8-6-2015

Estimating Equations for Spatial Extremes with Applications to Detection and Attribution Analysis of Changes in Climate Extremes

Zhuo Wang
zhuo.wang.uconn@hotmail.com

Estimating Equations for Spatial Extremes with Applications to Detection and Attribution Analysis of Changes in Climate Extremes

Zhuo Wang, Ph.D.
University of Connecticut, 2015

ABSTRACT

In inference for max-stable processes in regional frequency analysis, it is found that, when the dependence model is misspecified, the pairwise likelihood method leads to bias in estimating the shape parameter of the generalized extreme value (GEV) distribution. The bias can be serious when the dependence is strong. Motivated by the fact that the primary interest in many studies is the inference about marginal GEV parameters and that the spatial dependence is a nuisance, we propose a combined score equations (CSE) approach that does not need dependence assumptions beyond the univariate GEV distribution. The CSE method combines the score equations of GEV model at each site with an approximate correlation function of the scores to improve the estimation efficiency. Applied to fingerprinting of changes in climate extremes with a coordinate descent algorithm to estimate a large number of parameters, the CSE method provides a close analog to the optimal fingerprinting in detection and attribution of changes in

climate extremes. The approach is applied on extreme temperature in Australia under a perfect model setting and in Northern Europe with the dependence structure modeled by exponential correlation. The CSE approach with working independence reduces to the independence likelihood method, but the estimation is much faster. This approach is applied on the annual maximum daily maximum (TXx) and minimum (TNx) and annual minimum daily maximum (TXn) and minimum (TNn) temperatures during 1951–2010 over 17 subcontinents. In the single-signal analyses, anthropogenic and natural influence can be detected in all four indices separately, but the detection occurs more frequently in TNx. We also studied the two-signal analyses that both anthropogenic signal and natural signal are present in the model simultaneously, which has not been reported before under the extreme value framework due to methodological limitations. The anthropogenic signal is separable from natural signal in a few regions of the four indices. The 2006-2010 waiting time of the 1951-1955 20-yr return value of annual maximum extreme temperature was found to decrease while that of annual minimum extreme temperature was found to increase, implying substantially influence of the anthropogenic and natural forcing on extreme temperature.

Estimating Equations for Spatial Extremes with Applications to Detection and Attribution Analysis of Changes in Climate Extremes

Zhuo Wang

B.E., Actuarial Science, Central University of Finance and Economics, Beijing, China,

2011

A Dissertation
Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy
at the
University of Connecticut

2015

APPROVAL PAGE

Doctor of Philosophy Dissertation

Estimating Equations for Spatial Extremes with Applications to Detection and Attribution Analysis of Changes in Climate Extremes

Presented by

Zhuo Wang, B.E.

Major Advisor _____

Jun Yan

Associate Advisor _____

Zhiyi Chi

Associate Advisor _____

Dipak K. Dey

Associate Advisor _____

Xuebin Zhang

University of Connecticut

2015

Acknowledgements

I would like to express my deepest gratitude to my advisor, Dr. Jun Yan, for his invaluable guidance throughout my graduate studies, counsel and sharp criticism. He has led me into the statistics field of spatial extremes, and guided me through difficult moments in my research. I am also very grateful for his understanding and offering me the opportunity to be exchanged to a university in China. I feel very fortunate to have Dr. Yan as my advisor.

My sincere thanks also goes to my associate advisors Drs. Zhiyi Chi, Dipak Dey and Xuebin Zhang, for their valuable guidance in my academic study. Especially, I would like to thank Dr. Zhang for providing the data, computer resource and the expert opinion in climate change.

I am thankful for Dr. Chi's recommendation for an assistantship in the UConn Health Center, which provided a great opportunity to work with biostatistics problems.

I would also like to thank Dr. Ming-Hui Chen for facilitating the exchange program between our department and the math department at Shanghai Jiao Tong University.

Last but not least, I would like to thank my family for their support. My husband Zepeng Xie, who is always there for me, made the arduous journey of graduate school seem smooth and effortless.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Natural extremes, such as extreme rainfall or extreme temperature, have profound impact on both the environment and the society, so understanding the changes in climate extremes is necessary. Detection and attribution of changes in climate extremes have gained sharpened focus (e.g., Field et al., 2012; Seneviratne et al., 2012). Such analysis involves assessment of observed changes in relation to what are expected to have occurred in response to forcings external to the climate system. Examples of the latter are changes in the earth's energy budget due to increasing greenhouse gas concentrations that affect outgoing infrared radiation, or changes in incoming solar radiation or volcanic activity. Given the recent public attention to changes in climate extremes and the associated risk management (e.g., Field et al., 2012), detection and attribution of changes in climate extremes is of great importance for the understanding of climate change impacts on the environment and society.

Recent methods for spatial extremes modeling can be grouped into two major categories based on how spatial dependence is introduced. Methods in the first category introduce spatial dependence through a latent process, conditional on which independent

univariate extreme models are applied (Casson and Coles, 1999; Cooley et al., 2007; Sang and Gelfand, 2009; Fuentes et al., 2013). Statistical inferences and risk assessment are carried out in a Bayesian framework. Methods in the second category model spatial dependence directly through max-stable processes (de Haan, 1984; Smith, 1990a; Schlather, 2002). This approach is mathematically elegant as a max-stable process truly defines a random field over a continuous region. The joint density function, however, is intractable for any finite dimension greater than two or three, which makes the maximum likelihood approach infeasible. The composite likelihood approach based on pairwise marginal densities has been used for statistical inferences (Smith and Stephenson, 2009; Davison and Gholamrezaee, 2009; Padoan et al., 2010). Such a fully specified parametric model for spatial dependence may lead to more efficient estimation of the regression coefficients in marginal generalize extreme value (GEV) models, but the efficiency gain is at the cost of strong distributional assumptions. The pairwise likelihood requires that at least each pair of extreme observations follow a bivariate GEV distribution. When the model is misspecified, inconsistent estimators may result. In reality, a max-stable process model may very likely fail goodness-of-fit tests when the study region is of any practical size (Kojadinovic et al., 2015). A full parametric model may not be flexible enough to accommodate realities and, in situations where spatial dependence is of nuisance interest, not necessary.

A combined score equations approach is proposed to provide a general methodology for extreme value analysis in a spatial setting that exploits spatial dependence to increase

the efficiency without dependence assumptions beyond the univariate GEV distribution. The method is motivated by the fact that in the detection and attribution of the changes in climate extremes, the primary interest is the inference about the attribution parameter in the marginal GEV distributions and that the spatial dependence is a nuisance. At each site, the marginal GEV distribution leads to a score equation, which can be combined in some optimal way to improve efficiency by accounting the spatial correlation among them. Misspecification of the correlation structure does not affect the consistency of the CSE estimator. When the working correlation structure is closer to the truth than the independence structure, the resulting estimator is more efficient than that from the independence likelihood.

The first part of the dissertation is about incorporating the spatial dependence in regional frequency analysis through the max-stable process. The efficiency of regional frequency analysis is undermined by intersite dependence, which is usually ignored in parameter estimation. We propose a spatial index flood model where marginal GEV distributions are joined by an extreme-value copula characterized by a max-stable process for the spatial dependence. The parameters are estimated with a pairwise likelihood constructed from bivariate marginal GEV distributions. Through simulation, we compared the pairwise likelihood method with an L-moment method and an independence likelihood method under various spatial dependence models and dependence levels. The pairwise likelihood method was found to be the most efficient in mean squared error if the dependence model was correctly specified. However, when the dependence model was

not a max-stable model, the pairwise likelihood method led to serious bias in estimating the shape parameter and return levels, especially when the dependence was strong. The low estimation efficiency of the pairwise likelihood method in the misspecification of the dependence model motivates us to find another way to model the spatial dependence without requiring parametric specifications beyond univariate marginal GEV distributions.

In the second part of the dissertation, we propose a combined score equation method that combines the score equations of the GEV model at each grid box such that an approximate correlation function of the scores is used to improve the estimation efficiency. Under working independence, it reduces to the existing method, but the estimation is much faster with a coordinate descent algorithm. Unlike the pairwise likelihood method assuming max-stable processes, the CSE method does not need full specification of spatial dependence, and thus not suffer the consequence of the misspecification of the dependence model. The method is applied to extreme temperature in Australia under a perfect model setting and in Northern Europe with real data.

The third part of the thesis is an application of the CSE method with working independence on the regional detection and attribution analysis in extreme temperatures at 17 subcontinents. The CSE method with working independence is used in a coordinate descent algorithm. This coordinate descent algorithm is conceptually the same as the profile independence likelihood method used in Zwiers et al. (2011), however, the profile approach searches the optimal scaling factor among a large set of predetermined

values and is quite computing intensive: the single-signal analysis of the 16 regions in their paper took weeks on a computer cluster. Thus, extending the method in Zwiers et al. (2011) to two or more signals is unpractical. In contrast, the coordinate descent algorithm is computationally much more efficient and accurate. This approach makes it is possible to study the two-signal analyses in which two signals of external forcings are present in the location parameter of the GEV distribution at the same time. We study the two-signal detection analyses as well as single-signal detection analyses of four types of extreme temperatures at 17 subcontinents.

The dissertation is organized as follows. In Chapter 2, the max-stable process is introduced and applied with the pairwise likelihood method in regional frequency analysis. The combined score equation method is proposed in Chapter 3. Chapter 4 outlines the fingerprinting method. The CSE method with working exponential correlation structure is applied in the fingerprinting in a simulation study and two real data analyses. In Chapter 5, the CSE method with working independence is applied with a coordinate descent algorithm on the regional detection and attribution analysis in extreme temperatures at 17 subcontinents. Summary and future work are given in Chapter 6.

Chapter 2

Spatial Regional Frequency Analysis

2.1 Introduction

Regional frequency analysis (RFA) is widely-used in characterizing the frequency of the extreme events. It is a technique that based on the regionalization concept that trades space for time to obtain adequate estimation of model parameters based on data from a low-density network with short record length (see, e.g., Ouarda, 2013, for a recent review). It uses data from a number of sites that are identified to be in a homogeneous region in certain sense (see Section 2.2 for specific definition) to estimate the quantiles of the variables of interest at each site in the region. That is, short records from different sites within a homogeneous region are pooled to improve the estimation efficiency. Widely used in water resources research, the RFA approach has various models and has been extended to accommodate temporal nonstationarity (e.g., Ouarda et al., 2006; Cunderlik and Ouarda, 2006; Leclerc and Ouarda, 2007) and multivariate analysis (e.g., Ouarda et al., 2000; Javelle et al., 2002; Chebana and Ouarda, 2009). Nevertheless, our focus is the stationary index flood model with marginal GEV distributions. We improve the efficiency of this type of RFA by incorporating spatial dependence and

compare its performance with competing methods to better understand its advantages and limitations.

In an index flood model, the marginal distributions are identical apart from a site-specific scaling factor. Two popular methods are available for parameter estimation, neither of which needs to specify the intersite dependence. The L-moment method (e.g., Hosking and Wallis, 1997) estimates the parameters by solving the equations that match the sample L-moments with the population moments. It first estimates the site-specific scaling factor with at-site data, and then uses it to scale the data at each site. The scaleless data are then pooled to estimate the parameters of the shared scaleless distribution by matching the L-moments. Properties of the L-moments were studied by Hosking (1990). L-moments are more robust to sampling variability than conventional moments, and their existence only requires existence of the mean. The second method is the independence likelihood method that adds up the marginal loglikelihood from all sites, ignoring the intersite dependence, and then maximizes it (Smith, 1990b). The independence likelihood method in RFA gives the most efficient estimator for larger samples, and it can incorporate covariates into model parameters (e.g., Buishand, 1991; Northrop, 2004), which is necessary in many cases where temporal or spatial nonstationarity is present. Both the L-moment method and the independence likelihood method are robust to intersite dependence at the cost of low efficiency when the dependence is strong.

The impact of intersite dependence on many aspects of RFA is a fundamental issue that has been actively investigated. Intersite correlation does not introduce bias but increases the variance in predicting regional mean or moments (e.g., Matalas and Langbein, 1962; Stedinger, 1983). For index flood models, Hosking and Wallis (1988) reported similar findings in predicting flood quantiles. Intersite dependence was found in general to increase the variance of the estimator in other contexts such as estimation of regional exceeding probability of a flood level (Troutman and Karlinger, 2003) or a regional envelope curve (Castellarin et al., 2005). Intersite correlation is part of the model in probabilistic regional envelope curves (Castellarin, 2007; Viglione et al., 2012). It has been used in regression analysis with generalized least squares to estimate the parameters of a model of the target quantity as a function of basin characteristics (Griffis and Stedinger, 2007). For testing regional homogeneity with the heterogeneity measures of Hosking and Wallis (1993), intersite dependence reduces the power of the tests (Castellarin et al., 2008). Most of the existing simulation studies generated data from meta-Gaussian models, which essentially use the normal copula for the dependence structure (e.g., Hosking and Wallis, 1988). For extreme observations, however, the Pearson correlation coefficient may not be a good dependence measure (Embrechts et al., 2002) and the normal copula may not be a good dependence model (Genest and Favre, 2007; Gudendorf and Segers, 2010). Smith (1990b) reported a study where the intersite dependence was modeled with an extreme-value copula, but it was an exchangeable Gumbel copula which does not allow the dependence to weaken as the distance between

two sites increases.

Spatial extreme modeling has made progress recently in the statistics literature; see Davison et al. (2012) for a recent review. Max-stable processes extend the multivariate extreme value distribution to the infinite dimensional setting (de Haan, 1984), with marginal distribution of any dimension being multivariate extreme-value. These models provide a natural modeling framework for spatial extremes. A pairwise likelihood approach has been used in parameter estimation due to the unavailability of the joint multivariate density function (e.g., Padoan et al., 2010; Davison and Gholamrezaee, 2012). The pairwise likelihood approach has a robust feature that its validity in inference only needs the correct specification of the bivariate joint density of all the pairs, instead of the full joint density. A spatial index model retains the marginal GEV distributions and uses a max-stable process model for the dependence structure. The pairwise likelihood approach with pairwise bivariate generalized extreme value distributions can potentially increase the efficiency for marginal GEV parameter and return level estimation. Similar efficiency improvement with max-stable process model has recently been reported in detection of nonstationarity in precipitation extremes (Westra and Sisson, 2011).

The rest of this chapter is organized as follows. The index flood model with GEV distributions is reviewed in Section 2.2, along with two existing estimation methods: L-moment and independence likelihood. Spatial extreme models, their application in RFA with index flood model, and parameter estimation with a pairwise likelihood method are

introduced in Section 2.3. A large scale simulation study that compares the performance of proposed method with the L-moment method and the independence likelihood method is reported in Section 2.4. All three methods are illustrated in an example of Swiss annual maximum daily precipitation in Section 2.5. A discussion is given in Section 2.6.

2.2 Index Flood Model with GEV Distribution

2.2.1 Model

The index flood model is a widely-used RFA model with the homogeneity assumption being that all the sites have an identical distribution up to a site-specific scaling factor known as the index variable. It originated from applications to flood data in hydrology, but the method can be used with any kind of data (Hosking and Wallis, 1997, p.6). Examples of application to precipitation are Kysely and Picek (2007) and Ngongondo et al. (2011). Let $Q_s(\cdot)$ be the quantile function of the distribution at site s; i.e., $Q_s(u) = \inf\{y \in R : u \leq F_s(y)\}$, where F_s is the distribution function at site s. The index flood procedure assumes that for all site s in a homogeneous region, $Q_s(u) = c_s q(u)$, where c_s is a site specific index variable, and $q(\cdot)$ is called the regional growth curve, the scale-free quantile function shared by all sites. Within the region, the T-year return level at any site s which is the upper $1/T$-quantile, is proportional to the return level of the scale-free distribution: $Q_s(1 - 1/T) = c_s q(1 - 1/T)$.

The GEV distribution is often used to model the regional growth curve. It can

be obtained as the limit distribution of properly normalized maximum of a sequence of independent and identically distributed random variables. The probability density function of a GEV distribution is

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

where

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

The cumulative distribution function is

$$F(y;\, \mu, \sigma, \xi) = e^{-t(y)},$$

where μ, σ and ξ are the location, scale and shape parameters, respectively. Let GEV(μ, σ, ξ) denote this distribution. The shape parameter ξ determines the tail behavior; negative indicates light tail while positive indicates heavy tail. The distribution is known as the Gumbel distribution when $\xi = 0$. The case with $\xi > 0$ is of most interest since real data of extreme events often exhibits heavy tail. The quantile function of

GEV(μ, σ, ξ) is the inverse function of F in (2.2.1):

$$Q(u; \mu, \sigma, \xi) = \begin{cases} \mu + (\sigma/\xi)\{(-\ln u)^{-\xi} - 1\}, & \xi \neq 0; \\ \mu + \sigma\{-\ln(-\ln u)\}, & \xi = 0. \end{cases} \tag{2.1}$$

The T-year return level is then $Q(1 - 1/T)$.

When the GEV distribution is used in a index flood model, the location parameter can be used as the index variable. In particular, let $c_s = \mu_s$ and let Z be a GEV$(1, \gamma, \xi)$ variable. It is straightforward to show that the distribution of $Y_s = c_s Z$ is GEV$(\mu_s, \mu_s\gamma, \xi)$. Therefore, the homogeneity for this index flood model means that the ratio of the scale parameter to the location parameter of the GEV parameters is a constant (γ), and that the shape parameters at all sites are the same (e.g., Buishand, 1991; Hanel et al., 2009). Note that the index flood model only specifies the marginal GEV distributions; no spatial dependence is specified.

2.2.2 Existing Estimation Methods

Suppose that we observe annual maxima of a variable of interest at m sites over n years. Let Y_{ts}, $t = 1, \ldots, n$ and $s = 1, \ldots, m$, be the record in year t from site s. The data from year to year are assumed to be independent, but within the same year, spatial dependence exists across the sites. For ease of presentation, the notations are for balanced data where all sites have the same length of records, but the methods can be

easily adapted to use varying length of records. Let μ_s, $\sigma_s = \mu_s\gamma$, and ξ_s be the location, scale, and shape parameters, respectively, of the GEV distribution at site s. The parameters to be estimated are $\beta = (\mu_1, \ldots, \mu_m, \gamma, \xi)$. Two existing estimation methods are the L-moment method (Hosking and Wallis, 1988) and the independence likelihood method (Smith, 1990b; Hanel et al., 2009), neither of which requires the specification of the spatial dependence. As both methods target small samples, asymptotic variance estimator of the parameter estimator is not expected to work well, which is observed in our simulation studies in Section 2.4. A parametric bootstrap procedure with preserved spatial dependence is used to assess the uncertainty of the estimator.

L-Moment The L-moment method proceeds as follows. First, for each site s, estimate the GEV parameters (μ_s, σ_s, ξ_s) using data from this site with the L-moment method, and let $\hat{\mu}_s$ be the estimate of μ_s. Then use $\hat{\mu}_s$ to scale the data at each site s by letting $X_{ts} = Y_{ts}/\hat{\mu}_s$. Apply the L-moment method to the pooled, scaled data X_{ts}, $s = 1, \ldots, m$, $t = 1, \ldots, n$, to fit a GEV distribution with location 1, scale γ and shape ξ. The only extra difficulty in the last step is that the location parameter of the GEV distribution is restricted to be 1. With two unknown parameters (γ, ξ), the estimating equations match the first two sample L-moments (l_1, l_2) with their population counterparts:

$$l_1 = 1 - \gamma\{1 - \Gamma(1 - \xi)\}/\xi,$$

$$l_2 = -\gamma(1 - 2^{\xi})\Gamma(1 - \xi)/\xi,$$

where $\xi < 1$. The solutions to the equations are the L-moment estimates $(\hat{\gamma}, \hat{\xi})$. The implicit restriction $\xi < 1$ is required for the existence of the L-moments (finite mean), which makes the L-moment method more efficient than the likelihood method for small samples when $\xi < 1$ is true (e.g., Coles and Dixon, 1999). When the homogeneity assumption is not valid, bias may be introduced, but RFA may still be more accurate than single site analysis (Lettenmaier et al., 1987; Hosking and Wallis, 1997).

Independence Likelihood Denote $f_{ts}(\cdot; \beta)$ as the probability density function of the GEV distribution at site s in year t. The independence likelihood method estimates β by maximizing the log-likelihood function pretending that the sites are independent. That is, $\hat{\beta}$ is the maximizer of

$$L_{I,n}(\beta) = \sum_{t=1}^{n} \sum_{s=1}^{m} \log f_{ts}(Y_{ts}; \beta). \tag{2.2}$$

Similar to the L-moment method, this method only assumes correct specification of the marginal GEV distribution at each site. No spatial dependence is taken into account in the point estimation. For large samples, the variance of the estimator has a sandwich form under certain regularity conditions and can be consistently estimated by a sandwich estimator (Smith, 1990b). The sandwich variance adjusts for the unspecified spatial dependence. For small samples, however, a bootstrap procedure that preserves the spatial dependence can be used (Heffernan and Tawn, 2004). Modification of the likelihood method to improve its small sample performance has been obtained by adding

a penalty on the shape parameter (Coles and Dixon, 1999) or, equivalently, imposing a prior distribution on it (Martins and Stedinger, 2000). We do not consider them here because they introduce the complexity of penalty form selection or prior specification.

2.3 Spatial Index Flood Model

Spatial extreme models have gained much focus in the statistics literature. Can one exploit spatial extreme modeling in RFA to improve efficiency? If so, the spatial dependence can be used in a positive way for better efficiency instead of as nuisance that reduces the efficiency.

2.3.1 Max-stable Process

By Sklar's Theorem, the distribution function H of a p-dimensional continuous random vector $(X_1, \ldots, X_p)$ with marginal distribution $F_1, \ldots, F_p$, respectively, can be uniquely represented as

$$H(x_1, \ldots, x_p) = C\{F_1(x_1), \ldots, F_p(x_p)\}, \quad (x, \ldots, x_p) \in \mathbb{R}^p.$$

where $C : [0,1]^p \to [0,1]$, called a *copula*, is a p-dimensional distribution function with standard uniform marginals (Sklar, 1959). When H is a multivariate extreme value distribution, the corresponding copula C must be an extreme-value copula, which satisfies a max-stable property (Gudendorf and Segers, 2010). If all the margins are transformed

to unit Fréchet distribution with distribution function $G(x) = e^{-1/x}$, the max-stable property means

$$\Pr(Z_1 \leq kz_1, \ldots, Z_p \leq kz_p)^k = \Pr(Z_1 \leq z_1, \ldots, Z_p \leq z_p), \quad z_i > 0, \quad i = 1, \ldots, p, \quad k > 0.$$

Max-stability is a defining property for max-stable processes whose marginal copula in any dimension is an extreme-value copula.

In a recent review, Davison et al. (2012) gave a spectral characterization of max-stable processes that unifies the characterizations in de Haan (1984) and Schlather (2002). Consider a spatial domain $\mathcal{X} \subset \mathbb{R}^2$. Let $\{W(x) : x \in \mathcal{X}\}$ be a nonnegative stationary stochastic process on $\mathcal{X}$ with $E\{W(x)\} = 1$ and $W_1, W_2, \ldots$ be independent copies of W. Let $\zeta_1, \zeta_2, \ldots$ be the points of a Poisson process on $\mathbb{R}^+$ with intensity $s^{-2}\mathrm{d}s$. Then,

$$Z(x) = \max_j \zeta_j W_j(x), \tag{2.3}$$

is a stationary max-stable process on $\mathcal{X}$ with unit Fréchet marginal distributions. Different forms of $W(x)$ lead to different parametric max-stable models.

It is often desirable to measure the extremal dependence of m sites. The extremal coefficient is such a measure. Consider a max-stable process Z defined in (2.3) at sites

$\{x_1, \ldots, x_m \in \mathcal{X}\}$. The extremal coefficient of the m sites is

$$\Pr[Z(x_1) \leq z, \ldots, Z(x_m) \leq z] = \exp\left(-\frac{\theta_m}{z}\right). \tag{2.4}$$

It can be interpreted as the effective sample size of the m variables. The upper bound of θ_m is m, meaning complete independence, while the lower bound is 1, meaning complete dependence. Specifically, for two sites x_1 and x_2, a bivariate extremal coefficient function $\theta(h)$ can be defined as

$$\Pr[Z(x_1) \leq z, Z(x_2) \leq z] = \exp\left(-\frac{\theta(h)}{z}\right), \tag{2.5}$$

where $h = x_1 - x_2$ (Schlather and Tawn, 2003). In a spatial context, the bivariate extremal coefficient of two sites is often modeled to increase from 1 to 2 as the distance $\|h\|$ between the two sites increases from zero to infinity. When $\theta(h) = \theta(\|h\|)$ — the dependence measure depends only on distance instead of direction — the model is isotropic.

2.3.2 Parametric Max-stable Models

We consider three isotropic models that are used in the simulation study. The first model is obtained by taking $\mathcal{X} = \mathbb{R}^2$ and $W_j(x) = g(x - X_j)$, where g is a bivariate density function and $X_1, X_2, \ldots$ are the points of a homogeneous Poisson process with

unit rate on $\mathcal{X}$. The special case where g is the normal density with mean zero and covariance matrix Σ is known as the Smith model (Smith, 1990a). The bivariate marginal distribution function at two sites x_i and x_j is

$$\Pr[Z(x_i) \leq z_i, Z(x_j) \leq z_j] = \exp\left\{-\frac{1}{z_i}\Phi\left(\frac{a}{2} + \frac{1}{a}\log\frac{z_j}{z_i}\right) - \frac{1}{z_j}\Phi\left(\frac{a}{2} + \frac{1}{a}\log\frac{z_i}{z_j}\right)\right\}, \tag{2.6}$$

where Φ is the cumulative distribution function of the standard normal variable, $a^2 = \Delta^\top\Sigma^{-1}\Delta$, and $\Delta = x_i - x_j$. The bivariate density function can be obtained by differentiating the distribution function (e.g., Padoan et al., 2010). The bivariate extremal coefficient function is $\theta(h) = 2\Phi(\sqrt{h^\top\Sigma^{-1}h}/2)$, with range from 1 to 2, providing full range of dependence level. A limitation of the Smith model is that the storms generated from it have shapes that are too regular compared to the reality.

The second model we consider is the Schlather model (Schlather, 2002). It is obtained by taking $W(x) = \max\{0, 2\sqrt{\pi}\epsilon(x)\}$, where $\epsilon(x)$ is a stationary Gaussian process with unit variance and correlation function ρ, a function of the euclidean distance between two sites x_i and x_j. The bivariate marginal distribution function is

$$\Pr[Z(x_i) \leq z_i, Z(x_j) \leq z_j] = \exp\left\{-\frac{1}{2}\left(\frac{1}{z_i} + \frac{1}{z_j}\right)\left(1 + \sqrt{1 - 2[\rho(h)+1]\frac{z_i z_j}{(z_i+z_j)^2}}\right)\right\}, \tag{2.7}$$

where $h = \|x_i - x_j\|$. The bivariate extremal coefficient is $\theta(h) = 1 + \sqrt{(1-\rho(h))/2}$, with a range from 1 to $1 + \sqrt{1/2}$, or about 1.707. Therefore, the Schlather model does not

provide full range. It cannot be used to model sites that are completely independent. Models of correlation function are standard in spatial statistics (e.g., Banerjee et al., 2004, Table 2.1), and contain parameters that characterizing the strength of the spatial dependence.

The third model we consider is a geometric Gaussian process obtained by taking $W(x) = \exp\{\delta\epsilon(x) - \delta^2/2\}$, where $\epsilon(x)$ is again a stationary Gaussian process with unit variance and correlation function ρ, and $\delta^2 > 0$ is the variance of $W(x)$ on the log scale (Davison et al., 2012, p.172). The bivariate marginal distribution is the same as (2.6) for the Smith model, except that $a^2 = 2\delta^2\big(1 - \rho(h)\big)$. The bivariate extremal coefficient function is $\theta(h) = 2\Phi(\delta\sqrt{(1-\rho(h))/2})$. The range of $\theta(h)$ is from 1 to $2\Phi(\delta/\sqrt{2})$. The upper bound is 1.96, quite close to 2, if $\delta^2 = 8$. As $\delta \to \infty$, $\theta(h)$ approaches 2 for any ρ.

2.3.3 Application to RFA

Spatial Index Flood Model To incorporate the spatial dependence in the index flood model, we assume that the dependence structure among the sites is an extreme-value copula described by a max-stable process model. This assumption is in addition to the homogeneity assumption for the index flood model with GEV margins in Section 2.2. We keep using Y_{ts}, $s = 1, \ldots, m$, $t = 1, \ldots, n$, as the observed data at site s in year t. The spatial index flood model completely specifies the joint distribution of $\{Y_{ts} : s = 1, \ldots, m\}$ for each t. As seen from Sklar's theorem, it is sufficient to specify the marginal models and the spatial dependence structure. Under the setup of index flood model, the

marginal distribution at each site s is still GEV$(\mu_s, \mu_s\gamma, \xi)$. The extreme-value copula is specified by a parametric max-stable process with dependence parameter α. Although the model is fully specified and easy to understand, the joint density is unavailable except for lower dimensions ($m = 2, 3$) for certain parametric models. The bivariate marginal density of two sites depends on α in addition to the marginal GEV parameters. Let $F_s(\cdot; \beta)$ be the cumulative distribution function of the GEV distribution at site s, let G be the cumulative distribution function of the unit Fréchet distribution, and let G^{-1} be the inverse function of G. The bivariate density of site i and j, $f_{i,j}$, is

$$f_{i,j}(y_i, y_j; \beta, \alpha) = g_{i,j}\left(z_i, z_j; \alpha\right) |J(y_i, y_j; \beta)|,$$

where $g_{i,j}(z_i, z_j; \alpha)$ is the bivariate marginal density of the max-stable process model with unit Fréchet margins, $z_i = G^{-1}\{F_i(y_i; \beta)\}$, and

$$|J(y_i, y_j; \beta)| = \Big|\frac{\mathrm{d}}{\mathrm{d}y_i} G^{-1}\{F_i(y_i; \beta)\} \frac{\mathrm{d}}{\mathrm{d}y_j} G^{-1}\{F_j(y_j; \beta)\}\Big|.$$

Pairwise Likelihood Estimation Inferences about max-stable process models have been mostly based on the composite likelihood approach (Padoan et al., 2010; Davison and Gholamrezaee, 2012). The composite likelihood approach constructs an objective function, known as the composite likelihood, by putting together pieces of tractable likelihood, such as lower dimensional marginal densities (Lindsay, 1988). The composite

likelihood is maximized to give the maximum composite likelihood estimator (MCLE) as if it were a likelihood. Under mild conditions, correct specification of the pieces in the composite likelihood leads to consistency and asymptotic normality of the MCLE. It has wide applications where the full joint distribution is unavailable or intractable but lower-order marginal or conditional distributions are known (e.g., Varin, 2008; Varin et al., 2011). When the pieces in the composite likelihood are pairwise bivariate densities, the composite likelihood is also called pairwise likelihood. The independence likelihood in Smith (1990b) is also a composite likelihood constructed from the univariate marginal GEV distributions.

The dependence parameter α and marginal parameter β are estimated jointly in the pairwise likelihood method. The pairwise likelihood is constructed with the bivariate density of all the site-pairs within the same years:

$$L_{P,n}(\beta, \alpha) = \sum_{t=1}^{n} \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \log f_{i,j}(Y_{i,t}, Y_{j,t}; \beta, \alpha). \tag{2.8}$$

When the record lengths are different across sites, it can be constructed from all the available pairs within each year. Let $(\hat{\beta}_n, \hat{\alpha}_n)$ be the maximizer of the pairwise log-likelihood (2.8). Under certain regularity conditions, $(\hat{\beta}_n, \hat{\alpha}_n)$ is consistent to the true parameter vector and is asymptotically normally distributed (e.g., Padoan et al., 2010). The variance of $(\hat{\beta}_n, \hat{\alpha}_n)$ can be estimated by a sandwich estimator, which can only give valid inference when the sample size n is large. For small to moderate sample sizes, as

is often the case with RFA, a bootstrap variance estimator is preferred. Heffernan and Tawn (2004) proposed a bootstrap procedure that preserves the dependence structure for multivariate extremes. This procedure has been applied in a non-stationary index flood model (Hanel et al., 2009), and is used here.

2.4 Simulation Study

A simulation study was conducted to compare the performance of the three estimation methods for index flood model: L-moment, independence likelihood, and pairwise likelihood. Unlike the other two methods, which do not need to specify spatial dependence, the pairwise likelihood method incorporates spatial dependence through the extra specification of an extreme-value dependence model. It has the potential of being more efficient when the dependence model is correctly specified, but risks severe bias otherwise. The L-moment method has been found to be unbiased regardless of the spatial dependence (Hosking and Wallis, 1988). Nevertheless, existing studies all used normal copulas, which provide no extremal dependence, in generating data. The performance of the L-moment method for data with extremal dependence as generated from max-stable processes has not previously been assessed. Our simulation design reflects these needs.

2.4.1 Design

We considered data from m sites over n years in a study region $\mathcal{X} = [0, 10]^2$. Data from different years were independent, but within the same year, data from different sites were generated with spatial dependence. The center point $(5, 5)$ is included so that parameter estimates and return level estimates are compared at this point across scenarios. The additional $m-1$ sites were randomly generated in the region. The marginal distribution at site s is $\text{GEV}(\mu_s, \mu_s\gamma, \xi)$ with $\gamma = 0.3$, $\xi = 0.2$, $\mu_s = 37$ for $s = (5, 5)$, and μ_s for other sites randomly generated from a normal distribution $N(43.0, 4.4^2)$ and rounded to an integer. The parameters of this normal distribution were the sample mean and sample variance of the L-moment estimates of the μ_s's from an extreme rainfall data in Southern Ontario analyzed in Wang et al. (2014). The values of μ_s ranged from 33 to 51.

Four factors were considered in the experimental design: the spatial dependence model, the spatial dependence level, the number of sites m, and the length of the record n. Four spatial dependence models were used to generate data, including three extreme-value models and one non-extreme-value model. The three parametric isotropic extreme-value models were the Smith model, the Schlather model, and the geometric Gaussian model, abbreviated as SM, SC, and GG, respectively. The non-extreme-value model is a Gaussian copula, which is also known as meta-Gaussian model, abbreviated as GA. For each model, three levels of dependence were used: weak, moderate, and strong, abbreviated as W, M, and S, respectively. The SM model had $\Sigma = \tau I_2$, where I_2

is the identity matrix of dimension 2, with τ chosen to be 4, 16, and 64 for the W, M, and S dependence, respectively. Observations at two sites with distance over $2\tau^{1/2}$ would be close to independent. In our study region, these choices correspond to the cases where two sites are almost independent if their distance exceeds 4, 8, and 16, respectively. The SC model had a Gaussian correlation function $\rho(h) = \exp[-(\|h\|/\phi)^2]$ with range parameter $\phi > 0$ chosen such that the resulting bivariate extremal coefficient function matches that from the SM model as close as possible. It is a special case of the power exponential correlation family with smooth parameter fixed at 2 as in the R package `SpatialExtremes` (Ribatet and Singleton, 2013). Through nonlinear least squares, the values of ϕ were tuned to be 2.942, 5.910 and 13.153 for W, M, and S dependence, respectively. For the GG model, a bigger δ^2 offers fuller range of dependence level for two sites, but the data generating function for this model in the R package `SpatialExtremes` works well only for $\delta^2 < 10$. As a compromise, δ^2 was fixed at 8. The GG model also had a Gaussian correlation structure, and similarly through nonlinear least squares, the range parameter was set to be 7.134, 14.780, and 31.149 for W, M, and S dependence, respectively. For the GA model, an exponential correlation function $\rho(h) = \exp[-(\|h\|/\phi)]$ was used with range parameter ϕ to be 6, 12, and 20, which were chosen so that the fitted exponential correlation curves of the empirical correlation of the score functions of μ and ξ are close to those of SM model. Two levels of m were considered, $m \in \{10, 20\}$. When $m = 20$, 10 additional sites were generated and added to those sites used in the case of $m = 10$. Finally, two levels of n,

$n \in \{10, 25\}$, were considered. This design led to 48 scenarios.

For each scenario, we generated 1000 datasets. For each dataset, we estimated the GEV parameters and T-year return level Q_T for $T \in \{50, 100, 500\}$. The three methods, L-moment, independence likelihood, and pairwise likelihood, are abbreviated as LM, IL, and PL, respectively. In optimization, the IL estimator used the LM estimators as starting values, and the PL estimator used the IL estimators as starting values. Given the large number of parameters in the model, we maximized the likelihood with an iterative procedure that maximizes the objective function with respect to one parameter at a time while the other parameters are held constant. The procedure is iterated over all parameters until convergence. It was reported to give better estimation when the number of parameters is large (Blanchet and Davison, 2011). In contrast to the LM and IL method, the PL method needed to specify a dependence model, which may be correct or incorrect. We studied its performance under correct specification and misspecification of the dependence structure within the spatial extreme model and with non-extreme-value copula (GA model).

2.4.2 Results

The center site $[5, 5]$ which was presented in all replicates was used to do the comparison across various scenarios. As the results from the three spatial extreme models were similar, we use the GG model to represent the three extreme-value models. Since sample size $n = 10$ was too small for the likelihood methods to be numerically reliable, the results

for $n = 10$ were based on trimmed data where 2% from each tail were excluded in the summary. The results for $n = 25$ were stable and included all 1000 replicates. The PL method had correct specification of the dependence model for data generated from the extreme-value models (SM, SC, and GG). For data generated from the GA model, the PL method was obtained under the specification of a GG model with Gaussian correlation structure. For each method, we report the relative bias and the relative root mean squared error (RMSE) for the GEV parameters and return levels at the center point.

When PL Is Correctly Specified We first look at the results for data generated from extreme-value dependence models (Figure 1 for the GG model). The bias decreases for all methods as n goes from 10 to 25. At $n = 25$, the bias of the PL method is quite small, with the largest relative magnitude of 7.8%. The bias of the LM method, however, remains high, and it is bigger especially with stronger dependence level. More sites did not help, especially for strong dependence. The relative bias is 17.5% for ξ and 20.6% for Q_{500} under strong dependence level and $m = 20$. This behavior of the LM method is in contrast to the existing result that intersite dependence does not introduce bias in RFA (Hosking and Wallis, 1988). It may be explained by that the simulation here was done with data generated from max-stable processes, which ensures that all marginal copulas are extreme-value copulas. In existing studies, however, data were generated mostly with normal copulas, which is not an extreme-value copula.

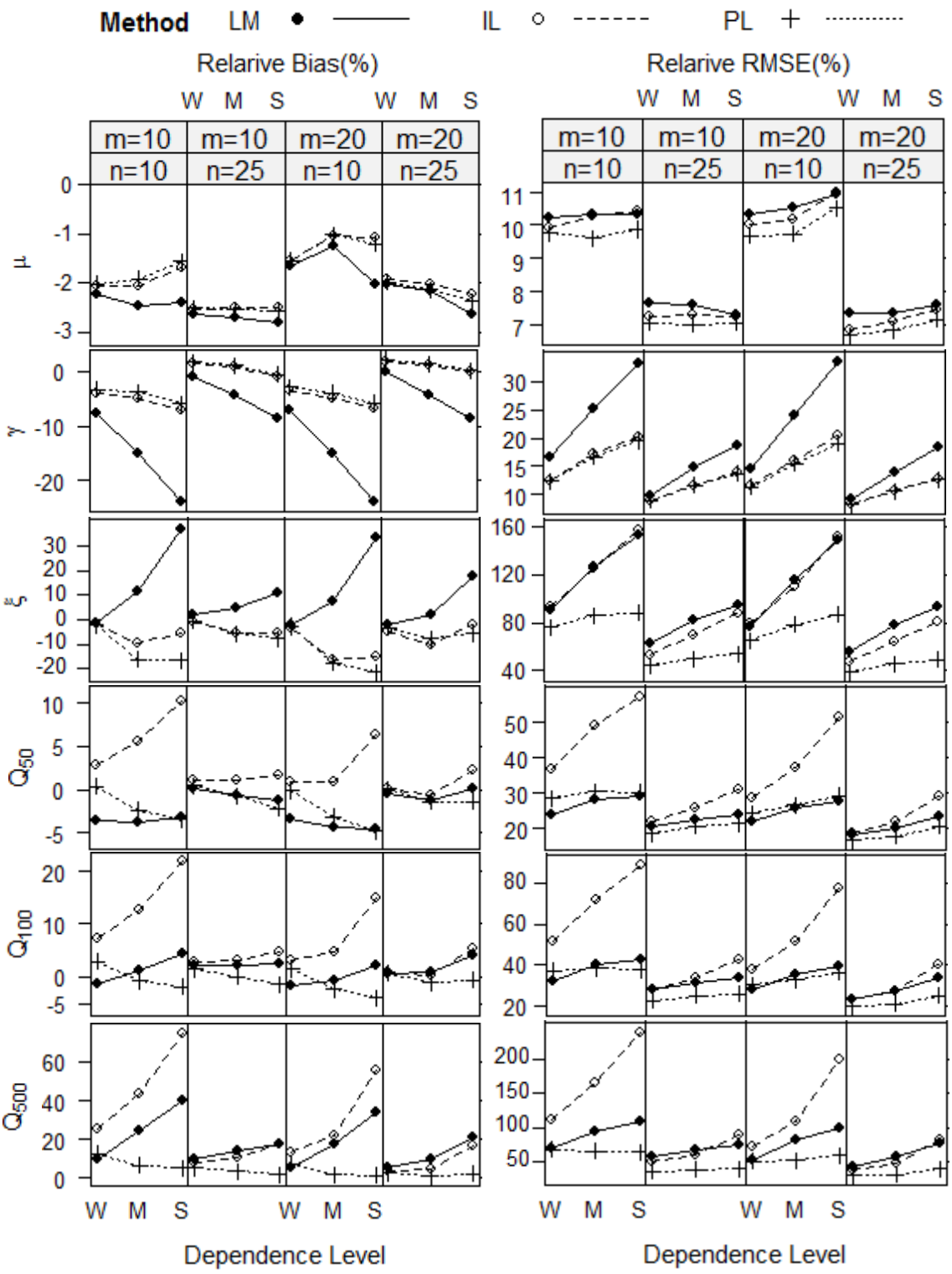

Figure 1: Relative bias (%) and relative RMSE (%) for three methods with data from the GG model.

The RMSE for all methods decreases as n increases as expected. Between $m = 10$ and $m = 20$, little difference was observed in RMSE. We have also tried $m = 5$ (not reported here) and found that the relative RMSEs did decrease when m increased from 5 to 10. This indicates that increasing the number of sites helps increase the efficiency for smaller m, but only up to a certain point, an observation consistent with the findings in Hosking and Wallis (1988). As the dependence gets stronger, the RMSE increases for all methods, but the magnitude of the change is the smallest for the PL method. This is because under correct specification, the PL method incorporates spatial dependence in the estimation while the other two methods do not. For both sample sizes, the PL method is a clear winner among the three. The comparison between the LM method and the IL method is mixed. For $n = 25$, the LM method is less efficient in parameter estimates but more efficient in some return level estimates than the IL method. This is possible because the return levels are non-linear transformations of the parameters; see equation (2.1). For $n = 10$, the LM performs better than the IL method in most of the return level estimates, even though it is less efficient in estimating γ and comparable in estimating μ and ξ. Further investigation revealed that the variance of the LM estimator is much smaller than that of the IL estimator, which compensates the larger bias (especially in ξ) of the LM estimator. This makes sense since it is known that L-moment estimator has a restriction $\xi < 1$, which makes it more efficient than the likelihood method for small samples when the restriction is true. In an earlier version under a slightly different simulation design, we had $n \in \{50, 100\}$ (not reported here)

and found that for large sample sizes, the efficiency order was PL, IL, and LM from the highest to the lowest. Among the three GEV parameters, the efficiency gain of the PL method relative to the IL method was always the greatest for the shape parameter ξ (in one case it was as large as 2.76); the RE for the μ and γ were close to 1. That is, the efficiency gain of the PL method is mostly realized in ξ, which controls the tail behavior of the GEV distribution. This leads to the efficiency gain of the PL method in estimating the return levels, especially for longer return periods such as 500 year.

Misspecification of PL within Spatial Extreme Models The efficiency gain in the PL method relative to the IL method comes with a cost: one needs to specify the spatial dependence model. We first look at the results for cases where misspecification is within the class of spatial extreme models; that is, one max-stable model is misspecified as another max-stable model. Figure 2 summarizes the relative efficiency (RE) in mean squared error of the PL method in estimating the GEV parameters and return levels, using the IL method as the reference (RE is the ratio of MSE of IL over MSE of PL), under both correct specification and misspecification within spatial extreme models for $n = 25$ and $m = 10$. When a SM model was misspecified as a GG model, or vice versa, the resulting estimator was almost as efficient as that under correct specification, and was much more efficient than the IL estimator. This is because the SM model and the GG model are very similar models, as evident from their similar bivariate distributions given in Section 2.3.2. When a SM or GG model was misspecified as a SC model,

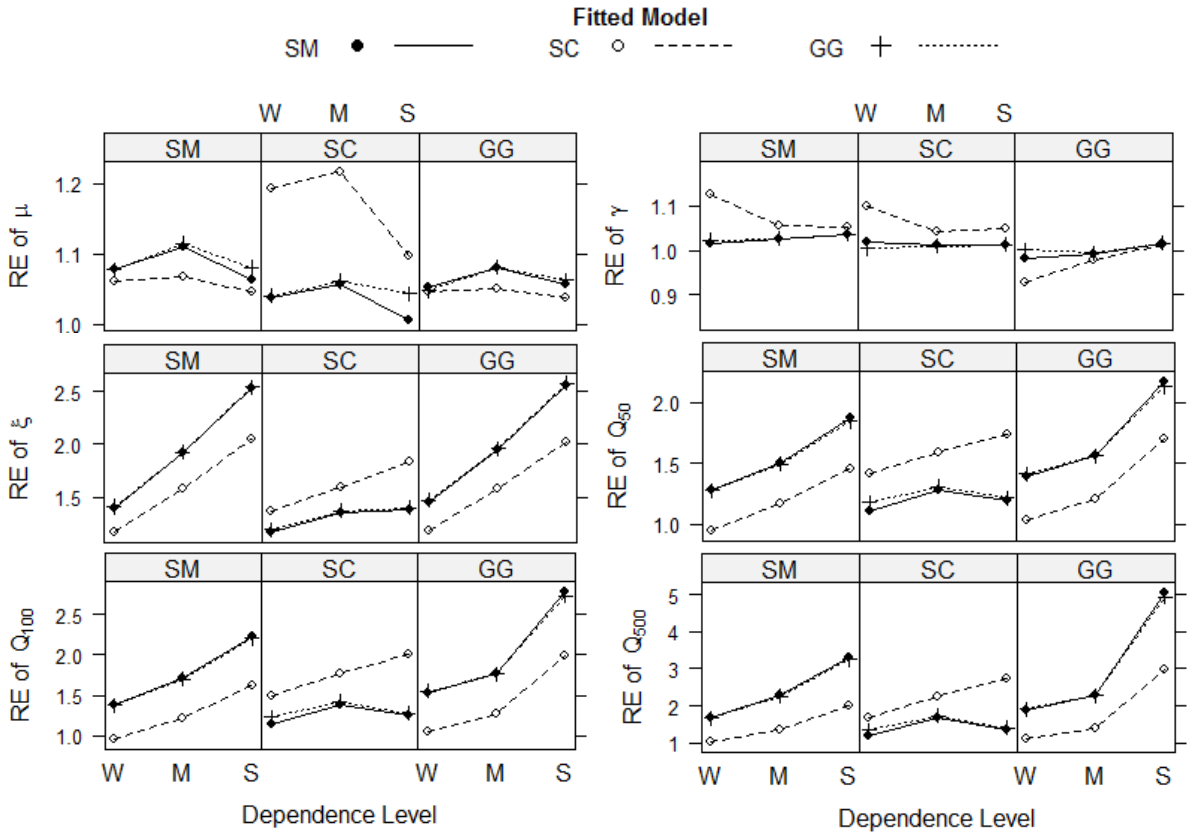

Figure 2: Relative efficiency (RE) of PL method (with the IL method as reference) under correct specification and misspecification within the class of extreme-value dependence models with $n = 25$ and $m = 10$. The grouped variable is the model that generated the data, and the line in each panel represents the corresponding fitted model.

the resulting PL estimator is comparable with the IL estimator for weak dependence, but more efficient than the IL estimator for moderate or strong dependence. This is as expected, because in contrast to the other two models, the SC model does not provide full range of dependence and, therefore, cannot accommodate weak dependence or close to independence. When a SC model was misspecified as a SM or GG model, the resulting PL estimator remained competitive compared to the IL estimator, especially for the cases with stronger spatial dependence. Among the three parametric models, both the SM model and GG model offer full range of dependence level and high efficiency under misspecification, but since the SM model gives too regular shapes of extreme observations to be observed in practice (e.g., Schlather, 2002), we recommend using the GG model. This is also why we presented the results for data generated from the GG model only in the main text.

Misspecification of PL under Non-Extreme-Value Model What if the true dependence model is a non-extreme-value copula but we fit an extreme-value dependence model? Figure 3 summarizes the results for data generated from the GA model, with the PL method specified under a GG model. The misspecified PL estimator has small bias in μ and γ, but large bias in ξ (as high as 50% for $m = 20$ and $n = 25$), which is much larger than that of the LM estimator or IL estimator. The bias increases as the dependence level gets stronger, and having more sites do not help. This large bias played a major part in the RMSE of the PL estimator; the RE for ξ and all return levels

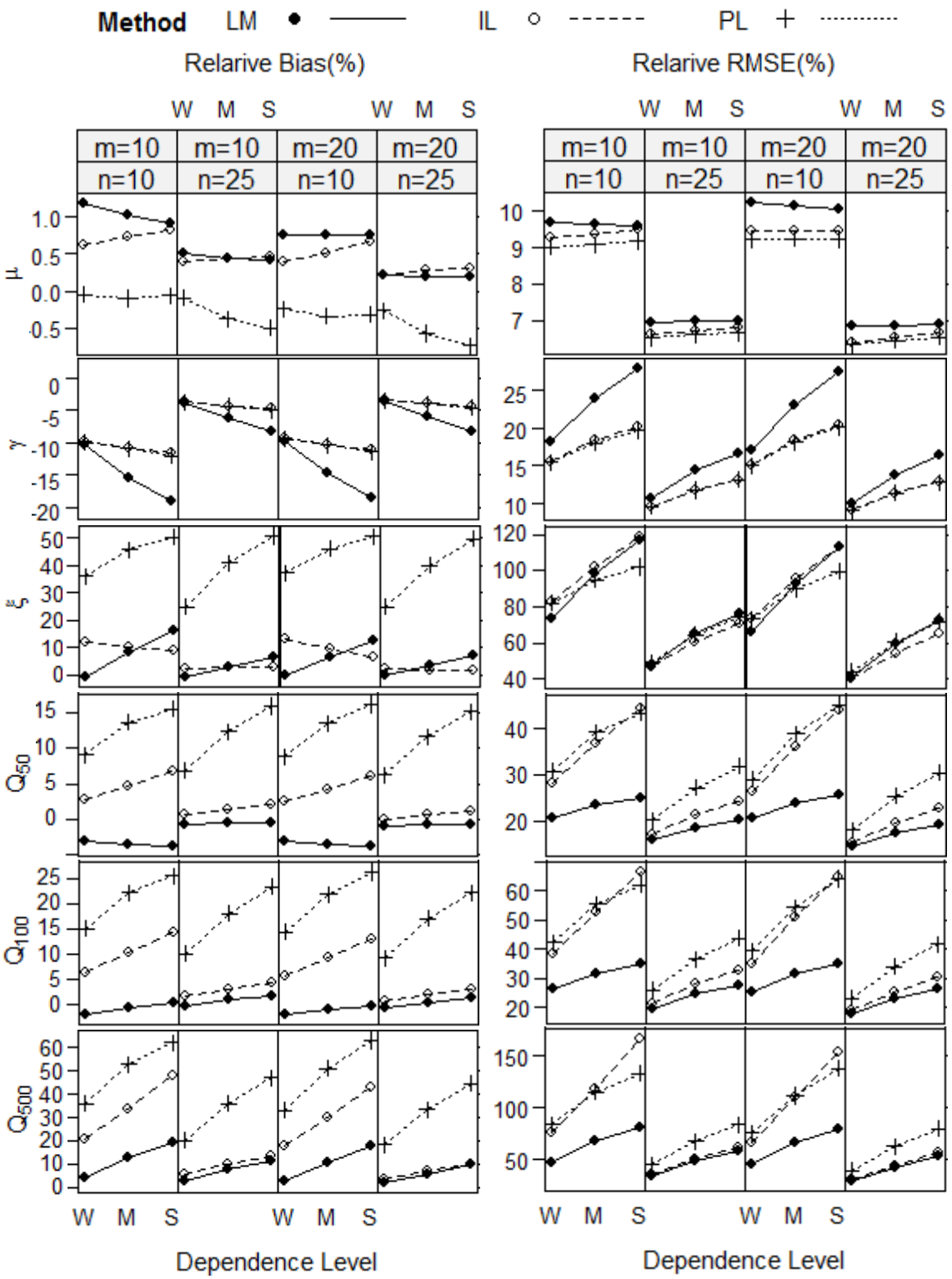

Figure 3: Relative Bias (%) and relative RMSE (%) for three methods with data from the GA model. The PL method using a GG model specification.

are as small as 0.51. The LM estimator is better than the IL estimator for all return levels. The performance of the LM method is similar to what was reported in Hosking and Wallis (1988) — its relative bias is only alarmingly noticeable (9.8% for Q_{500}) under the strong dependence level. This is reasonable because this scenario is the closest to the data generation scheme of Hosking and Wallis (1988), where a relative bias up to 5% for Q_{1000} was reported. Therefore, the spatial dependence does not affect the LM estimator under normal copula as much as it does under extreme-value copulas. For instance, the bias of the LM estimator in estimating Q_{500} with data from the GG model is about twice as much as that with data from the GA model: 33.6% v.s. 17.5% when $n = 10$ and 20.6% v.s. 9.8% when $n = 25$.

In summary, incorporating spatial dependence in the index flood model through max-stable processes may improve the efficiency of RFA, but at the cost of having to specify the dependence model. The LM method and the IL method do need to do so, which makes them attractive when no evidence supports extreme-value dependence. Misspecification of the PL method can lead undesired large bias. If, however, extreme-value copulas are known to correctly specify the dependence structure or provide adequate fit to the data through a goodness-of-fit test, then the PL method may be preferred by exploiting the dependence structure to give more efficient RFA. Essentially, it is still a story of bias-variance trade-off. To reap the potential efficiency gain in practice, one must check the goodness-of-fit of max-stable processes, which has been studied recently (Kojadinovic et al., 2015), in addition to the goodness-of-fit tests for the marginal GEV

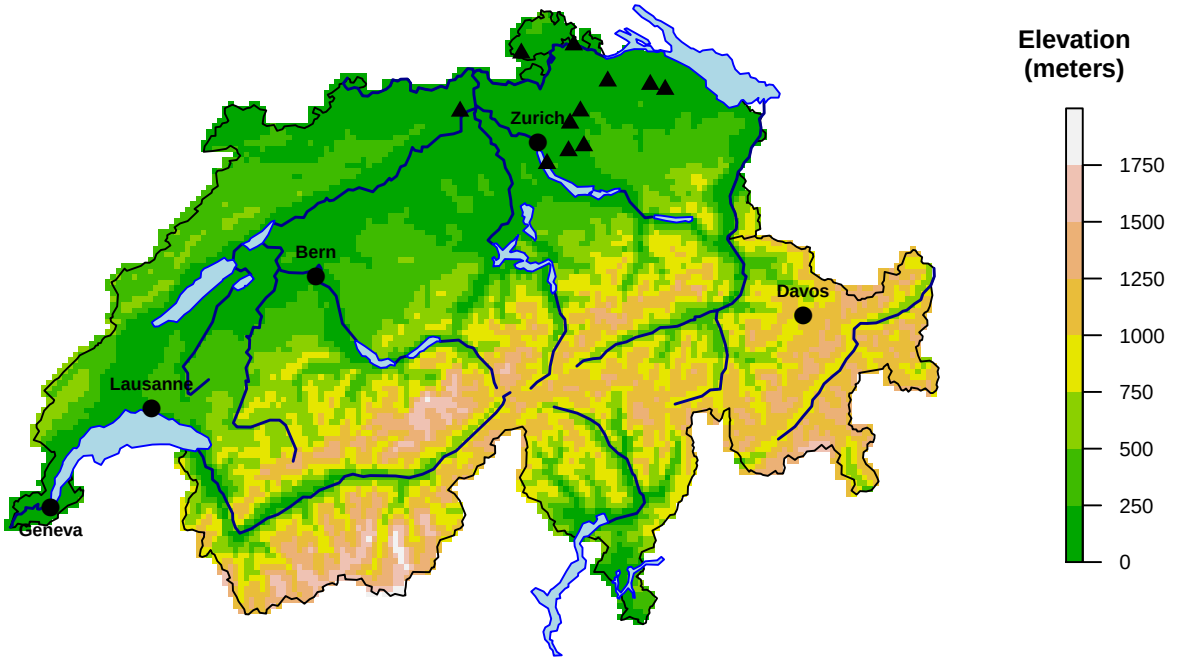

Figure 4: Elevation map of Switzerland with the 11 stations that were used in the Swiss rainfall analysis. The 11 stations are marked by triangles, and the dots represent cities in Switzerland.

models and the homogeneity assumption on the GEV parameters.

2.5 Illustration

For illustration, we applied the index flood model with all three estimation methods to the Swiss rainfall data that has been analyzed by many authors in modeling spatial extremes (e.g., Davison et al., 2012). The data consist of summer maximum daily precipitation (mm) for 51 stations over the years of 1962–2008 in the Plateau region of Switzerland; it is available in the R package `SpatialExtremes` (Ribatet and Singleton,

2013). To make a more realistic RFA, we used only the last 25 years of data from 1984 to 2008 ($n = 25$). We further filtered the sites in attempt to enhance the chance that the PL method gives reliable inferences. As we pointed in the discussion of the simulation study, the higher efficiency of the PL method is only achievable when the bivariate marginal distributions are correctly specified; otherwise, the PL method could lead to serious bias. The flood-index model we considered assumes that all the marginal distributions are GEV distributions with the same shape parameter and the same ratio of location parameter and scale parameter. These assumptions are shared by all three methods, but the PL method assumes additionally that the dependence structure can be captured by a max-stable (e.g., a geometric Gaussian) process. The goodness-of-fit of the geometric Gaussian process on this data has been checked graphically (Davison et al., 2012). A formal goodness-of-fit test for max-stable process models was not rejected for this data (Kojadinovic et al., 2015). The test, however, was applied on the whole region globally, and may have low power in detecting local lack-of-fit. Therefore, we further applied the test for bivariate extreme-value dependence (Kojadinovic et al., 2011) on all the pairs.

We filtered the sites by three tests on the model assumptions: 1) the goodness-of-fit of univariate GEV distribution was not rejected at any single site by a Kolmogorov–Smirnov test; 2) the homogeneity hypothesis (σ_s/μ_s and ξ_s are both constant) was not rejected by a nonparametric bootstrap test procedure which preserves the spatial dependence (Heffernan and Tawn, 2004); and 3) the hypothesis of bivariate extreme-value

dependence was not rejected for any pair of the sites by a nonparametric test proposed by Kojadinovic et al. (2011). Note that the first two are needed by all three methods, but the other one is only needed by the PL method. The 3rd test was an additional measure on model specification check given that the geometric Gaussian process has been known to fit this data well (Davison et al., 2012; Kojadinovic et al., 2015). For a different dataset, it will be necessary to run model diagnosis and global goodness-of-fit test too. This process ended up with 11 sites; see map in Figure 4.

We fitted the index flood model with marginal GEV distribution to the 25-year data of the 11 sites with all three methods. The PL method was carried out under the same GG model that was used in the simulation study; that is, it had a Gaussian correlation function with a single range parameter and $\delta^2 = 8$ was fixed. The standard errors of all the parameter estimates were obtained with a spatial-dependence-preserving bootstrap procedure (Heffernan and Tawn, 2004) with 1000 bootstrap samples.

Figure 5 summarizes the point estimate and 95% confidence interval (CI) for each parameter and three site-specific return levels (50, 100, and 500 year) from the three methods. The bounds of the 95% CI were the 2.5% and 97.5% percentiles of the 1000 bootstrap estimates, respectively. The point estimates from the three methods are similar for μ_s's, but quite different for γ and ξ. For γ, the estimates are 0.331 (s.e. 0.021), 0.392 (s.e. 0.018), and 0.389 (s.e. 0.018) for LM, IL, and PL, respectively. For ξ, the estimates are 0.345 (s.e. 0.063), 0.212 (s.e. 0.053), and 0.148 (s.e. 0.054) for LM, IL, and PL, respectively. The shape parameter is estimated to be significantly different from

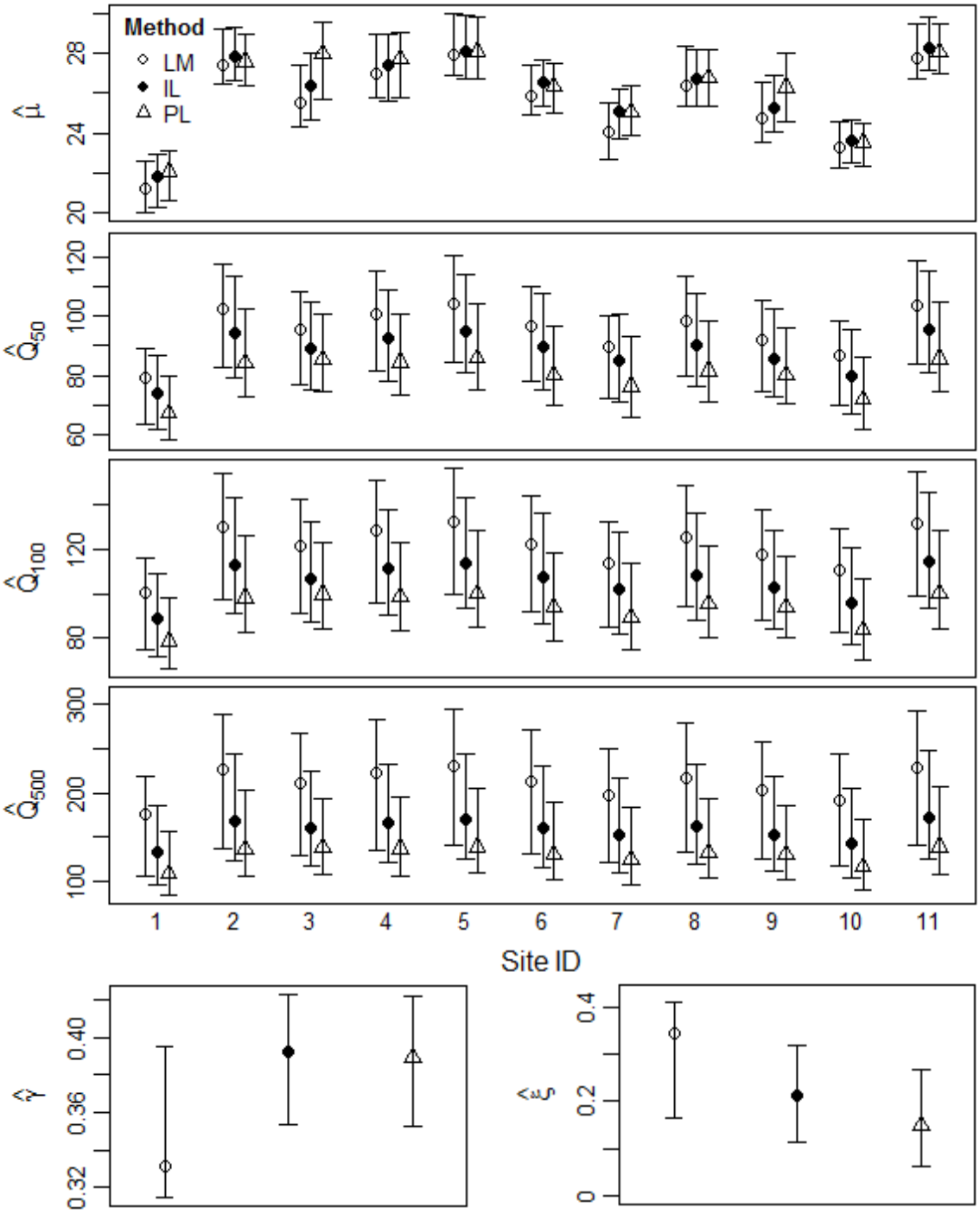

Figure 5: Estimated parameters and return levels (in mm) along with their 95% confidence intervals from the bootstrap procedure for the Swiss rainfall data. The PL method used a geometric Gaussian model with a Gaussian correlation function for the spatial dependence.

zero regardless of the method, suggesting the tails of the GEV distributions are heavier than the tail of the Gumbel distribution. The differences have a drastic effect on return level estimates. Consider for example the 11th site. The estimates of Q_{500} for this site are 228.39 (s.e. 40.17), 171.28 (s.e. 30.44), and 139.27 (s.e. 24.52), respectively, from the LM method, the IL method, and the PL method. The reduction in the standard error of the PL method is remarkable. The standard error of Q_{500} from the PL method is about 40% smaller than that from the LM method, and 20% smaller than that from the IL method. The bounds of the 95% bootstrap CIs are asymmetric around the point estimates, which are most notable for γ and ξ. The asymmetry appears to be in opposite direction for the LM method and the PL method, making the overlaps of the CIs to be bigger than those symmetric CIs constructed from the bootstrap standard errors. The CIs for the return levels have the shortest length from the PL method, followed by the IL method and then the LM method. We emphasize that the short CIs from the PL method does come at a cost — we had to specify the dependence model with a geometric Gaussian process with a Gaussian correlation structure. The estimate of the range parameter in the dependence model is $\hat{\phi} = 2377.44$ (s.e. 92.31). We have gone through the extra steps in filter the sites to be included in the analysis beyond the known model checking in the literature on this data, which turned out to be worth it.

The systematically lower return level estimates from the PL method than those from the LM method may be explained from two perspectives. First, the return level is a function of all three parameters of the GEV distribution but is most sensitive to

the shape parameter. The location parameter estimates are similar across different methods. The scale parameter estimates and the shape parameter estimates, however, tend to compensate each other: higher scale parameter estimate is accompanied with lower shape parameter estimate. The LM method has lower $\hat{\gamma}$, hence lower scale estimate, and higher $\hat{\xi}$, which led to the higher return level estimates. Second, the shape parameter seems to have the most room for efficiency improvement as seen in the simulation (Figure 2 in the manuscript). The efficiency gain of the PL method, assuming that the PL is correctly specified, is only on average if replicates were available. For a single dataset, the truth is unknown, and the lower point estimate of the shape parameter from the PL method than that from the IL method is quite likely, noting that the 95% bootstrap confidence intervals from the two methods overlap by large.

The analysis so far is based on the last 25 years data and the availability of the whole 47 years data enabled us to compare the performance of the three methods more thoroughly in other ways. We randomly selected 100 subsets of 25-year data, and for each subset we ran the same analysis as we did for the last 25 years. The same analysis was also repeated on the full 47 years of data. Table 1 summarizes results from these analyses. Site specific estimates are only presented for two sites with the smallest or the largest $\hat{\mu}_s$ from the full data analysis. Point estimates and bootstrap standard errors are reported for the full data analysis. For the 100 subset analyses, we reported the average of the point estimates, the average of the bootstrap standard errors, and the standard deviation of the point estimates. It is reassuring that the point estimates from

Table 1: Point estimate and bootstrap standard error for the full 47 years data analysis (abbreviated as Full), the average point estimate and average bootstrap standard error based on 100 subsets of 25 years (abbreviated as Ave), and the standard deviation of the 100 point estimates.

	Point Estimate						Standard Error						Standard Error of 100		
	LM		IL		PL		LM		IL		PL		point estimates		
	Full	Ave	Full	Ave	Full	Ave	Full	Ave	Full	Ave	Full	Ave	LM	IL	PL
μ	20.5	20.7	20.4	20.5	20.5	20.6	0.43	0.55	0.42	0.55	0.41	0.55	1.31	1.28	1.26
	26.5	26.6	26.5	26.6	26.4	26.5	0.49	0.65	0.46	0.61	0.45	0.60	1.70	1.61	1.53
Q_{50}	68.9	67.8	66.0	65.4	61.6	61.3	4.4	6.1	4.0	5.7	3.6	5.3	12.4	12.2	9.2
	88.9	87.3	85.7	84.8	79.2	79.0	5.8	8.1	5.4	7.7	4.9	7.1	16.8	15.4	11.6
Q_{100}	84.4	83.7	79.3	78.7	72.5	72.2	6.9	10.1	6.0	8.8	5.4	8.0	19.7	17.6	12.6
	108.9	107.8	102.8	102.1	93.2	93.0	9.2	13.2	8.1	11.8	7.2	10.6	26.3	22.5	16.1
Q_{500}	133.9	137.6	119.0	119.8	103.3	103.8	17.8	29.2	14.1	21.6	12.1	18.7	51.5	38.4	24.8
	172.8	177.5	154.4	155.3	132.9	133.7	23.5	38.3	18.8	28.6	16.0	24.7	68.0	49.3	31.9
γ	0.338	0.327	0.358	0.353	0.355	0.351	0.014	0.021	0.012	0.017	0.012	0.017	0.030	0.026	0.026
ξ	0.274	0.256	0.223	0.205	0.178	0.168	0.045	0.062	0.037	0.053	0.038	0.053	0.133	0.097	0.071

the full data analysis are very close to the average of those from the 100 subset analyses for all parameters and all three methods. For the estimates of the shape parameter and the return levels, the PL method always has the smallest standard error while the LM method always has the largest standard error, regardless of the full data analysis or the average of the subset analyses. This is consistent to the results from the last 25 years of data. The standard deviation of the 100 subset point estimates of the shape parameter and the return levels has the same pattern, but the difference in the magnitudes is even more obvious. For instance, the standard deviation of the 100 subset estimates Q_{500} is 68.0, 49.3, and 31.9 for LM, IL, and PL, respectively.

Another interesting finding is about the ratio of the standard errors from the full data analysis and those from the subset analyses. More data is associated with smaller standard errors in theory. For most parameter estimates and all methods, the ratios of the standard error from the full 47 years data to that from the average of the 25 years subsets are close to $\sqrt{25/47} \approx 0.73$. This is expected to happen for large sample, which suggests that the large sample results approximate the finite sample results quite well even for sample size of 25 in this data analysis. We also studied the ratios of the 95% bootstrap CI lengths, most of which were higher than 0.73 due to the asymmetry in the CIs.

2.6 Discussion

This paper explores the idea of incorporating intersite dependence in RFA with index flood models to improve the efficiency in estimation. The efficiency gain comes at the cost of having to specifying the dependence model in addition to the usual specifications such as marginal distributions and the regional homogeneity assumption. When the dependence model is correctly specified, smaller standard errors and narrower confidence intervals can be obtained for model parameter and return levels. Misspecification of the dependence model, however, may result in serious bias, especially when the true dependence model is not of extreme-value type and the dependence is strong. This makes it important to check the goodness-of-fit for the dependence structure, in addition to the usual check for marginal goodness-of-fit and regional homogeneity, to reap the efficiency gain. The L-moment method and the independence method may sometimes be preferable because they have no need for dependence model specification. The L-moment method implicitly constrains the shape parameter to be less than 1 for the existence of the L-moments, which gives efficient estimator for small samples when the constraint does hold. As extreme-value copula can be very different from non-extreme-value copula (e.g., normal copula), it has more effect on the bias of the L-moment estimator than does the normal copula for typical record lengths in RFA, a result that has not previously been reported. The independence likelihood method may have unreasonable estimates in small samples unless it is modified to impose a similar constraint, but for large samples,

it is more efficient than the L-moment method.

Spatial dependence in RFA is often a nuisance because the goal of an RFA is usually to estimate marginal return levels. With marginal GEV distributions, it is desirable to improve the efficiency without specifying a spatial dependence model. Specification and selection of a working dependence model can be avoided by a combined estimating equation approach based on data contrasts for clustered data (Stoner and Leroux, 2002). For spatial data, estimating equation approaches have also been applied for marginal models with no need to correctly specify the dependence structure (Yasui and Lele, 1997; Clayton and Lin, 2005; Lin, 2008, 2010). The marginal score equations can be combined in certain way to improve the efficiency (Nikoloulopoulos et al., 2011). In next chapter, we propose a combined score equations method that exploits spatial dependence to increase the efficiency without dependence assumptions beyond the univariate GEV distribution.

Chapter 3

Combined Score Equations

3.1 Introduction

The simulation results in Section 2.4 of Chapter 2 motivates us to find another way to account for the spatial dependence without requiring parametric specifications beyond univariate marginal GEV distributions when the primary interest is the inference about the marginal GEV distributions while the spatial dependence is a nuisance. We propose a combined score equations (CSE) approach which provides a general methodology for extreme value analysis in a spatial setting that exploits spatial dependence without dependence assumptions beyond the univariate GEV distribution. At each site, the marginal GEV distribution leads to a score equation, which can be combined in some optimal way to improve efficiency by accounting the spatial correlation among them. The idea is similar to the generalized estimating equations (GEE) (Liang and Zeger, 1986), which remains an active development area (see recent reviews and discussions, e.g., Sabo and Chaganty, 2010; Lee and Nelder, 2009; Chaganty and Joe, 2004; Lindsey and Lambert, 1998). The difference, however, is that the GEE only specifies marginal means while the CSE specifies marginal distributions. Estimating equations can be combined

for better efficiency in various ways (Qu et al., 2000; Stoner and Leroux, 2002). Bai et al. (2012) proposed joint composite estimating functions in spatiotemporal applications. In the longitudinal setting, Nikoloulopoulos et al. (2011) proposed a similar idea called weighted scores method, with the optimal weight matrix estimated based on a working normal copula. Our approach is a combined scores method for spatial extremes. A simple working correlation structure will be used to model the spatial correlation among the site-wise score functions.

The rest of this chapter is structured as follows. The CSE method is proposed in Section 3.2. Section 3.3 presents large scale simulation studies to further investigate the properties of the CSE method in comparison with competing methods under various multivariate extreme distributions. A discussion is concluded in Section 3.4.

3.2 Combined Score Equations

Suppose that we have block maxima for a collection of sites in a region over a certain period of years. Let Y_{ts} be the extreme observation of interest at site s in year t, $s = 1, \ldots, m,\ t = 1, \ldots, n$. The method can be easily adapted to unequal record length. Suppose that at each site s, Y_{ts} has density $f(\cdot; \theta_{ts})$ with parameter θ_{ts}. For ease of presentation, assume for now that θ_{ts} is a scalar; the multiparameter case will be discussed later. Let X_{ts} be a $p \times 1$ covariate vector for θ_{ts}. The parameter θ_{ts} is connected to a linear predictor $\eta_{ts} = X_{ts}^{\top}\beta$ via a known link function g: $g(\theta_{ts}) = \eta_{ts}$. We assume

that from year to year, the data are independent, but within the same year, spatial dependence exists. We further assume that the marginal distribution f at each site s is correctly specified, but no distributional assumption is made beyond the univariate marginal distributions.

Since the univariate marginal distribution at each site is correctly specified, the corresponding score functions can be derived. For each $t \in \{1, \ldots, n\}$, $s \in \{1, \ldots, m\}$, let $S_{ts} = \mathrm{d} \log f(Y_{ts}; \theta_{ts})/\mathrm{d}\theta_{ts}$. Correct model specification means $\mathrm{E}(S_{ts}) = 0$. The score equation for β at site s is

$$\sum_{t=1}^{n} X_{ts} \frac{\mathrm{d}\theta_{ts}}{\mathrm{d}\eta_{ts}} S_{ts} = 0. \tag{3.1}$$

To estimate β using data from all sites, a first solution is to take the sum of the equations in (3.1) over s. This combination can be expressed as

$$\sum_{t=1}^{n} X_t^\top A_t W_t^{-1} S_t = 0, \tag{3.2}$$

where $X_t^\top = (x_{t1}, \ldots, x_{tm})$, $A_t = \mathrm{diag}(\mathrm{d}\theta_{t1}/\mathrm{d}\eta_{t1}, \ldots, \mathrm{d}\theta_{tm}/\mathrm{d}\eta_{tm})$, W_t^{-1} is the weight matrix, and $S_t = (S_{t1}, \ldots, S_{tm})^\top$. When W_t is the identity matrix, the combined score equations are the derivative of the independence likelihood. More efficient estimator can be obtained from the family of estimating equations in (3.2) for other choices of W_t.

As in Nikoloulopoulos et al. (2011), the optimal W_t has the form

$$W_t = \Omega_t \Delta_t^{-1}, \tag{3.3}$$

where $\Omega_t = \mathrm{cov}(S_t)$ and

$$\Delta_t = -\mathrm{diag}\left\{ \mathrm{E}\left(\frac{\mathrm{d}^2 \log f(y_{t1}; \theta_{t1})}{\mathrm{d}\theta_{t1}^2} \right), \ldots, \mathrm{E}\left(\frac{\mathrm{d}^2 \log f(y_{tm}; \theta_{tm})}{\mathrm{d}\theta_{tm}^2} \right) \right\}.$$

Under independence, W_t reduces to the identity matrix as $\Omega = \Delta$, and the estimating equations (3.2) reduce to the score equations of the independence likelihood. The estimating equations can be solved with R package `nleqslv` (Hasselman, 2014).

In modeling spatial extremes, the marginal distributions at all sites are assumed to be the GEV distribution. The density function f in (3.2) is the density of F obtained by differentiating F with respect to y. In practice, the GEV distribution provides good approximation to distributions of univariate extreme observations such as annual maximum daily precipitation and annual maximum flood level. For component-wise sample maximum of multivariate data, the limiting distribution is a multivariate GEV distribution, which requires the converge of the copula structure to a special class of copulas known as extreme value copulas, in addition to the convergence of all univariate marginal distributions to GEV distributions. The convergence of the multivariate dependence structure may occur at a different rate than the convergence of the marginal distributions (Ledford and Tawn, 1996, 1997; Resnick, 2002). Therefore, it is preferable to avoid modeling spatial dependence explicitly when the marginal regression coefficients are of main interest, as in the case of detection and attribution analysis.

For the moment, consider a situation where the scale and shape parameters are all

known, and the location parameter θ_{ts} is modeled through $g(\theta_{ts}) = X_{ts}^{\top}\beta$. The question is how to find the optimal weight matrix W_t. In particular, since Δ_t is known from the marginal GEV distribution (Prescott and Walden, 1980), we just need to approximate the covariance matrix Ω_t of the score functions S_t. Our proposal is to apply the idea of GEE with some simple form of working spatial correlation structure. Because the univariate marginal GEV distributions are correctly specified, we have $\text{var}(S_{ts}) = \Delta_{ts}$ for $\xi > -0.5$ (Smith, 1985). Suppose all the clusters share the same correlation matrix, R, of the score function, then Ω_t can be written as $\Omega_t = \Delta_t^{1/2} R \Delta_t^{1/2}$. We need to find a reasonable correlation structure to describe the spatial dependence within a year. As the spatial dependence between the observations from two sites decays as their distance increases, we considered simple one-parameter working spatial correlation functions: exponential, spherical and Gaussian. For a given initial value of the parameter, the score functions can be evaluated and their empirical correlation can be used to fit the working correlation structure through for instance, non-linear least squares.

The GEV distribution has three parameters, each of which may incorporate covariate effects through a link function. Let μ_{ts}, σ_{ts}, and ξ_{ts} be the location, scale, and shape parameters, respectively, for the GEV distribution of Y_{ts}. Let $X_{\mu,ts}$, $X_{\sigma,ts}$, and $X_{\xi,ts}$ be the covariate vector for μ_{ts}, σ_{ts}, and ξ_{ts}, respectively. The GEV parameters are connected to covariates through known link functions:

$$g_\mu(\mu_{ts}) = X_{\mu,ts}^{\top}\beta_\mu, \qquad g_\sigma(\sigma_{ts}) = X_{\sigma,ts}^{\top}\beta_\sigma, \qquad g_\xi(\xi_{ts}) = X_{\xi,ts}^{\top}\beta_\xi,$$

where g_i's are link functions and β_i's are regression coefficients for $i \in \{\mu, \sigma, \xi\}$. This formulation is very general, covering most of the need in parameterization of the GEV model in practice. For instance, if the shape parameters are assumed to be the same across all years and all sites, we simply have $X_{\xi,ts} = 1$ and β_ξ as a scalar. To estimate all the parameters $\Theta = (\beta_\mu^\top, \beta_\sigma^\top, \beta_\xi^\top)^\top$, we use a block coordinate descent algorithm cycling through all three components (e.g., Tseng, 2001; Tseng and Yun, 2009). That is, we iteratively estimate one set of the parameters at a time while the other two sets are being fixed until convergence. Therefore, for each set of parameter estimation, we need to estimate a working correlation function for the score functions evaluated at the current parameter values.

To get an idea about the shape of the true correlation structure, we approximated it with the Monte Carlo method under two data generating scenarios used in our simulation study. The first scenario is an isotropic Smith model with moderate dependence level; see Section 3.3 for details. The second scenario is a 50–50 mixture of a geometric Gaussian model and a Gaussian copula model with moderate dependence level. That is, the dependence model is an extreme value copula in the first scenario but not in the second scenario. In both scenarios, we generated 1000 years of data and plotted the empirical pairwise correlation versus distance for the score functions for μ, σ, and ξ evaluated at the true parameter values; see Figure 3.2. Also shown are the fitted correlation functions from the three families: exponential, spherical, and Gaussian.

From Figure 3.2, the Gaussian curves provided the least favorable fit to the empirical

correlation, overestimating for small distance while underestimating for large distance. The spherical function and the exponential function both fitted the empirical correlations quite closely. In fact, the estimators from the two working correlation structures were very close in terms of efficiency. Their difference is that the spherical correlation becomes exactly zero after certain threshold, while the exponential correlation gets close to zero but never attains it. This difference means that the working spherical correlation function leads to a sparse approximation matrix of R, which can be exploited in computational speeding up. Besides the three correlation functions, we have also tried approximating the correlation via smooth curves models through basis splines with a couple of knots. Our exploratory simulation study suggested, however, that the extra flexibility does not necessarily lead to more efficient estimator, possibly because more parameters are estimated and more variability are introduced in the weight. Therefore, in our simulation study, we used the exponential correlation to construct the shared correlation matrix R across clusters. That is, the correlation between site j and site k is $\rho_{jk} = \exp(-d_{jk}/r)$, where d_{jk} is the pairwise distance and r is the parameter to be estimated through the empirical correlation of the standardized score function.

3.3 Simulation Studies

The performance of the CSE method in practical settings is investigated through a simulation study which had a general setting with data generated from various multivariate

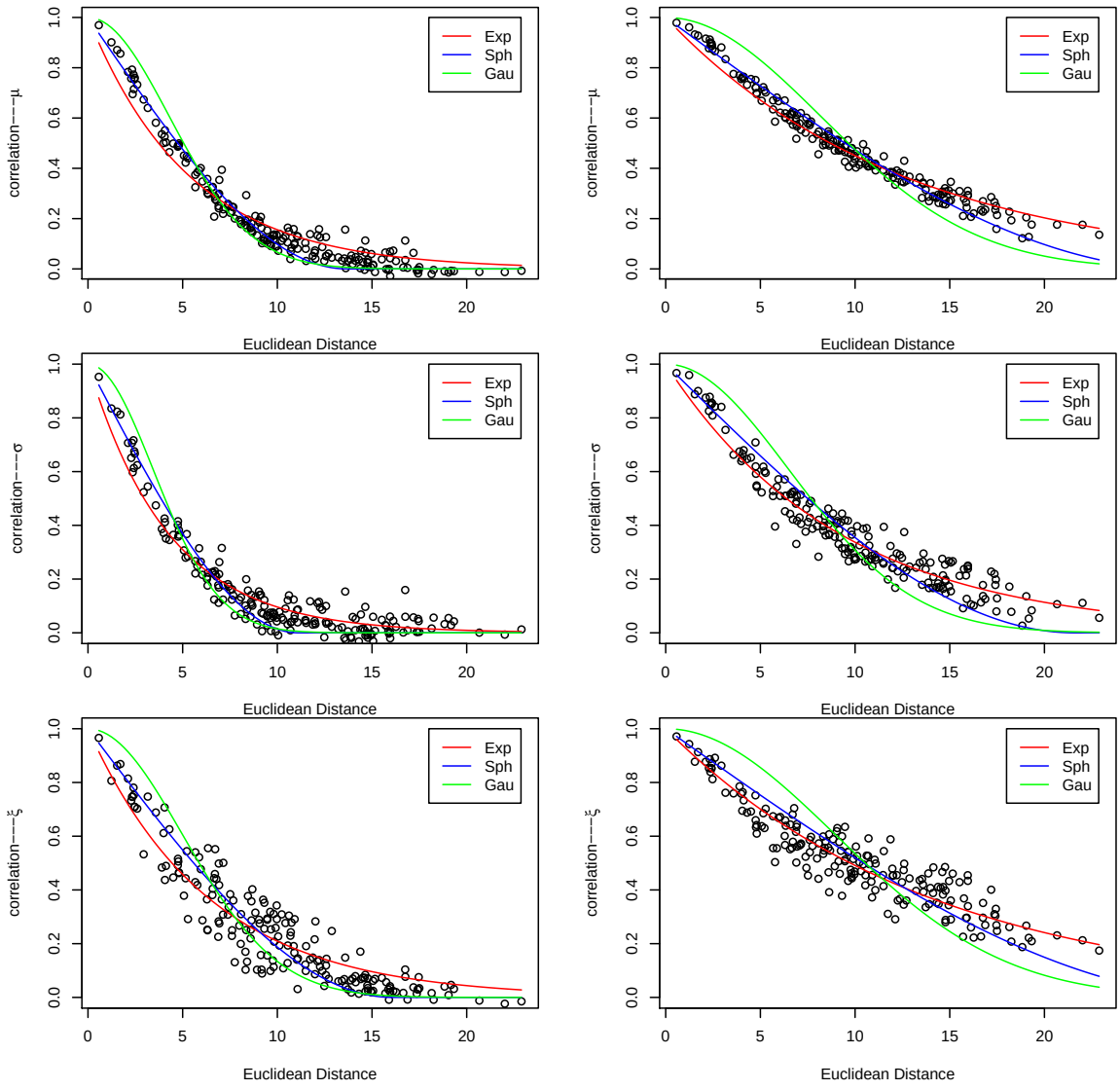

Figure 6: The empirical correlation of the standardized score function of μ, σ and ξ (points), and the corresponding non-linear least square fitted correlation curves from exponential (red), spherical (blue) and Gaussian (green) correlation function. Left: Data generated from an isotropic Smith max-stable process with $m = 20$, $n = 1000$, and moderate dependence level in region $[-10, 10]$. Right: Data generated from the mixture of a geometric Gaussian (50%) process and a multivariate normal model (50%) with $m = 20$, $n = 1000$, and strong dependence level in region $[-10, 10]$.

extreme distributions to compare the performance of three methods: CSE, IL, and PL.

Considered m sites uniformly generated over a square study region $[-c, c]^2$, with $c \in \{10, 20\}$. The length of the record was set at $n = 100$ years. The marginal model at each site s, $s = 1, \ldots, m$, was a GEV model with covariates latitude $X_1(s)$ and longitude $X_2(s)$:

$$\begin{cases} \mu_s = \beta_{\mu,0} + \beta_{\mu,1} X_1(s) + \beta_{\mu,2} X_2(s), \\ \sigma_s = \beta_{\sigma,0}, \\ \xi_s = \beta_{\xi,0}, \end{cases}$$

where $\beta_{\mu,0} = 15$, $\beta_{\mu,1} = -0.2$, $\beta_{\mu,2} = 0.25$, $\beta_{\sigma,0} = 4$ and $\beta_{\xi,0} = 0.2$. Data between different years were independent, while within the same year spatial dependence was imposed. Two data generating scenarios were considered for the spatial dependence: 1) an isotropic Smith (SM) max-stable model; and 2) a mixture of a geometric Gaussian (GG) max-stable model and a Gaussian copula (GA) model (See Section 2.3.1 for the detail of the SM model and GG model.) As the PL method depends on correct specification of the bivariate marginal distributions, the second scenario helps to assess its performance under misspecification.

3.3.1 Scenario 1: SM model

Three combinations of (c, m) were considered: $(10, 20)$, $(10, 80)$, and $(20, 80)$. Compared with the baseline $(c, m) = (10, 20)$, the scenario $(10, 80)$ contains more sites within the

same study region, while scenario $(20, 80)$ contains more sites in a larger study region with the same density. In other words, the number of sites m was increased by two ways: infill or domain expansion. Two levels of spatial dependence were considered, moderate and strong, abbreviated as M and S, respectively. The SM model had $\Sigma = \tau I_2$ in equation (2.6), where I_2 is the identity matrix of dimension 2, with τ chosen to 16 and 64 for the M and S dependence, respectively. This parameter determines the dependence level in a way that the two sites with distance larger than 2σ would be close to independent. For each setting, 1000 datasets were generated, each of which used a freshly generated set of m points. For the PL method, correct specification of the spatial dependence was used, while no dependence specification was needed for either the IL or the CSE method. The IL estimates were used as the starting values in the PL method and the CSE method. The CSE was solved by R package `nleqslv` (Hasselman, 2014). For each method, the average point estimate, root mean square error (RMSE), and the relative efficiency (RE) in terms of MSE using the IL estimate as the reference were summarized; see Table 2.

All three estimators appear to be unbiased in all settings. Comparison of MSE amounts to comparison among their variances. Increasing the number of sites in the same study region helps to reduce the variation, but the effect is much less compared to that from expanding the study region with the same site density, because the latter implies a larger effective sample size. The RE of the CSE estimator ranges from 1.152 to 1.477, indicating a clear advantage of the CSE estimator over the IL estimator. The

Table 2: Summaries of the simulation results for Scenario 1 with data from SM model. The relative efficiency (RE) was based on the MSE, with the IL estimate as reference.

					Estimate			RMSE			RE	
Dep	c	m	Par	True	IL	PL	CSE	IL	PL	CSE	PL	CSE
M	10	20	$\beta_{\mu,0}$	15	15.009	15.008	15.010	0.25	0.24	0.23	1.02	1.18
			$\beta_{\mu,1}$	−0.2	−0.200	−0.200	−0.200	0.02	0.02	0.02	1.01	1.15
			$\beta_{\mu,2}$	0.25	0.250	0.250	0.251	0.02	0.02	0.02	1.03	1.23
			$\beta_{\sigma,0}$	4	3.982	3.982	3.982	0.20	0.19	0.18	1.02	1.20
			$\beta_{\xi,0}$	0.2	0.199	0.199	0.198	0.04	0.04	0.04	1.32	1.18
	10	80	$\beta_{\mu,0}$	15	15.024	15.022	15.022	0.24	0.24	0.21	1.03	1.35
			$\beta_{\mu,1}$	−0.2	−0.200	−0.200	−0.200	0.02	0.02	0.02	1.02	1.44
			$\beta_{\mu,2}$	0.25	0.250	0.250	0.250	0.02	0.02	0.02	1.02	1.46
			$\beta_{\sigma,0}$	4	3.993	3.993	3.993	0.18	0.18	0.16	1.01	1.38
			$\beta_{\xi,0}$	0.2	0.196	0.197	0.196	0.04	0.04	0.04	1.34	1.37
	20	80	$\beta_{\mu,0}$	15	15.008	15.007	15.010	0.15	0.15	0.14	1.01	1.17
			$\beta_{\mu,1}$	−0.2	−0.200	−0.200	−0.200	0.01	0.01	0.01	1.03	1.31
			$\beta_{\mu,2}$	0.25	0.250	0.250	0.250	0.01	0.01	0.01	1.03	1.30
			$\beta_{\sigma,0}$	4	4.004	4.004	4.005	0.12	0.12	0.11	1.01	1.21
			$\beta_{\xi,0}$	0.2	0.199	0.199	0.199	0.03	0.02	0.02	1.10	1.22
S	10	20	$\beta_{\mu,0}$	15	15.034	15.032	15.030	0.35	0.35	0.32	1.04	1.24
			$\beta_{\mu,1}$	−0.2	−0.200	−0.200	−0.200	0.03	0.02	0.02	1.06	1.25
			$\beta_{\mu,2}$	0.25	0.248	0.248	0.249	0.02	0.02	0.02	1.04	1.23
			$\beta_{\sigma,0}$	4	4.001	4.002	3.998	0.27	0.27	0.24	1.00	1.19
			$\beta_{\xi,0}$	0.2	0.198	0.199	0.199	0.06	0.04	0.06	1.80	1.20
	10	80	$\beta_{\mu,0}$	15	15.041	15.038	15.028	0.35	0.34	0.30	1.06	1.42
			$\beta_{\mu,1}$	−0.2	−0.199	−0.199	−0.199	0.02	0.02	0.02	1.01	1.44
			$\beta_{\mu,2}$	0.25	0.250	0.250	0.250	0.02	0.02	0.02	1.01	1.48
			$\beta_{\sigma,0}$	4	3.993	3.994	3.983	0.26	0.26	0.22	1.01	1.36
			$\beta_{\xi,0}$	0.2	0.196	0.197	0.199	0.06	0.04	0.05	1.84	1.30
	20	80	$\beta_{\mu,0}$	15	15.024	15.022	15.022	0.24	0.24	0.21	1.03	1.35
			$\beta_{\mu,1}$	−0.2	−0.200	−0.200	−0.200	0.01	0.01	0.01	1.02	1.44
			$\beta_{\mu,2}$	0.25	0.250	0.250	0.250	0.01	0.01	0.01	1.02	1.46
			$\beta_{\sigma,0}$	4	3.993	3.992	3.993	0.18	0.18	0.16	1.01	1.38
			$\beta_{\xi,0}$	0.2	0.196	0.197	0.196	0.04	0.04	0.04	1.34	1.37

efficiency gain increases as the dependence level increases. The RE of the PL estimator ranges from 1.012 to 1.804, and as expected, more gains are observed under stronger dependence. A closer look reveals that most of the gains of the PL method are on the shape parameter, provided its correct dependence specification (which may not be realistic in practice); the RE for all other parameters has a maximum of 1.062, only a minimal improvement over the IL method. Since the shape parameter determines the tail behavior, it largely affects the estimating efficiency of return levels. In Table 2, the CSE method performs better in estimating ξ with $(c, m) \in \{(10, 80), (20, 80)\}$ under moderate dependence, and $(c, m) = (20, 80)$ under strong dependence. The highest RE of the PL method for ξ, 1.844, was obtained under strong dependence with $(c, m) = (10, 80)$. Nevertheless, with $(c, m) = (20, 80)$, the CSE is more efficient than or as efficient as the PL method under moderate or strong dependence. This suggests that for more sites in a larger region, the CSE method has the potential to be more efficient. Given that the PL estimator relies on correct specification of the dependence model, it only serves as a benchmark in this comparison in that in practice the correct specification may not be possible. The CSE estimator provides a quite impressive solution without the need for dependence model specification. The RE of the CSE estimator is larger than that of PL in estimating the location and scale parameters in all cases, and the difference can be as large as 0.4, i.e. 40% efficiency gain compared to PL, in some cases. This is an interesting finding since even when the PL method had the correct model specification, the CSE method can still be better. For example, with $(c, m) = (10, 80)$

under moderate dependence, the RE for $(\beta_{\mu,0}, \beta_{\mu,1}, \beta_{\mu,2}, \beta_{\sigma,0})$ is $(1.030, 1.023, 1.024, 1.015)$ for the PL method and $(1.350, 1.438, 1.464, 1.379)$ for the CSE method, a 30% to 40% efficiency gain relative to PL. This can be very useful in fingerprinting changes in climate extremes as the fingerprints are usually incorporated in the location parameter of the GEV distribution.

3.3.2 Scenario 2: Mixed Dependence Model

The second data generating scenario was designed to assess the performance of the CSE method relative to the PL method when the latter misspecifies the spatial dependence structure. The marginal models remain the same as in the first scenario. The dependence structure was contaminated: data generated from a GG max-stable process were contaminated with those from a GA model. The PL method specifies the dependence as the GG model, so that the severity of misspecification increases with the contamination rate p. Following the settings in Shang et al. (2015), the variance parameter of the GG process δ^2 was assumed to be known; it was selected to be 8 because the data generation function in R package `SpatialExtremes` (Ribatet and Singleton, 2013) provides good approximation only when $\delta^2 < 10$, and larger δ^2 can provide full range of dependence. The correlation function of the GG process was Gaussian $\rho(h) = \exp[-(\|h\|/\phi)^2]$, with range parameter $\phi \in \{14.27, 29.55\}$, corresponding to M and S dependence. These range parameter values were chosen through a non-linear least square method such that the bivariate extreme coefficient function of the GG model matches that of the SM model

closely. The GA model had an exponential correlation function $\rho(h) = \exp(-h/\tau)$, with $\tau \in \{12, 20\}$ tuned so that the empirical correlation of the scores from a large sample matches that from the GG model closely.

Under strong dependence, there were numerical issues in generating data from the GG model with $(c, m) = (10, 80)$ because frequently the Gaussian correlation structure with 80 points in $[-10, 10]^2$ led to a correlation matrix that is not of full rank. Therefore, only two configurations of (c, m) were considered for strong dependence: $\{(10, 20)$ and $(20, 80)\}$. Simulation was done with different contamination rate $p \in \{0, 0.1, 0.25, 0.5, 0.75, 1\}$. As in the first scenario, we ran 1000 replicates for each setting and each replicate had its own sites regenerated. The PL method was applied with the dependence model specified as GG with single dependence parameter ϕ; that is, PL had correct specification only for $p = 0$. The results for contamination rate $p \in \{0, 0.5, 1\}$ are summarized in Tables 3–5.

Both the IL estimator and the CSE estimator appear to have negligible bias in all cases in all the tables. The CSE estimator is uniformly more efficient than the IL estimator, with RE ranging from 1.046 to 1.353. Most of the RE of the CSE estimator are greater than 1.1. The efficiency is most noticeable for the coefficients of latitude and longitude in the location parameter, which attains 1.4 in a few occasions. When misspecified, the PL estimator shows only a little bias for the location and scale parameters, but its bias for the shape parameter is quite large. For example, in Table 4, with $(c, m) = (10, 20)$ and strong dependence, it is 25%. The bias is bigger under stronger

Table 3: Summaries of the simulation results for Scenario 2: mixed dependence model with $p = 0$. The relative efficiency (RE) was based on the MSE, with the IL estimate as reference.

					Estimate			RMSE			RE	
Dep	c	m	Par	True	IL	PL	CSE	IL	PL	CSE	PL	CSE
M	10	20	$\beta_{\mu,0}$	15	14.667	14.663	14.664	0.42	0.42	0.42	1.00	1.03
			$\beta_{\mu,1}$	−0.2	−0.200	−0.200	−0.200	0.02	0.02	0.02	1.02	1.22
			$\beta_{\mu,2}$	0.25	0.251	0.251	0.251	0.02	0.02	0.02	1.02	1.17
			$\beta_{\sigma,0}$	4	4.078	4.077	4.078	0.21	0.21	0.20	1.02	1.13
			$\beta_{\xi,0}$	0.2	0.195	0.197	0.196	0.04	0.04	0.04	1.46	1.12
	10	80	$\beta_{\mu,0}$	15	14.720	14.716	14.712	0.37	0.37	0.36	0.99	1.04
			$\beta_{\mu,1}$	−0.2	−0.201	−0.201	−0.200	0.02	0.02	0.02	1.02	1.25
			$\beta_{\mu,2}$	0.25	0.250	0.250	0.249	0.02	0.02	0.02	1.03	1.37
			$\beta_{\sigma,0}$	4	4.093	4.093	4.092	0.20	0.20	0.19	1.02	1.15
			$\beta_{\xi,0}$	0.2	0.193	0.195	0.192	0.04	0.03	0.04	1.43	1.18
	20	80	$\beta_{\mu,0}$	15	14.734	14.733	14.729	0.32	0.32	0.32	0.99	1.00
			$\beta_{\mu,1}$	−0.2	−0.200	−0.200	−0.200	0.01	0.01	0.01	1.02	1.24
			$\beta_{\mu,2}$	0.25	0.250	0.250	0.250	0.01	0.01	0.01	1.02	1.23
			$\beta_{\sigma,0}$	4	4.107	4.107	4.107	0.17	0.17	0.17	1.02	1.09
			$\beta_{\xi,0}$	0.2	0.192	0.193	0.192	0.03	0.03	0.03	1.33	1.09
S	10	20	$\beta_{\mu,0}$	15	14.681	14.675	14.662	0.47	0.47	0.46	1.01	1.04
			$\beta_{\mu,1}$	−0.2	−0.201	−0.201	−0.200	0.02	0.02	0.02	1.04	1.20
			$\beta_{\mu,2}$	0.25	0.250	0.250	0.250	0.02	0.02	0.02	1.05	1.23
			$\beta_{\sigma,0}$	4	4.069	4.067	4.062	0.26	0.26	0.24	1.01	1.20
			$\beta_{\xi,0}$	0.2	0.197	0.200	0.200	0.06	0.04	0.06	1.85	1.16
	20	80	$\beta_{\mu,0}$	15	14.703	14.700	14.698	0.38	0.38	0.37	0.99	1.05
			$\beta_{\mu,1}$	−0.2	−0.200	−0.200	−0.200	0.01	0.01	0.01	1.03	1.40
			$\beta_{\mu,2}$	0.25	0.250	0.250	0.250	0.01	0.01	0.01	1.03	1.41
			$\beta_{\sigma,0}$	4	4.086	4.085	4.084	0.20	0.20	0.18	1.01	1.17
			$\beta_{\xi,0}$	0.2	0.195	0.197	0.195	0.04	0.03	0.04	1.48	1.16

Table 4: Summaries of the simulation results for Scenario 2: mixed dependence model with $p = 0.5$. The relative efficiency (RE) was based on the MSE, with the IL estimate as reference.

					Estimate			RMSE			RE	
Dep	c	m	Par	True	IL	PL	CSE	IL	PL	CSE	PL	CSE
M	10	20	$\beta_{\mu,0}$	15	14.837	14.803	14.836	0.32	0.34	0.30	0.91	1.10
			$\beta_{\mu,1}$	−0.2	−0.200	−0.200	−0.200	0.02	0.02	0.02	1.04	1.20
			$\beta_{\mu,2}$	0.25	0.249	0.249	0.250	0.02	0.02	0.02	1.05	1.21
			$\beta_{\sigma,0}$	4	4.043	4.028	4.042	0.20	0.20	0.19	1.06	1.14
			$\beta_{\xi,0}$	0.2	0.195	0.217	0.194	0.04	0.04	0.04	1.06	1.13
	10	80	$\beta_{\mu,0}$	15	14.879	14.847	14.878	0.29	0.30	0.26	0.92	1.18
			$\beta_{\mu,1}$	−0.2	−0.200	−0.200	−0.200	0.02	0.02	0.02	1.04	1.21
			$\beta_{\mu,2}$	0.25	0.250	0.250	0.250	0.02	0.02	0.02	1.05	1.31
			$\beta_{\sigma,0}$	4	4.048	4.034	4.049	0.19	0.18	0.17	1.06	1.17
			$\beta_{\xi,0}$	0.2	0.193	0.213	0.192	0.04	0.03	0.03	1.17	1.18
	20	80	$\beta_{\mu,0}$	15	14.881	14.868	14.878	0.23	0.24	0.22	0.95	1.15
			$\beta_{\mu,1}$	−0.2	−0.200	−0.200	−0.200	0.01	0.01	0.01	1.03	1.19
			$\beta_{\mu,2}$	0.25	0.250	0.250	0.250	0.01	0.01	0.01	1.03	1.23
			$\beta_{\sigma,0}$	4	4.066	4.060	4.065	0.15	0.15	0.14	1.06	1.18
			$\beta_{\xi,0}$	0.2	0.193	0.202	0.193	0.03	0.02	0.02	1.30	1.16
S	10	20	$\beta_{\mu,0}$	15	14.820	14.751	14.811	0.40	0.43	0.39	0.88	1.08
			$\beta_{\mu,1}$	−0.2	−0.199	−0.199	−0.199	0.02	0.02	0.02	1.04	1.17
			$\beta_{\mu,2}$	0.25	0.249	0.249	0.250	0.02	0.02	0.02	1.08	1.21
			$\beta_{\sigma,0}$	4	4.013	3.986	4.010	0.25	0.25	0.24	1.04	1.14
			$\beta_{\xi,0}$	0.2	0.193	0.241	0.194	0.06	0.06	0.05	0.83	1.13
	20	80	$\beta_{\mu,0}$	15	14.866	14.833	14.866	0.28	0.29	0.27	0.91	1.09
			$\beta_{\mu,1}$	−0.2	−0.200	−0.200	−0.200	0.01	0.01	0.01	1.04	1.30
			$\beta_{\mu,2}$	0.25	0.250	0.250	0.250	0.01	0.01	0.01	1.05	1.31
			$\beta_{\sigma,0}$	4	4.051	4.036	4.050	0.18	0.18	0.17	1.07	1.13
			$\beta_{\xi,0}$	0.2	0.195	0.214	0.194	0.04	0.03	0.03	1.11	1.10

Table 5: Summaries of the simulation results for Scenario 2: mixed dependence model with $p = 1$. The relative efficiency (RE) was based on the MSE, with the IL estimate as reference.

					Estimate			RMSE			RE	
Dep	c	m	Par	True	IL	PL	CSE	IL	PL	CSE	PL	CSE
M	10	20	$\beta_{\mu,0}$	15	15.008	14.952	15.019	0.29	0.29	0.28	1.00	1.09
			$\beta_{\mu,1}$	−0.2	−0.200	−0.200	−0.200	0.02	0.02	0.02	1.04	1.12
			$\beta_{\mu,2}$	0.25	0.249	0.249	0.250	0.02	0.02	0.02	1.03	1.10
			$\beta_{\sigma,0}$	4	3.981	3.957	3.981	0.20	0.20	0.19	1.00	1.15
			$\beta_{\xi,0}$	0.2	0.200	0.236	0.199	0.04	0.05	0.03	0.52	1.10
	10	80	$\beta_{\mu,0}$	15	15.005	14.953	15.005	0.27	0.27	0.25	0.99	1.17
			$\beta_{\mu,1}$	−0.2	−0.201	−0.200	−0.201	0.02	0.02	0.02	1.04	1.23
			$\beta_{\mu,2}$	0.25	0.250	0.250	0.250	0.02	0.02	0.02	1.05	1.27
			$\beta_{\sigma,0}$	4	3.980	3.956	3.979	0.19	0.19	0.17	0.99	1.17
			$\beta_{\xi,0}$	0.2	0.201	0.234	0.200	0.03	0.04	0.03	0.50	1.11
	20	80	$\beta_{\mu,0}$	15	15.002	14.979	15.006	0.20	0.20	0.18	1.00	1.28
			$\beta_{\mu,1}$	−0.2	−0.200	−0.200	−0.200	0.01	0.01	0.01	1.03	1.30
			$\beta_{\mu,2}$	0.25	0.250	0.250	0.250	0.01	0.01	0.01	1.03	1.23
			$\beta_{\sigma,0}$	4	3.988	3.976	3.989	0.13	0.13	0.12	1.00	1.20
			$\beta_{\xi,0}$	0.2	0.200	0.215	0.200	0.02	0.02	0.02	0.69	1.10
S	10	20	$\beta_{\mu,0}$	15	15.011	14.924	15.021	0.34	0.34	0.32	0.99	1.09
			$\beta_{\mu,1}$	−0.2	−0.200	−0.200	−0.200	0.02	0.02	0.02	1.05	1.12
			$\beta_{\mu,2}$	0.25	0.249	0.249	0.250	0.02	0.02	0.02	1.04	1.10
			$\beta_{\sigma,0}$	4	3.976	3.947	3.974	0.24	0.24	0.23	1.01	1.16
			$\beta_{\xi,0}$	0.2	0.200	0.261	0.199	0.05	0.07	0.04	0.39	1.13
	20	80	$\beta_{\mu,0}$	15	15.006	14.962	15.007	0.25	0.25	0.23	0.99	1.21
			$\beta_{\mu,1}$	−0.2	−0.200	−0.200	−0.200	0.01	0.01	0.01	1.04	1.31
			$\beta_{\mu,2}$	0.25	0.250	0.250	0.250	0.01	0.01	0.01	1.05	1.24
			$\beta_{\sigma,0}$	4	3.984	3.963	3.982	0.17	0.17	0.16	0.99	1.18
			$\beta_{\xi,0}$	0.2	0.200	0.227	0.200	0.03	0.04	0.03	0.55	1.11

dependence level. Expanding the study region helps to reduce the bias. For the location and scale parameters, the RE of the PL estimator is marginally over 1 and never exceeds 1.1. For the shape parameter, however, the RE of the PL estimator heavily depends on the contamination rate p. It ranges from 1.477 in the correct specification case ($p = 0$) to 0.391 completely misspecified case ($p = 1$). This is in contrast to the CSE estimator, whose RE is stably greater than 1 and higher under strong dependence.

It is of interest to see the smallest contamination rate investigated at which the CSE method is more efficient than the PL method. Under moderate dependence, it is 0.5, 0.5, and 0.75 for (c, m) at (10, 20), (10, 80), and (20, 80), respectively. Under strong dependence, it is 0.25 and 0.5 for (c, m) at (10, 20) and (20, 80), respectively. This means that the room for the PL method to be preferred to the CSE method is rather limited under partly correct specification of the dependence, which may not be realistic in practice.

3.4 Discussion

The CSE method for estimating marginal regression parameters in spatial GEV model inherits its consistency, efficiency, and robustness to misspecification from the general GEE approach. The estimating equations that are jointly weighted are marginal score equations from univariate GEV models. The marginal GEV distributions are much easier to check and more realistic to assume than max-stable models, which additionally specify

the spatial dependence among the extreme observations. The optimal weight for the score equations is approximated by a simple one-parameter working spatial correlation function. Our simulation study shows that the CSE estimator is consistent as the IL estimator, but more efficient than the IL estimator. Unlike the PL estimator, it does not need to specify the full dependence model and, hence, does not suffer the consequence of misspecification as the PL estimator does.

Several directions from the proposed CSE method merit further investigation. First, incorporation of errors-in-variables would be of practical interest since the fingerprints are estimated as opposed to known for certain. It is well-known that measurement errors in covariates lead to biased regression coefficient estimation if they are left untreated. The bias may dominate the smaller variance from the CSE method. Second, the univariate Fisher information used in constructing the optimal weight matrix does not exist when the shape parameter is lower than -0.5, in which case, special treatment is needed in implementation. Thirdly, when the number of parameters is large, solving the estimating equations may have numerical difficulties; for example, iterations can be trapped between two points. Combining the estimating equations using the quadratic inference function approach (Qu et al., 2000; Bai et al., 2012) and minimizing an objective function is worth investigating.

Chapter 4

Fingerprinting Changes in Climate Extremes with CSE

4.1 Introduction

Detection and attribution of changes in climate extremes have gained sharpened focus (e.g., Field et al., 2012; Seneviratne et al., 2012). Such analysis involves assessment of observed changes in relation to what are expected to have occurred in response to forcings external to the climate system. One important issue with wide publicity is the observed warming globally since the late 19th century, and that most of the observed warming can be attributed to human influence (Bindoff et al., 2013). Human influence on climate has been detected in temperatures at continental, sub-continental, and even regional scales (e.g., Stott et al., 2010; Sun et al., 2014), and in other aspects of the climate system such as the distribution of global precipitation (Zhang et al., 2007; Marvel and Bonfils, 2013; Polson et al., 2013), humidity (Willett et al., 2007), and Arctic sea ice (Min et al., 2008).

For changes in climatic mean states, the fingerprint method based on the linear model has been widely used (e.g., Hasselmann, 1997). The fingerprint refers to the pattern of change in the climate that is expected in response to external forcing of the climate system, and is typically estimated from ensembles of climate model simulations. The method regresses observations onto the fingerprints to determine whether they are present in the observations. The residuals of the model, which are usually temporally and spatially correlated, contain information about the natural internal (chaotic) variability of the climate system. Inferences about the presence or absence of the fingerprints in the observations are made on the basis of the coefficient vector of scaling factors that adjust the amplitudes of fingerprints to provide best match to the observations. Most studies use a variant of this approach known as the optimal fingerprint method, in which the estimates of scaling factors are obtained with a generalized least squares approach, with the optimal weight chosen to be the inverse variance matrix of the residuals. The method gives the smallest variance of the scaling factor estimator. It is also interpreted to maximize the signal-to-noise ratio among all linear transformations of the response and the covariates (Hasselmann, 1997; Allen and Tett, 1999; Allen and Stott, 2003; Hegerl et al., 2007; Ribes et al., 2013). It should be noted that the variance-covariance matrix of the residuals is usually estimated from long control simulations performed with climate models since observed records are not enough for this purpose. The method is mature with good practice guidance (Hegerl et al., 2010); see also Hegerl et al. (2007), Stott et al. (2010) and Hegerl and Zwiers (2011) for recent reviews and references therein.

Widely used as the optimal fingerprint method is, it is not designed for changes in climate extremes. The principles of optimal detection and attribution can be applied to climate extremes, but it is more challenging because of sparsity of data, relatively low signal to noise ratio, and unique distributional properties of extreme values. A close analog of the classical optimal fingerprint method has not been established for changes in climate extremes with the GEV distribution. Some approaches have worked around it by transforming extremes (Min et al., 2011) or by applying the detection and attribution method to the parameters of fitted GEV distribution (Christidis et al., 2011b), or by averaging extreme values over large region (Wen et al., 2013; Zhang et al., 2013). Most recently, Zwiers et al. (2011) explicitly considered the distributional properties of extreme values in their detection and attribution analysis of extreme temperatures. Grid box specific signals were estimated as the location parameters the GEV distribution from outputs of climate model simulations. The signals were used as covariates in modeling the observed climate extremes, and the coefficients of signals were set to be the same across all grid boxes in fingerprinting. Spatial dependence discarded, the shared coefficient parameters were estimated by a profile independence likelihood (IL) approach. Block bootstrap that retains the spatial dependence was used to estimate the variance of the parameter estimator. The profile method combined with the bootstrap procedure is very computing intensive; the 16 regional studies in Zwiers et al. (2011) took weeks on a computer cluster. Extending to two or more signals is prohibitively expensive.

The CSE method proposed in Chapter 3 provides a general methodology for extreme value analysis in a spatial setting that does not require parametric specifications beyond univariate marginal GEV distributions. Misspecification of the correlation structure does not affect the consistency of the CSE estimator. When the working correlation structure is closer to the truth than the independence structure, the resulting estimator is more efficient than that from the independence likelihood. Moreover, the CSE approach provides a practical analog of the classic optimal fingerprinting for mean climate state in the context of detection and attribution of changes in climate extremes. With working independence, it is conceptually equivalent to the independence likelihood method of Zwiers et al. (2011), but computationally much more efficient and accurate. In combination with bootstrapping for inferences, the profile approach of Zwiers et al. (2011) is prohibitively expensive when two or more signals are in the model. The CSE method finds the parameter estimates quickly through a coordinate descent algorithm, with no need of grid specification. Similar to the GEE setting (Liang and Zeger, 1986), a non-independence working correlation may lead to narrower confidence intervals of the scale parameters of interests, and, hence, higher power in detection and attribution analysis.

The remainder of this chapter proceeds as follows. The methods of the signal estimation and detection analysis are presented in Section 4.2. Section 4.3 presents a simulation study to investigate the properties of the CSE method in comparison with competing methods under detection and attribution settings. In Section 4.4, the method

is applied to two detection and attribution analyses of changes in extreme temperature at two regions, one in Australia under a perfect model setting and the other an application to observations for the Northern Europe (NEU) region. In particular, the NEU analysis put the climate responses to both anthropogenic forcing and the natural forcing simultaneously in the fingerprinting model, which was unfeasible in Zwiers et al. (2011). A discussion concludes in Section 4.5.

4.2 Fingerprint Method

Following Zwiers et al. (2011), only the GEV location parameters are affected by the response to the external forcing, while the scale and shape parameters are unaffected. This is a reasonable assumption from the literature. Kharin and Zwiers (2005) examined the changes in temperature extremes in transient climate change simulations with the time dependent GEV location and scale parameters, and showed that changes are primarily associated with changes in the location parameter; no substantial changes in the shape parameter is found in most grid boxes. Brown et al. (2008) studied the nonstationary extremes with the Generalized Pareto (GP) distribution. They assumed that the three parameters of GP distribution linearly depend on time, and only found significant nonstationarity in location parameter, but not in scale or shape parameter. The detailed signal estimation and detection analysis are given below.

4.2.1 Signal Estimation

The signals in fingerprinting are estimated from the climate model simulation data of the external forcing. Usually, several climate models with multiple ensembles are available under the corresponding forcing. The GEV location parameters are used to represent the model-simulated changes in the climate extremes of corresponding forcing. The signal estimation procedure is similar to that in Zwiers et al. (2011) with the location parameters change every h-yr. The signals are estimated as the location parameters due to external forcing for individual climate model simulation with multiple ensembles. Suppose that there are totally m sites at a region with n-year record, and n is multiple times of h. Then the number of h-yr block is $B = n/h$. Also suppose there are totally l ensembles for one individual climate model; the estimation procedure is for one given climate model, and the results from different models will be aggregated later. Let Z_{tsu} be the model output at grid box s in year t from ensemble u for $s = 1, \ldots, m$, $t = 1, \ldots, n$, and $u = 1, \ldots, l$. A GEV model is assumed for Z_{tsu} with parameters

$$\begin{cases} \mu_{tsu} = \mu_{b(t),s}, \\ \sigma_{tsu} = \sigma_s, \\ \xi_{tsu} = \xi_s, \end{cases} \tag{4.1}$$

where $b(t) = \text{ceiling}(t/h)$, $b = 1, \ldots, B = n/h$. 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 h-yr. The parameters can be easily estimated through a maximum likelihood method at each grid box s, with ensembles treated as replicates. To reduce the variation in estimating the signal and thus may lead higher power in detection, we use a smoothed version $\tilde{\mu}_{d(t),s}$ of $\hat{\mu}_{d(t),s}$ from locally weighted scatterplot smoothing.

The estimation process is repeated for each climate model separately, and the averaged estimated location parameters over all climate models is used as the signals of the corresponding forcing.

4.2.2 Detection Analysis

To detect the contribution of the forcing of interest, a GEV distribution is fitted to the observational data at each grid box. The signals are used as the covariates corresponding to a common scaling factor in the GEV location paramter within the specific region; the GEV scale and shape parameters vary across different grid boxes. Suppose that extreme temperatures are available at m grid boxes over n years. The observed annual extreme climate variable Y_{ts} at grid box s in year t, $s = 1, \ldots, m$, $t = 1, \ldots, n$, is modeled by a GEV distribution with

$$\begin{cases} \mu_{ts} = \alpha_s + X_{ts}^{\top}\beta, \\ \sigma_{ts} = \sigma_s, \\ \xi_{ts} = \xi_s, \end{cases} \tag{4.2}$$

where X_{ts} be a $p \times 1$ vector of the relative signals of external forcing of interest at grid box s in year t, β is the scaling vector shared within the region, and α_s, σ_s, and ξ_s are grid box specific location, scale, and shape parameters, respectively. The relative signal is the estimated signal centered over the B h-yr blocks. The focus of the detection analysis is the inference about β. 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, so the conclusion of attribution is made. If the interval lies above unity, i.e., all the estimates of the scaling factor are larger than 1, then we can conclude that the observation is underestimated by the climate model. If the interval lies within zero and unity, i.e., all the estimates of the scaling factor are between 0 and 1, then the observation is overestimated by the climate model.

There are $3m+p$ unknown parameters in Model (4.2); simultaneously estimating such large number of parameters is challenging. Zwiers et al. (2011) fixed β within a selected list of values, then estimation becomes trivial since the grid boxes are independent of each other. They picked the β that minimizes the negative log-likelihood over all grid boxes within the region. We use a coordinate descent approach, estimating a small set of the unknown parameters each time with others being held. The CSE method can be easily applied to estimate β in (4.2) if $\zeta_s = (\alpha_s, \sigma_s, \xi_s)$'s were known. Application of the block coordinate descent algorithm leads to a two-step iterative process:

1. Given current estimate $\hat{\beta}_n$ of β, obtain the likelihood estimate $\hat{\zeta}_{s,n}$ of ζ_s separately at each grid box $s \in \{1, \ldots, m\}$.

2. Given current estimate $\hat{\zeta}_{s,n}$, obtain the CSE estimate $\hat{\beta}_n$ of β from solving (3.2) with an appropriately chosen working correlation structure.

The two steps iterate until $\hat{\beta}_n$ converges. This algorithm was reliable in our simulation study in Section 4.3.

The CSE method for fingerprinting has two advantages relative to the method of Zwiers et al. (2011). The first is efficiency in inference — non-independence working correlation may lead to more efficient estimator of β and higher power in detection and attribution analysis. The second advantage is computation efficiency and accuracy, even under working independence where the two methods are conceptually the same. In our analysis of NEU region in Section 4.4.2, when single signal is present, for each of the four outcome extreme temperatures, the profile method on a grid of 301 points took about 46–53 seconds on a laptop computer with an Intel Core 2.50GHZ CPU, while the CSE method took only 2–4 seconds. Further, the profile method requires that the predetermined grid covers the solution and that the grid is fine enough to approximate the solution. When multiple signals are present, which are often the cases in practical detection applications, the profile method becomes unpractical with high-dimensional grid, especially when combined with bootstrapping. In contrast, the CSE method only took 4–5 seconds when there are two β parameters in Model 4.2.

4.2.3 Uncertainty Assessment

To account for the uncertainty in the signal estimate and the point estimate of scaling factor, following the idea in Zwiers et al. (2011), a 32×32 block bootstrap that preserves both temporal and spatial dependence is performed. First, the ensemble data from each climate model are bootstrapped 32 times to obtain 32 signals, and then for each of the 32 signals, β is estimated from 32 bootstrap samples of the observational data.

The bootstrap sampling algorithm depends on the value of h in Model (4.1). In Section 4.3 simulation study and Section 4.4 real data analysis, the value of h is 10. The bootstrap sampling procedure of the signal estimate corresponding to $h = 10$ is the following.

(A) For each individual grid box of each ensemble, divide the n-year simulation data into $n/10$ nonoverlapping 10-yr blocks. Randomly sample 5-yr blocks data with replacement within the 10-yr blocks. All the grid boxes share the same sample order to keep the spatial dependence.

(B) All the ensembles of all climate models share the same sample order.

(C) Estimate the signals from the reordered data in step A.

(D) Repeat steps A to C 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.

(a) Subtracting the scaled signal $X_{ts}^{\top}\hat{\beta}$ from Y_{ts} in Model (4.2) to obtain the residuals.

(b) For each individual grid box, divide the residuals into nonoverlapping 5-yr blocks. Randomly reorder the 5-yr blocks residuals. All the grid boxes share the same sample order to keep the spatial dependence.

(c) Adding the scaled signal $X_{ts}^{\top}\hat{\beta}$ back to the reordered residuals, and denote it as $\tilde{Y}_{ts}$.

(d) Estimate the scaling factor from $\tilde{Y}_{ts}$ with the given signal.

(e) Repeat steps b to d 32 times for each of the 32 bootstrap samples of signal.

The above bootstrap procedure leads to 1024 bootstrap samples of the $\hat{\beta}$. Both the effects of the signal uncertainty and the natural internal variability have been considered, and the spatial-temporal dependence has been retained as much as possible. The 1024 samples give a sampling distribution of β, and the corresponding 5% quantile and 95% quantile leads to an approximate 90% confidence interval.

4.2.4 Goodness-of-fit Test

Goodness-of-fit test for the nonstationary GEV distribution in Model (4.2) can be done for each grid box separately. The nonstationary component needs to be removed first, and then test on the data without the nonstationary part. To assess the significance of the goodness-of-fit test, the uncertainty in parameter estimation needs to be account

for. Since β is shared by all grid boxes so that the test statistics at a given grid is also affected by data at other grids, we use the semiparametric bootstrap algorithm in Heffernan and Tawn (2004) to ensure the bootstrap samples not only replicate the marginal GEV distribution but also keep the feature of the spatial dependence of the original observed data. The procedure is the following.

i Transform the observational data into Gumbel residuals by the fitted GEV parameters in Model (4.2) at each grid box. So testing the goodness-of-fit of the GEV distribution of observed data equals to testing the goodness-of-fit of the standard Gumbel distribution of the Gumbel residuals.

ii Calculate the Kolmogorov–Smirnov test statistic at each grid box.

iii Apply the semiparametric bootstrap algorithm in Heffernan and Tawn (2004) on the observational data to generate a bootstrap sample of observed data. Apply steps i to ii on the sample data. Repeat this step 1000 times to obtain 1000 bootstrap samples of the Kolmogorov–Smirnov statistics at each grid box.

The 90% quantile of 1000 bootstrap samples of test statistics at each grid boxes is used as the critical value.

4.3 Simulation Study

To assess the performance of the CSE method in fingerprinting changes in extremes, data mimicking the daily maximum temperature setting in Australia in Section 4.4.1 were generated. Signals were estimated from 10 ensembles of climate model simulation in $n = 140$ years at $m = 29$ grid boxes in Australia; see Section 4.4.1 for details about the ensembles. Model (4.1) was fitted to the simulation data with $h = 10$. The estimated signals $\tilde{\mu}_{d(t),s}$ were used as input X_{ts} in the marginal model (4.2) to generate data. The location parameter α_s was set as the average of the smoothed signals $\tilde{\mu}_{d(t),s}$ over the decades for each s. The scale σ_s and shape ξ_s were set to be the corresponding estimates in model (4.1) for each s. Three levels of $\beta \in \{0, 0.5, 1\}$ were considered. The dependence model was a mixture of a GG model and a GA model as used in Section 3.3.2 with different range parameters. The level of spatial dependence was tuned based on the empirical dependence of the climate model simulations in the Australia data. From the 10 ensembles, the pairwise extremal coefficients were estimated for each pair of the 29 grid boxes. A GG model with Gaussian correlation was fitted to the pairwise extremal coefficients and the estimated range parameter was 18.17, which was used for ϕ as strong dependence in data generation. A weaker dependence was considered with $\phi = 12.11$, 2/3 of the ranger parameter value under strong dependence. The corresponding parameters of the GA model for the two dependence levels with correlation function were set at $\tau \in \{7, 10\}$ in the same way as in Section 3.3.2. Three mixing rates were considered:

$p \in \{0, 0.5, 1\}$, where p is the proportion of the GA model.

For each configuration, 1000 datasets were generated. Three estimation methods were applied for each dataset. The CSE method used the iterative process in Section 4.2.2; The IL method was just the CSE method with working independence. For the PL method, the GG dependence model was specified, and, hence, one additional dependence parameter needs to be estimated. Optimization with $3m + 2$ parameters simultaneously is challenging and the profile approach of Shang et al. (2015) is time consuming. A similar two-step iterative estimation procedure was used, with the second step estimating (β, ϕ) jointly with the PL method under the bivariate GEV distribution with GG spatial dependence structure. Constraints $\xi_s > -0.45$ were imposed during the estimation.

Table 6 summarizes the estimation results of the focus β in a fingerprinting application based on 1000 replicates. The bias is negligible in all the cases for all three methods. Stronger dependence leads to larger RMSE, but the magnitude of the increase is the smallest for the CSE method. The RE of the PL estimator and the CSE estimator are all above 1. The RE ranges from 1.04 to 1.21 for the PL estimator, and from 1.37 to 2.14 for the CSE estimator, indicating that the CSE estimator can be much more efficient than the PL estimator. Note that the edge of the CSE method relative to the PL method is evident even when the PL correctly specified the dependence. This result echoes the observations from the simulation study in Section 3.3, where the most efficiency gain in the CSE method is in estimating the location parameters. As expected, the RE of the PL method decreases as p increases because the dependence model is

Table 6: Summaries of the simulation results for mixed dependence model. The relative efficiency (RE) was based on the MSE, with the IL estimate as reference.

			Estimate			RMSE			RE	
p	ϕ	True	IL	PL	CSE	IL	PL	CSE	PL	CSE
0	12.11	0	−0.001	−0.001	0.001	0.120	0.114	0.103	1.10	1.37
		0.5	0.503	0.503	0.502	0.118	0.111	0.097	1.12	1.49
		1	1.005	1.005	1.005	0.119	0.112	0.098	1.13	1.48
	18.17	0	−0.001	−0.002	−0.004	0.153	0.140	0.104	1.19	2.14
		0.5	0.503	0.502	0.501	0.147	0.133	0.102	1.21	2.05
		1	1.007	1.007	1.002	0.146	0.134	0.103	1.19	1.99
0.5	12.11	0	0.004	0.003	0.001	0.116	0.112	0.094	1.08	1.51
		0.5	0.507	0.507	0.502	0.115	0.111	0.096	1.07	1.42
		1	0.997	0.997	1.000	0.115	0.112	0.097	1.06	1.40
	18.17	0	−0.004	−0.004	0.000	0.138	0.131	0.098	1.12	2.00
		0.5	0.500	0.500	0.499	0.144	0.136	0.100	1.13	2.08
		1	1.005	1.005	1.006	0.138	0.131	0.097	1.12	2.02
1	12.11	0	−0.001	−0.001	−0.002	0.110	0.108	0.091	1.04	1.46
		0.5	0.504	0.504	0.502	0.110	0.108	0.092	1.04	1.44
		1	0.997	0.997	1.000	0.112	0.110	0.093	1.04	1.44
	18.17	0	−0.001	0.000	−0.002	0.132	0.128	0.098	1.07	1.83
		0.5	0.505	0.505	0.502	0.133	0.129	0.099	1.07	1.80
		1	0.996	0.996	1.000	0.135	0.131	0.100	1.07	1.82

misspecified when $p \neq 0$. The RE is higher when the spatial dependence is stronger. The magnitude of the increase in RE ranges from 0.03 to 0.09 for the PL method and from 0.36 to 0.77 for the CSE method, suggesting that the latter has more potential with stronger spatial dependence.

4.4 Regional Applications to Extreme Temperatures

4.4.1 Perfect Model Detection for Australia

A "perfect model" detection approach was used to evaluate the time by which climate change may be detectable in extreme temperature and precipitation in the classic fingerprinting framework (Hegerl et al., 2004; Min et al., 2009), without involvement of extreme value analysis. By perfect model we mean that both the "observational" data and signals come from simulations of climate models. In this case, the "observational" data contains model simulated response as well as the characteristics of model simulated 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 between observations and model simulations.

Annual maxima of daily minimum temperature were extracted from simulations forced with the combined effect of anthropogenic and natural forcings, known as the ALL forcing, conducted with the Hadley Center Coupled Model, version 3 (HadCM3), for 29 grid boxes in Australia (lat: -45S to -11S; long: 110E to 155E). The signals were estimated from 10 ensembles of 140 years (1861–2000) simulation data with $h = 10$ in Model 4.1. In the perfect model setting, each ensemble was treated as an observation dataset and the rest nine ensembles as model simulations. This amounts to 10 detection and attribution analyses. Since the 10 ensembles were from the same climate model, the scaling factor β in (4.2) should be around 1 if the signal is strong enough to be detectable,

which is a feature of the perfect model setting. The IL method was a special case of the CSE method with W_t in equation 3.3 be the identity matrix. The CSE method used a working exponential correlation structure, which seemed appropriate from the empirical correlation function of the scores obtained with the IL estimate. The IL estimate was used as the starting value of the CSE method.

Table 7 summarizes the point estimates and 90% bootstrap confidence intervals for the 10 perfect model detection analyses. The point estimates from both methods are quite close to 1, which is what we expect to see in the perfect model detection setting. Although the two estimates are similar, the CSE estimates seem to be consistently lower in magnitude than the IL estimates. The discrepancy may be due to the errors in the signals, as the signals were estimated as opposed to known for certain. The error-in-variable might affect the CSE estimator more than it does the IL estimator; further investigation would be interesting. The 90% confidence intervals cover 1 in all the cases as expected. The confidence intervals from the CSE method are generally narrower than those from the IL method. The average reduction in interval length of CSE method is 11%, and the largest reduction is for ensemble 1: the lengths are 0.34 and 0.49, respectively, a 31% reduction.

4.4.2 Extreme Temperatures in Northern Europe (NEU)

The NEU region for detection with real data is bounded by 7.5°W to 37.5°E and by 50°N and 70°N. Four extreme temperatures were considered: annual maximum daily

Table 7: Summaries of the estimate of the scaling factor β, the corresponding 90% confidence interval and the interval length for the 10 perfect model detection analyses.

	Estimate		90% Confidence Interval		Interval Length	
Ensemble	IL	CSE	IL	CSE	IL	CSE
1	1.03	0.91	(0.78, 1.28)	(0.75, 1.09)	0.49	0.34
2	0.93	0.79	(0.68, 1.19)	(0.60, 0.99)	0.51	0.39
3	1.11	1.06	(0.90, 1.32)	(0.83, 1.30)	0.42	0.46
4	1.10	0.94	(0.86, 1.35)	(0.76, 1.15)	0.49	0.39
5	0.99	0.93	(0.75, 1.26)	(0.71, 1.15)	0.51	0.44
6	0.98	0.79	(0.73, 1.24)	(0.55, 1.03)	0.51	0.47
7	0.95	0.88	(0.77, 1.16)	(0.67, 1.10)	0.39	0.42
8	1.16	0.89	(0.94, 1.38)	(0.67, 1.09)	0.44	0.42
9	0.98	0.81	(0.67, 1.27)	(0.54, 1.07)	0.60	0.53
10	1.02	1.01	(0.81, 1.24)	(0.83, 1.20)	0.43	0.36

maximum (TXx) and minimum (TNx) and annual minimum daily maximum (TXn) and minimum (TNn). The observations were from the HadEX2 dataset (Donat et al., 2013) with 60 years (1951–2010) data at 67 grid boxes extracted. The climate model simulation data were matched with observed data, including a total of 22 ensembles from 6 climate models under the ALL (anthropogenic and natural) forcing and 26 ensembles from 6 climate models under the natural (NAT) forcing; see Table 8. The observed data were used to mask the simulated data, so that the simulated data have the same space/time coverage as the observed data. Since the GEV distribution is the limit distribution of the maximum of random variables, to fit the GEV distribution properly, we took the negative value of TXn and TNn in both the observed data and simulation data in our analysis. The NEU region was previously analyzed using different data in Zwiers

Table 8: List of models and ensemble sizes available 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	CM5A-MR	3	HadGEM2-ES	4
CanESM2	5	NorESM1-M	3	CNRM-CM5	6	IPSL-CM5A-LR	3

et al. (2011): 50 years data at 25 girds, totally 9 ensembles from 3 climate models for ALL forcing, and 14 ensembles from 4 climate models for anthropogenic (ANT) forcing. For each climate model, signal under each forcing was estimated by the procedure in Section 4.2.1 with $h = 10$ in Model 4.1, then the averaged signal over all models was used for the corresponding forcing. Since there is only a few climate models provide simulations with ANT forcing, in order to use as many as simulations we have, the ANT signal was calculated by subtracting NAT signal from ALL signal. We acknowledge that this difference may also be influenced by the uncertainties in climate models.

Besides conducting separate one-signal analyses for ALL and ANT, we also performed two-signal analyses in which both ANT and NAT signals are present in the model at the same time. That is, μ_{ts} in equation 4.2 becomes:

$$\mu_{ts} = \alpha_s + X_{A,ts}^{\top}\beta_A + X_{N,ts}^{\top}\beta_N, \tag{4.3}$$

where X_A and X_N are the ANT and NAT signals, respectively, and β_A and β_N are the

corresponding scaling factors. The CSE method used a working exponential correlation structure. The IL method was the CSE method with working independence, and was used to provide initial value for the CSE method with working exponential correlation function. The profile method of Zwiers et al. (2011), although equivalent to the IL method, would be computationally prohibitive. The point estimates and the 90% bootstrap confidence intervals of the scaling factors of the analyses are summarized in Table 9.

For the one-signal analyses, the CSE estimates of the scaling factors are smaller in magnitude than the IL estimates. This pattern is consistent with the observations from the perfect model analysis, but the differences appear to be larger, possibly due to smaller sample size and potential mismatch in the internal variability of observations and model simulations. The confidence intervals from the CSE method are shorter, in some cases drastically, than those from the IL method. For instance, for TXn with ANT forcing, the length of the 90% confidence intervals are 0.25 from the CSE method and 2.26 from the CSE method, a ratio of 1/9. Consequently, even though the point estimates from the CSE methods are smaller in magnitude, more signals may still be fingerprinted because of the much narrow confidence intervals. Influence of ALL and ANT were detected in TXx and TNx by both methods. For TNx and TNn, the IL method only detected ALL in TNn, while the CSE method additional suggested detection of ANT in TXn and ALL in TNn. Compared to the one-signal results of the NEU region reported in Zwiers et al. (2011), which were obtained with no constraints on the shape parameter of the GEV

Table 9: Summaries of the estimate of the scaling factor (est), the corresponding 90% confidence interval (CI), the interval length (len), and the number of grids fail the goodness of fit (gof) test at 10% significance level for NEU region.

			TXx				TNx			
Forcing	Me	Par	est	CI	len	gof	est	CI	len	gof
ALL	IL	β	0.97	(0.51, 1.46)	0.95	1	1.10	(0.73, 1.48)	0.75	2
	CSE	β	0.71	(0.39, 1.12)	0.72	2	0.69	(0.46, 0.95)	0.49	1
ANT	IL	β	1.02	(0.55, 1.51)	0.97	1	1.19	(0.77, 1.62)	0.85	2
	CSE	β	0.64	(0.32, 1.02)	0.69	2	0.52	(0.31, 0.74)	0.43	2
ANT&NAT	IL	β_A	1.02	(0.54, 1.50)	0.96	1	1.12	(0.75, 1.50)	0.76	2
		β_N	-0.09	$(-1.28, 1.05)$	2.32		0.91	$(-0.28, 2.07)$	2.35	
	CSE	β_A	0.70	(0.38, 1.12)	0.75	2	0.70	(0.47, 0.95)	0.48	1
		β_N	0.75	(0.37, 1.21)	0.84		0.59	(0.17, 1.01)	0.84	

			TXn				TNn			
Forcing	Me	Par	est	CI	len	gof	est	CI	len	gof
ALL	IL	β	0.45	$(-0.54, 1.46)$	2.00	0	0.84	$(-0.02, 1.76)$	1.78	0
	CSE	β	0.08	$(-0.10, 0.27)$	0.37	0	0.37	(0.16, 0.59)	0.44	0
ANT	IL	β	0.57	$(-0.57, 1.70)$	2.26	0	0.97	(0.01, 1.97)	1.96	0
	CSE	β	0.17	(0.04, 0.29)	0.25	0	0.27	(0.13, 0.42)	0.29	0
ANT&NAT	IL	β_A	0.60	$(-0.45, 1.66)$	2.12	0	0.94	(0.02, 1.89)	1.86	0
		β_N	-0.17	$(-2.10, 1.76)$	3.87		0.25	$(-1.92, 2.45)$	4.37	
	CSE	β_A	0.10	$(-0.08, 0.29)$	0.37	0	0.38	(0.17, 0.59)	0.42	0
		β_N	-0.17	$(-0.46, 0.09)$	0.56		0.30	(0.01, 0.56)	0.55	

distributions, we additionally detected ALL in TXx. Although both results detected ANT in TXx, our result provided stronger evidence since the lower bound of confidence interval was much further away above zero than theirs. No exact match was expected because the data used were different; we had more grid boxes, more ensembles and longer record length.

The results of two-signal analyses are novel contributions to the detection and attribution of changes in extreme temperatures as Zwiers et al. (2011) only did one-signal analyses. The CSE method detected both ANT and NAT in TXx, TNx, and TNn. For these three extreme temperatures, the IL method only detected ANT but not NAT, with much wider confidence intervals than the CSE method. Therefore, the influence of anthropogenic forcing had a detectable influence on extreme temperatures TXx, TNx, and TNn in the NEU region even when the natural forcing had been accounted for. These results provide more convincing evidence than the one-signal analyses in Zwiers et al. (2011) on the detectability of the anthropogenic influence on the extreme temperatures. They suggest that human influence has contributed to the observed changes in the warmest day temperature (TXx), the warmest night temperature (TNx), and the coldest night temperature (TNn).

Goodness of fit tests for the GEV distribution of the observed data were performed at each site for the models in Table 9. Only up to 2 grid boxes failed the goodness-of-fit tests in all the analyses reported in Table 9, which indicates that the GEV distribution with parameters specified by the detection model 4.2 are appropriate in general.

4.5 Discussion

Application of the CSE method in detection and attribution of changes in extremes provides an analog of the classical optimal fingerprinting. The weight that combines the marginal score equations has the same role as the weight in the generalized least squares in improving the efficiency of the regression parameter estimation. The CSE method proceeds under the assumption that extreme observations from year to year are independent, which means replicates of the response variable in years can be used to estimate the working weights. This is a small difference from the classical fingerprinting, where the weight (the inverse of the variance matrix of the error term) is obtained beforehand or estimated from ensembles of climate model simulations that represent the internal climate variability. An advantage of this assumption is that the covariance structure of the marginal score functions of the observed data does not need to be the same as that of the ensembles from climate model simulations. When temporal dependence exists, the current CSE method is still valid as it can be viewed as using working independence in the time direction. The block bootstrap method can still be used for interval estimation. A working spatial-temporal correlation function may help to further improve the efficiency if temporal dependence is not negligible. In the detection and attribution analyses of changes in extreme temperatures for the NEU region, the CSE method detected ALL in TXx, which was not detected in Zwiers et al. (2011), and allowed simultaneous detection of ANT and NAT in the same model, which has not been

done in the literature.

One problem in the CSE method is the measurement errors in covariates, i.e., the signals in this study. If left untreated, the measurement error in the signals can lead to biased estimate of the scaling factors, which will decrease the estimation efficiency of the CSE method. To avoid this problem, we will use the CSE approach with the working independence structure, which reduces to the independence likelihood method, in the detection and attribution analysis of climate changes in the next chapter. Though the CSE approach with the working independence and the profile independence likelihood method in Zwiers et al. (2011) are conceptually the same, the CSE approach is computationally feasible in multiple-signal detection analyses while the profile approach is unpractical.

Chapter 5

Detection and Attribution of Changes in Extreme Temperatures at Regional Scale

5.1 Introduction

IPCC (2014) reported that the increase in the frequency of warm temperature extremes and decrease in the frequency of cold temperature extremes are associated with human influence, and it is very likely (90–100% probability) that the observed changes in the frequency and intensity of daily extreme temperatures at global scale are attributed to human influence. Demonstration of anthropogenic and/or natural influence on changes in extreme temperature has gained sharpened focus, and many detection and attribution studies have reported anthropogenic influences at global and regional scales. Christidis et al. (2005) used a formal optimal detection method (Allen and Stott, 2003) to compare changes in the N warmest days and nights of a year ($N = 30, 10, 5, 1$) between

observations and climate model simulations at global scale with various forcings. They demonstrated the anthropogenic influence in extreme warm nights, and also cold nights and days but less robustly. Similar conclusions was obtained in Shiogama et al. (2006) with different climate models. Other than the optimal fingerprinting method, Morak et al. (2011) studied the global and regional changes in the number of warm nights annually by a least squares regression method, and concluded that the anthropogenic influence contributed to the increasing trend. Morak et al. (2012) further investigated changes in the frequency of warm and cold extremes in warm and cold seasons.

Though the optimal fingerprinting method can be applied on the extreme temperatures in the same way as for the climatic mean states, it does not take account the unique property of the extreme values. Most recently, several studies have applied the extreme value theory into the optimal detection, which has the potential to improve the signal-to-noise ratio by considering the unique distributional properties of extreme values. As discussed in Chapter 4, Zwiers et al. (2011) employed nonstationary extreme value theory in their analysis at both global and regional scale. A time evolving location parameter was also used as the signal in Christidis et al. (2011a) with a different extreme model. A threshold Generalized Pareto (GP) distribution with location parameter varies every decade was fitted at each grid and a formal optimal detection was applied. They detected the impacts from anthropogenic and natural signal in the extreme warm days. Min et al. (2013) employed the extreme value theory in standardizing indices: the extreme temperatures were converted into probability-based index through

GEV distribution; then the formal optimal fingerprinting method was applied on the standardized indices. They studied both single-signal and two-signal detection analyses, and confirmed the previous findings through a formal optimal fingerprinting method, and also first showed that ANT signals are separable from NAT signals at global and subcontinental scales. Later, Kim et al. (2015) updated the attribution analyses of extreme temperatures with the updated observations and multi-model simulations using the same analysis method in Min et al. (2013). Detection results are found to be generally insensitive to either different observed data coverage or different model samples.

This chapter is an extension of Zwiers et al. (2011). We also apply the independence method in fingerprinting, but instead of using the profile likelihood approach in the estimation, we use a coordinate descent approach. From Chapter 4, when modeling the spatial correlation by a simple working correlation structure in the CSE method, the measurement errors in covariates may lead to biased estimate. So we simply use the independence method in the analysis. The scaling factor is estimated by the CSE approach in equation (3.2) with the working independence. The profile approach fixes the value of the scaling factor within a predetermined range, and minimizes the negative log-likelihood function for each grid point. They search over the predetermined values and use the optimal one that minimizes the negative log-likelihood function as the estimate of the scaling factor. Searching among a large number of points is time consuming even for single-signal analysis: the analysis of 16 regions in Zwiers et al. (2011) took weeks on a computer cluster. Further, the profile method requires that the predetermined range

is large enough to cover the optimal solution of the scaling factor. When multiple signals are present, searching the best values within high dimensional predetermined values is computationally prohibitive for the profile approach. Thus extending the profile method to two or more signals is unpractical. In comparison, the coordinate descent approach has noticeable advantages on computation efficiency and accuracy over the profile approach. It does not need to be provided a predetermined range of possible values, and more importantly, it only takes a few seconds to solve the estimating equations to obtain the optimal solution of the scaling factor in both the single-signal and two-signal analysis, which makes analyzing the multiple signals at the same time be practical. The most important contribution of this study is the two-signal analysis. To examine whether the influence of human behavior and that of the natural factors on the extreme temperatures can be separated, we focus on the two-signal analyses in which both ANT and NAT signals are present in the location parameter of the GEV distribution at the same time, which has not been done in existing studies. Though Min et al. (2013) also studied the two-signal detection analyses, they used a standard optimal fingerprinting method, while we employ the extreme value theory such that the scaling factor is involved in the GEV parameters. Further, when using nonstationary extreme value theory to estimate the signal, the location parameters of the corresponding distributions vary for every h-yr, and most studies worked on $h = 10$ while we used $h = 5$ in our analysis. We apply the independence method on the extreme temperatures of period 1951–2010 over 17 subcontinents, analyzing whether the anthropogenic forcing and/or nature forcing

contributes to the changes in the extreme temperatures.

The outline of this chapter is as follows. The description of the observational and climate model simulation data are given in Section 5.2. Section 5.3 presents the methods of the signal estimation and detection analysis. The analysis results are discussed in detail in Section 5.4. Conclusions are given in Section 5.5.

5.2 Data Description

The same four types of extreme temperatures in Section 4.4.2 are considered: TXx, TNx, TXn, and TNn. The observations are from the HadEX2 dataset (Donat et al., 2013), with 3.75° longitude × 2.5° latitude grid, of period 1951–2010 over 17 different regions, see Table 10. The grid boxes at each region are provided that have at least 50-yr non-missing annual extreme values for the 60-yr period for each of the four types of temperature extremes. Annual extreme temperatures calculated from Coupled Model Intercomparison Project Phase 5 (CMIP5) multi-model datasets (Taylor et al., 2011) are first re-gridded to the spatial resolution of HadEX2, and then each simulation is masked by the observational coverage of each of the four types extremes to ensure consistent spatial and temporal coverage with observations. Grid boxes are then retrieved for each region using the same criteria for data availability as for the observations. We have 22 ensembles from 6 climate models under the ALL (anthropogenic and natural) forcing and 26 ensembles from 6 climate models under the NAT forcing; see Table 11. The

Table 10: The names and the corresponding acronyms of the 17 regions analyzed in the study.

Acronym	Region	Acronym	Region
ALA	Alaska	SAF	Southern Africa
CGI	Canada and Greenland	NAS	North Asia
WNA	Western North America	CAS	Central Asia
CNA	Central North America	TIB	Tibet
ENA	Eastern North America	EAS	East Asia
CAM	Central America and Mexico	SAS	South Asia
SSA	Southern South America	SEA	Southeast Asia
NEU	Northern Europe	AUS	Australia
SEU	Southern Europe		

Table 11: List of CMIP5 model analyzed in this study. The model names and ensemble sizes are given for ALL and NAT forcing. Bold font indicates ensemble sizes of extended model simulations that end in year 2010 and used for calculating signals.

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**	CM5A-MR	3	HadGEM2-ES	4
CanESM2	5	NorESM1-M	**3**	CNRM-CM5	6	IPSL-CM5A-LR	3

CMIP5 simulations typically end in 2005. In order to cover the more recent years, we use the extended simulations offered by some modelling centers at the cost of more limited availability of the simulations in two climate models.

5.3 Methods

5.3.1 Signal Estimation

The signal estimation procedure is given in Section 4.2.1. In Zwiers et al. (2011), the GEV location parameters change every 10-yr while ours change every 5-yr, i.e. $h = 5$ in model 4.1. We consider the 5-yr average because of the desire to identify the response from natural forcing predominately volcanic activity. The use of 5-yr average provides some smoothing to reduce noise and does not over smooth the NAT signal that is of intermittent nature. The 5-yr blocks are defined as 1951–1955, 1956–1960, ..., 2006–2010. The signals corresponding to the ALL forcing and NAT forcing of each of the four extreme temperatures are estimated by model 4.1. As for the signal of ANT forcing, since there is only a few climate models of ANT forcing available, to use as many as simulations, the ANT signal was obtained by subtracting NAT signal from ALL signal as has been done in other studies (e.g., Kim et al., 2015). However, simulations of daily temperatures are not always available under both the ALL and the NAT forcings for the same model. There is only one model, CanESM2, that produces extended simulations for both ALL and NAT forcings and that 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 modelling uncertainty in addition to that of ANT forcing.

5.3.2 Detection Analysis

The detection model is the same as model 4.2 in Section 4.2.2, but the solving method is different. The independence method is used in solving model (4.2). Since β is shared by all grid boxes within the region, the parameters needs to be estimated jointly. Let $\zeta_s = (\alpha_s, \sigma_s, \xi_s)$, then the negative independence log-likelihood function at a given region is:

$$L = -\sum_{s=1}^{m} l_s(\beta, \zeta_s) \tag{5.1}$$

where $l_s(\beta, \zeta_s)$ is the log-likelihood function at site s:

$$l_s(\beta, \zeta_s) = \sum_{t=1}^{n} \log f(Y_{ts};\, \alpha_s + X_{ts}^{\top}\beta, \sigma, \xi).$$

There are $3m + p$ unknown parameters in model (4.2); simultaneously estimating such large number of parameters is challenging. We use a coordinate descent approach, estimating a small set of the unknown parameters each time with others being held. The coordinate descent approach is a two-step iterative process: 1) Given current estimate $\hat{\beta}$ of β, the observed data at each grid box are independent, and $\zeta_s = (\alpha_s, \sigma_s, \xi_s)$ at each grid box can be estimated separately through likelihood approach. 2) Given current estimate $\hat{\zeta}_s$, β is the only unknown parameter; it can be estimated by minimizing the negative log-likelihood function in equation (5.1), which is equivalent to solving equation 3.2 with W_t be the identity matrix.

Zwiers et al. (2011) only studied the single-signal analyses, while in our detection analysis, besides the single-signal analyses of ALL forcing and ANT forcing, respectively, we also focus on the two-signal analyses: both ANT and NAT signals are present in the model at the same time. The two-signal analyses are conducted to examine whether the responses to different forcings can be separately detected. In two-signal detection, μ_{ts} in equation (4.2) becomes:

$$\mu_{ts} = \alpha_s + X_{A,ts}^{\top}\beta_A + X_{N,ts}^{\top}\beta_N, \tag{5.2}$$

where X_A and X_N are the relative signals of ANT and NAT forcings, respectively, and β_A and β_N are the corresponding scaling factors.

5.3.3 Uncertainty Assessment

Similar bootstrap sampling procedure in Section 4.2.3 is used to make inferences about β. However, since in signal estimation, the value of h in model 4.1 is 5 rather than 10, we need to use a different bootstrap sampling procedure of the signal than the procedure in Section 4.2.3. The following is the bootstrap sampling procedure of the signal estimate when $h = 5$.

(A) For each ensemble, transform the simulation data into Gumbel residuals by the fitted GEV parameters in model (4.1) at each grid box (Kharin and Zwiers, 2005).

(B) For each individual grid box, divide the data into nonoverlapping 5-yr blocks. Randomly sample 5-yr blocks data with replacement. All the grid boxes share the same sample order to keep the spatial dependence.

(C) Transform the reordered Gumbel residuals back into GEV distribution by the fitted GEV parameters in model (4.1), then estimate the signals from the transformed data.

(D) Repeat steps A to C 32 times.

The bootstrap sampling procedure of the observational data is the same as that in Section 4.2.3.

5.3.4 Waiting time

Zwiers et al. (2011) calculated the 1990s waiting time of the 1960s climate 20-yr return value. We calculate similar statistics, the 2006-2010 waiting time of the 1951-1955 climate 20-yr return value. The 1951-1955 climate 20-yr return value is a value of the temperature that will be exceeded once in every 20-yr on average based on the fitted GEV parameters in model (4.2) corresponding to 1951-1955 climate. The 2006-2010 waiting time of the 1951-1955 climate 20-yr return value is the inverse of the probability that the 1951-1955 climate 20-yr return value will be exceed once in every 20-yr on average based on the fitted GEV parameters in model (4.2) corresponding to 2006-2010 climate. In calculation, we first take the average of the exceeding probabilities from

all grid boxes in a region under consideration, and then the inverse of the averaged probability is used as the waiting time for this region. The difference between the 2006-2010 waiting time and 20-yr reflects the influence of the corresponding 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 recur once in 20 years) provides useful information for the assessment of impacts of climate change on extremes. Compared to 1951-1955, if the temperature in 2006-2010 increases, then we expect that the 2006-2010 waiting time will be longer than 20-yr for TXn and TNn, while shorter than 20-yr for TXx and TNx.

5.4 Results

Point estimates of the scaling factors and the corresponding 90% bootstrap confidence interval for single-signal detection of ALL forcing and ANT forcing, respectively, at 17 regions are summarized in Figure 7. From Figure 7, 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), which is in good agreement with 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.

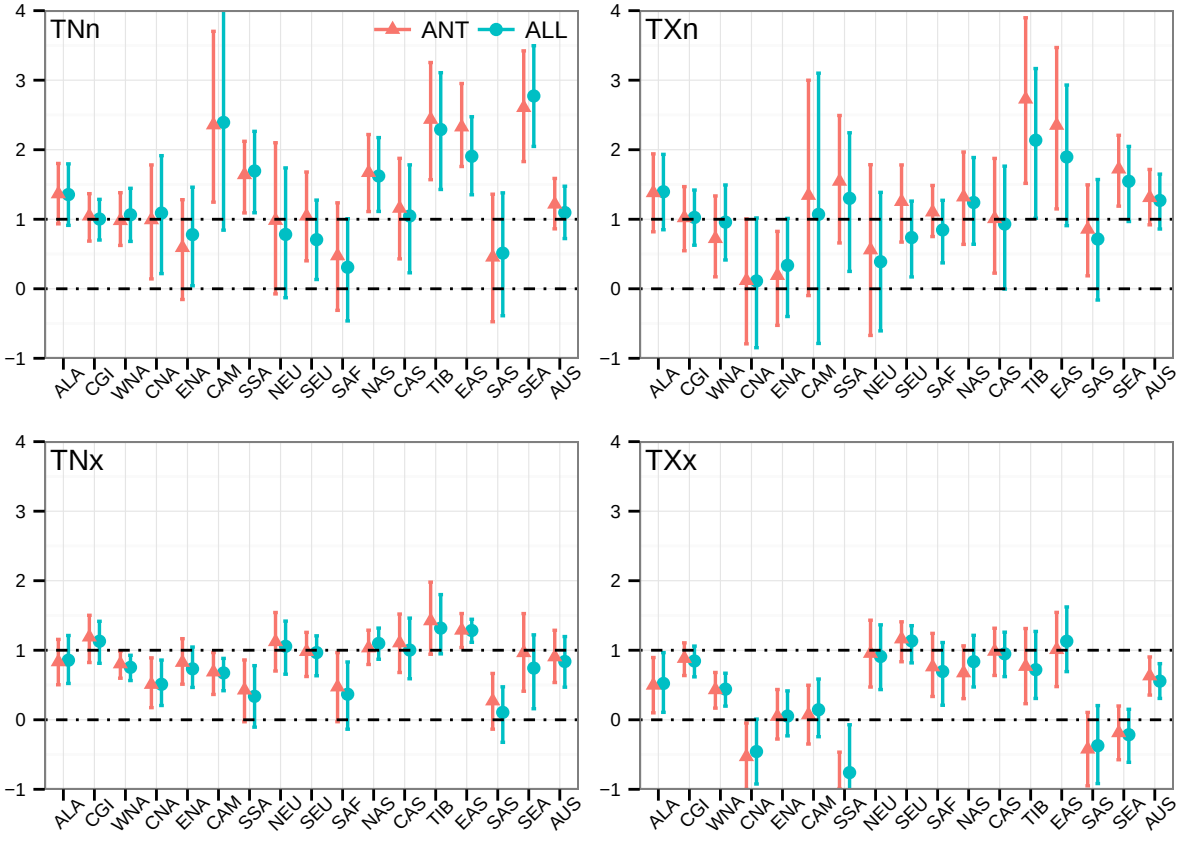

Figure 7: Point estimates of the scaling factors and the corresponding 90% confidence intervals for single-signal detection of ANT forcing and ALL forcing, respectively, at each region.

For the number of detected regions of each of the 4 types extreme temperatures, TNx is associated with most regions while TXx with least regions, which shows that the annual maximum of daily minimum temperature is the most detectable one for external forcings while the annual maximum daily maximum temperature is the least detectable One. This is in accord with the findings in Zwiers et al. (2011) and Min et al. (2013).

Figure 8 gives the results for two-signal (ANT and NAT) detection at 17 regions.

From Figure 8, the influence of ANT forcing is detected in more regions in two-signal analyses than that in single-signal analyses. The ANT signal is detected in TNx in all 17 regions. Compared to single-signal analyses, its influence is detected in two more regions for TNn, one more for TXn and one less for TXx. The influence of ANT forcing can be separable from that of NAT forcing in 4 to 6 regions for each of the 4 types of extreme temperatures. In both North Asia (NAS) and Tibet (TIB), the ANT signal is separable from NAT signal in 3 temperature indices — TNn, TNx and TXx. The detection results of two-signal analyses of TNx is in large agreement with the 2-signal detection results of TNx presented in Min et al. (2013): among our 6 detected regions and their 5 detected regions, 4 are in common — Central North America (CNA), Southern Europe (SEU), NAS and TIB.

The detection and attribution results of the single-signal and two-signal analyses for each type of extreme temperature at each region are summarized in Table 12. In Table 12, most confidence intervals that lie above zero also include unity, indicating the agreement of the fingerprints of external forcing with the observed changes. For those confidence intervals that lie above zero but not include unity, the confidence interval of TNn and TXn lies above unity, indicating that the climate models underestimate the observed annual minimum extreme temperatures; while confidence interval of TNx and TXx lies between zero and unity, implying that the climate models overestimate the annual maximum extreme temperatures. These findings are in accord with the findings in Zwiers et al. (2011), Min et al. (2013) and Kim et al. (2015).

Table 12: Analyses results for single-signal detection of ANT forcing and ALL forcing, respectively, and two-signal detection (ANT and NAT) at each region. 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; blank means the corresponding 90% confidence interval does not lie above zero.

	TNn				TXn				TNx				TXx			
Region	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	A	A	A		A	A	A	A	A	A	A		D	D	D	
CGI	A	A	A	A	A	A	A	D	A	A	A		A	A	A	
WNA	A	A	A		A	A	A	A	D	D	D		D	D	D	
CNA	A	A	A						D	D	D	A				
ENA	A		A	D			A	D	A	A	A					
CAM	A	D	A						D	D	D	A				
SSA	D	D	D	A	A	A	A				D					A
NEU			A						A	A	A		A	A	A	
SEU	A	A	A		A	A	A		A	A	A	A	A	A	A	A
SAF					A	A	A				A		A	A	A	
NAS	D	D	D	A	A	A	A		A	A	A	A	A	A	A	A
CAS	A	A	A			A	A		A	A	A		A	A	A	
TIB	D	D	D	A	D	D	D		A	A	A	D	A	A	A	A
EAS	D	D	D		A	D	A		D	D	A	D	A	A	A	D
SAS						A	A				D					
SEA	D	D	D	D	A	D	D		A	A	A					
AUS	A	A	A		A	A	A		A	A	A		D	D		

Table 13: Summaries of total grid boxes, the percentage of the grid boxes that fail the goodness-of-fit test at 10% significance level corresponding to ALL forcing, ANT forcing, and ANT and NAT forcing, respectively, at each region. The percentages that larger than 20% are in bold font.

	Total grids				ALL				ANT				ANT & NAT			
Region	TNn	TXn	TNx	TXx	TNn	TXn	TNx	TXx	TNn	TXn	TNx	TXx	TNn	TXn	TNx	TXx
ALA	73	73	73	73	2.7	0	0	2.7	4.1	0	0	2.7	2.7	0	0	2.7
CGI	64	61	31	41	4.2	0	0	0	1.8	0	0	0	3.0	2.4	0	0
WNA	29	29	25	28	0	0	0	0	0	0	0	0	0	0	0	0
CNA	85	85	62	71	0	0	0	0	0	0	0	0	0	0	0	0
ENA	167	167	95	114	0	16.1	0	0	0	19.4	0	0	0	6.5	0	0
CAM	43	43	43	43	0	3.4	8.0	14.3	0	3.4	8.0	10.7	0	3.4	12.0	7.1
SSA	94	95	89	91	9.3	2.3	7.7	0	7.0	0	3.8	0	7.0	2.3	7.7	0
NEU	31	31	31	31	4.5	3.0	3.0	11.9	3.0	7.5	1.5	14.9	0	0	3.0	6.0
SEU	277	277	265	246	0	0	2.2	3.9	0	0	2.2	2.0	0	0	4.3	3.9
SAF	67	67	67	67	7.1	7.4	11.1	10	7.1	3.7	11.1	10	7.1	7.4	11.1	20
NAS	28	27	18	20	0.7	3.6	0.4	0.4	1.4	2.9	1.1	0.4	1.4	3.2	0.4	1.2
CAS	48	48	24	33	2.4	1.2	1.6	2.8	2.4	1.2	1.6	1.4	3.5	1.2	1.6	4.2
TIB	19	19	15	17	19.0	14.3	16.1	4.9	19.0	12.7	8.9	6.6	**20.6**	14.3	10.7	3.3
EAS	64	64	46	51	3.2	0	4.5	2.2	6.4	2.1	3.4	2.2	6.4	1.1	5.6	1.1
SAS	43	43	26	32	**27.1**	18.8	**37.5**	12.1	**25.0**	**25.0**	**37.5**	12.1	**29.2**	**25.0**	**41.7**	15.2
SEA	63	63	56	61	21.1	0	**26.7**	0	**26.3**	0	**26.7**	0	15.8	0	20	5.9
AUS	75	75	75	75	1.6	1.6	0	0	1.6	1.6	0	0	3.1	3.3	0	0

To assess the fitness of GEV distribution in model 4.2, the goodness-of-fit test is performed at each grid box for single-signal analyses and two-signal analyses. The percentage of the grid boxes that fail the goodness-of-fit test at 10% significance level as well as the total number of grid boxes are summarized in Table 13. From Table 13, GEV distribution with the location parameters specified in model 4.2 is an appropriate distribution in general except for TNn, TXn and TNx in South Asia (SAS) for both single-signal and two-signal analyses, and TNn and TNx in Southeast Asia (SEA) for single-signal analyses.

Figure 9 gives the point estimates and the 90% confidence intervals of the 2006-2010 waiting time of the 1951-1955 climate 20-yr return value in two-signal analyses at 17 regions. Most confidence intervals of the waiting time do not cover 20-yr, indicating that the waiting times of most regions are estimated to have changed substantially between 1951-1955 and 2006-2010. The confidence intervals for TNx are significant in all regions, except for CAM, and the significant confidence intervals all lie below 20-yr, implying the decreases in waiting times for extreme annual maximum daily minimum temperature. For TXx, the waiting times are significantly less than 20-yr in 11 regions, and significantly larger than 20-yr in 2 regions. The confidence intervals lie above 20-yr in 11 regions for TXn and in 13 regions for TNn, which means the waiting times for the recurrence of the annual minimum extreme temperatures have increased. The increasing amounts are quite large in some regions, for example, the 90% confidence interval in ALA is (350, 32304) for TXn and (687, 31835) for TNn, indicating a huge

change in the extreme annual minimum temperatures that can be attributed to external forcings. The result of the waiting times corresponds to the reported global warming, and it shows that the changes in the climate extremes are substantially influenced by the ANT and NAT forcing, especially for the annual minimum extreme temperatures. The conclusions are in good agreement with Zwiers et al. (2011), except that their differences between the estimated waiting times and 20-yr are smaller than the differences in our case. This is because our waiting time is for 2006-2010 while theirs are for 1991-2000, and the global temperatures keep warming in the 21st century. The 10-yr blocks tend to smooth the changes more than 5-yr block. The differences also come from the fact that our waiting times are from two-signal analyses while theirs are from single-signal analyses.

5.5 Conclusions

This study compares the HadEX2 observation with the CMIP5 multi-model datasets in four extreme temperature indices during 1951-2010 at 17 subcontinents. 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 estimated signals of the ANT forcing, ALL forcing, and ANT forcing and NAT forcing. A coordinate descent algorithm is used in the estimation of the scaling factor from an independence likelihood. This study is an extension of Zwiers et al. (2011); the main difference is that we additionally

perform the two-signal detection analyses made possible by the coordinate descent algorithm. For signal analyses, the algorithm has higher computing effiency and accuracy than the profile likelihood approach used in Zwiers et al. (2011).

The results of single-signal analyses show that the ANT signal and the ALL signal are clearly detected in many subcontinents, with more frequently detection in TNx. The annual maximum extreme temperatures are found to be overestimated by the climate model, while the annual minimum extreme temperatures are underestimated by the climate model. These findings are in agreement with earlier studies (Zwiers et al., 2011; Min et al., 2013).

The two-signal detection analyses under the extreme value theory is the most important contribution of this study. Though other studies (Min et al., 2013; Kim et al., 2015) also performed the two-signal detection, their method was the classical fingerprinting method. Our results show that ANT signal is separable from NAT signal over 4-6 regions for each temperature indices. The reported detected regions in TNx is in large agreement with the results in Min et al. (2013). The results of the 2006-2010 waiting time of the 1951-1955 climate 20-yr return value in the two-signal analyses correspond to the trend of global warming. The waiting times for the temperature of coldest night and coldest day are found to have increased by a large amount over 20-yr. The waiting times for the temperature of the warmest night and warmest day are found to have dropped below 20-yr in a majority of the regions. From these evidences, the influence of human behavior and natural factor contributed to the observed changes in the annual extreme

temperatures, especially in the coldest night and day temperatures.

Spatial dependence has not been considered here as the formal detection analyses (e.g., Allen and Stott, 2003). Wang et al. (2015) has taken into account the spatial dependence to increase the estimation efficiency with a CSE method, which is a generalization of the coordinate descent approach we used in this study. The CSE method only assumes the marginal GEV distribution is correctly specified, The scores of the marginal GEV distribution at each site are combined in some optimal way to improve the estimation efficiency. A simple working correlation structure is be used to account for the spatial correlation. However, since the signals are estimated as opposed to known for certain, the measurement errors in covariates (the signals) lead to biased regression coefficient estimation, and this method needs further investigation.

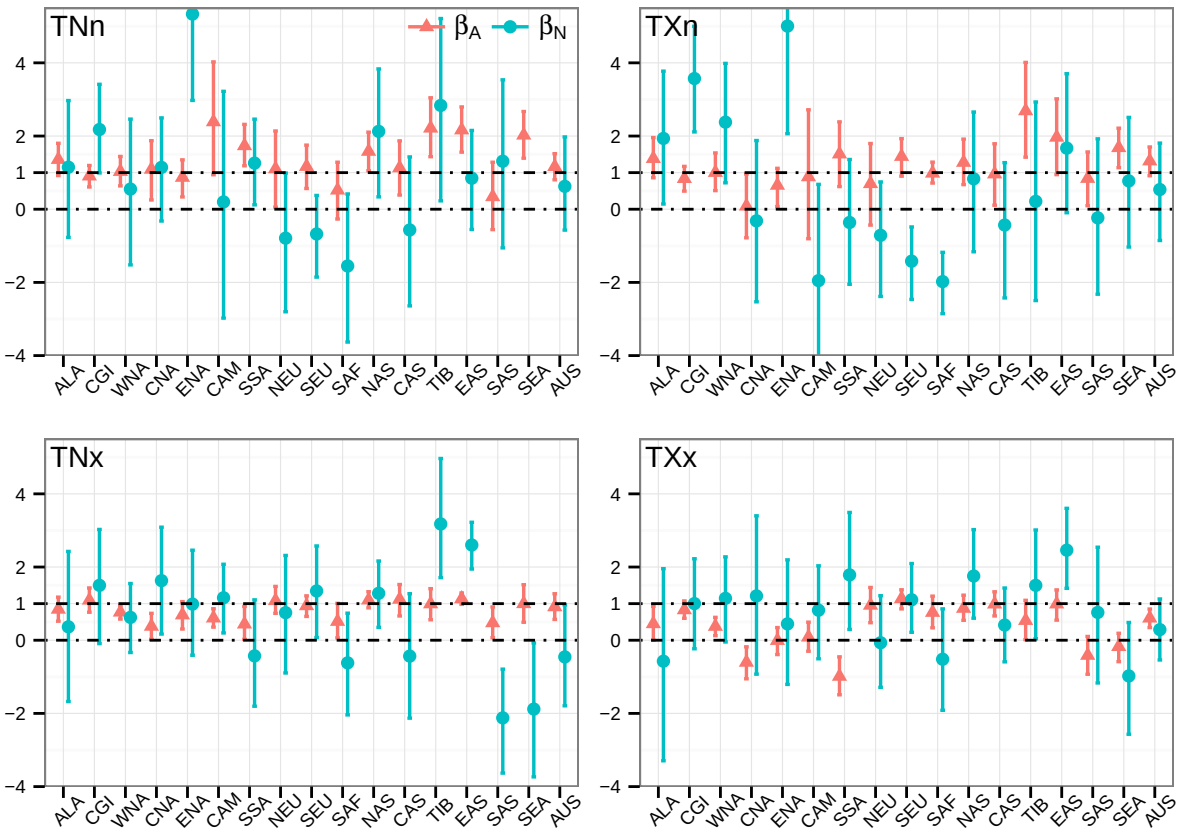

Figure 8: Point estimates of the scaling factors and the corresponding 90% confidence intervals for two-signal (ANT and NAT) detection at each region. Parameters β_A and β_N are the scaling factors corresponding to ANT forcing and NAT forcing, respectively. (Note: the 90% confidence interval of β_N for TNn in SEA region is (7.23, 14.90), which is too large to be plotted in the figure.)

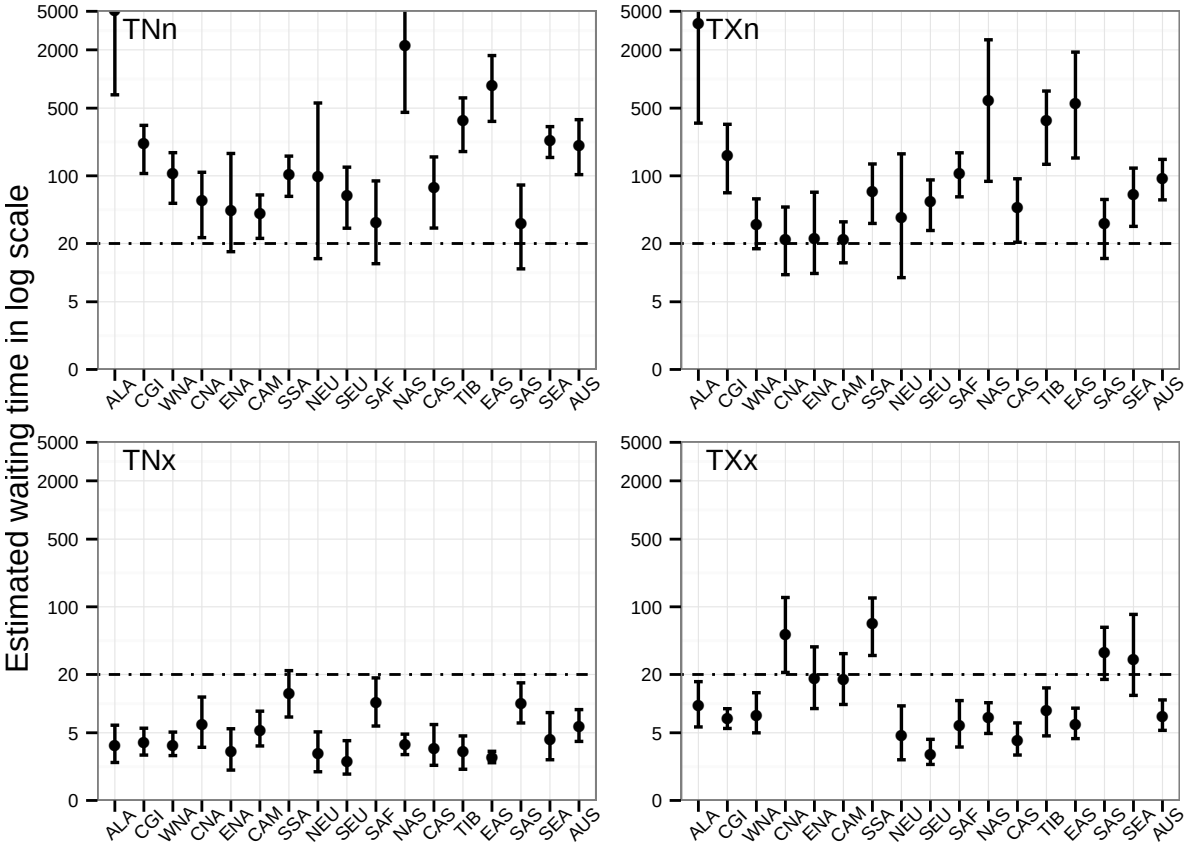

Figure 9: Point estimates and the 90% confidence intervals of the 2006-2010 waiting time of the 1951-1955 climate 20-yr return value in log scale for two-signal (ANT and NAT) detection at each region.

Chapter 6

Conclusions

6.1 Summary

This dissertation starts with the application of max-stable process with pairwise likelihood in RFA in Chapter 2. From the simulation study, when the dependence structure is correctly specified, the PL method has the highest estimation efficiency compared to the LM and IL method. However, when the spatial dependence is misspecified, the PL estimate has serious bias. In many studies, such as the RFA and the fingerprinting method in climate extremes, the primary interest is the inference about the marginal GEV distributions while the spatial dependence is a nuisance. Thus, we are motivated to find a way to account for the spatial dependence without requiring parametric specifications beyond univariate marginal GEV distributions; that is the CSE approach proposed in Chapter 3. The CSE method only assumes the marginal GEV distribution is correctly specified. The scores of the marginal GEV distribution at each site are combined in some optimal way to improve the estimation efficiency. A simple working correlation structure is used to model the spatial correlation among the site-wise score functions.

The CSE method with working exponential correlation structure is applied in the

detection and attribution of changes in extreme temperatures in Chapter 4. Compared to the IL method, the CSE method with spatial correlation has smaller estimation uncertainty in the simulation study, but it may introduce serious bias when the measurement errors exist in the covariates (signals), which is because the signals are estimated as opposed to known for certain. So in Chapter 5, we applied the CSE approach with a working independence structure, which reduces to the IL method, in the detection and attribution of the changes in four extreme temperature indices at 17 subcontinents. Though the CSE approach with working independence is conceptually the same as the profile independence likelihood method used in existing study, the CSE approach has higher computing efficiency and accuracy, and it is computationally feasible. The main contribution of Chapter 5 is the two-signal detection analyses under the extreme value theory with GEV location parameters proportional to the ANT signal and NAT signal at the same time, which has not been done by previous studies. The influences of the anthropogenic forcing and the combination of the anthropogenic and natural forcing are detected in regional scale separately. The anthropogenic signal can be separable from the natural signal in a few subcontinents.

6.2 Future Work

Applications of CSE in RFA. The CSE method can be applied in some other studies. It can be applied in RFA to model the intersite dependence across different

sites, since the goal of an RFA is usually to estimate marginal return levels and the dependence is a nuisance. An RFA with the GEV location parameter as the index variable would put constraints that $\mu_s/\sigma_s = \gamma$ at each site s, $s = 1, \ldots, m$, and ξ be fixed across the sites within the specific region. The regression model is a GEV distribution with parameters:

$$\begin{cases} \mu_{ts} = \gamma\sigma_s, \\ \sigma_{ts} = \sigma_s, \\ \xi_{ts} = \xi, \end{cases}$$

The above model can be solved by a two-step approach similar to Section 4.2.2. The common parameters γ and ξ can be estimated through CSE approach, and the sitewise parameter σ_s can be estimated through likelihood approach.

The L-moment method gives efficient estimator for small samples in RFA since it constrains the GEV shape parameter to be less than 1 for the existence of the L-moments. Similarly, we can put constraints on the shape parameter in the CSE approach to improve the estimation efficiency when the sample size is small. Coles and Dixon (1999) used a penalized likelihood approach to constrain the GEV shape parameter by a prior distribution, a Beta distribution. We can use similar prior distributions in the CSE approach to constrain the range of shape parameter.

Applications of optimal fingerprinting with precipitation extremes. Another important application of the CSE method is the detection and attribution of the extreme precipitation changes. Human behavior has been contributed to the increase in the extreme temperatures in our study, then consequently, the extreme precipitation is very likely to increase since precipitation closely relates to the temperature. The optimal fingerprinting has been used in the detection of human influence in the extreme precipitation. And the anthropogenic influence has been detected in many aspects of the precipitation, for example, the annual and seasonal precipitation over global land areas (Noake et al., 2012; Marvel and Bonfils, 2013) and the high-latitude precipitation (Min et al., 2011; Wan et al., 2014). However, none of the methods used in the existing studies employ the extreme value theory. Some studies (Zhang et al., 2013) only use the GEV distribution in transforming extreme precipitation into the probability-based index, and apply the classical optimal fingerprinting method on the probability-based index. It will be worth applying the CSE method in the detection of the changes in precipitation.

Methodological development. Several directions of the CSE approach needs further investigation. First, to account for the spatial dependence, the CSE method need to model the dependence by a one-parameter working correlation. Stoner and Leroux (2002) proposed a combined estimating equations approach for cluster data that avoids correlation parameter estimation. Several contrasts of the data are optimally weighted

and combined through a combination of estimating equations to increase the estimation efficiency relative to GEE. Second, in the CSE approach, the estimate is obtained by solving the equations. One numerical problem that may arise in solving the equation is the multiple root. Minimizing the objective function can avoid such problem. Qu et al. (2000) used a method of quadratic inference functions that is more efficient in parameter estimation than GEE when the working structure is misspecified. This method does not involve direct estimation of the correlation parameter; the inverse of the working correlation is modeled by a linear combination of basis matrices, such as the equicorrelated matrix and autoregressive correlation matrix. Bai et al. (2012) proposed a method that similar to the quadratic inference function. They propose a joint composite estimating function approach, combining three sets of estimating functions from spatial, temporal and spatiotemporal cross-pairs, to estimate the spatiotemporal covariance structures. Similar ideas can be used to combine the score equations and further improve the estimation efficiency.

Bibliography

Allen, M. R. and P. A. Stott (2003). Estimating signal amplitudes in optimal fingerprinting, part i: theory. *Climate Dynamics 21*, 477–491.

Allen, M. R. and S. F. B. Tett (1999). Checking for model consistency in optimal fingerprinting. *Climate Dynamics 15*, 419–434.

Bai, Y., P. X.-K. Song, and T. E. Raghunathan (2012). Joint composite estimating functions in spatiotemporal models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 74*(5), 799–824.

Banerjee, S., B. P. Carlin, and A. E. Gelfand (2004). *Hierarchical modeling and analysis for spatial data*. Chapman & Hall / CRC.

Bindoff, N. L., P. A. Stott, K. M. AchutaRao, M. R. Allen, N. Gillett, D. Gutzler, K. Hansingo, G. Hegerl, Y. Hu, S. Jain, I. I. Mokhov, J. Overland, J. Perlwitz, R. Sebbari, and X. Zhang (2013). Detection and attribution of climate change: From global to regional. In 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 (Eds.), *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, pp. 867–952. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

Blanchet, J. and A. C. Davison (2011). Spatial modeling of extreme snow depth. *Annals of Applied Statistics 5*(3), 1699–1725.

Brown, S. J., J. Caesar, and C. A. T. Ferro (2008). Global changes in extreme daily temperature since 1950. *Journal of Geophysical Research 113*.

Buishand, T. A. (1991). Extreme rainfall estimation by combining data from several sites. *Hydrological Sciences Journal 36*, 345–365.

Casson, E. and S. Coles (1999). Spatial regression models for extremes. *Extremes 1*, 449–468.

Castellarin, A. (2007). Probabilistic envelope curves for design flood estimation at ungauged sites. *Water Resources Research 43*(4), n/a–n/a.

Castellarin, A., D. Burn, and A. Brath (2008). Homogeneity testing: How homogeneous do heterogeneous cross-correlated regions seem? *Journal of Hydrology 360*(1-4), 67–76.

Castellarin, A., R. M. Vogel, and N. C. Matalas (2005). Probabilistic behavior of a regional envelope curve. *Water Resources Research 41*(6), n/a–n/a.

Chaganty, N. R. and H. Joe (2004). Efficiency of generalized estimating equations for binary responses. *Journal of the Royal Statistical Society, Series B: Statistical Methodology 66*(4), 851–860.

Chebana, F. and T. Ouarda (2009). Index flood–based multivariate regional frequency analysis. *Water Resources Research 45*(10), W10435.

Christidis, N., P. Stott, and S. J. Brown (2011a). The role of human activity in the recent warming of extremely warm daytime temperatures. *Journal of Climate 24*, 1922–1930.

Christidis, N., P. A. Stott, S. Brown, G. C. Hegerl, and J. Caesar (2005). Detection of changes in temperature extremes during the second half of the 20th century. *Geophysical Research Letters 32*(20), 1–4. L20716.

Christidis, N., P. A. Stott, and S. J. Brown (2011b). The role of human activity in the recent warming of extremely warm daytime temperatures. *Journal of Climate 24*(7), 1922–1930.

Clayton, M. K. and P.-S. Lin (2005). Analysis of binary spatial data by quasi-likelihood estimating equations. *The Annals of Statistics 33*(2), 542–555.

Coles, S. G. and M. J. Dixon (1999). Likelihood-based inference for extreme value models. *Extremes 2*, 5–23.

Cooley, D., D. Nychka, and P. Naveau (2007). Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association 102*(479), 824–840.

Cunderlik, J. M. and T. B. Ouarda (2006). Regional flood-duration–frequency modeling in the changing environment. *Journal of Hydrology 318*(1), 276–291.

Davison, A. C. and M. M. Gholamrezaee (2009). Geostatistics of extremes. Submitted manuscript, École polytechnique fédérale de Lausanne.

Davison, A. C. and M. M. Gholamrezaee (2012). Geostatistics of extremes. *Proceedings of the Royal Society A 468*, 581–608.

Davison, A. C., S. A. Padoan, and M. Ribatet (2012). Statistical modeling of spatial extremes. *Statistical Science 27*(2), 161–186.

de Haan, L. (1984). A spectral representation for max-stable processes. *The Annals of Probability 12*, 1194–1204.

Donat, M. G., L. V. Alexander, H. Yang, I. Durre, R. Vose, R. J. H. Dunn, K. M. Willett, E. Aguilar, M. Brunet, J. Caesar, B. Hewitson, C. Jack, A. M. G. Klein Tank, A. C. Kruger, J. Marengo, T. C. Peterson, M. Renom, C. Oria Rojas, M. Rusticucci, J. Salinger, A. S. Elrayah, S. S. Sekele, A. K. Srivastava, B. Trewin, C. Villarroel, L. A. Vincent, P. Zhai, X. Zhang, and S. Kitching (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*(5), 2098–2118.

Embrechts, P., A. McNeil, and D. Straumann (2002). Correlation and dependence in risk management: Properties and pitfalls. In M. A. H. Dempster (Ed.), *Risk management: value at risk and beyond*, pp. 176–223. Cambridge University Press, Cambridge.

Field, C. B., V. Barros, T. F. Stocker, D. Qin, D. J. Dokken, K. L. Ebi, M. D. Mastrandrea, K. J. Mach, G.-K. Plattner, S. K. Allen, M. Tignor, and P. M. Midgley (Eds.) (2012). *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change*. Cambridge, UK and New York, NY, USA: Cambridge University Press.

Fuentes, M., J. Henry, and B. Reich (2013). Nonparametric spatial models for extremes: Application to extreme temperature data. *Extremes 16*(1), 75–101.

Genest, C. and A.-C. Favre (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrological Engineering 12*, 347–368.

Griffis, V. and J. Stedinger (2007). The use of {GLS} regression in regional hydrologic analyses. *Journal of Hydrology 344*(1-2), 82 – 95.

Gudendorf, G. and J. Segers (2010). Extreme-value copulas. In P. Jaworski, F. Durante, W. K. Härdle, and T. Rychlik (Eds.), *Copula Theory and Its Applications*, pp. 127–145. Springer Media.

Hanel, M., T. A. Buishand, and C. Ferro (2009). A nonstationary index flood model for precipitation extremes in transient regional climate model simulations. *Journal of Geophysical Research 114*.

Hasselman, B. (2014). *nleqslv: Solve systems of non linear equations*. R package version 2.1.1.

Hasselmann, K. (1997). Multi-pattern fingerprint method for detection and attribution of climate change. *Climate Dynamics 13*, 601–611.

Heffernan, J. E. and J. A. Tawn (2004). A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 66*(3), 497–546.

Hegerl, G. and F. Zwiers (2011). Use of models in detection and attribution of climate change. *Wiley Interdisciplinary Reviews: Climate Change 2*(4), 570–591.

Hegerl, G., F. Zwiers, P. Stott, and V. Kharin (2004). Detectability of anthropogenic changes in annual temperature and precipitation extremes. *Journal of Climate 17*(19), 3683–3700.

Hegerl, G. C., 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. In T. F. Stocker, C. B. Field, D. Qin, V. Barros, G.-K. Plattner, M. Tignor, P. M. Midgley, and K. L. Ebi (Eds.), *Meeting Report of the Intergovernmental Panel on Climate Change Expert Meeting on Detection and Attribution of Anthropogenic Climate Change*, Bern, Switzerland. IPCC Working Group I Technical Support Unit, University of Bern.

Hegerl, G. C., F. W. Zwiers, P. Braconnot, N. P. Gillett, Y. Luo, J. A. Marengo Orsini, N. Nicholls, J. E. Penner, and P. Stott (2007). Understanding and attributing climate change. In S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, and H. L. Miller (Eds.), *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, pp. 663–745. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

Hosking, J. R. M. (1990). *L*-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society, Series B: Methodological 52*, 105–124.

Hosking, J. R. M. and J. R. Wallis (1988). The effect of intersite dependence on regional flood frequency analysis. *Water Resources Research 24*, 588–600.

Hosking, J. R. M. and J. R. Wallis (1993). Some statistics useful in regional frequency analysis. *Water Resources Research 29*(2), 271–281.

Hosking, J. R. M. and J. R. Wallis (1997). *Regional Frequency Analysis: an Approach Based on L-moments*. Cambridge University Press.

IPCC (2014). *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]*. IPCC, Geneva, Switzerland.

Javelle, P., T. B. Ouarda, M. Lang, B. Bobée, G. Galéa, and J.-M. Grésillon (2002). Development of regional flood-duration–frequency curves based on the index-flood method. *Journal of Hydrology 258*(1), 249–259.

Kharin, V. and F. Zwiers (2005). Estimating extremes in transient climate change simulations. *Journal of Climate 18*(8), 1156–1173.

Kim, Y.-H., S.-K. Min, X. Zhang, F. Zwiers, L. Alexander, M. Donat, and Y.-S. Tung (2015). Attribution of extreme temperature changes during 1951-2010. *Climate Dynamics*, 1–14.

Kojadinovic, I., J. Segers, and J. Yan (2011). Large-sample tests of extreme-value dependence for multivariate copulas. *Canadian Journal of Statistics 39*(4), 703–720.

Kojadinovic, I., H. Shang, and J. Yan (2015). A class of goodness-of-fit tests for spatial extremes models based on max-stable processes. *Statistics and Its Interfaces 8*(1), 45–62.

Kysely, J. and J. Picek (2007). Regional growth curves and improved design value estimates of extreme precipitation events in the Czech Republic. *Climate Research 33*(3), 243–255.

Leclerc, M. and T. B. Ouarda (2007). Non-stationary regional flood frequency analysis at ungauged sites. *Journal of Hydrology 343*(3), 254–265.

Ledford, A. W. and J. A. Tawn (1996). Statistics for near independence in multivariate extreme values. *Biometrika 83*, 169–187.

Ledford, A. W. and J. A. Tawn (1997). Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society, Series B: Methodological 59*, 475–499.

Lee, Y. and J. A. Nelder (2009). Reply to comments on "Likelihood Inference for Models with Unobservables: Another View" (Pkg: P255-302). *Statistical Science 24*(3), 294–302.

Lettenmaier, D. P., J. Wallis, and E. Wood (1987). Effect of regional heterogeneity on flood frequency estimation. *Water Resources Research 23*, 313–323.

Liang, K.-Y. and S. L. Zeger (1986). Longitudinal data analysis using generalized linear models. *Biometrika 73*, 13–22.

Lin, P.-S. (2008). Estimating equations for spatially correlated data in multidimensional space. *Biometrika 95*(4), 847–858.

Lin, P.-S. (2010). A working estimating equation for spatial count data. *Journal of Statistical Planning and Inference 140*(9), 2470–2477.

Lindsay, B. G. (1988). Composite likelihood methods. In N. U. Prabhu (Ed.), *Statistical Inference from Stochastic Processes*, pp. 221–239. American Mathematical Society.

Lindsey, J. K. and P. Lambert (1998). On the appropriateness of marginal models for repeated measurements in clinical trials. *Statistics in Medicine 17*, 447–469.

Martins, E. S. and J. R. Stedinger (2000). Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resources Research 36*(3), 737–744.

Marvel, K. and C. Bonfils (2013). Identifying external influences on global precipitation. *Proceedings of the National Academy of Sciences 110*(48), 19301–19306.

Matalas, N. C. and W. B. Langbein (1962). Information content of the mean. *Journal of Geophysical Research 67*(9), 3441–3448.

Min, S.-K., X. Zhang, F. Zwiers, P. Friederichs, and A. Hense (2009). Signal detectability in extreme precipitation changes assessed from twentieth century climate simulations. *Climate Dynamics 32*(1), 95–111.

Min, S.-K., X. Zhang, and F. W. Zwiers (2008). Human induced arctic moistening. *Science 320*, 518–520.

Min, S.-K., X. Zhang, F. W. Zwiers, and G. C. Hegerl (2011). Human contribution to more-intense precipitation extremes. *Nature 470*, 378–381.

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. *Nature 26*, 7430–7451.

Morak, S., G. C. Hegerl, and N. Christidis (2012). Detectable changes in the frequency of temperature extremes. *Journal of Climate 26*, 1561–1574.

Morak, S., G. C. Hegerl, and J. Kenyon (2011). Detectable regional changes in the number of warm nights. *Geophysical Research Letters 38*(17), 1–5.

Ngongondo, C. S., C.-Y. Xu, L. M. Tallaksen, B. Alemaw, and T. Chirwa (2011). Regional frequency analysis of rainfall extremes in Southern Malawi using the index rainfall and l-moments approaches. *Stochastic Environmental Research and Risk Assessment 25*(7), 939–955.

Nikoloulopoulos, A. K., H. Joe, and N. R. Chaganty (2011). Weighted scores method for regression models with dependent data. *Biostatistics 12*, 653–665.

Noake, K., D. Polson, G. Hegerl, and X. Zhang (2012). Changes in seasonal land precipitation during the latter twentieth-century. *Geophysical Research Letters 39*(3).

Northrop, P. (2004). Likelihood-based approaches to flood frequency estimation. *Journal of Hydrology 292*, 96–113.

Ouarda, T., J. Cunderlik, A. St-Hilaire, M. Barbet, P. Bruneau, and B. Bobée (2006). Data-based comparison of seasonality-based regional flood frequency methods. *Journal of Hydrology 330*(1), 329–339.

Ouarda, T. B., M. Haché, P. Bruneau, and B. Bobée (2000). Regional flood peak and volume estimation in Northern Canadian Basin. *Journal of Cold Regions Engineering 14*(4), 176–191.

Ouarda, T. B. M. J. (2013). Hydrological frequency analysis, regional. In A. El-Shaarawi and W. Piegorsch (Eds.), *Encyclopedia of Environmetrics* (2 ed.). John Wiley & Sons, Ltd.

Padoan, S. A., M. Ribatet, and S. A. Sisson (2010). Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association 105*(489), 263–277.

Polson, D., G. C. Hegerl, X. Zhang, and T. J. Osborn (2013). Causes of robust seasonal land precipitation changes. *Journal of Climate 26*(17), 6679–6697.

Prescott, P. and A. T. Walden (1980). Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. *Biometrika 67*(3), 723–724.

Qu, A., B. G. Lindsay, and B. Li (2000). Improving generalised estimating equations using quadratic inference functions. *Biometrika 87*(4), 823–836.

Resnick, S. (2002). Hidden regular variation, second order regular variation and asymptotic independence. *Extremes 5*, 303–336.

Ribatet, M. and R. Singleton (2013). *SpatialExtremes: Modelling Spatial Extremes*. R package version 1.9-1.

Ribes, A., S. Planton, and L. Terray (2013). Application of regularised optimal fingerprinting to attribution. Part I: Method, properties and idealised analysis. *Climate dynamics 41*(11-12), 2817–2836.

Sabo, R. T. and N. R. Chaganty (2010). What can go wrong when ignoring correlation bounds in the use of generalized estimating equations. *Statistics in Medicine 29*(24), 2501–2507.

Sang, H. and A. E. Gelfand (2009). Hierarchical modeling for extreme values observedx over space and time. *Environmental and Ecological Statistics 16*(3), 407–426.

Schlather, M. (2002). Models for stationary max-stable random fields. *Extremes 5*(1), 33–44.

Schlather, M. and J. A. Tawn (2003). A dependence measure for multivariate and spatial extreme values: Properties and inference. *Biometrika 90*(1), 139–156.

Seneviratne, S. I., N. Nicholls, D. Easterling, C. M. Goodess, S. Kanae, J. Kossin, Y. Luo, J. Marengo, K. McInnes, M. Rahimi, M. Reichstein, A. Sorteberg, C. Vera, , and X. Zhang (2012). Changes in climate extremes and their impacts on the natural physical environment. In C. B. Field, V. Barros, T. F. Stocker, D. Qin, D. J. Dokken, K. L. Ebi, M. D. Mastrandrea, K. J. Mach, G.-K. Plattner, S. K. Allen, M. Tignor, and P. M. Midgley (Eds.), *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change*, pp. 109–230. Cambridge, UK and New York, NY, USA: Cambridge University Press.

Shang, H., J. Yan, and X. Zhang (2015). A two-step approach to model california extreme precipitation. *The Annals of Applied Statistics 8*(1), 45–62.

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.

Sklar, A. W. (1959). Fonctions de répartition à n dimension et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris 8*, 229–231.

Smith, E. L. and A. G. Stephenson (2009). An extended Gaussian max-stable process model for spatial extremes. *Journal of Statistical Planning and Inference 139*(4), 1266–1275.

Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika 72*(1), 67–90.

Smith, R. L. (1990a). Max-stable processes and spatial extremes. Unpublished Manuscript.

Smith, R. L. (1990b). Regional estimation from spatially dependent data. Unpublished manuscript, University of Chapel Hill.

Stedinger, J. R. (1983). Estimating a regional flood frequency distribution. *Water Resources Research 19*(2), 503–510.

Stone, D. and M. Allen (2005). The end-to-end attribution problem: From emissions to impacts. *Climatic Change 71*(3), 303–318.

Stoner, J. A. and B. G. Leroux (2002). Analysis of clustered data: A combined estimating equations approach. *Biometrika 89*(3), 567–578.

Stott, P. A., N. P. Gillett, G. C. Hegerl, D. J. Karoly, D. A. Stone, X. Zhang, and F. Zwiers (2010). Detection and attribution of climate change: a regional perspective. *Wiley Interdisciplinary Reviews: Climate Change 1*(2), 192–211.

Sun, Y., X. Zhang, F. W. Zwiers, L. Song, H. Wan, T. Hu, H. Yin, and G. Ren (2014). Rapid increase in the risk of extreme summer heat in Eastern China. *Nature Climate Change 4*, 1082–1085.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl (2011). An overview of cmip5 and the experiment design. *Bull.amer.meteor.soc 93*(4), 485–498.

Troutman, B. M. and M. R. Karlinger (2003). Regional flood probabilities. *Water Resources Research 39*(4), n/a–n/a.

Tseng, P. (2001). Convergence of a block coordinate descent method for nondifferentiable minimization. *Journal of Optimization Theory and Applications 109*(3), 475–494.

Tseng, P. and S. Yun (2009). A coordinate gradient descent method for nonsmooth separable minimization. *Mathematical Programming 117*(1-2), 387–423.

Varin, C. (2008). On composite marginal likelihoods. *Advances in Statistical Analysis 92*(1), 1–28.

Varin, C., N. Reid, and D. Firth (2011). An overview of composite likelihood methods. *Statistica Sinica 21*(1), 5–42.

Viglione, A., A. Castellarin, M. Rogger, R. Merz, and G. Blöschl (2012). Extreme rainstorms: Comparing regional envelope curves to stochastically generated events. *Water Resources Research 48*(1), n/a–n/a.

Wan, H., X. Zhang, F. Zwiers, and S.-K. Min (2014). Attributing northern high-latitude precipitation change over the period 19662005 to human influence. *Climate Dynamics*, 1–14.

Wang, Z., J. Yan, and X. Zhang (2014). Incorporating spatial dependence in regional frequency analysis. *Water Resources Research 50*(12), 9570–9585.

Wang, Z., J. Yan, and X. Zhang (2015). Optimal fingerprinting in detection and attribution of changes in climate extremes with combined score equations. Manuscript submitted for publication.

Wen, Q. H., X. Zhang, Y. Xu, and B. Wang (2013). Detecting human influence on extreme temperatures in China. *Geophysical Research Letters 40*(6), 1171–1176.

Westra, S. and S. A. Sisson (2011). Detection of non-stationarity in precipitation extremes using a max-stable process model. *Journal of Hydrology 406*(1), 119–128.

Willett, K. M., N. P. Gillett, P. D. Jones, and P. W. Thorne (2007). Attribution of observed surface humidity changes to human influence. *Nature 449*(7163), 710–712.

Yasui, Y. and S. Lele (1997). A regression method for spatial disease rates: An estimating function approach. *Journal of the American Statistical Association 92*, 21–32.

Zhang, X., H. Wan, F. W. Zwiers, G. C. Hegerl, and S.-K. Min (2013). Attributing intensification of precipitation extremes to human influence. *Geophysical Research Letters 40*(19), 5252–5257.

Zhang, X., F. W. Zwiers, G. C. Hegerl, F. H. Lambert, N. P. Gillett, S. Solomon, P. A. Stott, and T. Nozawa (2007). Detection of human influence on 20th century precipitation trends. *Nature 448*, 461–465.

Zwiers, F. W., X. Zhang, and Y. Feng (2011). Anthropogenic influence on long return period daily temperature extremes at regional scales. *Journal of Climate 24*(3), 881–892.