as suggested from the numerical study and the perfect model detection study. Both 95% and 90% confidence intervals are plotted. The confidence intervals are much wider in the analyses of TXn and TNn than in the analyses of TXx and TNx. The differences in the lengths of the 95% and 90% confidence intervals are more obvious in the analyses of TXn and TNn than in the analyses of TXx and TNx. Since the 95% confidence intervals in the numerical study with ensemble size 50 only have 86–92% coverage empirically (Table 1), we recommend using the 95% confidence intervals to meet the need of 90% confidence intervals, which are frequently used in detection and attribution studies. This enlargement of the confidence intervals may change the qualitative conclusions in certain analyses. For example, for the scaling factor of the

ANT signal, the enlarged confidence intervals now cover zero in the analyses of TNn in Central Asia (CAS) and TXn in Tibet (TIB), which reverts the "detection" statements in these cases.

The detection results of the ANT signal are interpreted after accounting for the NAT forcing. The ANT signal is detected (with a confidence interval above zero), in TNn in most of the 13 regions except NEU and CAS; it is detected in TXn in a majority of the regions with the exception of CNA, ENA, NEU, CAS, and TIB. The confidence intervals in cases where the ANT signals is detected in TNn or TNx generally include one as a plausible scaling factor, with the exception of CGL, suggesting that the magnitudes of CanESM2-simulated changes in TNn and TXn are comparable to those observed. The influence of the ANT forcing is detected in TNx in all 13 regions; it is detected in TXx in almost all 13 regions except CNA, ENA, and SAS. The confidence intervals for scaling factors in regions where the ANT signal is detected in TNx or TXx are tighter than those in TNn and TXn, and below one, which suggests that the CanESM2 model responses to ANT in TNn and TXn are larger than observed in these regions. For the NAT signal, with a few exceptions, it is in general not detected due to the wide confidence intervals. Overall, the results are similar to but with fewer detected cases than those reported in the two-signal analyses of Wang et al. (2017). In addition to different climate models and grid box resolutions, their narrower than desired confidence intervals are an explanation too.

7. Discussion

Our main contribution is to advance the intuitive concept of fingerprinting in the framework of extreme value analyses as described in Zwiers, Zhang, and Feng (2011) to a more practical and better-understood method. The proposed marginal approaches bypass the specification of full dependence structure of the extremes, which, although affecting the estimation efficiency, is a nuisance in fingerprinting analyses relative to the primary goal of estimating the scaling factors. The computational efficiency of our method makes it feasible in more

general detection and attribution settings with multiple signals. The statistical properties of the method are examined including the efficiency of the scaling factor estimators with and without presence of the EIV issue, and the coverage rate of the bootstrap confidence intervals. The CSE method performs well when there is no EIV, that is, when the signals are known. When there are EIV, however, the advantage of CSE is compromised by its more severe bias relative to CSE-I, which in turn undermines the coverage of the confidence intervals. As a result, we recommend the use of CSE-I method in practical applications, especially when the climate model simulations are limited for the estimation of signal.

Our investigation on practical issues such as the coverage property of bootstrap confidence intervals and additivity between the ANT and NAT signals are of important practical value to the detection and attribution community. It is not until recently that the under-coverage of the confidence intervals in standard optimal fingerprinting was brought to attention (DelSole et al. 2019). In extremes modeling, similar under-coverage for the two-level bootstrap confidence intervals is observed even when there is no EIV (see the supplementary materials). Calibration of the confidence coefficient (e.g., Loh 1987) or more accurate procedures are needed. In presence of EIV, the under-coverage of the CSE estimator is worsened by its bias. The much less biased CSE-I method still needs some calibration since its 95% confidence intervals appear to provide about 90% coverage. This implies that the confidence intervals reported in existing works are narrower than necessary and, consequently, the detection conclusions might be over-stated. The additivity of the ANT and NAT signals in GEV location parameters assumed in earlier works (e.g., Wang et al. 2017) has not been checked against climate model simulations. Using the large ensembles under the ALL and NAT forcings from the CanESM2 model, our two-signal perfect model detection analyses over 13 subcontinental regions fill the gap. In spite of the slight downward bias of the CSE-I estimator due to EIV, no evidence was found against this assumption.

Several directions merit further investigation. Handling the EIV issue is of great practical interest in order to not only correct the bias but also improve the coverage rates of confidence intervals. Due to the special form of the GEV score equations, the bias cannot be handled by any existing parametric correction approaches (e.g., Huang and Wang 2001). The CSE method has great potential toward optimal fingerprinting for extremes if its bias due to EIV can be corrected. Incorporating multiple climate models in signal estimation is an important extension. We only used ensembles from one climate model, which makes the results dependent on this single model. Accounting the climate model uncertainty would lead to better usage of available model outputs and more reliable analyses. In the fingerprinting formulation, the GEV scale and shape parameters are different at each grid box but without any smoothness requirement. Imposing physically realistic levels of smoothness on the parameter surface might reduce the variation in signal estimation and bias in estimating the scaling factor. Finally, Bayesian hierarchical model provides an alternative paradigm (Berliner, Levine, and Shea 2000; Lee et al. 2005), which may address open methodological issues such as EIV (Katzfuss, Hammerling, and Smith 2017). Adaptation to the extreme value setting in combination with the widely used Bayesian approaches in assessing the outputs from multiple climate models (Schliep et al. 2010; Fix et al. 2018) would be of great value.

Supplementary Materials

Additional background materials and simulation studies: Working correlation function approximation; background of max-stable processes; numerical study when signals are observed in detection and attribution setting. (pdf file)

Computer code and data: R code for the proposed methods and the numerical study; observed data and climate model simulations of four temperature extremes used in the detection and attribution analysis of East Asia. (GNU zipped tar file)

Funding

References

Allen, M. R., and Stott, P. A. (2003), "Estimating Signal Amplitudes in Optimal Fingerprinting, Part I: Theory," *Climate Dynamics*, 21, 477–491. [1,2,3]

Berliner, L. M., Levine, R. A., and Shea, D. J. (2000), "Bayesian Climate Change Assessment," *Journal of Climate*, 13, 3805–3820. [11]

Bindoff, N. L., Stott, P. A., Achuta Rao, K. M., Allen, M. R., Gillett, N., Gutzler, D., Hansingo, K., Hegerl, G., Hu, Y., Jain, S., Mokhov, I. I., Overland, J., Perlwitz, J., Sebbari, R., and Zhang, X. (2013), "Detection and Attribution of Climate Change: From Global to Regional," in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Book Section 10), eds. T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley, Cambridge: Cambridge University Press, pp. 867–952. [2]

Blanchet, J., and Davison, A. C. (2011), "Spatial Modeling of Extreme Snow Depth," *Annals of Applied Statistics*, 5, 1699–1725. [3]

Bücher, A., and Segers, J. (2017), "On the Maximum Likelihood Estimator for the Generalized Extreme-Value Distribution," *Extremes*, 20, 839–872. [4]

Casson, E., and Coles, S. (1999), "Spatial Regression Models for Extremes," *Extremes*, 1, 449–468. [2]

Christidis, N., Stott, P. A., and Brown, S. J. (2011), "The Role of Human Activity in the Recent Warming of Extremely Warm Daytime Temperatures," *Journal of Climate*, 24, 1922–1930. [2]

Cooley, D., Nychka, D., and Naveau, P. (2007), "Bayesian Spatial Modeling of Extreme Precipitation Return Levels," *Journal of the American Statistical Association*, 102, 824–840. [2]

Cooley, D., and Sain, S. R. (2010), "Spatial Hierarchical Modeling of Precipitation Extremes From a Regional Climate Model," *Journal of Agricultural, Biological, and Environmental Statistics*, 15, 381–402. [5]

Davison, A. C., and Gholamrezaee, M. M. (2012), "Geostatistics of Extremes," *Proceedings: Mathematical, Physical and Engineering Sciences*, 468, 581–608. [2,3]

Davison, A. C., Padoan, S. A., and Ribatet, M. (2012), "Statistical Modeling of Spatial Extremes," *Statistical Science*, 27, 161–186. [2,3,5]

DelSole, T., Trenary, L., Yan, X., and Tippett, M. K. (2019), "Confidence Intervals in Optimal Fingerprinting," *Climate Dynamics*, 52, 4111–4126. [6,11]

Donat, M. G., Alexander, L. V., Yang, H., Durre, I., Vose, R., Dunn, R. J. H., Willett, K. M., Aguilar, E., Brunet, M., Caesar, J., Hewitson, B., Jack, C., Klein Tank, A. M. G., Kruger, A. C., Marengo, J., Peterson, T. C., Renom, M., Oria Rojas, C., Rusticucci, M., Salinger, J., Elrayah,

A. S., Sekele, S. S., Srivastava, A. K., Trewin, B., Villarroel, C., Vincent, L. A., Zhai, P., Zhang, X., and Kitching, S. (2013), "Updated Analyses of Temperature and Precipitation Extreme Indices Since the Beginning of the Twentieth Century: The HadEX2 Dataset," *Journal of Geophysical Research: Atmospheres*, 118, 2098–2118. [10]

Fix, M. J., Cooley, D., Sain, S. R., and Tebaldi, C. (2018), "A Comparison of U.S. Precipitation Extremes Under RCP8.5 and RCP4.5 With an Application of Pattern Scaling," *Climatic Change*, 146, 335–347. [11]

Fuentes, M., Henry, J., and Reich, B. (2013), "Nonparametric Spatial Models for Extremes: Application to Extreme Temperature Data," *Extremes*, 16, 75–101. [2]

Fyfe, J. C., Derksen, C., Mudryk, L., Flato, G. M., Santer, B. D., Swart, N. C., Molotch, N. P., Zhang, X., Wan, H., and Arora, V. K. (2017), "Large Near-Term Projected Snowpack Loss Over the Western United States," *Nature Communications*, 8, 14996. [3]

Giorgi, F., and Francisco, R. (2000), "Uncertainties in Regional Climate Change Prediction: A Regional Analysis of Ensemble Simulations With the HADCM2 Coupled AOGCM," *Climate Dynamics*, 16, 169–182. [6,8]

Hannart, A., Ribes, A., and Naveau, P. (2014), "Optimal Fingerprinting Under Multiple Sources of Uncertainty," *Geophysical Research Letters*, 41, 1261–1268. [2]

Hasselman, B. (2018), "nleqslv: Solve Systems of Nonlinear Equations," R Package Version 3.3.2. [5]

Hasselmann, K. (1997), "Multi-Pattern Fingerprint Method for Detection and Attribution of Climate Change," *Climate Dynamics*, 13, 601–611. [1]

Hegerl, G. C., Hoegh-Guldberg, O., Casassa, G., Hoerling, M. P., Kovats, R. S., Parmesan, C., Pierce, D. W., and Stott, P. A. (2010), "Good Practice Guidance Paper on Detection and Attribution Related to Anthropogenic Climate Change," in *Meeting Report of the Intergovernmental Panel on Climate Change Expert Meeting on Detection and Attribution of Anthropogenic Climate Change*, Bern, Switzerland, eds. T. F. Stocker, C. B. Field, D. Qin, V. Barros, G.-K. Plattner, M. Tignor, P. M. Midgley, and K. L. Ebi, IPCC Working Group I Technical Support Unit, University of Bern. [1]

Hegerl, G. C., Zwiers, F. W., Braconnot, P., Gillett, N. P., Luo, Y., Marengo Orsini, J. A., Nicholls, N., Penner, J. E., and Stott, P. (2007), "Understanding and Attributing Climate Change," in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, eds. S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, and H. L. Miller, Cambridge: Cambridge University Press, pp. 663–745. [2]

Hegerl, G. C., Zwiers, F. W., Stott, P. A., and Kharin, V. V. (2004), "Detectability of Anthropogenic Changes in Annual Temperature and Precipitation Extremes," *Journal of Climate*, 17, 3683–3700. [8]

Huang, Y. (2014), "Corrected Score With Sizable Covariate Measurement Error: Pathology and Remedy," *Statistica Sinica*, 24, 357. [6]

Huang, Y., and Wang, C. Y. (2001), "Consistent Functional Methods for Logistic Regression With Errors in Covariates," *Journal of the American Statistical Association*, 96, 1469–1482. [11]

Huntingford, C., Stott, P. A., Allen, M. R., and Lambert, F. H. (2006), "Incorporating Model Uncertainty Into Attribution of Observed Temperature Change," *Geophysical Research Letters*, 33, L05710. [2,4]

IPCC (2013a), "Annex I: Atlas of Global and Regional Climate Projections," in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, eds. T. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. Midgley, Cambridge: Cambridge University Press. [3]

——— (2013b), "Summary for Policymakers," in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, eds. T. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. Midgley, Cambridge: Cambridge University Press. [2]

Karl, T. R., and Knight, R. W. (1997), "The 1995 Chicago Heat Wave: How Likely Is a Recurrence?," *Bulletin of the American Meteorological Society*, 78, 1107–1119. [2]

Katzfuss, M., Hammerling, D., and Smith, R. L. (2017), "A Bayesian Hierarchical Model for Climate-Change Detection and Attribution," *Geophysical Research Letters*, 44, 5720–5728. [11]

Kharin, V. V., Zwiers, F. W., and Zhang, X. (2005), "Intercomparison of Near-Surface Temperature and Precipitation Extremes in AMIP-2 Simulations, Reanalyses, and Observations," *Journal of Climate*, 18, 5201–5223. [3]

Kojadinovic, I., Shang, H., and Yan, J. (2015), "A Class of Goodness-of-Fit Tests for Spatial Extremes Models Based on Max-Stable Processes," *Statistics and Its Interfaces*, 8, 45–62. [2,5]

Lai, T. L., and Small, D. (2007), "Marginal Regression Analysis of Longitudinal Data With Time-Dependent Covariates: A Generalized Method-of-Moments Approach," *Journal of the Royal Statistical Society*, Series B, 69, 79–99. [6]

Leadbetter, M. R., Lindgren, G., and Rootzén, H. (1983), *Extremes and Related Properties of Random Sequences and Processes*, New York: Springer-Verlag. [3]

Ledford, A. W., and Tawn, J. A. (1996), "Statistics for Near Independence in Multivariate Extreme Values," *Biometrika*, 83, 169–187. [5]

——— (1997), "Modelling Dependence Within Joint Tail Regions," *Journal of the Royal Statistical Society*, Series B, 59, 475–499. [5]

Lee, T. C. K., Zwiers, F. W., Hegerl, G. C., Zhang, X., and Tsao, M. (2005), "A Bayesian Climate Change Detection and Attribution Assessment," *Journal of Climate*, 18, 2429–2440. [11]

Liang, K.-Y., and Zeger, S. L. (1986), "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, 73, 13–22. [2]

Loh, W.-Y. (1987), "Calibrating Confidence Coefficients," *Journal of the American Statistical Association*, 82, 155–162. [7,11]

Martins, E. S., and Stedinger, J. R. (2000), "Generalized Maximum-Likelihood Generalized Extreme-Value Quantile Estimators for Hydrologic Data," *Water Resources Research*, 36, 737–744. [5]

Marvel, K., Schmidt, G. A., Shindell, D., Bonfils, C., LeGrande, A. N., Nazarenko, L., and Tsigaridis, K. (2015), "Do Responses to Different Anthropogenic Forcings Add Linearly in Climate Models?," *Environmental Research Letters*, 10, 104010. [6]

Min, S.-K., Zhang, X., Zwiers, F. W., Friederichs, P., and Hense, A. (2009), "Signal Detectability in Extreme Precipitation Changes Assessed From Twentieth Century Climate Simulations," *Climate Dynamics*, 32, 95–111. [8]

Min, S.-K., Zhang, X., Zwiers, F. W., and Hegerl, G. C. (2011), "Human Contribution to More-Intense Precipitation Extremes," *Nature*, 470, 378–381. [2]

Mitchell, J., and Karoly, D. (2001), "Detection of Climate Change and Attribution of Causes," in *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change* (Book Section 12), eds. G. Hegerl, F. Zwiers, M. Allen, and J. Marengo, Cambridge: Cambridge University Press, pp. 695–738. [2,3]

Nikoloulopoulos, A. K., Joe, H., and Chaganty, N. R. (2011), "Weighted Scores Method for Regression Models With Dependent Data," *Biostatistics*, 12, 653–665. [4]

Padoan, S., Ribatet, M., and Sisson, S. (2010), "Likelihood Inference for Max-Stable Processes," *Journal of the American Statistical Association*, 105, 263–277. [2,3]

Prescott, P., and Walden, A. T. (1980), "Maximum Likelihood Estimation of the Parameters of the Generalized Extreme-Value Distribution," *Biometrika*, 67, 723–724. [4]

Resnick, S. (2002), "Hidden Regular Variation, Second Order Regular Variation and Asymptotic Independence," *Extremes*, 5, 303–336. [5]

Ribatet, M., Dombry, C., and Oesting, M. (2016), "Spatial Extremes and Max-Stable Processes," in *Extreme Value Modeling and Risk Analysis: Methods and Applications*, eds. D. K. Dey and J. Yan, Boca Raton, FL: CRC Press, pp. 179–194. [2]

Ribes, A., Gillett, N. P., and Zwiers, F. W. (2015), "Designing Detection and Attribution Simulations for CMIP6 to Optimize the Estimation of Greenhouse Gas-Induced Warming," *Journal of Climate*, 28, 3435–3438. [5]

Ribes, A., Planton, S., and Terray, L. (2013), "Application of Regularised Optimal Fingerprinting to Attribution. Part I: Method, Properties and Idealised Analysis," *Climate Dynamics*, 41, 2817–2836. [2]

Risser, M. D., and Wehner, M. F. (2017), "Attributable Human-Induced Changes in the Likelihood and Magnitude of the Observed Extreme

Precipitation During Hurricane Harvey," *Geophysical Research Letters*, 44, 12,457–12,464. [3]

Sang, H., and Gelfand, A. E. (2009), "Hierarchical Modeling for Extreme Values Observed Over Space and Time," *Environmental and Ecological Statistics*, 16, 407–426. [2]

Schär, C., and Jendritzky, G. (2004), "Climate Change: Hot News From Summer 2003," *Nature*, 432, 559–560. [2]

Schlather, M., and Tawn, J. A. (2003), "A Dependence Measure for Multivariate and Spatial Extreme Values: Properties and Inference," *Biometrika*, 90, 139–156. [7]

Schliep, E. M., Cooley, D., Sain, S. R., and Hoeting, J. A. (2010), "A Comparison Study of Extreme Precipitation From Six Different Regional Climate Models via Spatial Hierarchical Modeling," *Extremes*, 13, 219–239. [11]

Shang, H., Yan, J., and Zhang, X. (2015), "A Two-Step Approach to Model Precipitation Extremes in California Based on Max-Stable and Marginal Point Processes," *The Annals of Applied Statistics*, 9, 452–473. [7]

Smith, E. L., and Stephenson, A. G. (2009), "An Extended Gaussian Max-Stable Process Model for Spatial Extremes," *Journal of Statistical Planning and Inference*, 139, 1266–1275. [2]

Smith, R. L. (1985), "Maximum Likelihood Estimation in a Class of Non-regular Cases," *Biometrika*, 72, 67–90. [4]

Tseng, P. (2001), "Convergence of a Block Coordinate Descent Method for Nondifferentiable Minimization," *Journal of Optimization Theory and Applications*, 109, 475–494. [4]

Tseng, P., and Yun, S. (2009), "A Coordinate Gradient Descent Method for Nonsmooth Separable Minimization," *Mathematical Programming*, Series B, 117, 387–423. [4]

Wang, Z., Jiang, Y., Wan, H., Yan, J., and Zhang, X. (2017), "Detection and Attribution of Changes in Extreme Temperatures at Regional Level," *Journal of Climate*, 30, 7035–7047. [3,6,10,11]

Wang, Z., Yan, J., and Zhang, X. (2014), "Incorporating Spatial Dependence in Regional Frequency Analysis," *Water Resources Research*, 50, 9570–9585. [5]

Wen, Q. H., Zhang, X., Xu, Y., and Wang, B. (2013), "Detecting Human Influence on Extreme Temperatures in China," *Geophysical Research Letters*, 40, 1171–1176. [2]

Zhang, X., Wan, H., Zwiers, F. W., Hegerl, G. C., and Min, S.-K. (2013), "Attributing Intensification of Precipitation Extremes to Human Influence," *Geophysical Research Letters*, 40, 5252–5257. [2]

Zhang, X., Wang, J., Zwiers, F. W., and Groisman, P. Y. (2010), "The Influence of Large-Scale Climate Variability on Winter Maximum Daily Precipitation Over North America," *Journal of Climate*, 23, 2902–2915. [3]

Zwiers, F. W., Hegerl, G. C., Zhang, X., and Wen, Q. (2014), "Quantifying the Human and Natural Contributions to Observed Climate Change," in *Statistics in Action* (Book Section 20), ed. J. F. Lawless, Boca Raton, FL: Chapman and Hall/CRC, pp. 349–370. [3]

Zwiers, F. W., Zhang, X., and Feng, Y. (2011), "Anthropogenic Influence on Long Return Period Daily Temperature Extremes at Regional Scales," *Journal of Climate*, 24, 881–892. [2,3,4,5,6,8,10]