# Max-stable Processes for Threshold Exceedances in Spatial Extremes

Soyoung Jeon

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Statistics and Operations Research.

Chapel Hill
2012

Approved by:

Richard L. Smith

Douglas G. Kelly

Jan Hannig

Jason Fine

Shankar Bhamidi

**ABSTRACT**
**Max-stable Processes for Threshold Exceedances in Spatial Extremes**
**(Under the direction of Richard L. Smith)**

The analysis of spatial extremes requires the joint modeling of a spatial process at a large number of stations. Multivariate extreme value theory can be used to model the joint extremal behavior of environmental data such as precipitation, snow depths or daily temperatures. Max-stable processes are the natural generalization of extremal dependence structures to infinite dimensions arising from the extension of multivariate extreme value theory. However, there have been few works on the threshold approach of max-stable processes.

Padoan, Ribatet and Sisson [2010] proposed the maximum composite likelihood approach for fitting max-stable processes to avoid the complexity and unavailability of the multivariate density function. We propose the threshold version of max-stable process estimation and we apply the pairwise composite likelihood method to it. We assume a strict form of condition, so called the second-order regular variation condition, for the distribution satisfying the domain of attraction. To obtain the limit behavior, we also consider the increasing domain structure with stochastic sampling design based on the setting and conditions in Lahiri [2003] and we then establish consistency and asymptotic normality of the estimator for dependence parameter in the threshold method of max-stable processes. The method is studied by simulation and illustrated by the application of temperature data in North Carolina, United States.

# Acknowledgments

I wish to express my deepest gratitude to my faculty advisor, Dr. Richard L. Smith, for his support and guidance over the past five years. A great deal of my enthusiasm for research stems from the opportunity to work with him. I have immense respect for him and deeply appreciate his inspiring suggestions, kind patience and trust throughout my graduate study.

I would also like to thank faculty members in Statistics Department for their help and guidance in the study of Statistics. I am also thankful for SAMSI of great opportunities working with research scientists in the field of spatial statistics, and especially I would like to thank to SAMSI working groups what I have been involved in, for very helpful comments and insights what I got.

A special thank you goes to my good fellow doctoral students. They provided me with lots of encouragement and empathy during the tough times of the research process. I am especially thankful to my dear friends for their sincere friendship and time sharing with me through the doctoral program.

Last but not least, I would like to express my loving gratitude to my parents and parents in law for their understanding and encouragement to follow my dreams. They instilled a drive in me that has helped me complete this five-year journey. I also would like to give special words to my husband and my sweet son. They always bring me a smile and make my life brilliant.

# Table of Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

Extreme value theory and its application are dealing with related methodologies to understand phenomena of rare events such as flooding, high temperatures and precipitations in environmental data. It is well known and widely publicized that extreme temperatures are relevant to the problem of sudden deaths from heatwaves and also, extreme levels of air pollution have strong influence on human health outcomes. Extreme environmental data involve in these statistical issues that have arisen in the study of human health and the application of extreme value analysis can be used to analyze the problems.

The behavior of rare events requires understanding of the tail distribution of quantity of interest. Extreme value theory has been studied for the univariate case in which extremes are observed as a single variable, during a few decades since Fisher and Tippett [1928] and Leadbetter, Lindgren and Rootzén [1983]. Smith [2003] and Beirlant et al. [2004] provides statistical methods in the analysis of extremes, and Coles [2001] is a very useful reference with the introduction of modeling and applications of extreme values. Multivariate extreme value theory has been developed to build the modeling of joint extremal behavior. Resnick [1987] reviewed relevant theories in the view of probability and measure theory for multivariate extremes. In a spatial context, a single quantity (e.g., sea level) is measured at multiple locations and the observed data are spatial variables which are distributed across the earth's surface. Therefore one ultimately requires the modeling of spatial extremes, and a spatial dependence among the different locations is of interest. Cooley et al. [2012] introduces several references which dealt with issues of spatial extremes.

It is natural to consider a stochastic process when the sample maxima are observed at each site of a spatial process. Max-stable processes have been developed as a class of stochastic processes suitable for studying spatial extremes. The first general characterization of max-stable processes was by de Haan [1984], and Smith [1990] has constructed a special case of max-stable processes which provides the useful interpretation of extreme rainfall models. Statistical techniques based on the Smith's max-stable model have been developed by Coles [1993] and Coles and Tawn [1996] and the well-known classes of max-stable processes are discussed further by Schlather [2002] and Kabluchko, Schlather and de Haan [2009].

Due to the complexity and unavailability of the full likelihood for the max-stable model, Padoan, Ribatet and Sisson [2010] developed the maximum composite likelihood approach to fit max-stable processes. However, the research on max-stable processes with exceedances over threshold has hardly been considered.

In this thesis, we are concerned with the development of a threshold approach using max-stable processes in spatial extremes. We review the background of extreme value theory, max-stable processes and spatial dependence measure in Chapter 2. In Chapter 3 we introduce our methodology to model exceedances over threshold using max-stable processes. Chapter 4 develops a theoretical framework and asymptotic properties, which are illustrated with a simulation study. In Chapter 5 the proposed approach is applied to the analysis of temperatures in North Carolina, and the conclusion and discussion are drawn in Chapter 6.

# Chapter 2

# Background

## 2.1 Extreme Value Theory

In this section, we outline the background of univariate extreme sequences of independent and identically distributed (i.i.d.) random variables. Univariate extreme value theory can be extended to multivariate extremes. Fundamental theory and practice of univariate and multivariate extremes have been well-established.

Let $X_1, \cdots, X_n$ be i.i.d. random variables with the same probability distribution $F$ and let $M_n = \max(X_1, \cdots, X_n)$ be the maximum. If $M_n$ converges under renormalization to some nondegenerate limit, then the limit must be a member of the parametric family, i.e. there exist suitable normalizing constants $a_n > 0$, $b_n$ and the distribution $\widetilde{G}$ such that

$$P\left\{\frac{M_n - b_n}{a_n} \le x\right\} = F^n(a_n x + b_n) \longrightarrow \widetilde{G}(x), \qquad \text{as } n \to \infty \tag{2.1}$$

where $\widetilde{G}$ is a nondegenerate distribution function. The distribution functions $\widetilde{G}$ which are possible limit laws for maxima of i.i.d. sequences form the class of so-called *max-stable* distributions. It is said that a nondegenerate function $\widetilde{G}$ is *max-stable* if, there are constants $A_N > 0$ and $B_N$ such that $\widetilde{G}^N(x) = \widetilde{G}(A_N x + B_N)$ for each $N = 2, 3, \cdots$.

Every max-stable distribution $\widetilde{G}$ has one of the following three parametric forms, called *three Extreme Value Distributions* (EVD).

$$
\begin{aligned}
\text{Type I } (\textit{Gumbel})\text{: } \widetilde{G}(x) &= \exp(-e^{-x}), \quad -\infty < x < \infty; \\
\text{Type II } (\textit{Fréchet})\text{: } \widetilde{G}(x) &= \begin{cases} 0, & x \le 0, \\ \exp(-x^{-\alpha}), & x > 0, \alpha > 0; \end{cases} \\
\text{Type III } (\textit{Weibull})\text{: } \widetilde{G}(x) &= \begin{cases} \exp\left(-(-x)^{\alpha}\right), & x \le 0, \alpha > 0, \\ 1, & x > 0. \end{cases}
\end{aligned}
$$

The *Three Types Theorem* was originally stated by Fisher and Tippett [1928] and derived rigorously by Gnedenko [1943]. Leadbetter, Lindgren and Rootzén [1983] showed that a distribution function is max-stable if and only if it is of the same type as one of the three extreme value distributions listed.

The three types of EVD can be represented as $G$ combining into a single parametric family distribution, which is called the *Generalized Extreme Value* (GEV) distribution:

$$G(x; \mu, \psi, \xi) = \exp\left\{ -\left(1 + \xi \frac{x - \mu}{\sigma}\right)_{+}^{-1/\xi} \right\},$$

where $y_+ = \max(0, y)$, $\mu$ is a location parameter, $\sigma > 0$ is a scale parameter and $\xi$ is a shape parameter which determines the tail behavior. The Generalized Extreme Value distribution $G$ has a max-stable property: if $X_1, \cdots, X_N$ are i.i.d. from $G$, then $\max(X_1, \cdots, X_N)$ also has the same distribution, i.e.

$$G^N(x) = G(A_N x + B_N) \text{ for existing constants } A_N > 0, B_N.$$

The form of the limiting distribution is invariant under monotonic transformation. Therefore, without loss of generality we can transform the GEV distribution into a specific form and consider the Fréchet form for convenience:

$$P\{X \leq x\} = \exp(-x^{-\alpha}), \quad x > 0$$

where $\alpha > 0$. The case $\alpha = 1$ is called *unit Fréchet*. For our current application, we use the GEV distribution transformed into the unit Fréchet distribution,

$$P\left\{\left(1 + \xi \frac{M_n - \mu}{\sigma}\right)_+^{1/\xi} \leq z\right\} = P(Z \leq z) = \exp(-1/z), \quad z > 0,$$

and note that the unit Fréchet form is a distribution which has the max-stable property.

Multivariate extreme value theory is concerned with the joint distribution of extremes of two or more random variables. Suppose we have i.i.d. observations from a $K$-dimensional random vector $(X_{i1}, \cdots, X_{iK})$, $i = 1, 2, \cdots$, and let $\mathbf{M}_n = (M_{n1}, \cdots, M_{nK})$ denote the $K$-dimensional vector of componentwise maxima, $M_{nk} = \max(X_{1k}, \cdots, X_{nk})$, $k = 1, \cdots, K$. A limit distribution for $\mathbf{M}_n$ is said to exist if there exist $a_{nk} > 0$ and $b_{nk}$ for $k = 1, \cdots, K$ such that

$$\lim_{n \to \infty} P\left\{\frac{M_{n1} - b_{n1}}{a_{n1}} \leq x_1, \cdots, \frac{M_{nK} - b_{nK}}{a_{nK}} \leq x_K\right\} = G(x_1, \cdots, x_K). \tag{2.2}$$

Then $G$ is a multivariate extreme value distribution and if (2.2) holds, then $G$ is max-stable if there exist $A_{Nk} > 0$ and $B_{Nk}$, $k = 1, \cdots, K$, for any $N > 1$ such that

$$G^N(x_1, \cdots, x_K) = G(A_{N1}x_1 + B_{N1}, \cdots, A_{NK}x_K + B_{NK}).$$

If $G$ is a multivariate EVD, the marginal distribution must be represented by the GEV distribution and each marginal GEV distribution can be transformed into unit Fréchet margin, which has the max-stable property.

The finite-dimensional framework of multivariate extreme distribution is extended to an infinite-dimensional generalization with spatial processes. The infinite-dimensional extremes has quite analogous extension to the theory of max-stable random vector. Let $\mathcal{S}$ be a study region and denote $s$ as a location in the study region. If there exist normalizing sequences $a_n(s)$ and $b_n(s)$ for all $s \in \mathcal{S}$ such that the sequence of stochastic processes

$$\max_{i=1,\cdots,n} \frac{X_i(s) - b_n(s)}{a_n(s)} \xrightarrow{\mathcal{D}} Y(s) \tag{2.3}$$

where $Y(s)$ is non-degenerate for all $s$, then the limit process $Y(s)$ is a max-stable process. A finite sample $\{Y(s_1), \cdots, Y(s_D)\}$ can be concerned as a realization of a spatial process $Y(s)$ for more realistic setting.

## 2.2 Dependence of Spatial Extremes: Extremal Coefficient

In the analysis of spatial extremes, one can be interested with measuring spatial dependence among locations. Quantifying spatial dependence has been studied in the field of geostatistics and one of metrics is the variogram which is typically used in the geostatistics. Let $Y(s)$ be a stationary stochastic process and suppose

$$\text{Var}[Y(s) - Y(s')] = 2\gamma(s - s') \quad \text{for all } s, s' \in \mathcal{S}.$$

The quantity $2\gamma(\cdot)$, so-called *variogram*, depends on the increments $s - s'$ and $\gamma(\cdot)$ has been called *semivariogram* which determines the degree of spatial dependence of $Y(\cdot)$ (see Matheron [1987] and Cressie [1993]). However the (semi-)variogram is an inadequate tool to analyze spatial dependence of extreme data, since the traditional geostatistics does not deal with the tail distribution.

In this section, we focus on another metric, *extremal coefficient*, to characterize the tail dependence. Suppose a $d$-dimensional random variable $\mathbf{X}$ has the common marginal distributions $F(x)$. The extremal coefficient $\theta_d$ can be defined by the relation

$$Pr\{\max(X_1, \cdots, X_d) \le x\} = F^{\theta_d}(x).$$

Assuming the standard form of unit Fréchet distribution on each margin, we can characterize the dependence among the components of marginal distribution independently. Let **Z** be $d$-dimensional maxima with unit Fréchet margins and whose multivariate extreme value distribution is expressed as

$$Pr\{Z_1 \le z_1, \cdots, Z_d \le z_d\} = \exp\{-V(z_1, \cdots, z_d)\}, \tag{2.4}$$

where the exponent measure $V$ is a homogeneous function of order $-1$. Due to the homogeneity of $V$, the extremal dependence can be measured by $V$ which implies

$$\begin{aligned}&\text{complete dependence if } V(z_1, \cdots, z_d) = \max\left(\tfrac{1}{z_1}, \cdots, \tfrac{1}{z_d}\right),\\ &\text{complete independence if } V(z_1, \cdots, z_d) = \tfrac{1}{z_1} + \cdots + \tfrac{1}{z_d}.\end{aligned}$$

The relationship between the extremal coefficient $\theta_d$ and the exponent measure $V$ is drawn from

$$Pr\{Z_1 \le z, \cdots, Z_d \le z\} = \exp\left(-\frac{\theta_d}{z}\right), \tag{2.5}$$
$$\theta_d = V(1, \cdots, 1)$$

where $1 \le \theta_d \le d$ with the lower and upper bounds corresponding to complete dependence and complete independence, respectively.

We consider a pairwise extremal coefficient as a special case of (2.5) in the spatial domain. Let $Y(s)$ be a spatial process with unit Fréchet margin for all $s \in \mathcal{S}$ and then extremal dependence between different sites $s$ and $s'$ is obtained by,

$$Pr\{Y(s) \le y, Y(s') \le y\} = \exp\left(-\frac{\theta(s-s')}{y}\right).$$

A naive estimator of the pairwise extremal coefficient is proposed by Smith [1990];

$$\widehat{\theta}(s - s') = \frac{n}{\sum_{i=1}^{n} \min\{Y_i(s)^{-1}, Y_i(s')^{-1}\}}.$$

Schlather and Tawn [2003] investigated theoretical properties of the extremal coefficients and proposed self-consistent estimators of $\theta$ (i.e. estimators that satisfy the properties of extremal coefficients) for the multivariate and spatial case.

## 2.3 Max-stable Processes

### 2.3.1 Models of Max-stable Processes

Now consider the max-stable processes as an infinite dimensional generalization of extreme value theory. Suppose $X(s), s \in \mathcal{S}$ is a stochastic process, where $\mathcal{S} \subseteq \mathbb{R}^d$ is an arbitrary index set. We can interpret $X(\cdot)$ as a spatial process with an appropriate generalization of (2.2) as following: for each $n \geq 1$, there exist continuous functions $a_n(s)$ positive and $b_n(s)$ real, for $s \in \mathcal{S}$ such that

$$Pr^n\left\{\frac{X(s_j) - b_n(s_j)}{a_n(s_j)} \leq x(s_j), j = 1, \cdots, K\right\} \longrightarrow G_{s_1,\cdots,s_K}(x(s_1), \cdots, x(s_K)). \qquad (2.6)$$

Then $G_{s_1,\cdots,s_K}$ is a multivariate extreme value distribution and the limiting process is *max-stable* if (2.6) holds for all possible subsets $s_1, \cdots, s_K \in \mathcal{S}$. Note that this is equivalent to the expression in equation (2.3).

We are interested in modeling and estimation using max-stable processes for extremes observed at each site of a spatial process. A general representation of max-stable processes was first given by de Haan [1984]. The conceptual idea of max-stable processes can be constructed by two components: a stochastic process $\{W(s)\}$ and a Poisson process $\Pi$ with intensity $d\zeta/\zeta^2$ on $(0, \infty)$. If $\{W_i(s)\}_{i \in \mathbb{N}}$ is independent copies of $W(s)$ with $E[W(s)] = 1$ for all $s$ and $\zeta_i \in \Pi$, $i \geq 1$, is points of the Poisson process, then

$$Y(s) = \max_{i \geq 1} \zeta_i W_i(s), \quad s \in \mathcal{S}$$

is a max-stable process with unit Fréchet margins. The joint distribution function for max-stable processes is given as

$$P(Y(s) \leq y(s), s \in \mathcal{S}) = \exp\left( -E\left[ \sup_{s \in \mathcal{S}} \left\{ \frac{W(s)}{y(s)} \right\} \right] \right), \tag{2.7}$$

or practically, it can be rewritten to the equivalent equation (2.4) on the set $\{s_1, \cdots, s_D\} \subset \mathcal{S}$, where

$$V(y_1, \cdots, y_D) = E\left[ \sup_{d=1,\cdots,D} \left\{ \frac{W(s_d)}{y(s_d)} \right\} \right]. \tag{2.8}$$

The construction of different max-stable processes can be differentiated from different choices of the $W(s)$ process and the well-known classes of max-stable processes are discussed by Smith [1990], Schlather [2002] and Kabluchko, Schlather and de Haan [2009].

### The Smith Model

Smith [1990] proposed new max-stable stochastic processes under the following construction. Let $\{(\zeta_i, s_i), i \geq 1\}$ denote the points of a Poisson process on $(0, \infty) \times \mathbb{R}^d$ with intensity measure $\zeta^{-2} d\zeta ds$. Define a non-negative function $\{f(x)\}$ on $\mathbb{R}^d$ such that $\int f(x)dx = 1$ and

$$Y(s) = \max_{i \geq 1} \zeta_i f(s - s_i).$$

Then a max-stable process $Y(\cdot)$ can be obtained with unit Fréchet margins. The Smith's max-stable process proposed a useful interpretation for modeling of rainfall-storms. Rainfall amounts can be measured by observing the shape of a storm centered at location $s_i$ as $f$, and the magnitude of the storm as $\zeta_i$. Then the max-stable process $Y(s)$ represents maximum rainfal amounts taking over all storms for each site in $\mathcal{S}$.

Smith considered a specific setting, so-called a *Gaussian extreme value process*, where $f(x) = (2\pi)^{-d}|\Sigma|^{-1/2}\exp\big(-\frac{1}{2}x^T\Sigma^{-1}x\big)$ is a multivariate normal density with covariance matrix $\Sigma$. Then the joint distribution at two sites is obtained in a closed form,

$$P\big(Y(s_1) \le y_1, Y(s_2) \le y_2\big) \\ = \exp\left\{-\frac{1}{y_1}\Phi\left(\frac{a}{2}+\frac{1}{a}\log\frac{y_2}{y_1}\right)-\frac{1}{y_2}\Phi\left(\frac{a}{2}+\frac{1}{a}\log\frac{y_1}{y_2}\right)\right\} \tag{2.9}$$

where $a = \sqrt{(s_1-s_2)^T\Sigma^{-1}(s_1-s_2)}$ and $\Phi$ is the standard normal cumulative distribution function. The positive value $a$ represents the spatial dependence according to the distance between two sites. The limits $a \to 0$ and $a \to \infty$ correspond to complete dependence and independence, respectively. One may write out the pairwise extremal coefficients explicitly as

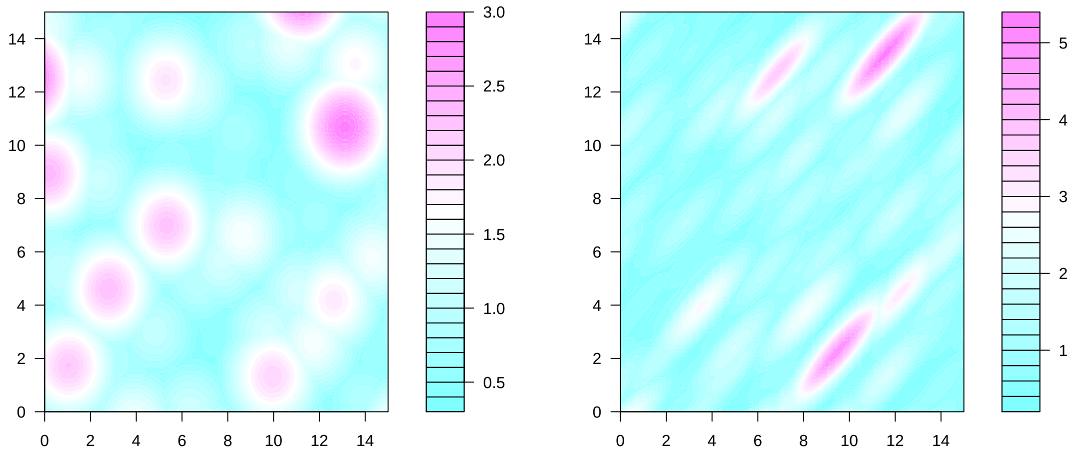
$$\theta(h) = 2\Phi\left(\frac{\sqrt{(s_1-s_2)^T\Sigma^{-1}(s_1-s_2)}}{2}\right)$$

where $h$ is the Euclidean distance, $\|s_1 - s_2\|$, between two stations. Realizations of the Gaussian extreme value process are shown in Figure 2.1.

### The Schlather Model

More recently, Schlather [2002] suggested a new class of max-stable processes based on a stationary random field with finite expectation. Let $W_i(s), i = 1, 2, \cdots$ be i.i.d. stochastic processes on $\mathbb{R}^d$, and let $\mu = E[\max(0, W_i(s))] < \infty$ and $\{\zeta_i, i \ge 1\}$ denote the points of a Poisson process on $(0,\infty)$ with intensity measure $\mu^{-1}\zeta^{-2}d\zeta$. Then a stationary max-stable process with unit Fréchet margins can be obtained by:

$$Y(s) = \max_{i\ge 1}\zeta_i \max(0, W_i(s))$$



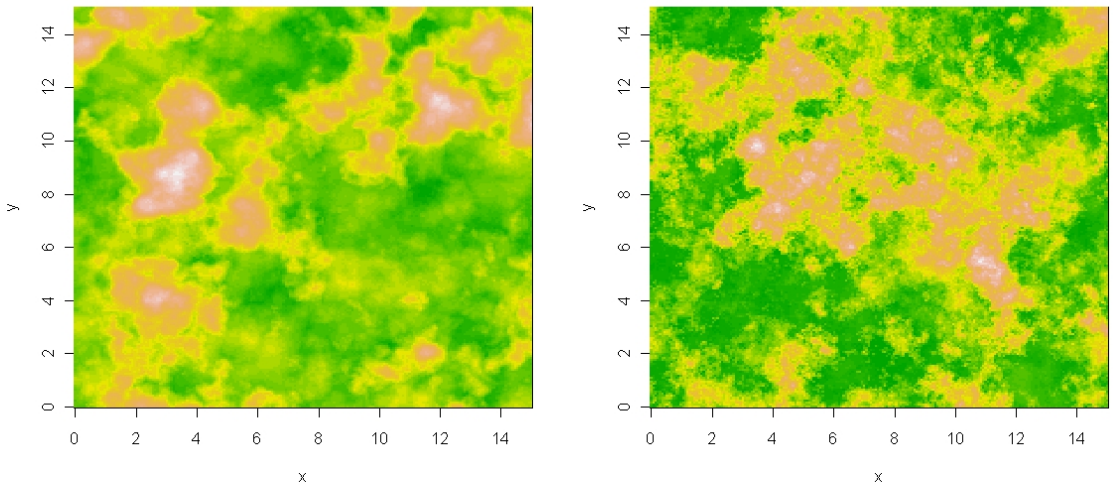
(a) $\sigma_{11} = \sigma_{22} = 9/8$ and $\sigma_{12} = 0$ (b) $\sigma_{11} = \sigma_{22} = 9/8$ and $\sigma_{12} = 1$

Figure 2.1: Realization of the Smith model with different covariance matrices from SpatialExtremes R package

where the $W_i$ are i.i.d. copies of $W(s)$ for all $i$. The max-stable process provides a more flexible class of max-stable processes by taking a stationary random process $W_i(s)$ and it also gives the interpretation of spatial storm modeling. The spatial rainfall events are explained by the structure of spatial dependence but this process represents more general case than Smith's model. The shape of storms is deterministic with $f(\cdot)$ in Smith's model while the storms may have a random shape in Schlather's model.

Schlather specified a model for a stationary Gaussian random process. Let $W_i$ be a stationary Gaussian random field with unit variance, correlation $\rho(\cdot)$ and $\mu^{-1} = \sqrt{2\pi}$, then the process $Y(s)$ is called as an *extremal Gaussian process* and the bivariate marginal distributions are given explicitly by

$$P\big(Y(s_1) \leq y_1, Y(s_2) \leq y_2\big) = \exp\left\{ -\frac{1}{2}\left(\frac{1}{y_1} + \frac{1}{y_2}\right)\left(1 + \sqrt{1 - 2(\rho(h)+1)\frac{y_1 y_2}{(y_1+y_2)^2}}\right)\right\} \tag{2.10}$$




(a) sill=range=smooth=1 (b) sill=smooth=1 and range=1.5

Figure 2.2: Realization of the Schlather model with different correlation functions from SpatialExtremes R package; (a) Whittle-Matérn correlation and (b) Powered exponential correlation

where $h$ is the Euclidean distance between station $s_1$ and $s_2$. The pairwise extremal coefficient is obtained by

$$\theta(h) = 1 + \left( \frac{1 - \rho(h)}{2} \right)^{1/2}.$$

Figure 2.2 shows the realizations of the extremal Gaussian process with different correlation functions.

The Schlather model cannot attain the case of independence for extremes as the distance $h$ increases, since the extremal coefficient $\theta(h)$ is in the interval $[1, 1.838]$. To overcome the problem, the process $W_i(s)$ can be restricted to a random set $\mathcal{B}$, i.e.,

$$Y(s) = \max_i \zeta_i W_i(s) I_{\mathcal{B}_i}(s - S_i)$$

where $I_{\mathcal{B}}$ is the indicator function of a compact random set $\mathcal{B} \subset \mathcal{S}$ and $S_i$ are the points of a Poisson process. If $W_i$ is a Gaussian process, the bivariate marginal distribution is

$$P\big(Y(s_1) \le y_1, Y(s_2) \le y_2\big)$$
$$= \exp\left\{ -\left(\frac{1}{y_1} + \frac{1}{y_2}\right)\left[1 - \frac{\alpha(h)}{2}\left(1 - \sqrt{1 - 2(\rho(h)+1)\frac{y_1 y_2}{(y_1+y_2)^2}}\right)\right]\right\}$$

where $\alpha(h) = E\{|\mathcal{B} \cap (h + \mathcal{B})|\}/E(|\mathcal{B}|) \in [0,1]$. One possible choice for the set $\mathcal{B}$ is a disc of radius $r$ and it leads to take $\alpha(h) \doteq \{1 - |h|/(2r)\}_+$, which equals to 0 representing the independence of extremes in the case of $|h| > 2r$ (see details in Davison and Gholamrezaee [2012]). The extremal coefficient is

$$\theta(h) = 2 - \alpha(h)\left\{1 - \left(\frac{1 - \rho(h)}{2}\right)^{1/2}\right\},$$

which accounts independent extremes by taking any value in the interval $[1,2]$.

**The Brown-Resnick Process**

The original Brown-Resnick process was introduced with the Brownian motion for max-stable process by Brown and Resnick [1977]. Kabluchko, Schlather and de Haan [2009] has constructed a more general class of max-stable processes, so-called the *Brown-Resnick process*, by replacing the Brownian motion by other stochastic processes.

Kabluchko, Schlather and de Haan [2009] proposed the same basic structure by Schlather but an alternative specification for the $W_i$ processes. Let $W_i(s) = \exp\{e_i(s) - \frac{1}{2}\sigma^2(s)\}$ where $e_i(s)$ is a Gaussian process with stationary increments, assuming a weaker condition than second-order stationarity, and $\sigma^2(s) = Var\{e(s)\}$. Then the process defined above can be a very general class of max-stable processes and it takes advantage of the connection with standard geostatistics by allowing the use of variogram. The bivariate CDF is the same with the Smith model where the dependence parameter $a^2 = \gamma(h)$ and $\gamma(\cdot)$ is the variogram of $e(\cdot)$. The closed form of the bivariate distributions for the Brown-Resnick process associated to the variogram $\gamma$ is given by

$$P\big(Y(s_1) \le y_1, Y(s_2) \le y_2\big)$$
$$= \exp\left\{ -\frac{1}{y_1}\Phi\left(\frac{\sqrt{\gamma(h)}}{2} + \frac{1}{\sqrt{\gamma(h)}}\log\frac{y_2}{y_1}\right) - \frac{1}{y_2}\Phi\left(\frac{\sqrt{\gamma(h)}}{2} + \frac{1}{\sqrt{\gamma(h)}}\log\frac{y_1}{y_2}\right)\right\} \quad (2.11)$$

where $\Phi$ is the standard normal distribution function and $h$ is the Euclidean distance between location $s_1$ and $s_2$. The pairwise extremal coefficients are given as

$$\theta(h) = 2\Phi\left(\frac{\sqrt{\gamma(h)}}{2}\right).$$

### 2.3.2 Fitting Max-stable Processes

We are interested with the analysis of spatial extremes at a large number of stations and the standard methods of estimation, such as MLE and Bayes methods, require a full likelihood. However the full likelihood for the max-stable processes may not be available analytically, because there are difficulties to achieve the expression of differentiation for the joint distribution function (2.7) and to calculate the exponent measure (2.8) due to the complexity of its analytic form. With the lack of an explicit form of the joint distribution, Padoan, Ribatet and Sisson [2010] developed a pairwise composite likelihood approach to fit max-stable processes, based on a composite likelihood method by Lindsay [1988].

For a parametric statistical model with density function family $\{f(\mathbf{y};\psi) : \mathbf{y} \in \mathcal{Y} \subseteq \mathbb{R}^K, \psi \in \Psi \subseteq \mathbb{R}^d\}$ and a set of marginal or conditional events $\{\mathcal{I}_k : k \in \mathcal{K}\}$ subset of some sigma algebra on $\mathcal{Y}$, the composite log-likelihood is defined by

$$l_{\mathcal{C}}(\psi;\mathbf{y}) = \sum_{k\in\mathcal{K}} w_k \log f(\mathbf{y} \in \mathcal{I}_k;\psi),$$

where $\log f(\mathbf{y} \in \mathcal{I}_k;\psi)$ is the log-likelihood associated with event $\mathcal{I}_k$ and $\{w_k\}_{k\in\mathcal{K}}$ are nonnegative weights.

A composite score function $\mathrm{D}_\psi l_\mathcal{C}(\psi;\mathbf{y})$ is defined by first-order partial derivatives of $l_\mathcal{C}(\psi;\mathbf{y})$ with respect to $\psi$ and then the maximum composite likelihood estimator of $\psi$, if it is unique, is obtained by solving $\mathrm{D}_\psi l_\mathcal{C}(\widehat{\psi}_{MCLE};\mathbf{y}) = 0$. Second-order partial derivatives of the composite score function yield the Hessian matrix $\mathrm{H}_\psi l_\mathcal{C}(\psi;\mathbf{y})$. Under appropriate conditions based on Lindsay [1988] and Cox and Reid [2004], the maximum composite likelihood estimator may have consistency and asymptotically normality as

$$\widehat{\psi}_{MCLE} \sim N(\psi, \tilde{I}(\psi)^{-1}) \text{ with } \tilde{I}(\psi) = H(\psi)J(\psi)^{-1}H(\psi),$$

where $H(\psi) = \mathbb{E}\{-\mathrm{H}_\psi l_\mathcal{C}(\psi;\mathbf{Y})\}$ is the expected second order derivatives of the score function, and the covariance matrix of the score function is $J(\psi) = \mathbb{V}\{\mathrm{D}_\psi l_\mathcal{C}(\psi;\mathbf{Y})\}$, which are analogues of the expected information matrix and the variance of the score vector. The maximum composite likelihood estimator may not be asymptotically efficient in that $\tilde{I}(\psi)^{-1}$, the inverse of the Godambe information matrix, may not attain the Cramér-Rao bound although it can be unbiased.

### Pairwise Composite Likelihoods in Spatial Extremes

Assume $M$ i.i.d. replications of a stochastic process with bivariate densities $f(y_i, y_j;\psi)$, $1 \le i, j \le K$, in a spatial region with $K$ locations. Then the pairwise composite log-likelihood is defined by

$$l_\mathcal{P}(\psi;\mathbf{Y}) = \sum_{m=1}^{M}\sum_{i=1}^{K-1}\sum_{j=i+1}^{K} w_{ij}\log f(y_{mi}, y_{mj};\psi), \tag{2.12}$$

where $(i,j)$ is a pair of stations and $w_{ij}$ is nonnegative weight functions. One may set the weight as an indicator function, i.e., $w_{ij} = 1$ if $\parallel s_1 - s_2 \parallel \le \delta$, and 0 otherwise. The *maximum pairwise composite likelihood estimator* (MCLE), $\hat{\psi}$, is chosen to maximize (2.12).

Padoan, Ribatet and Sisson [2010] stated the asymptotic properties of MCLE based on the joint estimation, which maximizes the pairwise composite likelihood instead of the full likelihood. For the estimation, a pairwise composite log-likelihood is constructed as the form (2.12) and we consider the bijection $(Y_i, Y_j) = g(Z_i, Z_j)$, where $g$ is some monotonic increasing transformation to the unit Fréchet. Then by change of variables, we represent the bivariate density over GEV margins as the form,

$$f_{Y_i,Y_j}(y_i, y_j) = f_{Z_i,Z_j}[g^{-1}(y_i, y_j)]|J(y_i, y_j)|,$$

where $f_{Z_i,Z_j}(z_i, z_j)$ denotes the joint density of max-stable processes and the determinant of the Jacobian is given by

$$|J(y_i, y_j)| = \frac{1}{\sigma_i \sigma_j} \left(1 + \frac{\xi_i(y_i - \mu_i)}{\sigma_i}\right)_+^{1/\xi_i - 1} \left(1 + \frac{\xi_j(y_j - \mu_j)}{\sigma_j}\right)_+^{1/\xi_j - 1}.$$

GEV marginal parameters and the dependence parameters can be estimated in a unified framework by the change of variable technique. Variances of parameter estimates are provided through the inverse of the Godambe information matrix, with estimates of the matrices $H(\psi)$ and $J(\psi)$ given by

$$\hat{H}(\hat{\psi}_{MCLE}) = -\sum_{m=1}^{M} \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} \mathrm{H}_\psi \log f(y_{mi}, y_{mj}; \hat{\psi}_{MCLE})$$

and

$$\hat{J}(\hat{\psi}_{MCLE}) = \sum_{m=1}^{M} \left\{ \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} \mathrm{D}_\psi \log f(y_{mi}, y_{mj}; \hat{\psi}_{MCLE}) \right\} \times \left\{ \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} \mathrm{D}_\psi \log f(y_{mi}, y_{mj}; \hat{\psi}_{MCLE}) \right\}^T.$$

In practice, the matrix $\hat{H}$ is obtained through the numerical maximization routine and the explicit form of $\hat{J}$ is also derived [Padoan, Ribatet and Sisson, 2010, Appendix A.5].

A general explicit expression of higher-order ($p \geq 3$) maximum composite likelihood has not been developed, but a closed form for triple-wise composite likelihood function of a Gaussian extreme value process was derived by Genton, Ma and Sang [2011] and the efficiency gain was obtained in inference of triplewise composite likelihood comparing with pairwise one in a spatial domain $\mathbb{R}^2$.

# Chapter 3

# Threshold Approach of Max-stable Processes

## 3.1 Introduction to Threshold Approach

Consider the distribution of all observations $X$ over a high threshold $u$ and let $Y = X - u > 0$, then

$$F_u(y) = Pr\{Y \le y | Y > 0\} = \frac{F(u+y) - F(u)}{1 - F(u)}.$$

As $u \to x_0 = \sup\{x : F(x) < 1\}$, we can find a limit $H$ called *Generalized Pareto Distribution* (GPD)

$$F_u(y) \approx H(y; \sigma_u, \xi) = 1 - \left(1 + \xi \frac{y}{\sigma_u}\right)_+^{-1/\xi}. \tag{3.1}$$

Smith [1987] described the bias versus variance tradeoff in the choice of threshold $u$ of univariate case. If the threshold $u$ increases, the variance of estimators will be high due to small $N$ (number of exceedances) while the estimates are biased due to the poor approximation of $F_u(\cdot)$ by $H(\cdot)$ if $u$ is too small. Thus limit theorems on the threshold approach in the literature are presented as $N \to \infty$ and $u \equiv u_N \to x_0$ simultaneously.

Pickands [1975] established the rigorous connection between the classical extreme value theory and the generalized Pareto distribution and proved that the limit of the form (3.1) exists if and only if there exist normalizing constants and the limiting form of $H$ such that the classical extreme value limit (2.1) holds. Thus the limit result for exceedances over thresholds is equivalent to the limit distribution for maxima in this sense.

As an another statistical approach for threshold exceedances, the Point Process Approach is introduced by Smith [1989]. This approach considers a process based on a two dimensional plot of exceedance times and exceedance values, which has been developed from the point process viewpoints of extreme values by Leadbetter, Lindgren and Rootzén [1983].

Under the proper normalization, the asymptotic theory of threshold exceedances proved that the process behaves like a *nonhomogeneous Poisson process.* A nonhomogeneous Poisson process on a domain $\mathcal{D}$ is denoted by an intensity $\lambda(x), x \in \mathcal{D}$, such that if $A$ is a measurable subset of $\mathcal{D}$ and $N(A)$ is the number of points in $A$, then $N(A)$ has a Poisson distribution with mean

$$\Lambda(A) = \int_A \lambda(x)dx.$$

For the present application, we denote $(T_i, Y_i)$ as the time of the $i$th exceedance of the threshold and the observed excess value $Y_i > u$, then the probability of observing an exceedance in an infinitesimal region $t < T_i < t + dt, y < Y_i < y + dy$ can be written as

$$\frac{1}{\sigma}\left(1 + \xi\frac{y - \mu}{\sigma}\right)_+^{-1/\xi - 1} dydt, y > \mu. \tag{3.2}$$

To fit the model, if a nonhomogeneous process with intensity $\lambda(t, y)$ is observed on a domain $\mathcal{D}$ and if there are the $N$ observed exceedances over $u$ through a time interval of $T$ units, then the likelihood associated with the events $(T_i, Y_i), 1 \le i \le N$, is

$$\prod_{i=1}^{N} \lambda(T_i, Y_i) \cdot \exp\left\{ - \int_{\mathcal{D}} \lambda(t, y) dt dy \right\}$$
$$= \prod_{i=1}^{N} \left\{ \frac{1}{\sigma} \left(1 + \xi \frac{Y_i - \mu}{\sigma}\right)_+^{-1/\xi - 1} \right\} \cdot \exp\left\{ - T \left(1 + \xi \frac{u - \mu}{\sigma}\right)^{-1/\xi} \right\} \tag{3.3}$$

and (3.3) is maximized with respect to unknown parameters $(\mu, \sigma, \xi)$.

For the inhomogeneous case the parameters $\mu$, $\sigma$ and $\xi$ are all allowed to be time-dependent, denoted by $\mu_t$, $\sigma_t$ and $\xi_t$. Thus (3.2) is extended by the form

$$\frac{1}{\sigma_t} \left(1 + \xi_t \frac{y - \mu_t}{\sigma_t}\right)_+^{-1/\xi_t - 1} dy dt, y > \mu_t.$$

and if we also allow the threshold $u_t$ to depend on time $t$, the likelihood associated with (3.3) is now

$$\prod_i \left\{ \frac{1}{\sigma_{T_i}} \left(1 + \xi_{T_i} \frac{Y_i - \mu_{T_i}}{\sigma_{T_i}}\right)_+^{-1/\xi_{T_i} - 1} \right\} \cdot \exp\left\{ - \int_0^T \left(1 + \xi_t \frac{u_t - \mu_t}{\sigma_t}\right)^{-1/\xi_t} dt \right\}.$$

## 3.2 Methodology for Exceedances over Threshold

As in the univariate case, the threshold method has been developed in the multivariate case as well. Let $(x_0, y_0)$ denote the upper endpoint of $F$, where $(x_0, y_0) = \sup\{(x, y) : F(x, y) < 1\}$, and define the conditional distribution of $(X - u, Y - v)$ given $X > u$ or $Y > v$,

$$F_{u,v}(x, y) = \frac{F(u + x, v + y) - F(u, v)}{1 - F(u, v)}. \tag{3.4}$$

Then the conditional distribution of bivariate exceedances converges to $H$ where $H$ is a multivariate generalized Pareto distribution by Rootzén and Tajvidi [2006].

In this dissertation we develop an alternative methodology for threshold exceedances using max-stable processes with unit Frechét margins. We suggest the modeling of the bivariate threshold exceedances by assuming that the asymptotic distribution holds exactly above a threshold and it leads to a simplified dependence structure for max-stable processes as we characterize the dependence among the components of bivariate marginal distribution in (2.9), (2.10) and (2.11).

The likelihood representation for this threshold method is also developed to fit the model and this has a similar idea by Smith, Tawn and Coles [1997] which establishes a joint distribution for Markov chains where the bivariate distributions were assumed to be of bivariate extreme value distribution form above a threshold.

Suppose we have annual maxima $\{Y_{t^*s}, t^* = 1, \cdots, T^*, s = 1, \cdots, D\}$ where $Y_{t^*s}$ is the value at site $s$ in year $t^*$. We assume the vectors $\{Y_{t^*s}\}$ are independent for different $t^*$ with joint densities given by a max-stable process, i.e., an explicit expression for its bivariate joint distribution is known and the marginal distributions are unit Fréchet for each $t^*$ and $s$. Then the joint bivariate distribution of the annual maxima, $F_{AM}$ is written by

$$F_{AM}(y_{t^*s}, y_{t^*s'}; \theta) = Pr\{Y_{t^*s} \leq y_{t^*s}, Y_{t^*s'} \leq y_{t^*s'}; \theta\},$$

where $\theta$ is the dependence parameter which can be estimated by the max-stable model. Now suppose that the daily data are $\{X_{ts}, t = 1, \cdots, T, s = 1, \cdots, D\}$ and the joint bivariate distribution function is $F_{DA}(x_{ts}, x_{ts'}; \theta)$. Assume that the daily data $X_{ts}$ form i.i.d. random processes and the annual maxima are $Y_{t^*s}$. Then the relationship between their bivariate distributions is

$$\begin{aligned} F_{DA}(x_{ts}, x_{ts'}; \theta) &= Pr\{X_{ts} \leq x_{ts}, X_{ts'} \leq x_{ts'}; \theta\} \\ &= F_{AM}(x_{ts}, x_{ts'}; \theta)^{1/M} \end{aligned} \tag{3.5}$$

where $M$ is the number of days in a year. We can have a closed form for $F_{AM}$ from the max-stable theory and also get an expression for $F_{DA}$ from the above representation.

In practice, we would expect to apply some notion of thresholding. Suppose we fix the threshold $u$ and we assume that the same threshold for all locations for convenience. Then we observe exceedances $\{X_{ts}\}$ such that $X_{ts} > u$. Let $\delta_s = I(X_{ts} > u)$ where $I$ is the indicator function. We can obtain the following joint distribution of $(\delta_s, X_{ts}, \delta_{s'}, X_{ts'})$ from four possible regions by including or excluding the interval over threshold $u$ (see Figure 3.1),

$$\begin{aligned} Pr\{\delta_s = 0, \delta_{s'} = 0\} &= F_{DA}(u, u) \\ Pr\{\delta_s = 1, \delta_{s'} = 0, X_{ts} < x_{ts}\} &= F_{DA}(x_{ts}, u) \\ Pr\{\delta_s = 0, \delta_{s'} = 1, X_{ts'} < x_{ts'}\} &= F_{DA}(u, x_{ts'}) \\ Pr\{\delta_s = 1, \delta_{s'} = 1\} &= F_{DA}(x_{ts}, x_{ts'}). \end{aligned}$$

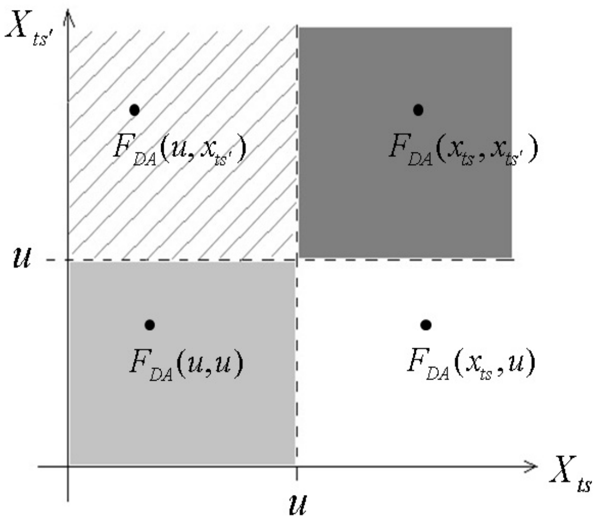



Figure 3.1: Likelihood contribution according to four possible restrictions

We extend the threshold version of max-stable processes and apply the maximum composite likelihood method on it. The likelihood contribution of the pair $(x_{ts}, x_{ts'})$ derived from the joint bivariate density can be obtained by

$$L(X_{ts}, X_{ts'}; \boldsymbol{\theta}, \boldsymbol{\eta}) = \begin{cases} F_{DA}(u, u) & \text{if } x_{ts} \le u, x_{ts'} \le u, \\ \frac{\partial}{\partial x_{ts}} F_{DA}(x_{ts}, u) & \text{if } x_{ts} > u, x_{ts'} \le u, \\ \frac{\partial}{\partial x_{ts'}} F_{DA}(u, x_{ts'}) & \text{if } x_{ts} \le u, x_{ts'} > u, \\ \frac{\partial^2}{\partial x_{ts} \partial x_{ts'}} F_{DA}(x_{ts}, x_{ts'}) & \text{if } x_{ts} > u, x_{ts'} > u. \end{cases}$$

where $\boldsymbol{\theta}$ is the dependence parameter vector and $\boldsymbol{\eta}$ is a vector of marginal GEV parameter. Combining the above likelihood representation with a pairwise likelihood, we assume $T$ i.i.d. replications of a stochastic process with bivariate densities of the unit Frechét margins $L(X_{ts}, X_{ts'}; \boldsymbol{\theta}, \boldsymbol{\eta}), 1 \leq s, s' \leq D$. Then the pairwise composite log-likelihood for a thresholded process is

$$l(\boldsymbol{\theta}, \boldsymbol{\eta}) = \sum_{t=1}^{T} \sum_{s=1}^{D-1} \sum_{s'=s+1}^{D} w_{ss'} \log L(X_{ts}, X_{ts'}; \boldsymbol{\theta}, \boldsymbol{\eta}) = \sum_{t=1}^{T} l_t(\boldsymbol{\theta}, \boldsymbol{\eta}) \tag{3.6}$$

where $l_t(\boldsymbol{\theta}, \boldsymbol{\eta}) = \sum_{s=1}^{D-1} \sum_{s'=s+1}^{D} w_{ss'} \log L(X_{ts}, X_{ts'}; \boldsymbol{\theta}, \boldsymbol{\eta}_0)$, $(s, s')$ is a pair of different stations and $T$ is a number of observations. In practice, the marginal parameter $\boldsymbol{\eta}$ will be estimated but we let $\boldsymbol{\eta}$ be the true value $\boldsymbol{\eta}_0$ to simplify theoretical justification. Thus we fix the marginal GEV parameters $\boldsymbol{\eta} = \boldsymbol{\eta}_0$ and estimate the dependence parameter $\boldsymbol{\theta}$. A dependence parameter $\boldsymbol{\theta}$ can be estimated by maximizing the pairwise composite likelihood function (3.6) with the known value $\boldsymbol{\eta}_0$.

Suppose $\mathbf{X}^{(t)} = (X_{ts}, X_{ts'})$ and denote the composite score functions by pairwise log-likelihood derivatives as

$$\begin{aligned}
&\mathrm{D}(\boldsymbol{\theta}; \mathbf{X}^{(t)}) = \frac{\partial l_t(\boldsymbol{\theta}, \boldsymbol{\eta}_0)}{\partial \boldsymbol{\theta}}, \\
&\mathbb{D}(\boldsymbol{\theta}; \mathbf{X}^{(1)}, \cdots, \mathbf{X}^{(T)}) = \mathbb{D}_{\boldsymbol{\theta}_0} l(\boldsymbol{\theta}, \boldsymbol{\eta}_0; \mathbf{X}^{(1)}, \cdots, \mathbf{X}^{(T)}) = \sum_{t=1}^{T} \mathrm{D}(\boldsymbol{\theta}; \mathbf{X}^{(t)}).
\end{aligned}$$

Then the estimating equations

$$\mathbb{D}(\widehat{\boldsymbol{\theta}}; \mathbf{X}^{(1)}, \cdots, \mathbf{X}^{(T)}) = \mathbb{D}_{\boldsymbol{\theta}} l(\widehat{\boldsymbol{\theta}}, \boldsymbol{\eta}_0; \mathbf{X}^{(1)}, \cdots, \mathbf{X}^{(T)}) = 0.$$

The parameter estimator $\widehat{\boldsymbol{\theta}}$ is a root to solve above estimating equations and we now start to describe the theoretical framework with more strict conditions to obtain asymptotic properties of the estimator in Chapter 4.

# Chapter 4

# Asymptotic Behavior of Estimates for Dependence Parameters

## 4.1 Theoretical Framework of Threshold Approach

Suppose that $(X_i, Y_i)$, $i = 1, \cdots, n$, is a sequence of i.i.d. random vectors and $F$ be the common distribution of $(X_i, Y_i)$ with marginal distributions $F_1$ and $F_2$. A distribution function $F$ is said to be in the *domain of attraction* of a distribution function $G$, shortly $F \in D(G)$, if

$$\lim_{n\to\infty} F^n(a_n x + b_n, c_n y + d_n) = G(x, y), \; a_n, c_n > 0 \text{ and } b_n, d_n \in \mathbb{R} \tag{4.1}$$

for all $x$ and $y$. The two marginals of $G(x, \infty)$ and $G(\infty, y)$ are one-dimensional extreme value distributions satisfying

$$\begin{aligned} \lim_{n\to\infty} F_1^n(a_n x + b_n) &= \exp\{-(1 + \xi_1 x)^{-1/\xi_1}\}, \\ \lim_{n\to\infty} F_2^n(c_n y + d_n) &= \exp\{-(1 + \xi_2 y)^{-1/\xi_2}\} \end{aligned}$$

where $\xi_1$ and $\xi_2$ are real parameters.

Let $(x_0, y_0)$ denote the upper endpoint of $F(x, y)$ and the conditional distribution of $(X - u, Y - v)$ given $X > u$ or $Y > v$ is defined as in (3.4). The equation (4.1) by taking logarithms can be expressed as

$$\lim_{t\to\infty} t\big\{1 - F(a_t x + b_t, c_t y + d_t)\big\} = -\log G(x,y) =: \Phi(x,y) \tag{4.2}$$

and it is checked easily that (4.2) implies that

$$\begin{aligned}\lim_{t\to\infty} F_{b_t,d_t}(a_t x, c_t y) &= \lim_{t\to\infty}\left(1 - \frac{t\big\{1 - F(a_t x + b_t, c_t y + d_t)\big\}}{t\big\{1 - F(b_t, d_t)\big\}}\right)\\ &= 1 - \frac{-\log G(x,y)}{-\log G(0,0)} =: H(x,y)\end{aligned}$$

where $H$ is a bivariate generalized Pareto distribution. It has been illustrated that $H$ is a good approximation of $F_{b_t,d_t}$ in the sense that

$$\lim_{t\to\infty} \sup_{0<(a_t x, c_t y)<(x_0 - b_t, y_0 - d_t)} |F_{b_t,d_t}(a_t x, c_t y) - H(x,y)| = 0,$$

if and only if $F$ is in the maximum domain of attraction of the corresponding extreme value distribution $G$ (Rootzén and Tajvidi [2006]).

### 4.1.1 Second-order Regular Variation Condition

To obtain a limiting distribution of $F_{u,v}$ we assume a strict form of condition, so called the *second-order regular variation condition*, for the distribution satisfying the domain of attraction. The ideas of second-order regular variation have been applied to the statistics of extremes. Asymptotic properties of estimators in univariate extreme value theory have been investigated with the second-order regular variation (see Smith [1987], de Haan and Stadtmüller [1996], and Drees [1998]), and the second-order regular variation condition was studied for bivariate extremes by de Haan and Ferreira [2006].

**Definition 1.** *A function $f(x)$ is regular varying with index $\tau_1$ if for some $\tau_1 \in \mathbb{R}$,*

$$\lim_{t\to\infty} \frac{f(tx)}{f(t)} = x^{\tau_1}, \quad x > 0.$$

*The function $f(x)$ is second-order regular varying with the first order $\tau_1$ and the second order $\tau_2$ if there exists a function $q(t) \to 0$ as $t \to \infty$ such that*

$$\lim_{t\to\infty} \frac{\frac{f(tx)}{f(t)} - x^{\tau_1}}{q(t)} = x^{\tau_2}, \quad x > 0.$$

Just as in the univariate case, the representation of bivariate regular variation exists: the function $f(x,y) : \mathbb{R}_+^2 \to \mathbb{R}_+$ is regular varying of index $\tau$ if

$$\lim_{t\to\infty} \frac{f(tx,ty)}{f(t,t)} = r(x,y)$$

where $r(\lambda x, \lambda y) = \lambda^\tau r(x,y)$ for some $\lambda > 0$. See Resnick [2007] for the related discussion of multivariate regular variation.

Suppose that the following second-order regular variation condition holds (de Haan and Ferreira [2006]): there exists a positive or negative function $\alpha$ with $\lim_{t\to\infty} \alpha(t) = 0$ and a function $Q$ not a multiple of $\Phi$ such that

$$\lim_{t\to\infty} \frac{t\{1 - F(a_t x + b_t, c_t y + d_t)\} - \Phi(x,y)}{\alpha(t)} = Q(x,y) \tag{4.3}$$

locally uniformly for $(x,y) \in (0,\infty] \times (0,\infty]$. Define $U_i$ as the inverse function of $1/(1 - F_i)$, $i = 1,2$ and it is known that for $x, y > 0$,

$$\lim_{t'\to\infty} \frac{U_1(t'x) - U_1(t')}{a(t')} = \frac{x^{\gamma_1} - 1}{\gamma_1},$$
$$\lim_{t'\to\infty} \frac{U_2(t'y) - U_2(t')}{c(t')} = \frac{y^{\gamma_2} - 1}{\gamma_2}.$$

For $t \neq t'$, define $a_t$, $b_t$, $c_t$ and $d_t$ such that $a(t') \equiv a_t, U_1(t') \equiv b_t, c(t') \equiv c_t$, and

$U_2(t') \equiv d_t$ respectively. Let

$$x_t := \frac{U_1(tx) - b_t}{a_t},$$
$$y_t := \frac{U_2(ty) - d_t}{c_t},$$

and we could rewrite the form (4.2) as

$$\lim_{t\to\infty} t\left\{1 - F(U_1(tx), U_2(ty))\right\} = -\log G\left(\frac{x^{\gamma_1} - 1}{\gamma_1}, \frac{y^{\gamma_2} - 1}{\gamma_2}\right) =: \Phi_0(x, y).$$

It follows the similar form of the second-order condition (4.3),

$$\lim_{t\to\infty} \frac{\frac{1-F\left(U_1\left(\frac{x}{1-F_1(b_t)}\right),U_2\left(\frac{y}{1-F_2(d_t)}\right)\right)}{1-F(b_t,d_t)} - \frac{\Phi_0(x,y)}{\Phi_0(1,1)}}{\alpha\left(\frac{1}{1-F(b_t,d_t)}\right)} = Q\left(\frac{x^{\gamma_1} - 1}{\gamma_1}, \frac{y^{\gamma_2} - 1}{\gamma_2}\right). \tag{4.4}$$

We can rewrite the condition (4.4) and the following second order condition holds for $F_{b_t,d_t}$.

**Condition 1.** *There exists a positive or negative function $A(\cdot)$ such that*

$$F_{b_t,d_t}(a_t x, c_t y) = H(x, y) + A(t)\Psi(x, y) + R_t(x, y), \text{ for all } t \text{ and } x, y > 0 \tag{4.5}$$

*where either*

*(i)* $\Psi \equiv 0$, $A(t) = o(1)$ *and* $R_t(x, y) = o(A(t))$ *as* $t \to \infty$, *or*

*(ii)* $\Psi$ *is continuous and not a multiple of* $H$, $A(t) = o(1)$ *and* $R_t(x, y) = o(A(t))$ *as* $t \to \infty$.

The second order regular variation condition implements the domain of attraction condition as a special asymptotic expansion of the conditional distribution $F_{b_t,d_t}$ near infinity. The asymptotic behavior of tail distribution turns out to depend on how the regular variation condition behaves.

More precisely, suppose that $(X, Y)$ are i.i.d. from $F_{b_t,d_t}$, not from bivariate generalized Pareto distribution $H$. We can determine a remainder $A$ with the second-order condition (ii) such that

$$\sup_{0<(a_t x, c_t y)<(x_0-b_t, y_0-d_t)} |F_{b_t,d_t}(a_t x, c_t y) - H(x,y)| = O(A(t))$$

where $A(t) \to 0$ as $(b_t, d_t) \to (x_0, y_0)$. In that case we expect that the remainder function $A$ will produce a bias in limit distribution.

Suppose $N \to \infty$, $(b_t, d_t) = (b(t_N), d(t_N)) \to (x_0, y_0)$, and $A(t_N) = O\big(\frac{1}{\sqrt{N}}\big)$. Then under some mild conditions, we obtain the limiting distribution of estimator $\widehat{\boldsymbol{\theta}}$ of a dependence parameter,

$$\sqrt{N}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbb{H}^{-1}\boldsymbol{b}, \mathbb{H}^{-1}\mathbb{V}\mathbb{H}^{-1})$$

for some $\boldsymbol{b}$ where $N^{-1}\mathbb{E}[-\mathrm{H}(\boldsymbol{\theta}_0)] \to \mathbb{H}$ whose H is analogous of the Hessian matrix, and $N^{-1}\mathbb{E}[\mathrm{D}(\boldsymbol{\theta}_0)\mathrm{D}(\boldsymbol{\theta}_0)^T] \to \mathbb{V}$. If the second-order condition (i) holds and $\sqrt{N}A(t_N) \to 0$, then we will have $\boldsymbol{b} = \mathbf{0}$ which implies no bias.

In order to obtain asymptotic properties for $\hat{\boldsymbol{\theta}}$, we need to understand the behavior of $\mathbb{E}\mathrm{D}(\boldsymbol{\theta}_0)$ given the second-order regular variation, where D is the score functions of pairwise composite likelihood. The following defines the statement on how integrals of the score functions behave corresponding to the second-order condition.

**Proposition 1.** *Let $g_t(x,y)$ be any measurable function. Suppose $F_{b_t,d_t}$ satisfies the condition with (i) or (ii) with function $A$. Define $f_{b_t,d_t} = \frac{d^2 F_{b_t,d_t}}{dxdy}$, $h(x,y) = \frac{d^2 H(x,y)}{dxdy}$ and $\psi(x,y) = \frac{d^2\Psi(x,y)}{dxdy}$. If*

$$\left| g_t(x,y) \left\{ \frac{f_{b_t,d_t}(a_t x, c_t y) - h(x,y)}{A(t)} - \psi(x,y) \right\} \right| \leq K(x,y) \tag{4.6}$$

*which $K(x,y)$ is integrable, then in case of (i)*

$$\int_E g_t(x,y)dF_{b_t,d_t}(a_t x, c_t y) = \int_E g_t(x,y)dH(x,y) + O(A(t))$$

*and in case of (ii)*

$$\int_E g_t(x,y)dF_{b_t,d_t}(a_t x, c_t y) = \int_E g_t(x,y)dH(x,y) + A(t)\int_E g_t(x,y)d\Psi(x,y) + o(A(t)).$$

*Proof.* As $t \to \infty$, we have to prove that

$$\int_{(0,\infty]^2} g_t(x,y)\left\{\frac{f_{b_t,d_t}(a_t x, c_t y) - h(x,y)}{A(t)} - \psi(x,y)\right\}dxdy \longrightarrow 0.$$

and by dominated convergence theorem, it is sufficient to show that

$$\left|g_t(x,y)\left\{\frac{f_{b_t,d_t}(a_t x, c_t y) - h(x,y)}{A(t)} - \psi(x,y)\right\}\right| \leq K(x,y).$$

where $K(x,y)$ is an integrable function. □

Let $g_t(x,y)$ be the score functions from the pairwise composite likelihood of a max-stable process. Note that $\int g_t(x,y)dH(x,y) = 0$ since $g_t(x,y)$ is the score function. Limit distribution of estimator for dependence parameter can be determined by **Condition 1** and **Proposition 1** implies that condition (4.6) should be satisfied for the limit behavior. We end this section with an example to demonstrate how the proposition works. Here we focus on the example with a certain type of $g_t(x,y)$, the score function obtained from the composite likelihood of Brown-Resnick process, and we intend to show that the condition (4.6) holds assuming that $F$ is a bivariate normal distribution.

**Example (bivariate normal distribution):** Suppose that $(X,Y)$ are i.i.d. from a bivariate normal distribution $F$ with mean 0, variance 1 and correlation coefficient $\rho$. First we would like to prove that bivariate normal distribution satisfies (4.5). We consider $G$ in (4.1) as a bivariate extreme value distribution with Gumbel margins and suppose the limiting form of bivariate normal $G(x,y) = \exp\{-e^{-x} - e^{-y}\}$ in the case of the independence. A max-stable process with unit Fréchet margins will be fitted and the transformations $X' = \log X$ and $Y' = \log Y$ can be made from unit Fréchet to Gumbel.

Mills ratio for a normal density implies that

$$\begin{aligned} \frac{1-\Phi(x)}{\phi(x)} &\sim \left\{\frac{1}{x} - \frac{1}{x^3} + \frac{1\cdot 3}{x^5} - \frac{1\cdot 3\cdot 5}{x^7} + \cdots\right\}, \\ \frac{P(X>x, Y>y)}{\phi(x,y)} &\sim \frac{(1-\rho^2)^2}{(x-\rho y)(y-\rho x)} \times \\ &\qquad \left\{1-(1-\rho^2)\left(\frac{1}{(x-\rho y)^2} - \frac{\rho}{(x-\rho y)(y-\rho x)} + \frac{1}{(y-\rho x)^2}\right) + \cdots\right\} \end{aligned}$$

(see Ruben [1964] for the bivariate normal density). From the fact that

$$1 - F(x,y) = 1 - \Phi(x) + 1 - \Phi(y) - P(X>x, Y>y),$$

we could set the lower bound and upper bound for $\frac{1-F(x,y)}{\phi(x,y)}$ such that

$$\begin{aligned} &\left(\frac{1-F(x,y)}{\phi(x,y)}\right)^L \le \frac{1-F(x,y)}{\phi(x,y)} \le \left(\frac{1-F(x,y)}{\phi(x,y)}\right)^U, \\ \text{where } &\left(\frac{1-F(x,y)}{\phi(x,y)}\right)^L = \frac{1}{x} + \frac{1}{y} - \frac{1}{x^3} - \frac{1}{y^3} - \frac{(1-\rho^2)^2}{(x-\rho y)(y-\rho x)}, \\ &\left(\frac{1-F(x,y)}{\phi(x,y)}\right)^U = \frac{1}{x} + \frac{1}{y} - \frac{(1-\rho^2)^2}{(x-\rho y)(y-\rho x)} + \frac{(1-\rho^2)^3}{(x-\rho y)(y-\rho x)} \times \\ &\qquad \left(\frac{1}{(x-\rho y)^2} - \frac{\rho}{(x-\rho y)(y-\rho x)} + \frac{1}{(y-\rho x)^2}\right). \end{aligned}$$

From the well-known results of extreme value theory, define $b_t$ by $1 - \Phi(b_t) = \frac{1}{t}$

and $a_t = 1/b_t$. Or we might set normalized constants

$$
\begin{aligned}
a_t &= \frac{1}{\sqrt{2\log t}} \\
b_t &= \sqrt{2\log t} - \frac{\frac{1}{2}(\log\log t + \log 4\pi)}{\sqrt{2\log t}}.
\end{aligned}
$$

Conditional distribution of exceedances over threshold is written as

$$
F_{b_t,d_t}(a_t x, c_t y) = 1 - \frac{t\{1 - F(a_t x + b_t, c_t y + d_t)\}}{t\{1 - F(b_t, d_t)\}}
$$

and we now concentrate on $\frac{1-F(a_t x+b_t, c_t y+d_t)}{1-F(b_t,d_t)}$,

$$
\begin{aligned}
&\frac{1 - F(a_t x + b_t, c_t y + d_t)}{1 - F(b_t, d_t)} = \frac{1 - F(a_t x + b_t, c_t y + d_t)}{\phi(a_t x + b_t, c_t y + d_t)} \cdot \frac{\phi(b_t, d_t)}{1 - F(b_t, d_t)} \cdot \frac{\phi(a_t x + b_t, c_t y + d_t)}{\phi(b_t, d_t)} \\
&\geq \left\{ \frac{b_t}{x + b_t^2} + \frac{d_t}{y + d_t^2} - \frac{b_t^3}{(x + b_t^2)^3} - \frac{b_t^3}{(y + b_t^2)^3} - \frac{(1-\rho^2)^2 b_t^2}{\big(x - \rho y + b_t^2(1-\rho)\big)\big(y - \rho x + b_t^2(1-\rho)\big)} \right\} \\
&\quad \times \left\{ \frac{b_t^4}{2b_t^3 - (1+\rho)^2 b_t^2 + \frac{(1+\rho)^3(2-\rho)}{1-\rho}} \right\} \frac{\phi(a_t x + b_t, c_t y + d_t)}{\phi(b_t, d_t)} \\
&\sim \left\{ \frac{2b_t^2 - (1+\rho)^2 b_t - 2}{b_t^3} \right\} \left\{ \frac{b_t^4}{2b_t^3 - (1+\rho)^2 b_t^2 + \frac{(1+\rho)^3(2-\rho)}{1-\rho}} \right\} \frac{\phi(x/b_t + b_t, y/b_t + b_t)}{\phi(b_t, b_t)}
\end{aligned}
$$

and also,

$$
\begin{aligned}
&\frac{1 - F(a_t x + b_t, c_t y + d_t)}{1 - F(b_t, d_t)} \\
&\quad \leq \Bigg[ \frac{b_t}{x + b_t^2} + \frac{d_t}{y + d_t^2} - \frac{(1-\rho^2)^2 b_t^2}{\big(x - \rho y + b_t^2(1-\rho)\big)\big(y - \rho x + b_t^2(1-\rho)\big)} \Bigg\{ 1 - (1-\rho^2) \times \\
&\quad \Bigg( \frac{b_t^2}{\big(x - \rho y + b_t^2(1-\rho)\big)^2} + \frac{b_t^2}{\big(y - \rho x + b_t^2(1-\rho)\big)^2} \\
&\quad - \frac{\rho b_t^2}{\big(x - \rho y + b_t^2(1-\rho)\big)\big(y - \rho x + b_t^2(1-\rho)\big)} \Bigg) \Bigg\} \Bigg] \left( \frac{b_t^3}{2b_t^2 - (1+\rho)^2 b_t - 2} \right) \\
&\quad \times \frac{\phi(a_t x + b_t, c_t y + d_t)}{\phi(b_t, d_t)} \\
&\quad \sim \left\{ \frac{2b_t^3 - (1+\rho)^2 b_t^2 + \frac{(1+\rho)^3(2-\rho)}{1-\rho}}{b_t^4} \right\} \left\{ \frac{b_t^3}{2b_t^2 - (1+\rho)^2 b_t - 2} \right\} \frac{\phi(x/b_t + b_t, y/b_t + b_t)}{\phi(b_t, b_t)}.
\end{aligned}
$$

Thus

$$F_{b_t,d_t}(a_t x, c_t y) - H(x,y) = -\frac{1-F(a_t x + b_t, c_t y + d_t)}{1-F(b_t,d_t)} + (e^{-x} + e^{-y})$$
$$\sim \left\{ -\frac{2b_t^3 - (1+\rho)^2 b_t^2 - 2b_t}{2b_t^3 - (1+\rho)^2 b_t^2 + \frac{(1+\rho)^3(2-\rho)}{1-\rho}} + 1 \right\} \left\{ \frac{\phi(x/b_t + b_t, y/b_t + b_t)}{\phi(b_t,b_t)} + e^{-x} + e^{-y} \right\}$$

and

$$-\frac{2b_t^3 - (1+\rho)^2 b_t^2 - 2b_t}{2b_t^3 - (1+\rho)^2 b_t^2 + \frac{(1+\rho)^3(2-\rho)}{1-\rho}} + 1 = \frac{2b_t + \frac{(1+\rho)^3(2-\rho)}{1-\rho}}{2b_t^3 - (1+\rho)^2 b_t^2 + \frac{(1+\rho)^3(2-\rho)}{1-\rho}}.$$

We obtain the formation of (4.5)

$$\lim_{t\to\infty} \frac{F_{b_t,d_t}(a_t x, c_t y) - H(x,y)}{A(t)} = \Psi(x,y)$$

where $A(t) = \frac{1}{b_t^2} = \frac{1}{2\log t}$ and $\Psi(x,y) = \exp\left\{ -\frac{x+y}{1+\rho} \right\} + e^{-x} + e^{-y}$.

Next,

$$f_{b_t,d_t}(a_t x, c_t y) = \frac{a_t c_t}{1-F(b_t,d_t)} \cdot \frac{1}{2\pi\sqrt{1-\rho^2}} \times$$
$$\exp\left\{ -\frac{(a_t x + b_t)^2 + (c_t y + d_t)^2 - 2\rho(a_t x + b)(c_t y + d_t)}{2(1-\rho^2)} \right\}$$
$$= a_t c_t \frac{\phi(b_t,d_t)}{1-F(b_t,d_t)} \cdot \frac{\phi(a_t x + b_t, c_t y + d_t)}{\phi(b_t,d_t)} \doteq a_t c_t \frac{\phi(b_t,d_t)}{1-F(b_t,d_t)} \cdot V_t(x,y) \qquad (4.7)$$

where $\phi(x,y)$ is a bivariate normal density with correlation $\rho$.

$\frac{\phi(x,y)}{1-F(x,y)}$ as a factor of $f_{b_t,d_t}(a_t x, c_t y)$ in the equation (4.7) has the lower and upper bounds that

$$\left( \frac{\phi(x,y)}{1-F(x,y)} \right)^L \le \frac{\phi(x,y)}{1-F(x,y)} \le \left( \frac{\phi(x,y)}{1-F(x,y)} \right)^U,$$

where

$$\left(\frac{\phi(x,y)}{1-F(x,y)}\right)^L = \left\{\frac{1}{x}+\frac{1}{y}-\frac{(1-\rho^2)^2}{(x-\rho y)(y-\rho x)}+\frac{(1-\rho^2)^3}{(x-\rho y)(y-\rho x)}\times\right.$$
$$\left.\left(\frac{1}{(x-\rho y)^2}-\frac{\rho}{(x-\rho y)(y-\rho x)}+\frac{1}{(y-\rho x)^2}\right)\right\}^{-1},$$
$$\left(\frac{\phi(x,y)}{1-F(x,y)}\right)^U = \left\{\frac{1}{x}+\frac{1}{y}-\frac{1}{x^3}-\frac{1}{y^3}-\frac{(1-\rho^2)^2}{(x-\rho y)(y-\rho x)}\right\}^{-1}.$$

Since $f_{b_t,d_t}(a_t x, c_t y) = a_t c_t \frac{\phi(b_t,d_t)}{1-F(b_t,d_t)} \cdot V_t(x,y)$, using above normalized constants and assuming $b_t = d_t$

$$a_t c_t\left(\frac{\phi(b_t,d_t)}{1-F(b_t,d_t)}\right)^L = \frac{b_t^2}{2b_t^3-(1+\rho)^2 b_t^2+\frac{(1+\rho)^3(2-\rho)}{1-\rho}}$$
$$a_t c_t\left(\frac{\phi(b_t,d_t)}{1-F(b_t,d_t)}\right)^U = \frac{b_t}{2b_t^2-(1+\rho)^2 b_t-2}$$
$$V_t(x,y) = \frac{\phi(x/b_t+b_t, y/b_t+b_t)}{\phi(b_t,b_t)} = \exp\left\{-\frac{x^2+y^2-2\rho xy}{2(1-\rho^2)b_t^2}-\frac{x+y}{1+\rho}\right\}.$$

Thus we could get the following form of bounds

$$f^L_{b_t,d_t}(a_t x, c_t y) = \frac{b_t^2}{2b_t^3-(1+\rho)^2 b_t^2+\frac{(1+\rho)^3(2-\rho)}{1-\rho}}\frac{\phi(x/b_t+b_t, y/b_t+b_t)}{\phi(b_t,b_t)}$$
$$f^U_{b_t,d_t}(a_t x, c_t y) = \frac{b_t}{2b_t^2-(1+\rho)^2 b_t-2}\frac{\phi(x/b_t+b_t, y/b_t+b_t)}{\phi(b_t,b_t)}.$$

Meanwhile

$$h(x,y) = \frac{\partial^2 H(x,y)}{\partial x\partial y} = -\frac{1}{\log G(0,0)}\cdot\frac{\partial^2}{\partial x\partial y}\log G(x,y) = 0.$$

Therefore

$$\begin{aligned}
f_{b_t,d_t}(a_t x, c_t y) - h(x,y) &\geq \{f_{b_t,d_t}(a_t x, c_t y) - h(x,y)\}^L \\
&= \frac{b_t^2}{2b_t^3 - (1+\rho)^2 b_t^2 + \frac{(1+\rho)^3(2-\rho)}{1-\rho}} \frac{\phi(x/b_t + b_t, y/b_t + b_t)}{\phi(b_t, b_t)}, \\
f_{b_t,d_t}(a_t x, c_t y) - h(x,y) &\leq \{f_{b_t,d_t}(a_t x, c_t y) - h(x,y)\}^U \\
&= \frac{b_t}{2b_t^2 - (1+\rho)^2 b_t - 2} \frac{\phi(x/b_t + b_t, y/b_t + b_t)}{\phi(b_t, b_t)}.
\end{aligned}$$

Define $A(t) = \frac{1}{2\log t}$ ($A(t) \to 0$ as $t \to \infty$) and $\psi(x,y) = -\frac{\rho}{2(1-\rho^2)}$ to satisfy the condition (4.5). Then we could show that

$$\begin{aligned}
\frac{f_{b_t,d_t}(a_t x, c_t y) - h(x,y)}{A(t)} - \psi(x,y) &\geq \frac{f_{b_t,d_t}(a_t x, c_t y)^L - h(x,y)}{1/(2\log t)} - \psi(x,y) \\
&\sim \exp\left(-\frac{x+y}{1+\rho}\right)\left\{\frac{b_t}{2}\exp\left(-a_t^2\frac{x^2+y^2-2\rho xy}{2(1-\rho^2)}\right) - \frac{1}{(1+\rho)^2}\right\}, \\
\frac{f_{b_t,d_t}(a_t x, c_t y) - h(x,y)}{A(t)} - \psi(x,y) &\leq \frac{f_{b_t,d_t}(a_t x, c_t y)^U - h(x,y)}{1/(2\log t)} - \psi(x,y) \\
&\sim \exp\left(-\frac{x+y}{1+\rho}\right)\left\{\frac{b_t}{2}\exp\left(-a_t^2\frac{x^2+y^2-2\rho xy}{2(1-\rho^2)}\right) - \frac{1}{(1+\rho)^2}\right\}.
\end{aligned}$$

This limit for bounds of $\frac{f_{b_t,d_t}-h}{A(t)} - \psi(x,y)$ will be used to prove that the product of a function $g_t(x,y)$ and $\frac{f_{b_t,d_t}-h}{A(t)} - \psi(x,y)$ is bounded by an integrable function as shown in (4.6). Suppose that $g_t(x,y) = \frac{\partial}{\partial\theta}\log f_{DA}(x,y;\theta)$ where $f_{DA} = \frac{\partial^2 F_{DA}(x,y)}{\partial x \partial y}$. Any max-stable process can be fitted for modeling annual maxima of data and we can obtain the score function by our threshold method with the composite likelihood approach. We arbitrarily choose the Brown-Resnick process with Gumbel margins to obtain the joint bivariate distribution of annual data, $F_{AM}$, and a joint bivariate distribution of daily data, $F_{DA}(x,y)$, is determined by the relation (3.5).

$$F_{AM}(x,y;\theta) = \exp\{B(x,y;\theta)\},$$

where $B(x,y;\theta) = \Big\{ -\frac{1}{x}\Phi\Big(\frac{\sqrt{\gamma(h;\theta)}}{2} + \frac{1}{\sqrt{\gamma(h;\theta)}}\log\frac{y}{x}\Big) - \frac{1}{y}\Phi\Big(\frac{\sqrt{\gamma(h;\theta)}}{2} + \frac{1}{\sqrt{\gamma(h;\theta)}}\log\frac{x}{y}\Big)\Big\}$ and

$$\log f_{DA}(x,y;\theta) = \frac{1}{M}B(x,y;\theta) + \log J(x,y;\theta),$$

where $J(x,y;\theta) = \frac{1}{M}\frac{\partial^2 B(x,y;\theta)}{\partial x \partial y} + \frac{1}{M^2}\frac{\partial B(x,y;\theta)}{\partial x}\cdot\frac{\partial B(x,y;\theta)}{\partial y}$. Therefore,

$$g_t(x,y) = \frac{1}{M}\frac{\partial B(x,y;\theta)}{\partial \theta} + J(x,y;\theta)^{-1}\Bigg(\frac{\partial J(x,y;\theta)}{\partial\theta}\Bigg) \tag{4.8}$$

where $\frac{\partial J(\theta)}{\partial\theta} = \frac{1}{M}\frac{\partial}{\partial\theta}\Big(\frac{\partial^2 B(x,y;\theta)}{\partial x\partial y}\Big) + \frac{1}{M^2}\frac{\partial}{\partial\theta}\Big(\frac{\partial B(x,y;\theta)}{\partial x}\Big)\cdot\frac{\partial B(x,y;\theta)}{\partial y} + \frac{1}{M^2}\frac{\partial B(x,y;\theta)}{\partial x}\cdot\frac{\partial}{\partial\theta}\Big(\frac{\partial B(x,y;\theta)}{\partial y}\Big)$. With some calculations, the derivatives of $J(x,y;\theta)$ and $B(x,y;\theta)$, shortly $J$ and $B$, can be obtained as in Appendix A and the boundness of the product is of interest:

$$\begin{aligned}&\Bigg|g_t(x,y)\Bigg\{\frac{f_{b_t,d_t}(a_t x, c_t y) - h(x,y)}{A(t)} - \psi(x,y)\Bigg\}\Bigg| \\ &\qquad \le \Bigg|g_t(x,y)\exp\Bigg(-\frac{x+y}{1+\rho}\Bigg)\Bigg\{\frac{b_t}{2}\exp\Bigg(-a_t^2\frac{x^2+y^2-2\rho xy}{2(1-\rho^2)}\Bigg) - \frac{1}{(1+\rho)^2}\Bigg\}\Bigg|. \end{aligned} \tag{4.9}$$

<u>**Case (i):** $x = y$</u>

$$\begin{aligned}
\frac{\partial B}{\partial\theta} &= \Bigg(\frac{\partial\gamma}{\partial\theta}\Bigg)\Bigg\{-e^{-x}\Bigg(\frac{1}{2\sqrt{\gamma}}\Bigg)\phi\Bigg(\frac{\sqrt{\gamma}}{2}\Bigg)\Bigg\},\\
J(\theta) &= \frac{\sqrt{\gamma}}{M}e^{-x}\phi\Bigg(\frac{\sqrt{\gamma}}{2}\Bigg) + \frac{1}{M^2}e^{-2x}\Bigg\{\Phi^2\Bigg(\frac{\sqrt{\gamma}}{2}\Bigg) - \frac{2}{\gamma}\phi^2\Bigg(\frac{\sqrt{\gamma}}{2}\Bigg)\Bigg\},\\
\frac{\partial J}{\partial\theta} &= \Bigg(\frac{\partial\gamma}{\partial\theta}\Bigg)\Bigg\{-\frac{1}{M}\Bigg(\frac{1}{8\sqrt{\gamma}} + \frac{1}{2\sqrt{\gamma^3}}\Bigg)e^{-x}\phi\Bigg(\frac{\sqrt{\gamma}}{2}\Bigg) + \frac{1}{M^2}\Bigg(\frac{1}{2\sqrt{\gamma}}\Bigg)e^{-2x}\phi\Bigg(\frac{\sqrt{\gamma}}{2}\Bigg)\Phi\Bigg(\frac{\sqrt{\gamma}}{2}\Bigg)\Bigg\}.
\end{aligned}$$

Then

$$g_t(x,y) \le \Bigg(\frac{\partial\gamma}{\partial\theta}\Bigg)\Bigg\{-\frac{1}{M}\Bigg(\frac{1}{2\sqrt{\gamma}}\Bigg)\Bigg(1 - \frac{\Phi\big(\frac{\sqrt{\gamma}}{2}\big)}{\sqrt{\gamma}\phi\big(\frac{\sqrt{\gamma}}{2}\big) + \frac{e^{-x}}{M}\big\{\Phi^2\big(\frac{\sqrt{\gamma}}{2}\big) - \frac{2}{\gamma}\phi^2\big(\frac{\sqrt{\gamma}}{2}\big)\big\}}\Bigg)\phi\Bigg(\frac{\sqrt{\gamma}}{2}\Bigg)\Bigg\}e^{-x}$$

and therefore, for some constants $C_i$

$$
\begin{aligned}
&\left| g_t(x,y)\left\{ \frac{f_{b_t,d_t}(a_t x, c_t y) - h(x,y)}{A(t)} - \psi(x,y) \right\} \right| \\
&\qquad \le C_1 \left( \frac{\partial \gamma}{\partial \theta} \right) e^{-x} e^{-\frac{2x}{1+\rho}} \left\{ b_t \exp\left( - \frac{x^2}{(1+\rho) b_t^2} \right) - \frac{2}{(1+\rho)^2} \right\} \\
&\qquad \le C_2 \phi\left( \frac{\sqrt{2}x}{\sqrt{1+\rho} b_t} + \frac{b_t(3+\rho)}{\sqrt{2(1+\rho)}} \right),
\end{aligned}
$$

which implies that (4.9) is bounded by an integrable function.

**Case (ii): $y = x + k$ and $x \to \infty$**

Let

$$
\begin{aligned}
\frac{\sqrt{\gamma}}{2} + \frac{1}{\sqrt{\gamma}}(y - x) &= \frac{\sqrt{\gamma}}{2} + \frac{k}{\sqrt{\gamma}} = a, \\
\frac{\sqrt{\gamma}}{2} + \frac{1}{\sqrt{\gamma}}(x - y) &= \frac{\sqrt{\gamma}}{2} - \frac{k}{\sqrt{\gamma}} = b, \\
\frac{1}{4\sqrt{\gamma}} - \frac{1}{2\sqrt{\gamma^3}}(x - y) &= \frac{1}{2\gamma}\left( \frac{\sqrt{\gamma}}{2} + \frac{y - x}{\sqrt{\gamma}} \right) = \frac{1}{2\gamma} a, \\
\frac{1}{4\sqrt{\gamma}} - \frac{1}{2\sqrt{\gamma^3}}(y - x) &= \frac{1}{2\gamma}\left( \frac{\sqrt{\gamma}}{2} + \frac{x - y}{\sqrt{\gamma}} \right) = \frac{1}{2\gamma} b.
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial B}{\partial \theta} &= \left( \frac{\partial \gamma}{\partial \theta} \right) \left\{ - e^{-x} \left( \frac{1}{2\gamma} \right) \big( b\phi(a) + e^{-k} a \phi(b) \big) \right\}, \\
J(\theta) &= \frac{1}{M} e^{-x} \big( b\phi(a) + e^{-k} a\phi(b) \big) \\
&\quad + \frac{1}{M^2} e^{-2x} \left\{ e^{-k} \left( \Phi(a)\Phi(b) + \frac{1}{\sqrt{\gamma}} \Phi(a)\phi(b) + \frac{1}{\sqrt{\gamma}} \phi(a)\Phi(b) \right) \right. \\
&\quad \left. - \frac{\phi(a)}{\sqrt{\gamma}} \left( \Phi(a) + \frac{\phi(a)}{\sqrt{\gamma}} \right) - e^{-2k} \frac{\phi(b)}{\sqrt{\gamma}} \left( \Phi(b) + \frac{\phi(b)}{\sqrt{\gamma}} \right) \right\}, \\
\frac{\partial J}{\partial \theta} &= \left( \frac{\partial \gamma}{\partial \theta} \right) \left[ \frac{1}{M} e^{-x} \big( \phi(a) k_3(k) + e^{-k} \phi(b) k_3(-k) \big) \right. \\
&\quad + \frac{e^{-2x}}{M^2} \big\{ \phi(a) k_1(k) + e^{-k} \phi(b) k_2(-k) \big\} \left\{ e^{-k} \left( \Phi(b) + \frac{\phi(b)}{\sqrt{\gamma}} \right) - \frac{\phi(a)}{\sqrt{\gamma}} \right\} \\
&\quad \left. + \frac{e^{-2x}}{M^2} \big\{ \phi(a) k_2(k) + e^{-k} \phi(b) k_1(-k) \big\} \left\{ \left( \Phi(a) + \frac{\phi(a)}{\sqrt{\gamma}} \right) - e^{-k} \frac{\phi(b)}{\sqrt{\gamma}} \right\} \right],
\end{aligned}
$$

where $k_1,\ k_2$ and $k_3$ are defined in Appendix A. Then for some constants $K_i$,

$$g_t(x,y) \le \left(\frac{\partial\gamma}{\partial\theta}\right)\left\{-\frac{1}{M}\left(\frac{1}{2\gamma}\right)\left(K_1 b\phi(a) + K_2 a\phi(b)e^{-k}\right)\right\}e^{-x}$$

and therefore, for some constants $C_i$

$$\begin{aligned}
&\left|g_t(x,y)\left\{\frac{f_{b_t,d_t}(a_t x, c_t y) - h(x,y)}{A(t)} - \psi(x,y)\right\}\right| \\
&\quad \le C_1\left(\frac{\partial\gamma}{\partial\theta}\right)e^{-x}e^{-\frac{2x+k}{1+\rho}}\cdot\frac{b_t}{2}\exp\left\{-\frac{2(1-\rho)x^2+2(1-\rho)kx}{2(1-\rho^2)b_t^2}\right\} \\
&\quad \le C_2\phi\left(\frac{2x+(3+\rho)b_t^2+k}{\sqrt{2(1+\rho)}b_t}\right)
\end{aligned}$$

which implies that (4.9) is bounded by an integrable function.

For the general case of $x \to \infty$ and $y \to \infty$, the boundness can be obtained. In (4.8), the first term $\frac{1}{M}\frac{\partial B}{\partial\theta}$ consists of the components; $-e^{-x}\phi\left(\frac{\sqrt{\gamma}}{2}+\frac{y-x}{\sqrt{\gamma}}\right)$ and $-e^{-y}\phi\left(\frac{\sqrt{\gamma}}{2}+\frac{x-y}{\sqrt{\gamma}}\right)$. In the second term of $g_t(x,y)$, $J(x,y;\theta)^{-1}\left(\frac{\partial J(x,y;\theta)}{\partial\theta}\right)$ is also dominated by $e^{-x}\phi\left(\frac{\sqrt{\gamma}}{2}+\frac{y-x}{\sqrt{\gamma}}\right)$ and $e^{-y}\phi\left(\frac{\sqrt{\gamma}}{2}+\frac{x-y}{\sqrt{\gamma}}\right)$. Then (4.9) is bounded by a function of $\phi(C_1x, C_2y)$ for a constant $C_i$, which is integrable. □

### 4.1.2 Spatial Structure and Sampling Design

Asymptotic results have been proved for spatial processes which are observed at finitely many locations in the sampling region. Central Limit Theorems for spatial data have been studied on infill domain and increasing domain structure under two types of sampling designs, a class of fixed (regular) lattice and stochastic (irregular) designs, in existing literature. Infill domain structure assumes that the sampling region is bounded and locations of data fill in increasingly and densely, while the sampling region is unbounded in the increasing domain structure. Lahiri [2003] is concerned with more complex spatial structure, called mixed asymptotic structure, as a mixture of infill- and increasing domain assumption. In the mixed asymptotic structure, the sampling region is unbounded and sites fill in densely over the region. Covariance parameters are not always consistently estimable if the spatial domain is bounded (Zhang [2004]), while the same parameters are estimable under the increasing domain structure (Mardia and Marshall [1984]). Here we focus on the increasing-domain case under stochastic design based on setting and conditions in Lahiri [2003]. Increasing domain structure takes advantage of dealing with asymptotic properties of estimators easily rather than the infill asymptotic structure. We could take account of more realistic setting under the stochastic sampling design than the fixed lattice design.

Suppose that the stationary random field $\{Z(\boldsymbol{s}); \boldsymbol{s} \in \mathbb{R}^d\}$ is observed at many stations $\boldsymbol{s}$ in the sampling region $R_n$. Under the increasing domain structure, $R_n$ is unbounded with $n$ and there is a minimum distance separating any two sites for all $n$. We assume that the sampling region $R_n$ is inflated by the factor $\lambda_n$ from the set $R_0$, i.e.,

$$R_n = \lambda_n R_0.$$

For the stochastic designs of sampling sites, we assume that the sampling sites $\{\boldsymbol{s}_1, \cdots, \boldsymbol{s}_n\}$ are obtained from a random vectors $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$ by

$$\boldsymbol{s}_i = \lambda_n \boldsymbol{x}_i, \quad 1 \le i \le n$$

where $\boldsymbol{x}_i$ is a sequence of i.i.d. random vectors from a continuous probability density function $f(\boldsymbol{x})$ and its realization $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$ are in $R_0$. In this stochastic design, the sample size $n$ is determined by the growth rate $\lambda_n$ by the relation $n \sim C\lambda_n^d$.

We now consider our threshold approach. Note that we assume the marginal GEV parameter $\boldsymbol{\eta}$ is known as the simplest case, though we would like to address the case $\boldsymbol{\eta}$ unknown as well. Assuming that $\boldsymbol{\eta}$ is known as $\boldsymbol{\eta}_0$, we can rewrite (3.6) and partial derivatives with the temporal domain fixed, as

$$
\begin{aligned}
l(\boldsymbol{\theta}) &= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \sum_{t=1}^{T} w_{ij} \log L(X_{ti}, X_{tj}; \boldsymbol{\theta}) \\
&= \sum_{i<j} w_{ij} \log L_{ij}(\boldsymbol{\theta}),
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= \sum_{i<j} \frac{w_{ij}}{L_{ij}(\boldsymbol{\theta})} \cdot \frac{\partial L_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \\
\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} &= \sum_{i<j} \frac{w_{ij}}{L_{ij}^2(\boldsymbol{\theta})} \left\{ \frac{\partial^2 L_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \cdot L_{ij}(\boldsymbol{\theta}) - \frac{\partial L_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \left( \frac{\partial L_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T \right\}
\end{aligned}
$$

where $w_{ij}$ is the weight function on the $(i, j)$th pair which does not take any values outside $R_n$, $L_{ij} = F_{ij}(x_i, x_j) I_{\{x_i > u, x_j > u\}} + F_i(x_i, u) I_{\{x_i > u, x_j \le u\}} + F_j(u, x_j) I_{\{x_i \le u, x_j > u\}} + F_{DA}(u, u) I_{\{x_i \le u, x_j \le u\}}$, and $F_{ij} = \frac{\partial^2 F_{DA}}{\partial x_i \partial x_j}$. Here $u$ is the threshold, not a fixed constant, which varies as the sample size goes to infinity.

We concentrate on the first term of $L_{ij}$ which is the case that both exceed the threshold. Let us define notations related with the first term by

$$
\begin{aligned}
Q_K(\boldsymbol{\theta}) &= \sum_{i<j}^{K} w_{ij} \log F_{ij}(\boldsymbol{\theta}) I_{\{x_i > u, x_j > u\}} \\
\frac{\partial Q_K(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= \sum_{i<j}^{K} \frac{w_{ij}}{F_{ij}(\boldsymbol{\theta})} \cdot \frac{\partial F_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} I_{\{x_i > u, x_j > u\}} \\
\frac{\partial^2 Q_K(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} &= \sum_{i<j}^{K} w_{ij} \left\{ \frac{1}{F_{ij}(\boldsymbol{\theta})} \frac{\partial^2 F_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} - \frac{1}{F_{ij}^2} \left( \frac{\partial F_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \left( \frac{\partial F_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T \right\} I_{\{x_i > u, x_j > u\}}
\end{aligned}
$$

where $K$ is the number of all combination of pairs.

Next we denote the form of the strong mixing assumption to deal with dependence through pairs. Let $X(\boldsymbol{s}_i) = Z(\boldsymbol{s}_i) I(Z(\boldsymbol{s}_i) > u_n)$ and $\mathcal{F}(\mathcal{G}_k)$ be $\sigma$-field generated by $\{(X(\boldsymbol{s}_i), X(\boldsymbol{s}_j)); \boldsymbol{s}_i, \boldsymbol{s}_j \in \mathcal{G}_k, 1 \le i, j \le n, k = 1, \cdots, K\}$. For any two subsets $A$ and $B$ of $\mathbb{R}^d$, the mixing condition is defined by

$$
\tilde{\alpha}(\mathcal{G}_1, \mathcal{G}_2) = \sup\{|P(A \cap B) - P(A)P(B)| : A \in \mathcal{F}(\mathcal{G}_1), B \in \mathcal{F}(\mathcal{G}_2)\}
$$

and let

$$d(\mathcal{G}_1, \mathcal{G}_2) = \inf\{|\boldsymbol{s} - \boldsymbol{s}'| : \boldsymbol{s} \in \mathcal{G}_1, \boldsymbol{s}' \in \mathcal{G}_2\}$$

which is the minimum distance from element of a pair $\mathcal{G}_1$ to element of another pair $\mathcal{G}_2$. Then the strong mixing coefficient is defined as

$$\alpha(a,b) = \sup\{\tilde{\alpha}(\mathcal{G}_1, \mathcal{G}_2) : d(\mathcal{G}_1, \mathcal{G}_2) \geq a, \quad \mathcal{G}_1, \mathcal{G}_2 \in \mathcal{R}_3(b)\}$$

where $\mathcal{R}_3(b) \equiv \{\cup_{i=1}^3 D_i : \sum_{i=1}^3 |D_i| \leq b\}$, the the collection of all disjoint unions of three cubes $D_1, D_2$ and $D_3$ in $\mathbb{R}^d$, and it specifies the general form of the sets $\mathcal{G}_1$ and $\mathcal{G}_2$ that are bounded. Assume that there exist a nonincreasing function $\alpha_1(\cdot)$ such that $\lim_{a\to\infty} \alpha_1(a) = 0$ and a nondecreasing function $\beta(\cdot)$ satisfying

$$\alpha(a,b) \leq \alpha_1(a)\beta(b), \quad a, b > 0.$$

In our approach, what we are interested in is the bivariate function,

$$\frac{\partial}{\partial \boldsymbol{\theta}} \log F_{ij}(x_i, x_j; \boldsymbol{\theta}) I(x_i > u, x_j > u) \doteq g_k\big(X(\boldsymbol{s}_i), X(\boldsymbol{s}_j)\big) \doteq Z_k(\boldsymbol{s}^k), \tag{4.10}$$

where $Z_k$ is obviously different from the original process $Z$. Let $\sigma(\cdot)$ denote the auto covariance function of the process $Z_k$ such that for all $\boldsymbol{s}_i, \boldsymbol{s}_j, \boldsymbol{h_1}, \boldsymbol{h_2} \in \mathbb{R}^d$,

$$\sigma(\boldsymbol{h}) = Cov(Z_k, Z_l) = Cov\Big[g_k\big(X(\boldsymbol{s}_i), X(\boldsymbol{s}_j)\big), g_l\big(X(\boldsymbol{s}_i + \boldsymbol{h}_1), X(\boldsymbol{s}_j + \boldsymbol{h}_2)\big)\Big].$$

To simplify the notation, let $s_{1K}^2 = Ew_K^2(\lambda_n \mathbf{X}_1) = \int\int w_{ij}^2(\lambda_n \mathbf{X}_{ij}) f(\boldsymbol{x}_i, \boldsymbol{x}_j) d\boldsymbol{x}_i d\boldsymbol{x}_j$ where $\mathbf{X}_{ij} = (\boldsymbol{x}_i, \boldsymbol{x}_j)$, $M_k = \{\sup|w_K(\boldsymbol{h})|;\ \boldsymbol{h} \in \mathbb{R}^{2d}\}$ and $\gamma_{1k}^2 = \frac{M_k^2}{s_{1K}^2}$.

We will use the following conditions which are similar with (S.1)-(S.5) in Lahiri (2003) to prove the asymptotic distribution of process.

(A′1) $\int\int |\sigma(\boldsymbol{x})| d\boldsymbol{x} < \infty$

(A′2) Let $R_0$ be a Borel set satisfying $R_0^* \subset R_0 \subset \bar{R}_0^*$ and $R_0^*$ be an open connected subset of $(-1/2, 1/2]^d$. The pdf $f(x)$ is continuous, everywhere positive with support $\bar{R}_0$, the closure of the set $R_0 \subset R^d$.

(A′3) Locations $\boldsymbol{x}_i$ are i.i.d. from $f$ over $R_0$ and suppose that $f(\boldsymbol{x}_i, \boldsymbol{x}_j) = f(\boldsymbol{x}_i) f(\boldsymbol{x}_j)$. The joint pdf $f(\boldsymbol{x}_i, \boldsymbol{x}_j) \in [m_f, M_f]$ where $m_f$ and $M_f$ are constants in $(0, \infty)$.

(A′4) $\frac{\int\int w_{ij}\left(\lambda_n(\boldsymbol{x}_i,\boldsymbol{x}_j)\right) w_{pq}\left(\lambda_n(\boldsymbol{x}_i,\boldsymbol{x}_j)+\boldsymbol{h}\right) f^2(\boldsymbol{x}_i,\boldsymbol{x}_j) d\boldsymbol{x}_i d\boldsymbol{x}_j}{\int\int w_{ij}^2\left(\lambda_n(\boldsymbol{x}_i,\boldsymbol{x}_j)\right) f(\boldsymbol{x}_i,\boldsymbol{x}_j) d\boldsymbol{x}_i d\boldsymbol{x}_j} \to Q_1(\boldsymbol{h})$ for all $i \neq p, j \neq q, \boldsymbol{h} \in \mathbb{R}^{2d}$

(A′5) $\frac{\int\int w_{ij}\left(\lambda_n(\boldsymbol{x}_i,\boldsymbol{x}_j)\right) w_{iq}\left(\lambda_n(\boldsymbol{x}_i,\boldsymbol{x}_j)+(0,\boldsymbol{h})\right) f^2(\boldsymbol{x}_i,\boldsymbol{x}_j) d\boldsymbol{x}_i d\boldsymbol{x}_j}{\int\int w_{ij}^2\left(\lambda_n(\boldsymbol{x}_i,\boldsymbol{x}_j)\right) f(\boldsymbol{x}_i,\boldsymbol{x}_j) d\boldsymbol{x}_i d\boldsymbol{x}_j} \to Q_2(\boldsymbol{h})$ for all $i = p, j \neq q, \boldsymbol{h} \in \mathbb{R}^d$.

(A′6) $\gamma_{1k}^2 = \frac{M_k^2}{s_{1K}^2} = O(K^a)$ for some $a \in [0, 1/8)$

(A′7) There exist sequences $\{\lambda_{1n}\}$, $\{\lambda_{2n}\}$ with $\{\lambda_{1n}\} \geq \{\lambda_{2n}\} \geq \log\{\lambda_n\}$ such that

- (i) $\gamma_{1k}^2 (\log n)^2 \left[\frac{\lambda_{1n}}{\lambda_n} + \frac{\lambda_{2n}}{\lambda_{1n}}\right] = o(1)$
- (ii) $\gamma_{1k}^4 (\log n)^4 \left(\frac{\lambda_{1n}^d}{\lambda_n^d}\right) \sum_{k=1}^{\lambda_{1n}} k^{2d-1} \alpha_1(k) = o(1)$
- (iii) $\frac{\lambda_{1n}^d}{\lambda_n^d} \alpha_1(\lambda_{2n}) \beta(\lambda_n^d) = o(1)$
- (iv) $\gamma_{1k}^2 [\lambda_{1n}^d \alpha_1(\lambda_{2n}) + \sum_{k=\lambda_{1n}}^{\infty} k^{d-1} \alpha_1(k)] \beta(\lambda_{1n}^d) = o(1)$

**Theorem 1.** *Assume that conditions (A′1)-(A′7) hold. Suppose that $Z_k(\boldsymbol{s}^k)$ in (4.10) is a stationary stochastic process such that $E|Z_k(0)|^{2+\delta} < \infty$ and $\int t^{d-1} \alpha_1(t)^{\frac{\delta}{2+\delta}} dt < \infty$ for some $\delta > 0$. If $n/\lambda_n^d \to C_1 \in (0, \infty)$ as $n \to \infty$, then*

$$(K s_{1K}^2)^{-1/2} \sum_{k=1}^{K} w_K(\boldsymbol{s}_k) Z_k(\boldsymbol{s}^k) \xrightarrow{d} N\left(0, \sigma(\boldsymbol{0}) + C_1 \int \sigma\big((0,\boldsymbol{h})\big) Q_2(\boldsymbol{h}) d\boldsymbol{h} + C_1^2 \int \sigma(\boldsymbol{h}) Q_1(\boldsymbol{h}) d\boldsymbol{h}\right).$$

Proof of Theorem 1 is shown in Appendix B.

## 4.2 Asymptotic Properties: Asymptotic Normality and Consistency

We use the following regularity conditions to obtain an asymptotic behavior of estimates of dependence parameters.

(A1) The support $\boldsymbol{\chi}$ of the bivariate density function of the data does not depend on $\boldsymbol{\theta} \in \Theta$ and the parameter space $\Theta$ is an open subset of $\mathbb{R}^p$ with identifiable parametrization.

(A2) The pairwise composite log likelihood is at least twice continuously differentiable in $\boldsymbol{\theta}$.

(A3) (smoothness of composite likelihood) $Q_T(\boldsymbol{\theta})$ exists and is continuous and $\mathbb{H}(\boldsymbol{\theta})$ is also continuous in a neighborhood $\Theta^*$ of $\boldsymbol{\theta}_0$.

(A4) For all $\boldsymbol{\theta}_0 \in \Theta$, there exists an integrable function $M(x,y)$ such that

$$\sup_{\boldsymbol{\theta} \in \Theta^*} \left| \frac{\partial^2 Q_T(\boldsymbol{\theta}; x, y)}{\partial \theta_i \partial \theta_j} \right| \leq M(x,y), \ \ i, j = 1, \cdots, p.$$

(A5) The third partial derivatives of the composite likelihood are bounded by integrable functions.

(A6) (equivalent condition of Proposition 1) The score function of composite likelihood $D$ satisfies that

$$\left| D(x,y) \left\{ \frac{f_{b_t,d_t}(a_t x, c_t y) - h(x,y)}{A(t)} - \psi(x,y) \right\} \right| \leq K(x,y)$$

which $K(x,y)$ is integrable.

**Theorem 2.** *(Asymptotic Normality) Suppose that condition (4.5) with (i) or (ii) is satisfied and conditions of Theorem 1 hold. Suppose* $N \to \infty$, $(b_k, d_k) = (b(k)_N, d(k)_N) \to$

$(x_0, y_0)$, *and* $A(k_N) = O\Big(\frac{1}{\sqrt{Ns_{1N}^2}}\Big)$. *If*

$$\sqrt{Ns_{1N}^2}A(k_N) \longrightarrow \lambda \in [0, \infty),$$

*and either* $\lambda = 0$ *and (i) holds, then the the solutions of likelihood equations verify*

$$\sqrt{N}(s_{1N}^2)^{-1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N\big(\mathbf{0}, \mathbb{H}(\boldsymbol{\theta}_0)^{-1}\mathbb{V}(\boldsymbol{\theta}_0)\mathbb{H}(\boldsymbol{\theta}_0)^{-1}\big), \tag{4.11}$$

*or (ii) holds, then*

$$\sqrt{N}(s_{1N}^2)^{-1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N\big(\mathbb{H}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{b}, \mathbb{H}(\boldsymbol{\theta}_0)^{-1}\mathbb{V}(\boldsymbol{\theta}_0)\mathbb{H}(\boldsymbol{\theta}_0)^{-1}\big), \tag{4.12}$$

*where* $\mathbb{H}(\boldsymbol{\theta}_0) = \mathbb{E}[-\mathrm{D}'(\boldsymbol{\theta}_0)]$, $\boldsymbol{b} = \lim_{N\to\infty}\big(Ns_{1N}^2\big)^{-1/2}\mathbb{E}\big\{\sum_{k=1}^N \mathrm{D}(\boldsymbol{\theta}_0; \boldsymbol{X}^{(k)})\big\}$ *(defined below) and* $\mathbb{V}(\boldsymbol{\theta}_0) = \mathbb{E}[\mathrm{D}(\boldsymbol{\theta}_0)\mathrm{D}(\boldsymbol{\theta}_0)^T]$.

*Proof.* Denote that

$$\begin{aligned}
Q_T(\boldsymbol{\theta}) &= \sum_{i<j}^K \sum_t w_{ij} \log L_{ij}(\boldsymbol{\theta}) \\
&= \sum_{i<j}^K \sum_t w_{ij} \log\big\{F_{ij}(x_i, x_j)I_{\{x_i>u, x_j>u\}} + F_i(x_i, u)I_{\{x_i>u, x_j\le u\}} \\
&\qquad + F_j(u, x_j)I_{\{x_i\le u, x_j>u\}} + F_{DA}(u, u)I_{\{x_i\le u, x_j\le u\}}\big\} \\
&:= \sum_{i<j}^K \sum_t w_{ij} \log\big\{L_{1ij}(\boldsymbol{\theta}; x_i, x_j) + L_{2ij}(\boldsymbol{\theta}; x_i) + L_{3ij}(\boldsymbol{\theta}; x_j) + L_{4ij}(\boldsymbol{\theta})\big\},
\end{aligned}$$

$$\begin{aligned}
\frac{\partial Q_T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= \sum_{i<j}^K \sum_t \frac{w_{ij}}{L_{ij}(\boldsymbol{\theta})} \cdot \frac{\partial L_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \\
&= \sum_{i<j}^K \sum_t \frac{w_{ij}}{L_{ij}(\boldsymbol{\theta})} \cdot \frac{\partial}{\partial \boldsymbol{\theta}}\big\{L_{1ij}(\boldsymbol{\theta}; x_i, x_j) + L_{2ij}(\boldsymbol{\theta}; x_i) + L_{3ij}(\boldsymbol{\theta}; x_j) + L_{4ij}(\boldsymbol{\theta})\big\} \\
&:= \sum_{k=1}^N D\big(\boldsymbol{\theta}; (X_{ti}^{(k)}, X_{tj}^{(k)})\big) = \sum_{k=1}^N \big\{D_1(\boldsymbol{\theta}) + D_2(\boldsymbol{\theta}) + D_3(\boldsymbol{\theta}) + D_4(\boldsymbol{\theta})\big\}
\end{aligned}$$

where

$$D_1(\boldsymbol{\theta}) = \sum_{i<j}^{K}\sum_{t}\frac{w_{ij}}{L_{ij}(\boldsymbol{\theta})}\cdot\frac{\partial}{\partial\boldsymbol{\theta}}L_{1ij}(\boldsymbol{\theta};x_i,x_j),$$
$$D_2(\boldsymbol{\theta}) = \sum_{i<j}^{K}\sum_{t}\frac{w_{ij}}{L_{ij}(\boldsymbol{\theta})}\cdot\frac{\partial}{\partial\boldsymbol{\theta}}L_{2ij}(\boldsymbol{\theta};x_i),$$
$$D_3(\boldsymbol{\theta}) = \sum_{i<j}^{K}\sum_{t}\frac{w_{ij}}{L_{ij}(\boldsymbol{\theta})}\cdot\frac{\partial}{\partial\boldsymbol{\theta}}L_{3ij}(\boldsymbol{\theta};x_j),$$
$$D_4(\boldsymbol{\theta}) = \sum_{i<j}^{K}\sum_{t}\frac{w_{ij}}{L_{ij}(\boldsymbol{\theta})}\cdot\frac{\partial}{\partial\boldsymbol{\theta}}L_{4ij}(\boldsymbol{\theta}),$$

and $K$ is the number of all combination of pairs. We now consider $N$ the number of exceedances as a primary role in deriving the asymptotic behavior. By notations and condition (A2), we have Taylor expansion about $\boldsymbol{\theta}_0$ as follows.

$$\begin{aligned}
0 &= \left.\frac{\partial Q_T(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} \\
&= \left.\frac{\partial Q_T(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} + (\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)^T\left.\frac{\partial^2 Q_T(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^2}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} + \frac{1}{2}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)^T\left.\frac{\partial^3 Q_T(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^3}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) \\
&\quad \text{where } \boldsymbol{\theta}^* \text{ lies between } \widehat{\boldsymbol{\theta}} \text{ and } \boldsymbol{\theta}_0 \\
&= \sum_{k=1}^{N}\mathrm{D}(\boldsymbol{\theta}_0;(X_{ti}^{(k)},X_{tj}^{(k)})) + (\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)^T\sum_{k=1}^{N}\mathrm{D}'(\boldsymbol{\theta}_0;(X_{ti}^{(k)},X_{tj}^{(k)})) \\
&\qquad + \frac{1}{2}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)^T\sum_{k=1}^{N}\mathrm{D}''(\boldsymbol{\theta}^*;(X_{ti}^{(k)},X_{tj}^{(k)}))(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)
\end{aligned}$$

Then we rewrite the equation as

$$\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\mathrm{D}(\boldsymbol{\theta}_0;\mathbf{X}^{(k)}) \\
&\quad = \left\{-\frac{1}{N}\sum_{k=1}^{N}\mathrm{D}'(\boldsymbol{\theta}_0;\mathbf{X}^{(k)}) - \frac{1}{2}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)^T\frac{1}{N}\sum_{k=1}^{N}\mathrm{D}''(\boldsymbol{\theta}^*;\mathbf{X}^{(k)})\right\}\sqrt{N}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0),
\end{aligned}$$

and then

$$\sqrt{\frac{N}{s_{1N}^2}}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$$
$$= \Bigg\{ \underbrace{-\frac{1}{N}\sum_{k=1}^{N} \mathrm{D}'(\boldsymbol{\theta}_0) - \frac{1}{2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \frac{1}{N}\sum_{k=1}^{N} \mathrm{D}''(\boldsymbol{\theta}^*)}_{\text{(a)}} \Bigg\}^{-1} \underbrace{\frac{1}{\sqrt{N s_{1N}^2}} \sum_{k=1}^{N} \mathrm{D}(\boldsymbol{\theta}_0)}_{\text{(b)}}. \quad (4.13)$$

We establish the following for separate terms in equation (4.13):

(I) By the consistency of $\widehat{\boldsymbol{\theta}}$ and condition (A5), expectation of the last term in parentheses can be ignored. Since $\widehat{\boldsymbol{\theta}}$ is consistent, $\widehat{\boldsymbol{\theta}} \in \Theta^*$ with $P_{\boldsymbol{\theta}_0}$-probability 1. Let $B \subset \Theta^*$ be a closed ball with the center $\boldsymbol{\theta}_0$. By the condition (A4),

$$\sup_{\widehat{\boldsymbol{\theta}} \in B} \left\| \frac{\partial^2 Q_T(\boldsymbol{\theta}_0; x, y)}{\partial\theta_i \partial\theta_j} - \frac{\partial^2 Q_T(\widehat{\boldsymbol{\theta}}; x, y)}{\partial\theta_i \partial\theta_j} \right\|$$

is bounded and then, for large $N$,

$$\limsup_N \left\| \frac{1}{N}\sum_{k=1}^{N} \mathrm{D}'(\boldsymbol{\theta}_0) - \frac{1}{N}\sum_{k=1}^{N} \mathrm{D}'(\widehat{\boldsymbol{\theta}}) \right\| \le \varepsilon$$

in probability (see details in Guyon [1995]). $\frac{1}{N}\sum_{k=1}^{N} \mathrm{D}'(\boldsymbol{\theta}_0) - \mathbb{E}[-\mathrm{D}'(\boldsymbol{\theta}_0)]$ converges to 0 by the law of large numbers, and hence (a) converges to $\mathbb{E}[-\mathrm{D}'(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0)]$ in probability.

(II) First consider that $(X_{ti}^{(k)}, X_{tj}^{(k)})$, $k = 1, \cdots, N$ are i.i.d. from exact multivariate GPD distribution $H$.

$$\mathbb{E}\mathrm{D} := \mathbb{E}\left\{ \frac{1}{\sqrt{N s_{1N}^2}} \sum_{k=1}^{N} \mathrm{D}(\boldsymbol{\theta}_0; \mathbf{X}^{(k)}) \right\}$$
$$= \frac{1}{\sqrt{N s_{1N}^2}} \mathbb{E} \sum_{k=1}^{N} \frac{\partial}{\partial \boldsymbol{\theta}} w_{ij} \log L_{ij}(X_{ti}^{(k)}, X_{tj}^{(k)}; \boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = 0 \text{ (no bias)}.$$

Then by **Theorem 1**, (b) converges in distribution to $N(0, \mathbb{V}(\boldsymbol{\theta}_0))$ where

$$
\begin{aligned}
\mathbb{V}(\boldsymbol{\theta}_0) &= \mathbb{E}[\mathrm{D}(\boldsymbol{\theta}_0)\mathrm{D}(\boldsymbol{\theta}_0)^T] \\
&= \mathbb{E}\big[(\mathrm{D}_1 + \mathrm{D}_2 + \mathrm{D}_3 + \mathrm{D}_4)(\mathrm{D}_1 + \mathrm{D}_2 + \mathrm{D}_3 + \mathrm{D}_4)^T\big] \\
&= \mathrm{Var}[\mathrm{D}_1{\mathrm{D}_1}^T] + \mathrm{Var}[\mathrm{D}_2{\mathrm{D}_2}^T] + \mathrm{Var}[\mathrm{D}_3{\mathrm{D}_3}^T] + \mathrm{Var}[\mathrm{D}_4{\mathrm{D}_4}^T], \\
\mathrm{Var}[\mathrm{D}_1{\mathrm{D}_1}^T] &= \sigma(\boldsymbol{\theta}_0; \mathbf{0}) + C_1 \int \sigma\big(\boldsymbol{\theta}_0; (0, \boldsymbol{h})\big) Q_2(\boldsymbol{h}) d\boldsymbol{h} + C_1^2 \int \sigma(\boldsymbol{\theta}_0; \boldsymbol{h}) Q_1(\boldsymbol{h}) d\boldsymbol{h}
\end{aligned}
$$

and $\mathrm{Var}[\mathrm{D}_2{\mathrm{D}_2}^T]$, $\mathrm{Var}[\mathrm{D}_3{\mathrm{D}_3}^T]$ and $\mathrm{Var}[\mathrm{D}_4{\mathrm{D}_4}^T]$ have similar forms with the variance of $\mathrm{D}_1{\mathrm{D}_1}^T$. Note that the event $\{x_i > u, x_j > u\}$ of $D_1$ is uncorrelated with the event $\{x_i > u, x_j \le u\}$ of $D_2$, and $\mathrm{Cov}(D_i, D_j) = 0$ for $i \ne j$.
Now suppose that $(X_{ti}^{(k)}, X_{tj}^{(k)})$, $k = 1, \cdots, N$ are from $F_{b_k,d_k}$ not $H$. If $F \in D(G)$, there exists the exceedance level $(b_k, d_k)$ such that $F_{b_k,d_k}$ converges to $H$ as $(b_k, d_k) \to (x_0, y_0)$. The bivariate generalized pareto distribution $H$ preserves under the suitable change of exceedance levels (Rootzén and Tajvidi [2006]). The second-order condition (4.5) describes the difference between $F_{b_k,d_k}$ and $H$ with the remainder function $A(k)$, i.e., as $k \to \infty$, with the second order condition (ii)

$$
\limsup_{k \to \infty} |F_{b_k,d_k}(a_k x_i, c_k x_j) - H(x_i, x_j)| = O(A(k)).
$$

**Proposition 1** (ii) results from the condition (A6), and by the property of score function

$$
\begin{aligned}
\mathbb{E}\left\{\frac{1}{\sqrt{N s_{1N}^2}} \sum_{k=1}^{N} \mathrm{D}(\boldsymbol{\theta}_0; \mathbf{X}^{(k)})\right\} &= \frac{1}{\sqrt{N s_{1N}^2}} \int \sum \mathrm{D}(\boldsymbol{\theta}_0; \mathbf{X}^{(k)}) dF_{b_k,d_k}(a_k x_i, c_k x_j) \\
&= \sqrt{N s_{1N}^2} A(k) \cdot \frac{1}{N s_{1N}^2} \int \sum \mathrm{D}(\boldsymbol{\theta}_0; \mathbf{X}^{(k)}) d\Psi(x_i, x_j) + o(A(k)) \to \lambda\mu,
\end{aligned}
$$

where $\mu = \lim_{N \to \infty} \frac{1}{N s_{1N}^2} \int \sum \mathrm{D}(\boldsymbol{\theta}_0; \mathbf{X}^{(k)}) d\Psi(x_i, x_j)$.

Then for some finite vector $\boldsymbol{b}$

$$\left(Ns_{1N}^2\right)^{-1/2}\mathbb{E}\left\{\sum_{k=1}^{N}\mathrm{D}(\boldsymbol{\theta}_0;\mathbf{X}^{(k)})\right\} \to \boldsymbol{b},$$

and (b) converges in distribution to $N(\boldsymbol{b},\mathbb{V}(\boldsymbol{\theta}_0))$. Therefore the limit distribution of $\widehat{\boldsymbol{\theta}}$, (4.12) follows by Slutsky's Theorem. If the second-order condition (i) holds and $\sqrt{Ns_{1N}^2}A(k_N) \to 0$, $\boldsymbol{b} = \mathbf{0}$ which implies no bias and then (4.11) holds.

□

To prove consistency, we describe the theorem of Amemiya [1985].

**Theorem 3.** *(Amemiya [1985]) Assume the following:*

*(B1) $\Theta$ is an open subset of Euclidean p-space (the true value $\theta_0$ is an interior point of $\Theta$),*

*(B2) The criterion function $S_N(\theta)$ is a measurable function for all $\theta_0 \in \Theta$, and $\nabla S_N$ exists and is continuous in an open neighborbood of $\theta_0$,*

*(B3) $\frac{1}{N}S_N(\theta)$ converges in probability uniformly to a non-stochastic function $S(\theta)$ in an open neighborhood of $\theta_0$, and $S(\theta)$ attains a strict local maximum at $\theta_0$.*

*Then there exists a sequence $\epsilon_N \to 0$ such that*

$$P\{\exists\theta^* \text{ such that } |\theta^* - \theta_0| < \epsilon_N, \nabla S_N(\theta^*) = 0\} \to 1, \text{ as } N \to \infty.$$

**Theorem 4.** *(Consistency) Let $\mathbf{X}_t^{(k)} = (X_{ti}^{(k)}, X_{tj}^{(k)})$, $k = 1,\cdots,N$ be i.i.d. random variables with bivariate distribution $F$. Let $\hat{\boldsymbol{\theta}}$ be the maximum pairwise composite log-likelihood estimator such that*

$$\nabla S_N(\hat{\boldsymbol{\theta}}) := \sum_{i<j}^{K}\sum_{t=1}^{T} w_{ij}\frac{\partial}{\partial\boldsymbol{\theta}}\log L(X_{ti},X_{tj};\boldsymbol{\theta})\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 0.$$

*If the second moment condition of composite score function is satisfied and conditions (A1), (B1) and (B2) hold, then as $N \to \infty$ and $(b_k, d_k) = (b(k)_N, d(k)_N) \to (x_0, y_0)$, there exists $\hat{\boldsymbol{\theta}}$ such that $|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0| < \epsilon_N$ and $\nabla S_N(\hat{\boldsymbol{\theta}}) = 0$ for any sequence $\epsilon_N \to 0$.*

*Proof.* Assumptions *(B1)* and *(B2)* in **Theorem 3** are satisfied by our criterion functions and assumptions. Jensen's inequality implies

$$\int \log\left\{\frac{f(x;\theta)}{f(x;\theta_0)}\right\} f(x;\theta_0)dx \le \log \int f(x;\theta)dx = 0. \tag{4.14}$$

We rewrite it as

$$E_{\theta_0}\left[\log\frac{f(x;\theta)}{f(x;\theta_0)}\right] \le 0 \Leftrightarrow \theta_0 = \arg\max_{\theta\in\Theta} E_{\theta_0}\left[\log\frac{f(x;\theta)}{f(x;\theta_0)}\right].$$

Here a sum of pairwise log-likelihoods can be considered. Let

$$S_N(\boldsymbol{\theta}) = \sum_{i<j}^{K}\sum_{t=1}^{T} w_{ij} \log L(X_{ti}, X_{tj}; \boldsymbol{\theta}).$$

We know that by the law of large numbers,

$$\begin{aligned}\frac{1}{N}S_N(\boldsymbol{\theta}_0) &= \frac{1}{N}\sum_{k=1}^{N} w_{ij}^{(k)} \log L(X_{ti}^{(k)}, X_{tj}^{(k)}; \boldsymbol{\theta}_0) = \frac{1}{N}\sum_{k=1}^{N} w^{(k)} \log L(\boldsymbol{X}_t^{(k)}; \boldsymbol{\theta}_0)\\ &\longrightarrow E_{\boldsymbol{\theta}_0}\big(w^{(1)} \log L(\boldsymbol{X}_t^{(1)}; \boldsymbol{\theta}_0)\big) =: S(\boldsymbol{\theta}_0).\end{aligned}$$

By the moment condition of $\nabla S_N(\boldsymbol{\theta})$, we have that $E|\nabla S_N(\boldsymbol{\theta}^*)|^2 < C_0$ for some $C_0$. Using a Taylor's expansion,

$$\begin{aligned}\left\|\Big(\frac{1}{N}S_N(\boldsymbol{\theta}) - S(\boldsymbol{\theta})\Big) - \Big(\frac{1}{N}S_N(\boldsymbol{\theta}_0) - S(\boldsymbol{\theta}_0)\Big)\right\|^2 &= \left\|\Big(\frac{1}{N}\nabla S_N(\boldsymbol{\theta}^*) - \nabla S_N(\boldsymbol{\theta}^{**})\Big)(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\right\|^2\\ &\le \left(\frac{1}{N}E|\nabla S_N(\boldsymbol{\theta}^*)|^2 + E|\nabla S_N(\boldsymbol{\theta}^{**})|^2\right)||\boldsymbol{\theta} - \boldsymbol{\theta}_0||^2\\ &\le \left(\frac{C_0}{N} + C_0\right)||\boldsymbol{\theta} - \boldsymbol{\theta}_0||^2\\ &\longrightarrow C_0||\boldsymbol{\theta} - \boldsymbol{\theta}_0||^2\end{aligned} \tag{4.15}$$

for some $\boldsymbol{\theta}^*$ and $\boldsymbol{\theta}^{**}$ between $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}$. By the moment condition of $\nabla S_N(\boldsymbol{\theta})$, the right hand side of (4.15) converges to 0 uniformly over a sequence of $||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| < \epsilon_N$ as $\epsilon_N \to 0$. Also we have that $\frac{1}{N} S_N(\boldsymbol{\theta}_0) - S(\boldsymbol{\theta}_0) \xrightarrow{p} 0$ by the law of large numbers and $\frac{1}{N} S_N(\boldsymbol{\theta})$ converges in probability uniformly to $S(\boldsymbol{\theta})$ on a neighborhood of $\boldsymbol{\theta}_0$.

Now we claim that $S(\boldsymbol{\theta})$ attains a local maximum at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. The previous result (4.14) implies that

$$E_{\boldsymbol{\theta}_0}\left[\log \frac{\prod_k L(\boldsymbol{X}_t^{(k)}; \boldsymbol{\theta})}{\prod_k L(\boldsymbol{X}_t^{(k)}; \boldsymbol{\theta}_0)}\right] \leq \log E_{\boldsymbol{\theta}_0}\left[\frac{\prod_k L(\boldsymbol{X}_t^{(k)}; \boldsymbol{\theta})}{\prod_k L(\boldsymbol{X}_t^{(k)}; \boldsymbol{\theta}_0)}\right] = 0$$

and for any $\boldsymbol{\theta}$,

$$E_{\boldsymbol{\theta}_0}\big(\log \prod_k L(\boldsymbol{X}_t^{(k)}; \boldsymbol{\theta}_0)\big) \geq E_{\boldsymbol{\theta}_0}\big(\log \prod_k L(\boldsymbol{X}_t^{(k)}; \boldsymbol{\theta})\big).$$

where the equality holds with (A1), the identifiability assumption of parameter. $E_{\boldsymbol{\theta}_0}\big[\log \frac{\prod_k L(\boldsymbol{X}_t^{(k)}; \boldsymbol{\theta})}{\prod_k L(\boldsymbol{X}_t^{(k)}; \boldsymbol{\theta}_0)}\big] \leq 0$ holds for any distribution of $\boldsymbol{X}_t^{(k)}$ with finite second moments of score function, and the maximum of $E_{\boldsymbol{\theta}_0}[\log \prod_k L(\boldsymbol{X}_t^{(k)}; \boldsymbol{\theta})]$ over $\boldsymbol{\theta}$ is attained at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Thus we prove the *(B3)* of the **Theorem 3**. □

## 4.3 Simulation Study

We conduct some simulation studies to illustrate the asymptotic behavior of the estimators described in Section 4.2. The simulation is examined for the daily max-stable process with unit Fréchet margins with $T = 1000$ days during 10 years (i.e., $M = 100$ in equation [3.5]). We consider the Gaussian extreme value processes with two different spatial dependence structures of the covariance matrix:

$$\Sigma = \begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix}$$

(i) the Gaussian extreme value process with $\Sigma_1$ ($\alpha = 2$, $\beta = 0$ and $\gamma = 3$);

(ii) the Gaussian extreme value process with $\Sigma_2$ ($\alpha = 2$, $\beta = 1.5$ and $\gamma = 3$).

We generate $n = 20$ stations from the uniform density function $f(\cdot)$ over $R_0 = (-1/2, 1/2]^d$ and determine the growth rate $\lambda_n = \sqrt{n}$ in case of $d = 2$ to satisfy the relation $n \sim C\lambda_n^d$ in the spatial structure and stochastic sampling design of sites. To adjust the threshold approach based on the pairwise composite likelihood, we consider a weight function such that for some constant $\delta_0$,

$$w(h) = \begin{cases} 1 & \text{if } h \leq \delta_0 \\ 0 & \text{if } h > \delta_0. \end{cases}$$

where $h$ is a distance between two stations. Here $\delta_0$ is selected by $\sqrt{2n}/2$, the half diagonal of sampling region, which satisfies the condition (A′6) on growth rate of weight function for the asymptotic result.

To illustrate the asymptotic performance of estimates for dependence parameter $\boldsymbol{\theta} = (\alpha, \beta, \gamma)$, the averages of the estimators are compared to the asymptotic mean of $\hat{\boldsymbol{\theta}}$. In each model, the estimation of dependence parameters is based on 500 replications, and the classical Monte Carlo integration is used to implement the theoretical bias and variance of the estimators as the number of exceedances $N$ increases.

Theoretical bias and average bias of estimators $\hat{\boldsymbol{\theta}}$ for Smith model (i) are plotted in Figure 4.1. As the number of exceedances increases, bias of estimators (gray curve) tends to decrease towards the theoretical bias (solid curve) though each estimator shows the different slope on the decay. The bias of $\hat{\alpha}$ goes on with the pattern of decay of theoretical one, while bias of $\hat{\beta}$ and $\hat{\gamma}$ decreases as theoretical bias goes up to the line of zero bias.

This irregular pattern of each dependence parameter estimation might be caused by the interaction between parameters in estimating them as components of covariance matrix. Now we plot the extremal coefficient curves with the parameter estimators and compare them with those estimated directly. One can expect the problem

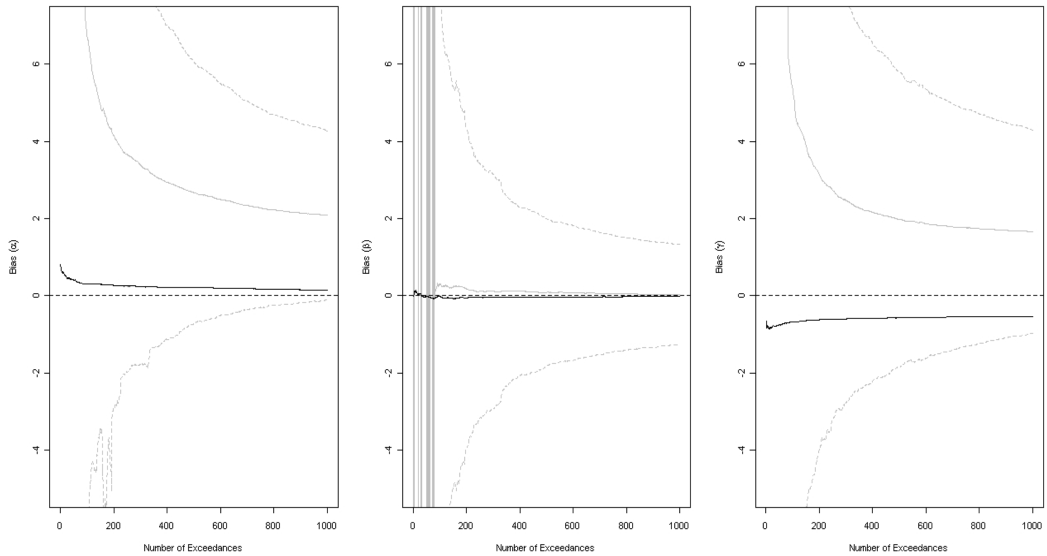



Figure 4.1: Graphical summary of asymptotic behaviors of $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\gamma}$ for Smith model (i) from left to right. Gray curve is the average bias of estimators, gray dashed curves are the boundary of 95% confidence interval, and black solid curve is the theoretical bias.

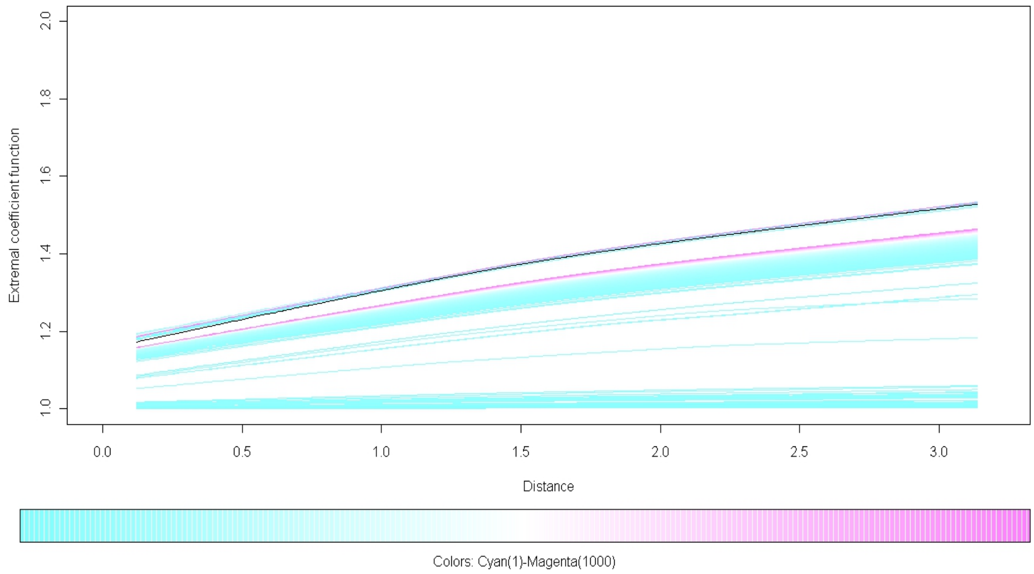



Figure 4.2: Extremal coefficient functions for the Smith model (i). Upper thin color layer is based on theoretical mean of estimates and lower thick color layer is based on average estimates. In a layer, each line represent a extremal coefficient curve at each $N$ and the line changes the color from cyan ($N$ = 1) to magenta ($N$ = 1000). Black solid line is the true extremal coefficient curve.

to be reduced when working with the extremal coefficient.

Figure 4.2 shows estimated extremal coefficient functions by $\hat{\boldsymbol{\theta}}$. As the number of exceedances increases, the color changes from cyan to magenta. Extremal coefficient by the asymptotic bias overlapped almost with the true coefficient function (black solid curve). As the number of exceedances increases, The extremal coefficient curve measured by dependence estimators approximates the theoretical extremal coefficient curve. However, there still exists a gap between the theoretical extremal coefficient and estimated one and the gap gets broader as the distance between two locations is larger.

Theoretical bias and average bias of estimators $\hat{\boldsymbol{\theta}}$ for Smith model (ii) are shown in Figure 4.3. As the number of exceedances increases, bias of estimates tends to go towards the pattern of theoretical bias. There is some gaps between theoretical bias and estimated bias though the estimation of dependence parameter is much more stable comparing with that in model (i).

Figure 4.4 shows estimated extremal coefficient functions by $\hat{\boldsymbol{\theta}}$. As the number of exceedances increases, The extremal coefficient curve measured by dependence estimators approximates the theoretical extremal coefficient curve. Unlike the gap in Figure 4.3, the estimated extremal coefficient is catching up with the theoretical one along by a little gap. However, the quality of asymptotic approximation seems dependent on the degree of correlation $\beta$ since Figure 4.4 shows the poor approximation to the true extremal coefficient curve comparing with Figure 4.2.

Suggestion on the choice of the threshold point is discussed further now. For the simplicity, the threshold can be selected as the value of the 95th percentile of distribution function in practice. However finding an optimal threshold is another important issue and we suggest an optimal threshold minimizing the mean squared error, which incorporates both the bias of the estimator and its variance based on the asymptotic normality in Section 4.2.

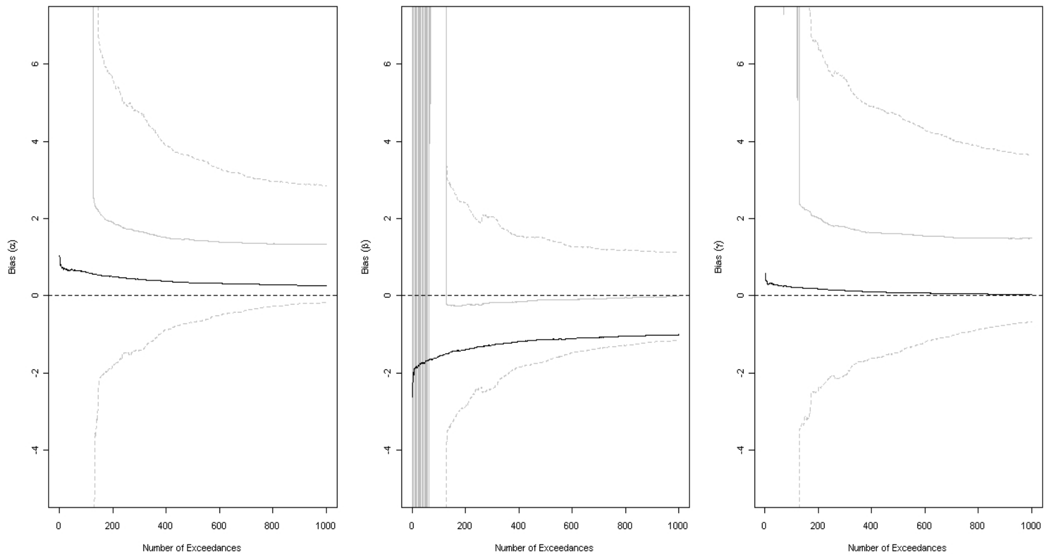



Figure 4.3: Graphical summary of asymptotic behaviors of $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\gamma}$ for Smith model (ii) from left to right. Gray curve is the average bias of estimators, gray dashed curves are the boundary of 95% confidence interval, and black solid curve is the theoretical bias.

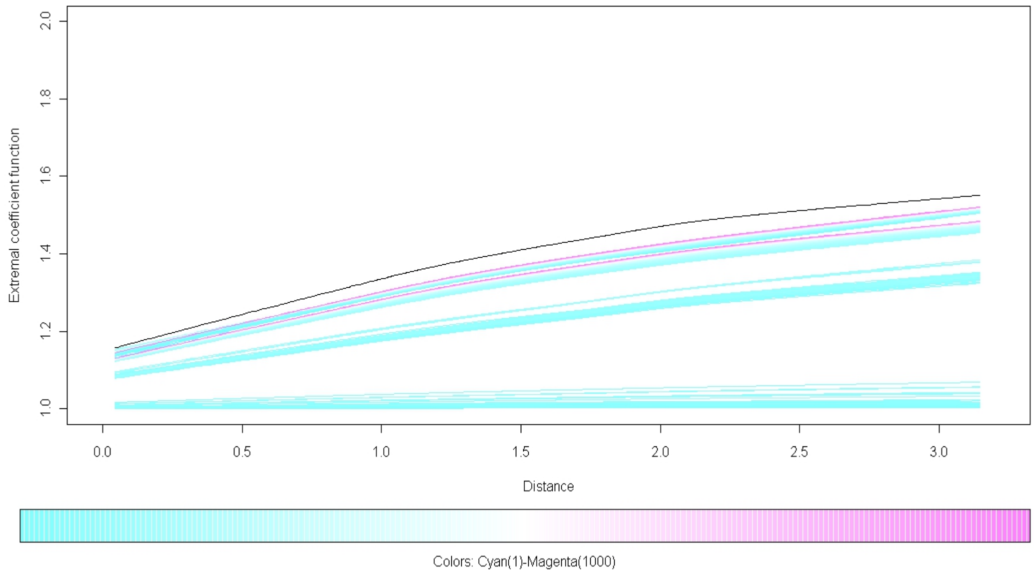



Figure 4.4: Extremal coefficient functions for the Smith model (ii). Upper thin color layer is based on theoretical mean of estimates and lower thick color layer is based on average estimates. In a layer, each line represent a extremal coefficient curve at each $N$ and the line changes the color from cyan ($N$ = 1) to magenta ($N$ = 1000). Black solid line is the true extremal coefficient curve.

Figure 4.5 shows the mean squared error for each estimator in Smith model (i). The mean squared errors of $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$ are decreasing rapidly against $N$ and show the stability between $N = 3500$ and $N = 5000$. As shown in Figure 4.2, the theoretical extremal coefficient has a nice approximation to the true coefficient function, and the increases of squared bias seem to be less effective than variance decreases on the selection of threshold to minimize the MSE.

The mean squared error for each estimator of Smith model (ii) is shown in Figure 4.6. The mean squared errors of $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$ are decreasing rapidly as $N$ increases to 1000, and have the minimum between $N = 1500$ and $N = 2000$. In Figure 4.4, the theoretical extremal coefficient shows the poor approximation to the true coefficient function. Thus calculation of MSE is affected by the increase of bias as number of exceedances becomes greater than 1500. The threshold point is suggested as the value between 90th and 95th percentile.

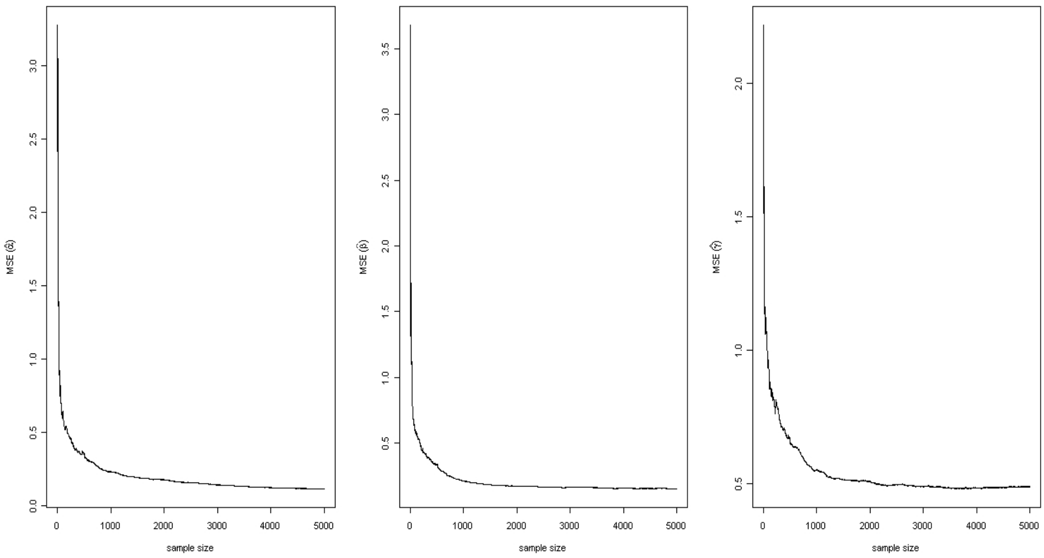



Figure 4.5: Mean squared error of $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\gamma}$ for Smith (i) from left to right

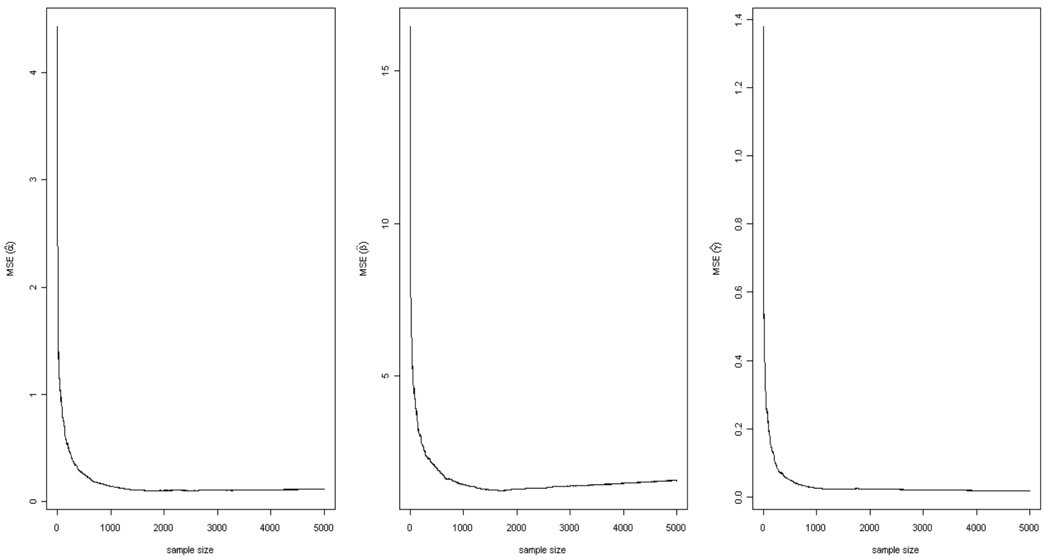



Figure 4.6: Mean squared error of $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\gamma}$ for Smith (ii) from left to right

# Chapter 5

# Application

The methods are illustrated by application to temperature data of North Carolina, U.S. and the modeling procedures are examined via simulation. The different max-stable models investigated are the following.

(i) Smith model with covariance matrix:

$$\Sigma = \begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix}$$

(ii) Brown-Resnick process with power law variogram:

$$\gamma(h) = \begin{cases} 0 & \text{if } h = 0 \\ a + bh^{\lambda} & \text{if } h > 0, a > 0, b > 0, 0 \leq \lambda < 2 \end{cases}$$

(iii) Schlather model with powered exponential correlation function:

$$\rho(h) = c \exp\Big\{ - \Big(\frac{h}{R}\Big)^{\kappa} \Big\}, 0 \leq c \leq 1, R > 0, 0 < \kappa \leq 2$$

(iv) Schlather model with Matérn correlation function:

$$\rho(h) = c \frac{2^{1-\nu}}{\Gamma(\nu)} \Big(\frac{h}{R}\Big)^{\nu} K_{\nu}\Big(\frac{h}{R}\Big), 0 \leq c \leq 1, R > 0, \nu > 0$$

We fit max-stable processes to annual maxima of temperature in Section 5.1 and focus on the threshold approach for daily temperature using max-stable processes in Section 5.2.

## 5.1 Max-stable Processes for Annual Maxima of Temperature

We apply max-stable processes in an analysis of North Carolina temperature data. Annual maxima of temperature, measured in Fahrenheit, in North Carolina, U.S. are observed over a period of 52 years, from 1957 to 2008, at 25 stations (see Figure 5.1). Plots of annual maxima given the time period at each station are shown in Figure 5.2. We assume that spatial processes of annual maxima of temperature are stationary from the figure.

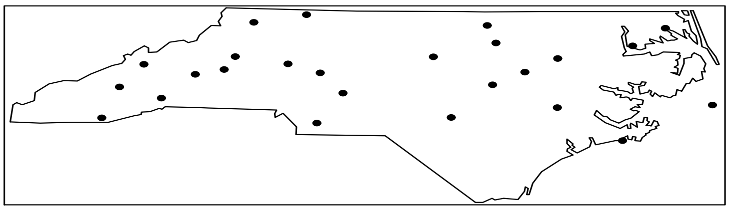


Figure 5.1: Map of 25 stations in North Carolina, United States

The GEV parameters for each marginal distribution at a site can be estimated by MLE and we can transform the marginal distribution into unit Fréchet form. Then we characterize the dependence structure with max-stable processes. Table 5.1 summarizes the results for the different models. According to the maximum composite log-likelihood (MCL) criterion, the Brown-Resnick process shows the best performance, and the Smith model gives relatively poor performance comparing with other two Schlather's models. Both Schlather models are also competitive and give similar values of the MCL though we use two different correlation functions.

The pairwise log likelihood (2.12) by Padoan, Ribatet and Sisson [2010] allows

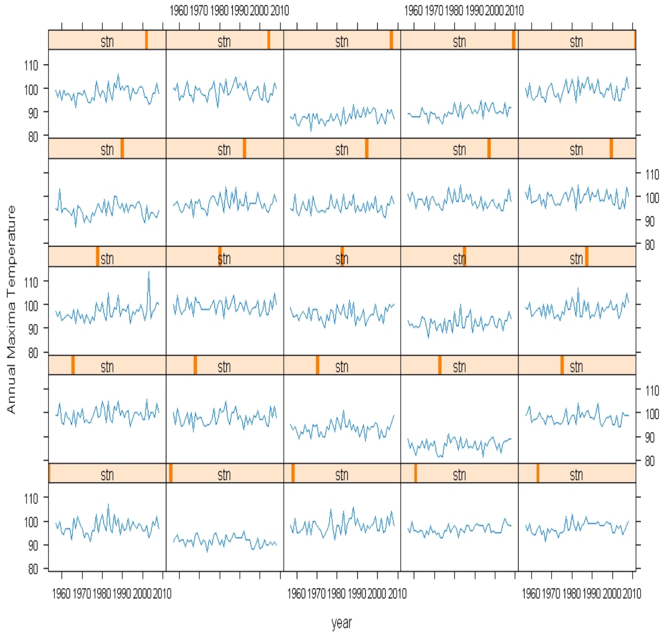



Figure 5.2: Annual maximum temperature from 1957 to 2008 at 25 stations

simultaneous estimation of the spatial dependence parameters between pairs of sites and also the three parameters of GEV marginal distribution at each location. We now examine the influence of permitting the separate use of GEV marginal parameters and dependence parameters via simulation studies. For the simulation, the true process $Z$ is generated from the max-stable process with unit Fréchet margins and $Y$ over

Table 5.1: Fitting max-stable processes and their corresponding maximized composite log-likelihood

| Model | Parameter | | | MCL |
|---|---|---|---|---|
| **Smith** | $\hat{\alpha} \approx 1.940$, | $\hat{\beta} \approx -0.032$, | $\hat{\gamma} \approx 0.097$ | -66137.36 |
| **Brown-Resnick** (power law) | $\hat{a} \approx 1.794$, | $\hat{b} \approx 0.078$, | $\hat{\lambda} \approx 1.949$ | -64400.76 |
| **Schlather** (powered exp.) | $\hat{c} \approx 0.598$, | $\hat{R} \approx 4.270$, | $\hat{\kappa} \approx 1.560$ | -64499.46 |
| **Schlather** (Matérn) | $\hat{c} \approx 0.601$, | $\hat{R} \approx 2.181$, | $\hat{\nu} \approx 1.296$ | -64500.29 |

GEV margins is obtained using GEV parameters for each station.

$$Y = \mu_s + \psi_s \frac{Z^{\xi_s} - 1}{\xi_s}, \quad s = 1, 2, \cdots, S$$

and we can transform $Y$ into $\widehat{Z}$, which is an estimate of $Z$, by fitting GEV parameters to $Y$. We can fit max-stable processes to the original $Z$ and get a simulation result. A simulation is also obtained for $\widehat{Z}$ to find MCLE of max-stable processes. We would check the approximation for the true process, by comparing the result of $Z$ and that of $\widehat{Z}$.

Table 5.2 summarizes the performances of max-stable processes for $Z$, $N = 100$ at 9 stations which is generated from the Smith model with covariance function ($\sigma_{11} = 1.5$, $\sigma_{12} = -0.05$, $\sigma_{22} = 0.1$). Table 5.3 shows the simulation result for $\widehat{Z}$ at 9 stations from the Smith model with the same covariance function. The Smith model fits well in both simulations because the true model is considered as Smith's and the Brown-Resnick process is relatively competitive. Comparing with the simulation for $Z$ and $\widehat{Z}$, the mean squared errors of parameters in the Smith model for original observation $Z$ is always smaller than those for $\widehat{Z}$. Thus the estimators for $\widehat{Z}$ are more variable than those of fitted max-stable processes with the true process.

Table 5.4 summarizes the performances of max-stable processes for $Z$, $N = 100$ at 9 stations which is generated from the Schlather model with powered exponential correlation function (sill= 0.5985, range= 1.4309, smooth= 1.5610). Table 5.5 shows the simulation result for $\widehat{Z}$ at 9 stations from the Schlather model with the same correlation function. The estimators of range parameter were transformed to the value by logarithm due to their large variation. The Schlather model fits well in both simulations of $Z$ and $\widehat{Z}$ because $Z$ is generated from Schlather's. The Brown-Resnick process is also relatively competitive. Comparing both simulations, the mean squared errors of parameters in Schlather model for original observation $Z$ is always smaller than those for $\widehat{Z}$ with the interpretation that the estimators for $\widehat{Z}$ have larger variation than those of the true process.

Table 5.2: Composite MLE's based on $N = 100$ simulations for $Z$ generated from Smith model. Standard deviations (in parentheses) are obtained and mean squared error (MSE) are reported in bold and parentheses.

| Model | | Parameter (s.d., **MSE**) | | MCL |
|---|---|---|---|---|
| True | $\sigma_{11} = 1.5$, | $\sigma_{12} = -0.05$, | $\sigma_{22} = 0.1$ | |
| Smith | $\hat{\alpha} \approx 1.6870$, | $\hat{\beta} \approx -0.0579$, | $\hat{\gamma} \approx 0.1032$ | -7457.72 |
| | (0.3313) | (0.0570) | (0.0195) | |
| | **(0.1447)** | **(0.0033)** | **(0.0004)** | |
| Brown-Resnick | $\hat{a} \approx 0.0635$, | $\hat{b} \approx 2.5526$, | $\hat{\lambda} \approx 1.4183$ | -7591.74 |
| | (0.0943) | (0.4204) | (0.2135) | |
| Schlather | $\hat{c} \approx 0.9218$, | $\hat{R} \approx 0.9838$, | $\hat{\kappa} \approx 1.8789$ | -7606.01 |
| | (0.0378) | (0.1304) | (0.1696) | |

Table 5.3: Composite MLE's based on $N = 100$ simulations for $\widehat{Z}$ generated from Smith model. Standard deviations (in parentheses) are obtained and mean squared error (MSE) are reported in bold and parentheses.

| Model | | Parameter (s.d., **MSE**) | | MCL |
|---|---|---|---|---|
| True | $\sigma_{11} = 1.5$, | $\sigma_{12} = -0.05$, | $\sigma_{22} = 0.1$ | |
| Smith | $\hat{\alpha} \approx 1.6512$, | $\hat{\beta} \approx -0.0582$, | $\hat{\gamma} \approx 0.0994$ | -7547.82 |
| | (0.5918) | (0.0687) | (0.0274) | |
| | **(0.3731)** | **(0.0047)** | **(0.0007)** | |
| Brown-Resnick | $\hat{a} \approx 0.0753$, | $\hat{b} \approx 2.7170$, | $\hat{\lambda} \approx 1.4082$ | -7679.79 |
| | (0.1189) | (0.6257) | (0.2127) | |
| Schlather | $\hat{c} \approx 0.9208$, | $\hat{R} \approx 0.9611$, | $\hat{\kappa} \approx 1.8812$ | -7689.91 |
| | (0.0425) | (0.1617) | (0.1612) | |

Table 5.4: Composite MLE's based on $N = 100$ simulations for $Z$ generated from Schlather model. Standard deviations (in parentheses) are obtained and mean squared error (MSE) are reported in bold and parentheses.

| Model | | Parameter (s.d., **MSE**) | | MCL |
|---|---|---|---|---|
| True | sill= 0.5985, | log(range)= 1.4309, | smooth= 1.5610 | |
| Schlather | $\hat{c} \approx 0.6802$, | $\log(\hat{R}) \approx 1.8081$, | $\hat{\kappa} \approx 1.3290$ | -7546.48 |
| | (0.1691) | (1.1124) | (0.7538) | |
| | **(0.0353)** | **(1.3796)** | **(0.6221)** | |
| Brown-Resnick | $\hat{a} \approx 1.3995$, | $\hat{b} \approx 0.9445$, | $\hat{\lambda} \approx 1.0555$ | -7658.43 |
| | (0.8848) | (0.8411) | (0.8624) | |
| Smith | $\hat{\alpha} \approx 0.3301$, | $\hat{\beta} \approx 0.0323$, | $\hat{\gamma} \approx 0.0877$ | -7822.61 |
| | (0.1205) | (0.0582) | (0.0317) | |

Table 5.5: Composite MLE's based on $N = 100$ simulations for $\widehat{Z}$ generated from Schlather model. Standard deviations (in parentheses) are obtained and mean squared error (MSE) are reported in bold and parentheses.

| Model | | Parameter (s.d., **MSE**) | | MCL |
|---|---|---|---|---|
| True | sill= 0.5985, | log(range)= 1.4309, | smooth= 1.5610 | |
| Schlather | $\hat{c} \approx 0.6883$, | $\log(\hat{R}) \approx 1.7249$, | $\hat{\kappa} \approx 1.1647$ | -7636.70 |
| | (0.2000) | (1.3785) | (0.7875) | |
| | **(0.0481)** | **(1.9867)** | **(0.7771)** | |
| Brown-Resnick | $\hat{a} \approx 1.5606$, | $\hat{b} \approx 0.9070$, | $\hat{\lambda} \approx 1.1820$ | -7720.98 |
| | (0.9470) | (0.9719) | (0.8461) | |
| Smith | $\hat{\alpha} \approx 0.3164$, | $\hat{\beta} \approx 0.0251$, | $\hat{\gamma} \approx 0.0868$ | -7884.35 |
| | (0.1265) | (0.0624) | (0.0381) | |

## 5.2 Threshold Approach for Daily Temperature

An application of the threshold approach was conducted with the previous North Carolina temperature data. The maximum daily temperatures are recorded at 16 stations through the same period, 1957-2008. The seasonal effect is avoided by restricting the data to the summer season, which consists of June, July and August. Therefore $M$

in the equation (3.5) of the relation between daily data and annual maxima is 92, the number of days per year in the summer season.

The choice of threshold is made to get approximately the same proportion, for example 0.05, of exceedances over the threshold in each margin. Thus $u = 0.62738$ is selected as a fixed threshold through the stations, which is the transformed value into unit Fréchet. The 16 stations are shown in Figure 5.3.

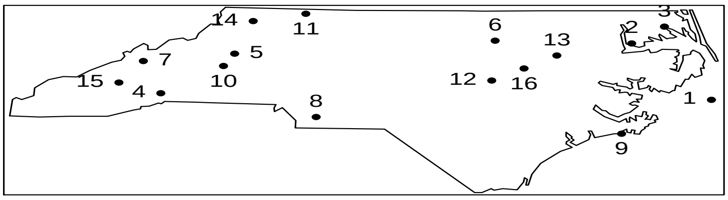



Figure 5.3: Map of 16 stations for the threshold approach

### 5.2.1 Modeling and Parameter Estimation

The previous models investigated for the annual maxima temperature were used, but we did not consider Schlather's model with Matérn correlation function because it plays a similar role with the powered exponential correlation. Table 5.6 shows the results of fitting the three different max-stable models. The Brown-Resnick process gives the best performance, according to the MCL criterion. Note is that the Smith model for daily temperature fits better than the Schlather's unlike the previous result for annual maxima.

### 5.2.2 Spatial Dependence of Thresholded Exceedances

To ensure proper fitting of the generalized Pareto distribution as the asymptotic distribution of exceedances over the threshold, a summary of the marginal analysis is shown in Table 5.7. Approximately the same proportion is kept with 0.05 and the parameter $\sigma_j$ in each margin refers the scale parameter of generalized Pareto

Table 5.6: Fitting max-stable processes for threshold exceedances of daily temperature data and their corresponding maximized composite log-likelihood, $u = .62738$

| Model | Parameter | | | MCL |
|---|---|---|---|---|
| **Smith** | $\hat{\alpha} \approx 4.120$, | $\hat{\beta} \approx -0.066$, | $\hat{\gamma} \approx 0.165$ | -363081.4 |
| **Brown-Resnick** (power law) | $\hat{a} \approx 0.789$, | $\hat{b} \approx 1.483$, | $\hat{\lambda} \approx 0.603$ | -355808.8 |
| **Schlather** (powered exp.) | | $\hat{R} \approx 0.651$, | $\hat{\kappa} \approx 0.777$ | -367309.4 |

Table 5.7: Summary of the marginal analysis with $u = .62738$. Region where each station belongs to is identified with a letter among M(mountain), P(piedmont) and C(costal).

| site $j$ | region | number of exceedances | proportion of exceedances | $\sigma_j$ ($\xi_j = -0.05, fixed$) |
|---|---|---|---|---|
| 1 Cape Hatteras | C | 241 | 0.050 | 1.178 |
| 2 Edenton | C | 224 | 0.047 | 1.425 |
| 3 Elizabeth City | C | 331 | 0.069 | 1.648 |
| 4 Hendersonville | M | 367 | 0.077 | 1.740 |
| 5 Lenoir | M | 436 | 0.091 | 2.054 |
| 6 Louisburg | P | 208 | 0.043 | 2.340 |
| 7 Marshall | M | 405 | 0.085 | 1.731 |
| 8 Monroe | P | 347 | 0.073 | 2.502 |
| 9 Morehead City | C | 206 | 0.043 | 2.373 |
| 10 Morganton | M | 372 | 0.078 | 2.138 |
| 11 Mount Airy | M | 302 | 0.063 | 2.309 |
| 12 Smithfield | C | 311 | 0.065 | 1.647 |
| 13 Tarboro | C | 261 | 0.055 | 2.489 |
| 14 Transou | M | 313 | 0.065 | 1.690 |
| 15 Waynesville | M | 430 | 0.089 | 1.454 |
| 16 Wilson | C | 260 | 0.054 | 2.360 |

distribution. The shape parameters $\xi_j$ were fixed as $-0.05$, which is the maximum likelihood estimator of GPD.

As shown in Figure 5.4, North Carolina is divided into three major geographic regions: the West or Mountains formed mostly by the Blue Ridge and Great Smoky Mountains, the Middle or Piedmont Plateau, and the Eastern also known as the Coastal Plain. The result of Table 5.7 shows different scales which vary at different sites and several locations with the scale parameter greater than 2 are mostly in the mountains and coastal region. Those stations are shown in Figure 5.5.


Figure 5.4: Three North Carolina regions: Mountains, Piedmont and Coastal Plain, (http://thomaslegion.net/threenorthcarolinageographicregionscoastalplain thepiedmontandthemountainsmaps.html)

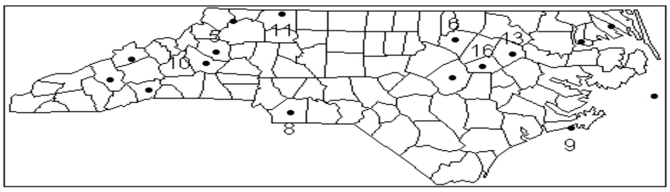



Figure 5.5: Numbered stations having the scale parameter such that $\sigma_j > 2$

Now we consider the dependence with the proportion of common exceedances between two sites. In terms of the common exceedances of two independent sites, the proportion would be .0025, the product of each marginal proportion (.05) of exceedances. Thus we expect that the proportion of common exceedances is closer to .0025 where two sites further apart are considered less dependent than sites closer

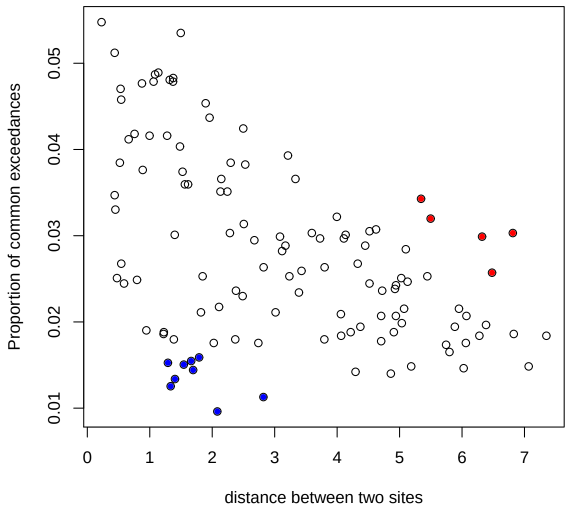



Figure 5.6: Dependence between two sites through the proportion of common exceedances

together. Figure 5.6 plots the proportion of common exceedances according to the distance between two sites. All pairs are above .0025 and we can conclude that the bivariate distributions are dependent. Therefore, it might be suitable to fit a bivariate distribution with spatial dependence structure and the max-stable processes we considered would be a possible statistical model. The proportion plot satisfies the overall pattern that the proportion is also decreasing as the distance of pairs is increasing. However, some points did not match with the pattern and the pairs of site according to the points are identified in Table 5.8 and Table 5.9.

The set of sites in Table 5.8 indicates the lower left dots in Figure 5.6 showing independence even though the pair has a shorter distance. Most points are correlated with station 9, which is Morehead City, centrally located on North Carolina's coast. The city is independent with other sites in the coastal plain region (stations 2, 3, 6, 12, 13 and 16, see Figure 5.3). Station 1 in Cape Hatteras appears independent with stations 2 and 6. We also conclude that stations located on the coast and stations in the coastal plain are independent because the area directly on the coast may be strongly influenced by its proximity to the ocean.

The upper right points violating the general pattern in Figure 5.6 are identified in Table 5.9. The points show the dependence but the long distance between two sites and station 3 is mostly related with this set of sites. Station 3 is Elizabeth City where its location is at the narrowing of the Pasquotank River and on the Intracoastal Waterway. It shows dependence with other sites in the mountain region, station 4, 5, 7, 10 and 15 (see Figure 5.3). Thus there might exist some dependence between several sites in the mountain area and site 3 due to geographical reasons.

Table 5.8: Pairs of site with short distance and independence

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| site1 | 1 | 1 | 6 | 9 | 9 | 9 | 9 | 9 | 9 |
| site2 | 6 | 2 | 2 | 2 | 3 | 6 | 12 | 13 | 16 |

Table 5.9: Pairs of site with long distance and dependence

| | | | | | |
|---|---|---|---|---|---|
| site1 | 3 | 3 | 3 | 3 | 3 |
| site2 | 4 | 5 | 7 | 10 | 15 |

# Chapter 6

# Discussion

## 6.1 Conclusion

Our work proposed the threshold version of max-stable process estimation and we have applied the pairwise composite likelihood method on it. More specifically, we have suggested the modeling of the bivariate exceedances over threshold and it leads to a simplified dependence structure for max-stable processes. An important motivation of this methodology is the possibility of threshold approach to construct approximation of the joint distribution, by assuming an asymptotic distribution of exceedances over a given threshold. We have derived our simulation results under two Smith models to examine the asymptotic property of estimates.

The threshold approach takes advantage of avoiding the loss of information which is caused when we are concerned with only maxima of data. Our method is expected to become one promising tool to characterize the dependence structure in spatial extremes. The proposed approach was successfully applied to the analysis of temperatures in North Carolina and the results give us quite meaningful interpretations of temperature data under North Carolina's geographical conditions.

Moreover, we have also investigated an optimal threshold to minimize the mean squared error based on the asymptotic behavior of the estimator for dependence parameter. The choice of optimal threshold would be an open topic itself for further research. It provides very valuable information in the field of environmental statistics.

When we are interested in flooding, for example, which may be considered as extreme events, choosing the adequate threshold to avoid the risk of flooding might be useful for quantifying the spatial extremal dependence.

## 6.2 Future research

### The issue of "joint" versus "marginal" estimation

Shi, Smith and Coles [1992] presented the issue of "joint" versus "marginal" estimation for bivariate extremes. Bivariate joint distributions are separated from two parts of parameter estimation; one in the marginal distributions and the other defining the dependence between variables transformed into unit Fréchet margins. *Joint estimation* is the approach maximizing the full likelihood with respect to all parameters in a single optimization. Instead of joint estimation, we have used an alternative approach which estimates the GEV parameters first and the dependence parameters separately with fixed GEV parameters (called the *marginal estimation*). Considering the margin and dependence parameters seperately, we would like to provide variances of parameter estimates through the Fisher information matrix by pairwise composite likelihood as an alternative to full likelihood.

Shi, Smith and Coles [1992] set up a separate notation for the marginal estimation. Let $\tilde{\theta}$ denote the column vector of parameters $(\theta, \alpha)$ separating from GEV parameters and dependence parameters, where $\theta^T = (\mu_i, \psi_i, \xi_i, \mu_j, \psi_j, \xi_j)$ at two sites $i$ and $j$, and $\alpha$ consists of the dependence parameters. Let $l_n(\tilde{\theta}) = l_n(\theta, \alpha)$ denote the log likelihood based on $n$ observations, and let $l_n^*(\theta) = l_n(\theta, \alpha_0)$ denote the log likelihood for $\theta$ assuming $Y_i$ and $Y_j$ are independent. The marginal estimator $\hat{\theta}$ is the value of $\theta$ that maximizes $l_n^*(\theta)$ and $\hat{\alpha}$ maximizes $l_n(\hat{\theta}, \alpha)$ with respect to $\alpha$ for the fixed $\hat{\theta}$. Then $\text{Cov}(\hat{\theta})$, $\text{Cov}(\hat{\theta}, \hat{\alpha})$ and $\text{Cov}(\hat{\alpha})$ are derived from the Fisher information matrix based on each $l_n^*$ and $l_n$. Combining those covariance matrices gives the full asymptotic covariance matrix of $(\hat{\theta}, \hat{\alpha})$.

This marginal approach could result in a significant difference with the properties of the estimation, see Shi, Smith and Coles [1992] and Shi [1995]. The asymptotic covariance matrix of the estimators under the marginal method can be compared with the corresponding matrix for the joint method, and simulations supported the result that the marginal method may be inefficient even though it is simpler computationally.

The same issue of "joint" versus "marginal" estimation arises for the threshold approach. We would compare two methods via simulations with calculating the information matrix numerically. Like the simulation of comparison with $Z$ and $\widehat{Z}$ in Section 3.1, we could also examine the comparison of parameter estimation for true process over unit Fréchet margins ($Z$), and that of process ($\widehat{Z}$), transformed from GEV margins, permitting the marginal estimation under the different setting of max-stable processes. Also we consider how to improve the parameter estimation of max-stable processes for $\widehat{Z}$ without any information about true process $Z$, as suggested in the simulation study of $Z$ and $\widehat{Z}$.

**Dependence of more than two sites**

We would like to explore the dependence structure of more than two sites, like triple stations. The optimal way to see the dependence between multiple sites will also be considered. The dependence from max-stable processes such as Smith and Schlather model can be examined via simulation, and we will conclude which model better explains the dependence between sites. If there exists some discrepancy with the dependence plot in practice, we could draw statistically meaningful results from the discrepancy.

# Appendix

## Appendix A. Example

$$B(x,y;\theta) = \left\{ - e^{-x}\Phi\left(\frac{\sqrt{\gamma}}{2} + \frac{y-x}{\sqrt{\gamma}}\right) - e^{-y}\Phi\left(\frac{\sqrt{\gamma}}{2} + \frac{x-y}{\sqrt{\gamma}}\right)\right\} :$$

$$\begin{aligned}
\frac{\partial B}{\partial \theta} &= \left(\frac{\partial \gamma}{\partial \theta}\right)\left\{ - e^{-x}\phi\left(\frac{\sqrt{\gamma}}{2} + \frac{y-x}{\sqrt{\gamma}}\right)\left(\frac{1}{4\sqrt{\gamma}} - \frac{y-x}{2\sqrt{\gamma^3}}\right)\right. \\
&\quad \left. - e^{-y}\phi\left(\frac{\sqrt{\gamma}}{2} + \frac{x-y}{\sqrt{\gamma}}\right)\left(\frac{1}{4\sqrt{\gamma}} - \frac{x-y}{2\sqrt{\gamma^3}}\right)\right\}, \\
\frac{\partial B}{\partial x} &= e^{-x}\Phi\left(\frac{\sqrt{\gamma}}{2} + \frac{y-x}{\sqrt{\gamma}}\right) + \frac{e^{-x}}{\sqrt{\gamma}}\phi\left(\frac{\sqrt{\gamma}}{2} + \frac{y-x}{\sqrt{\gamma}}\right) - \frac{e^{-y}}{\sqrt{\gamma}}\phi\left(\frac{\sqrt{\gamma}}{2} + \frac{x-y}{\sqrt{\gamma}}\right), \\
\frac{\partial B}{\partial y} &= e^{-y}\Phi\left(\frac{\sqrt{\gamma}}{2} + \frac{x-y}{\sqrt{\gamma}}\right) + \frac{e^{-y}}{\sqrt{\gamma}}\phi\left(\frac{\sqrt{\gamma}}{2} + \frac{x-y}{\sqrt{\gamma}}\right) - \frac{e^{-x}}{\sqrt{\gamma}}\phi\left(\frac{\sqrt{\gamma}}{2} + \frac{y-x}{\sqrt{\gamma}}\right).
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial \theta}\left(\frac{\partial B}{\partial x}\right) &= \left(\frac{\partial \gamma}{\partial \theta}\right)\left\{ e^{-x}\phi\left(\frac{\sqrt{\gamma}}{2} + \frac{y-x}{\sqrt{\gamma}}\right)\left(\frac{1}{8\sqrt{\gamma}} - \frac{1}{2\sqrt{\gamma^3}} - \frac{y-x}{2\sqrt{\gamma^3}} + \frac{(y-x)^2}{2\gamma\sqrt{\gamma^3}}\right)\right. \\
&\quad \left. + e^{-y}\phi\left(\frac{\sqrt{\gamma}}{2} + \frac{x-y}{\sqrt{\gamma}}\right)\left(\frac{1}{8\sqrt{\gamma}} + \frac{1}{2\sqrt{\gamma^3}} - \frac{(x-y)^2}{2\gamma\sqrt{\gamma^3}}\right)\right\} \\
&\doteq \left(\frac{\partial \gamma}{\partial \theta}\right)\left\{ e^{-x}\phi\left(\frac{\sqrt{\gamma}}{2} + \frac{y-x}{\sqrt{\gamma}}\right)k_1(y-x) + e^{-y}\phi\left(\frac{\sqrt{\gamma}}{2} + \frac{x-y}{\sqrt{\gamma}}\right)k_2(x-y)\right\} \\
\text{where } k_1(x) &= \frac{1}{8\sqrt{\gamma}} - \frac{1}{2\sqrt{\gamma^3}} - \frac{x}{2\sqrt{\gamma^3}} + \frac{x^2}{2\gamma\sqrt{\gamma^3}} \\
\text{and } k_2(x) &= \frac{1}{8\sqrt{\gamma}} + \frac{1}{2\sqrt{\gamma^3}} - \frac{x^2}{2\gamma\sqrt{\gamma^3}}.
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial}{\partial\theta}\left(\frac{\partial B}{\partial x}\right) &= \left(\frac{\partial\gamma}{\partial\theta}\right)\left\{e^{-y}\phi\left(\frac{\sqrt{\gamma}}{2}+\frac{x-y}{\sqrt{\gamma}}\right)\left(\frac{1}{8\sqrt{\gamma}}-\frac{1}{2\sqrt{\gamma^3}}-\frac{x-y}{2\sqrt{\gamma^3}}+\frac{(x-y)^2}{2\gamma\sqrt{\gamma^3}}\right)\right.\\
&\quad\left.+e^{-x}\phi\left(\frac{\sqrt{\gamma}}{2}+\frac{y-x}{\sqrt{\gamma}}\right)\left(\frac{1}{8\sqrt{\gamma}}+\frac{1}{2\sqrt{\gamma^3}}-\frac{(y-x)^2}{2\gamma\sqrt{\gamma^3}}\right)\right\}\\
&\doteq \left(\frac{\partial\gamma}{\partial\theta}\right)\left\{e^{-y}\phi\left(\frac{\sqrt{\gamma}}{2}+\frac{x-y}{\sqrt{\gamma}}\right)k_1(x-y)+e^{-x}\phi\left(\frac{\sqrt{\gamma}}{2}+\frac{y-x}{\sqrt{\gamma}}\right)k_2(y-x)\right\}\\
\frac{\partial^2 B}{\partial x\partial y} &= e^{-x}\phi\left(\frac{\sqrt{\gamma}}{2}+\frac{y-x}{\sqrt{\gamma}}\right)\left(\frac{1}{2\sqrt{\gamma}}-\frac{y-x}{\sqrt{\gamma^3}}\right)+e^{-y}\phi\left(\frac{\sqrt{\gamma}}{2}+\frac{x-y}{\sqrt{\gamma}}\right)\left(\frac{1}{2\sqrt{\gamma}}-\frac{x-y}{\sqrt{\gamma^3}}\right),\\
\frac{\partial}{\partial\theta}\left(\frac{\partial^2 B}{\partial x\partial y}\right) &= \left(\frac{\partial\gamma}{\partial\theta}\right)\left[e^{-x}\phi\left(\frac{\sqrt{\gamma}}{2}+\frac{y-x}{\sqrt{\gamma}}\right)\times\right.\\
&\qquad\left\{-\frac{1}{16\sqrt{\gamma}}-\frac{1}{4\sqrt{\gamma^3}}+\left(\frac{1}{8\sqrt{\gamma^3}}+\frac{3}{2\gamma\sqrt{\gamma^3}}\right)(y-x)+\frac{(y-x)^2}{4\gamma\sqrt{\gamma^3}}-\frac{(y-x)^3}{2\gamma^2\sqrt{\gamma^3}}\right\}\\
&\quad+e^{-y}\phi\left(\frac{\sqrt{\gamma}}{2}+\frac{x-y}{\sqrt{\gamma}}\right)\times\\
&\qquad\left.\left\{-\frac{1}{16\sqrt{\gamma}}-\frac{1}{4\sqrt{\gamma^3}}+\left(\frac{1}{8\sqrt{\gamma^3}}+\frac{3}{2\gamma\sqrt{\gamma^3}}\right)(x-y)+\frac{(x-y)^2}{4\gamma\sqrt{\gamma^3}}-\frac{(x-y)^3}{2\gamma^2\sqrt{\gamma^3}}\right\}\right]\\
&\doteq \left(\frac{\partial\gamma}{\partial\theta}\right)\left[e^{-x}\phi\left(\frac{\sqrt{\gamma}}{2}+\frac{y-x}{\sqrt{\gamma}}\right)k_3(y-x)+e^{-y}\phi\left(\frac{\sqrt{\gamma}}{2}+\frac{x-y}{\sqrt{\gamma}}\right)k_3(x-y)\right],\\
\text{where } k_3(x) &= -\frac{1}{16\sqrt{\gamma}}-\frac{1}{4\sqrt{\gamma^3}}+\left(\frac{1}{8\sqrt{\gamma^3}}+\frac{3}{2\gamma\sqrt{\gamma^3}}\right)x+\frac{x^2}{4\gamma\sqrt{\gamma^3}}-\frac{x^3}{2\gamma^2\sqrt{\gamma^3}}.
\end{aligned}
$$

Let $a = \frac{\sqrt{\gamma}}{2}+\frac{y-x}{\sqrt{\gamma}}$ and $b = \frac{\sqrt{\gamma}}{2}+\frac{x-y}{\sqrt{\gamma}}$.

$$
\begin{aligned}
J(x,y;\theta) &= \frac{1}{M}\frac{\partial^2 B}{\partial x\partial y}+\frac{1}{M^2}\frac{\partial B}{\partial x}\cdot\frac{\partial B}{\partial y}\\
&= \frac{1}{M}\left\{e^{-x}\phi(a)b+e^{-y}\phi(b)a\right\}\\
&\quad+\frac{1}{M^2}\left\{e^{-x}e^{-y}\left(\Phi(a)\Phi(b)+\frac{1}{\sqrt{\gamma}}\Phi(a)\phi(b)+\frac{1}{\sqrt{\gamma}}\phi(a)\Phi(b)\right)\right.\\
&\qquad\left.-e^{-2x}\frac{\phi(a)}{\sqrt{\gamma}}\left(\Phi(a)+\frac{\phi(a)}{\sqrt{\gamma}}\right)-e^{-2y}\frac{\phi(b)}{\sqrt{\gamma}}\left(\Phi(b)+\frac{\phi(b)}{\sqrt{\gamma}}\right)\right\}.
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial J}{\partial \theta} &= \frac{1}{M}\frac{\partial}{\partial \theta}\Big(\frac{\partial^2 B}{\partial x \partial y}\Big) + \frac{1}{M^2}\frac{\partial}{\partial \theta}\Big(\frac{\partial B}{\partial x}\Big)\cdot\frac{\partial B}{\partial y} + \frac{1}{M^2}\frac{\partial B}{\partial x}\cdot\frac{\partial}{\partial \theta}\Big(\frac{\partial B}{\partial y}\Big) \\
&= \left(\frac{\partial \gamma}{\partial \theta}\right)\left[\frac{1}{M}\left\{e^{-x}\phi(a)k_3(y-x) + e^{-y}\phi(b)k_3(x-y)\right\}\right. \\
&\quad + \frac{1}{M^2}\left\{e^{-x}\phi(a)k_1(y-x) + e^{-y}\phi(b)k_2(x-y)\right\}\left\{e^{-y}\left(\Phi(b) + \frac{\phi(b)}{\sqrt{\gamma}}\right) - e^{-x}\frac{\phi(a)}{\sqrt{\gamma}}\right\} \\
&\quad \left. + \frac{1}{M^2}\left\{e^{-y}\phi(b)k_1(x-y) + e^{-x}\phi(a)k_2(y-x)\right\}\left\{e^{-x}\left(\Phi(a) + \frac{\phi(a)}{\sqrt{\gamma}}\right) - e^{-y}\frac{\phi(b)}{\sqrt{\gamma}}\right\}\right].
\end{aligned}
$$

## Appendix B. Proof of Theorem 1

*Proof.* WLOG, assume $w_K = w_{ij}\big((\mathbf{s}_i, \mathbf{s}_j)\big) = 0 \quad \forall \mathbf{s} \in R_n^c$.

$$\begin{aligned}\sigma_K^2 &= \sum_i \sum_{j>i} \sum_p \sum_{q>p} w_{ij}\big(\lambda_n(\mathbf{x}_i, \mathbf{x}_j)\big) w_{pq}\big(\lambda_n(\mathbf{x}_p, \mathbf{x}_q)\big)\sigma\big(\lambda_n(\mathbf{x}_i, \mathbf{x}_j), \lambda_n(\mathbf{x}_p, \mathbf{x}_q)\big) \\ &\equiv \sum_i \sum_{j>i} \sum_p \sum_{q>p} h_K(\mathbf{X}_{ij}, \mathbf{X}_{pq}), \quad \mathbf{X}_{ij} = (\mathbf{x}_i, \mathbf{x}_j)\end{aligned}$$

Assume that $f(\mathbf{x}_i, \mathbf{x}_j) = f(\mathbf{x}_i) f(\mathbf{x}_j) \in [m_f, M_f]$ where $m_f$ and $M_f$ are constants.

$$\begin{aligned}&\left|\frac{\int\int w_{ij}\big(\lambda_n(\mathbf{x}_i, \mathbf{x}_j)\big) w_{pq}\big(\lambda_n(\mathbf{x}_i, \mathbf{x}_j) + \mathbf{h}\big) f^2(\mathbf{x}_i, \mathbf{x}_j) d\mathbf{x}_i d\mathbf{x}_j}{\int\int w_{ij}^2\big(\lambda_n(\mathbf{x}_i, \mathbf{x}_j)\big) f(\mathbf{x}_i, \mathbf{x}_j) d\mathbf{x}_i d\mathbf{x}_j}\right| \\ &\quad \le \frac{M_f^2 \int\int w_{ij}\big(\lambda_n(\mathbf{x}_i, \mathbf{x}_j)\big) w_{pq}\big(\lambda_n(\mathbf{x}_i, \mathbf{x}_j) + \mathbf{h}\big) d\mathbf{x}_i d\mathbf{x}_j}{m_f \int\int w_{ij}^2\big(\lambda_n(\mathbf{x}_i, \mathbf{x}_j)\big) d\mathbf{x}_i d\mathbf{x}_j} \\ &\quad \le \left(\frac{M_f^2}{m_f}\right) \sqrt{\frac{\int\int w_{pq}^2\big(\lambda_n(\mathbf{x}_i, \mathbf{x}_j) + \mathbf{h}\big) d\mathbf{x}_i d\mathbf{x}_j}{\int\int w_{ij}^2\big(\lambda_n(\mathbf{x}_i, \mathbf{x}_j)\big) d\mathbf{x}_i d\mathbf{x}_j}} \text{ (by C-S inequality) } \le \frac{M_f^2}{m_f} < \infty.\end{aligned}$$

$$\begin{aligned}E\sigma_K^2 &= K(K-1) E w_K(\lambda_n \mathbf{X}_{ij}) w_K(\lambda_n \mathbf{X}_{pq}) \sigma\big(\lambda_n(\mathbf{X}_{ij} - \mathbf{X}_{pq})\big) + K E w_K(\lambda_n \mathbf{X}_{ij})^2 \sigma(\mathbf{0}) \\ &= \frac{n(n-1)(n-2)(n-3)}{4} E w_K(\lambda_n \mathbf{X}_{ij}) w_K(\lambda_n \mathbf{X}_{pq}) \sigma\big(\lambda_n(\mathbf{X}_{ij} - \mathbf{X}_{pq})\big) \\ &\quad + n(n-1)(n-2) E w_K(\lambda_n \mathbf{X}_{ij}) w_K(\lambda_n \mathbf{X}_{iq}) \sigma\big(\lambda_n(\mathbf{X}_{ij} - \mathbf{X}_{iq})\big) \\ &\quad + \frac{n(n-1)}{2} E w_K(\lambda_n \mathbf{X}_{ij})^2 \sigma(\mathbf{0}) \\ &= \frac{n(n-1)(n-2)(n-3)}{4} \lambda_n^{-2d} \int \sigma(\mathbf{h}) \int w_{ij}(\lambda_n \mathbf{X}_{ij}) w_{pq}(\lambda_n \mathbf{X}_{ij} + \mathbf{h}) \times \\ &\quad f(\mathbf{X}_{ij}) f(\mathbf{X}_{ij} + \lambda_n^{-1}\mathbf{h}) d\mathbf{X}_{ij} d\mathbf{h} \\ &\quad + n(n-1)(n-2)\lambda_n^{-d} \times \\ &\quad \int \sigma\big((0, \mathbf{h})\big) \int w_{ij}(\lambda_n \mathbf{X}_{ij}) w_{iq}\big(\lambda_n \mathbf{X}_{ij} + (0, \mathbf{h})\big) f(\mathbf{X}_{ij}) f\big(\mathbf{X}_{ij} + \lambda_n^{-1}(0, \mathbf{h})\big) d\mathbf{X}_{ij} d\mathbf{h} \\ &\quad + \frac{n(n-1)}{2} E w_K(\lambda_n \mathbf{X}_{ij})^2 \sigma(\mathbf{0})\end{aligned}$$

$$
\begin{aligned}
&\longrightarrow Kn^2\lambda_n^{-2d}Ew_K^2(\lambda_n\mathbf{X}_1)\int\sigma(\mathbf{h})Q_1(\mathbf{h})d\mathbf{h}\\
&\quad + Kn\lambda_n^{-d}Ew_K^2(\lambda_n\mathbf{X}_1)\int\sigma\big((0,\mathbf{h})\big)Q_2(\mathbf{h})d\mathbf{h} + KEw_K(\lambda_n\mathbf{X}_1)^2\sigma(\mathbf{0})\\
&= (Ks_{1k}^2)\bigg(\sigma(\mathbf{0}) + C_1\int\sigma\big((0,\mathbf{h})\big)Q_2(\mathbf{h})d\mathbf{h} + C_1^2\int\sigma(\mathbf{h})Q_1(\mathbf{h})d\mathbf{h}\bigg),
\end{aligned}
$$

as $n\to\infty$ $(K\to\infty)$, by (A′4), (A′5) and dominated convergence theorem.

$$
\begin{aligned}
\sigma_K^2 &= \sum_{a=1}^{K}\sum_{b=1}^{K}h_K(\mathbf{X}_a,\mathbf{X}_b),\\
h_{1K}(\mathbf{x}) &= Eh_K(\mathbf{x},\mathbf{X}_1),\ \ \mathbf{x}\in\mathbb{R}^{2d}\\
\text{(Eq. (5.6), Lahiri (2003))}\ \sigma_K^2 - E\sigma_K^2 &= \sum_{a=1}^{K}\big[h_K(\mathbf{X}_a,\mathbf{X}_a) - Eh_K(\mathbf{X}_1,\mathbf{X}_1)\big]\\
&\quad + \sum_{b=1}^{K-1}(K-b)\big[h_{1K}(\mathbf{X}_b) - Eh_{1K}(\mathbf{X}_b)\big]\\
&\quad + \sum_{a=2}^{K}\sum_{b=1}^{a-1}\big[h_K(\mathbf{X}_a,\mathbf{X}_b) - Eh_K(\mathbf{X}_b,\mathbf{X}_1)\big]\\
&\doteq D_{1K} + D_{2K} + D_{3K}\\
\big|Eh_K(\mathbf{x},\mathbf{X}_1)^r\big| &= \bigg|\int\int w_K(\lambda_n\mathbf{x})^r w_K(\lambda_n\mathbf{s})^r\sigma^r(\lambda_n\mathbf{x},\lambda_n\mathbf{s})f(\mathbf{s})d\mathbf{s}\bigg|\\
&\le (M_k^2)^r M_f\lambda_n^{-2d}\int|\sigma(\mathbf{s})|^r d\mathbf{s},\ \text{by (A'1)}.\\
\big|Eh_K(\mathbf{X}_1,\mathbf{X}_2)^r\big| &\le E\big|Eh_K(\mathbf{X}_1,\mathbf{X}_2)^r|\mathbf{X}_1)\big| \le C\big(M_f,\sigma(\cdot)\big)M_k^{2r}\lambda_n^{-2d}.
\end{aligned}
$$

Then

$$
\begin{aligned}
&\sum_{K=1}^{\infty}E\big(\sigma_K^2 - E\sigma_K^2\big)^4\big/\big(K^2\lambda_n^{-2d}Ew_K^2(\lambda_n\mathbf{X}_1)\big)^4\\
&\qquad\le C(M_f,\sigma(\cdot),C_1)\sum_{K=1}^{\infty}\left(\frac{M_k^2}{Ew_K^2(\lambda_n\mathbf{X}_1)}\right)^4\left(\frac{K^6\lambda_n^{-8d}}{K^8\lambda_n^{-8d}}\right)\\
&\qquad= C(M_f,\sigma(\cdot),C_1)\sum_{K=1}^{\infty}(\gamma_{1K}^2)^4\left(\frac{K^6\lambda_n^{-8d}}{K^8\lambda_n^{-8d}}\right)\\
&\qquad= C(M_f,\sigma(\cdot),C_1)\sum_{K=1}^{\infty}\frac{(\gamma_{1K}^2)^4}{K^2} < \infty,\quad \text{by (A'6)}
\end{aligned}
$$

since

$$
\begin{aligned}
ED_{1n}^4 &\le C\big\{KEh_K(\mathbf{X}_1,\mathbf{X}_1)^4 + K^2\big(Eh_K(\mathbf{X}_1,\mathbf{X}_1)^2\big)^2\big\} \\
&\le C\sigma(\mathbf{0})^4 K^2 (s_{1K}^2\gamma_{1k}^2)^4\lambda_n^{-4d} \le C\sigma(\mathbf{0})^4 KEw_K(\lambda_n\mathbf{X}_1)^8, \\
ED_{2n}^4 &\le C\Big[\sum_{b=1}^{K}(K-b)^4 Eh_{1K}(\mathbf{X}_1)^4 + \Big\{\sum_{a=1}^{K}(K-a)^2 Eh_{1K}(\mathbf{X}_1)^2\Big\}^2\Big] \\
&\le C(M_f,\sigma(\cdot))K^6 M_k^8\lambda_n^{-8d}, \\
ED_{3n}^4 &\le CK\sum_{a=2}^{K} E\Big\{\sum_{b=1}^{a-1}\big(h_K(\mathbf{X}_a,\mathbf{X}_b) - h_{1K}(\mathbf{X}_b)\big)\Big\}^4 \\
&\le CK\sum_{a=2}^{K}\Bigg[E\Big\{(a-1)E\big[(h_K(\mathbf{X}_a,\mathbf{X}_1) - h_{1K}(\mathbf{X}_1))^4|\mathbf{X}_a\big] \\
&\qquad + \big((a-1)E[(h_K(\mathbf{X}_a,\mathbf{X}_1) - h_{1K}(\mathbf{X}_1))^2|\mathbf{X}_a]\big)^2\Big\} \\
&\qquad + E\big[(a-1)(h_{1K}(\mathbf{X}_a) - Eh_{1K}(\mathbf{X}_1))\big]^4\Bigg] \\
&\le C(M_f,\sigma(\cdot))M_k^8[K^3\lambda_n^{-2d} + K^4\lambda_n^{-4d} + K^6\lambda_n^{-8d}]
\end{aligned}
$$

(see details in Eq. (5.7)-(5.9), Lahiri (2003)). It follows the analogous result by Lemma 5.2 (i) in Lahiri (2003). If $n/\lambda_n^d \to C_1 \in (0,\infty)$ and (A$'$1), (A$'$4) and (A$'$5) hold, then

$$
\big(K\cdot Ew_K^2(\lambda_n\mathbf{X}_1)\big)^{-1}\sigma_K^2 \to \Big(\sigma(\mathbf{0}) + C_1\int\sigma\big((0,\mathbf{h})\big)Q_2(\mathbf{h})d\mathbf{h} + C_1^2\int\sigma(\mathbf{h})Q_1(\mathbf{h})d\mathbf{h}\Big).
$$

Let $\xi_k \triangleq \xi_k(\mathbf{s}^k) = Z_k(\mathbf{s}^k) - EZ_k(\mathbf{s}^k)$. Define for $c > 0$,

$$
\begin{aligned}
\eta_k &= \xi_k I(|\xi_k| \le c) - E\xi_0 I(|\xi_0| \le c) \\
\gamma_k &= \xi_k I(|\xi_k| > c) - E\xi_0 I(|\xi_0| > c)
\end{aligned}
$$

where $\xi_0 = \xi_k(\mathbf{0})$. Let $S_K^{1*} = \sum_{k=1}^K w_k\eta_k,\ S_K^{2*} = \sum_{k=1}^K w_k\gamma_k$, and

$$\sigma_1^*(\mathbf{x};c) = Cov(\xi_k I(|\xi_k| \le c), \xi_0 I(|\xi_0| \le c)),$$
$$\sigma_2^*(\mathbf{x};c) = Cov(\xi_k I(|\xi_k| > c), \xi_0 I(|\xi_0| > c)).$$

We separate the sum of centered processes into two parts,

$$\begin{aligned} S_K &\equiv \sum_{k=1}^K w_k(\mathbf{s}^k)\xi_k(\mathbf{s}^k) \\ &= \sum_{k=1}^K w_k\eta_k + \sum_{k=1}^K w_k\gamma_k = S_K^{1*} + S_K^{2*}. \end{aligned}$$

By the moment condition on $Z_k(\cdot)$ and the strong mixing condition,

$$\begin{aligned} \max_{j=1,2}\int\int|\sigma_j^*(\mathbf{x};c)|d\mathbf{x} &\le \int\int\left(E|\xi_k(\mathbf{0})^{2+\delta}|\right)^{2/(2+\delta)}\alpha(|\mathbf{x}|;1)^{\delta/(2+\delta)}d\mathbf{x} \\ &\le C\big(d,\delta,E|\xi_k(\mathbf{0})|^{2+\delta},\beta(1)\big)\int_0^\infty t^{d-1}\alpha_1(t)^{\delta/(2+\delta)}dt < \infty \\ \Rightarrow \int\int|\sigma_j^*(\mathbf{x};c)|d\mathbf{x} < \infty \quad \forall c>0, j=1,2. & \qquad (6.1) \end{aligned}$$

Since $|Q_1(\mathbf{x})| \le 1$, we obtain that for all $\mathbf{x}$ and $c$,

$$\begin{aligned} &\left|\int\int\sigma(\mathbf{x})Q_1(\mathbf{x})d\mathbf{x} - \int\int\sigma_1^*(\mathbf{x};c)Q_1(\mathbf{x})d\mathbf{x}\right| \\ &\le \int\int\Big\{|Cov\big(\xi_k(\mathbf{x})I(|\xi_k(\mathbf{x})|>c),\xi_k(\mathbf{0})\big)| \\ &\qquad + |Cov\big(\xi_k(\mathbf{x})I(|\xi_k(\mathbf{x})|\le c),\xi_k(\mathbf{0})I(|\xi_k(\mathbf{0})|>c)\big)|\Big\}d\mathbf{x} \\ &\le C(d)\big(E|\xi_k(\mathbf{0})|^{2+\delta}\big)^{\frac{1}{2+\delta}}\big(E|\xi_k(\mathbf{0})I(|\xi_k(\mathbf{0})|>c)|^{2+\delta}\big)^{\frac{1}{2+\delta}}\int_0^\infty t^{d-1}\alpha_1(t)^{\delta/(2+\delta)}dt \\ &\longrightarrow 0 \quad \text{as } c\to\infty \\ \Rightarrow &\int\int\sigma(\mathbf{x})Q_1(\mathbf{x})d\mathbf{x} - \int\int\sigma_1^*(\mathbf{x};c)Q_1(\mathbf{x})d\mathbf{x} = o(1). \qquad (6.2) \end{aligned}$$

The similar one can be applied to the form with $Q_2$.

$$P\left(\lim_{K\to\infty}\left[ES_K^{2*}(c)-\tilde{\sigma}_{2,K}^2(c)\right]/(Ks_{1K}^2)=0\right)=1 \tag{6.3}$$

where

$$\tilde{\sigma}_{2,K}^2(c)\equiv K(K-1)Ew_K(\lambda_n\mathbf{X}_1)w_K(\lambda_n\mathbf{X}_2)\sigma_2^*\big(\lambda_n(\mathbf{X}_1-\mathbf{X}_2);c\big)+KEw_K^2(\lambda_n\mathbf{X}_1)\sigma_2^*(\mathbf{0};c).$$

From the previous proof of the asymptotic variance, we can obtain the result such that for any $c>0$,

$$\begin{aligned}\tilde{\sigma}_{2,K}^2(c) = (Ks_{1K}^2)&\left\{C_1^2\int\sigma_2^*(\mathbf{h};c)Q_1(\mathbf{h})d\mathbf{h}+C_1\int\sigma_2^*\big((0,h);c\big)Q_2(h)dh\right\}\\ &+Ks_{1K}^2\sigma_2^*(\mathbf{0};c)\end{aligned}$$

as $n\to\infty$. Since $|\int\sigma_2^*(\mathbf{h};c)Q_1(\mathbf{h})d\mathbf{h}|+|\int\sigma_2^*\big((0,h);c\big)Q_2(h)dh|+|\sigma_2^*(\mathbf{0};c)|=o(1)$ as $c\to\infty$, then by (6.1),(6.2) and (6.3),

$$P\left(\lim_{c\to\infty}\limsup_{n\to\infty}ES_K^{2*}(c)/(Ks_{1K}^2)=0\right)=1.$$

Now we apply a classical Bernstein blocking technique for the proof of asymptotic normality. Notations for the blocking technique of Bernstein are same with those of Lahiri. Let $\{\lambda_{1n}\}$ and $\{\lambda_{2n}\}$ be two sequences satisfying the condition (A′6) and $\{\lambda_{3n}\}=\{\lambda_{1n}\}+\{\lambda_{2n}\}$. Then the partition of the region $R_n$ is denoted by

$$\Gamma_n(l;\epsilon)\equiv I_1(\epsilon_1)\times\cdots I_d(\epsilon_d),\quad \epsilon=(\epsilon_1,\cdots,\epsilon_d)'\in\{1,2\}^d,$$

where $I_j(\epsilon_j)=(l_j\lambda_{3n},l_j\lambda_{3n}+\lambda_{1n}]$, if $\epsilon_j=1$ and $I_j(\epsilon_j)=(l_j\lambda_{3n}+\lambda_{1n},(l_j+1)\lambda_{3n},]$, if

$\epsilon_j = 2$. Note that with $q(\epsilon) \equiv [\{1 \le j \le d : \epsilon_j = 1\}]$,

$$|\Gamma_n(l;\epsilon)| = \lambda_{1n}^{q(\epsilon)} \lambda_{2n}^{d-q(\epsilon)}$$

for all $l$ and $\epsilon$. Let $\epsilon_0 = (1, \cdots, 1)'$. Then

$$|\Gamma_n(l;\epsilon)| = o(|\Gamma_n(l;\epsilon_0)|).$$

Let $L_{1n} = \{l : \Gamma_n(l;\mathbf{0}) \subset R_n\}$ be the index set of all hypercubes $\Gamma_n(l;\mathbf{0})$ that are contained in $R_n$, and let $L_{2n} = \{l : \Gamma_n(l;\mathbf{0}) \cap R_n \neq 0, \Gamma_n(l;\mathbf{0}) \cap R_n^c \neq \varnothing\}$ be the index set of boundary hypercubes. With the notation above, $S_K^{1*}$ can be separated into the sum of big blocks and small blocks and the sum of remaining variables. Here we consider only the case that station elements $i$ and $j$ are in the same block. If sums of pair whose elements are in different block, the joint probability of exceeding over the threshold would be zero as the sampling region is growing. Thus as $n \to \infty$, sums of pair would converge to 0 and it could be negligible in consideration of our sum of processes.

$$\begin{aligned}
S_K^{1*}/\sigma_K &= \sum_{k=1}^{K} w_k(\mathbf{s}^k)\eta_k(\mathbf{s}^k)/\sigma_K \\
&= \sum_{l \in L_{1n}} S_K^{1*}(l;\epsilon_0) + \sum_{\epsilon \neq \epsilon_0} \sum_{l \in L_{1n}} S_K^{1*}(l;\epsilon) + \sum_{l \in L_{2n}} S_K^{1*}(l;\mathbf{0}) \\
&= \sum_{q=1}^{|L_{1n}|} \sum_{k \in J_q} w_k\eta_k/\sigma_K + \sum_{q=1}^{|L_{1n}|} \sum_{k \in H_q} w_k\eta_k/\sigma_K + \sum_{k \in L_{2n}} w_k\eta_k/\sigma_K \\
&\triangleq \sum_{q=1}^{|L_{1n}|} S'_{1Kq} + \sum_{q=1}^{|L_{1n}|} S'_{2Kq} + \sum_{k \in L_{2n}} w_k\eta_k/\sigma_K \\
&= S'_{1K} + S'_{2K} + S'_{3K} \\
&\quad \text{(big blocks + little blocks + leftover)}
\end{aligned}$$

where $\sigma_K^2 = Var\big(\sum_{k=1}^{K} w_K(\mathbf{s}_k)\xi_k(\mathbf{s}_k)\big)$, $S'_{1Kq} = \sum_{k \in J_q} w_k\eta_k/\sigma_K$ and $S'_{2Kq} = \sum_{k \in H_q} w_k\eta_k/\sigma_K$.

Two big blocks $\Gamma(l_1;\epsilon_0)$ and $\Gamma(l_2;\epsilon_0)$ are separated by the distance

$$d(\Gamma(l_1;\epsilon_0),\Gamma(l_2;\epsilon_0)) \geq [(|l_1 - l_2| - d)_+\lambda_{3n}] + \lambda_{2n}.$$

By the strong mixing condition,

$$\left|E\exp(itS'_{1K}) - \prod_{l\in L_{1n}} E\exp\left(itS_K(l;\epsilon_0)\right)\right| \leq C|L_{1n}|\alpha(\lambda_{2n};\lambda_n^d).$$

Therefore the asymptotic behavior can be shown with the independence of $S'_{1Kq}$. Using Lemma A.1 in Lahiri(2003), we show that with probability one,

$$\sum_{q=1} E{S'_{1Kq}}^4\sigma_K^4 = o([K^2\lambda_n^{-2d}s_{1K}^2]^2), \tag{6.4}$$

$$Var(S'_{2K}\sigma_K) = o(K^2\lambda_n^{-2d}s_{1K}^2), \tag{6.5}$$

$$Var(S'_{3K}\sigma_K) = o(K^2\lambda_n^{-2d}s_{1K}^2). \tag{6.6}$$

Now we have to show that

$$\sum_{q=1} E{S'_{1Kq}}^2\sigma_K^2 - \sigma_K^2 = o(K^2\lambda_n^{-2d}s_{1K}^2). \tag{6.7}$$

To prove above equation, we use Lemma 5.1 in Lahiri(2003) and (6.4)-(6.6).

$$
\begin{aligned}
&\left|\sum_{q=1} E{S'_{1Kq}}^2\sigma_K^2 - \sigma_K^2\right| \\
&\le \left|\sum_{q=1} E{S'_{1Kq}}^2\sigma_K^2 - E(S'_{1K}\sigma_K)^2\right| \\
&\qquad + 2\sigma_K^2\Big(E(S'_{2K}+S'_{3K})^2\Big)^{1/2}E({S'_{1K}}^2)^{1/2} + E(S'_{2K}+S'_{3K})^2\sigma_K^2 \\
&\le C\bigg[\sum C_0^2 M_n^2(\lambda_{1n}^{2d}n^2\lambda_n^{-2d} + \log n)^2\alpha([(|l_1-l_2|-d)_+\lambda_{3n}] + \lambda_{2n};\lambda_{1n}^d)\bigg] + o(K^2\lambda_n^{-2d}s_{1K}^2) \\
&\le C(d,C_0)M_n^2(\lambda_{1n}^{2d}n^2\lambda_n^{-2d} + \log n)^2(\lambda_n/\lambda_{3n})^{2d}\times \\
&\qquad \Bigg(\alpha(\lambda_{2n};\lambda_{1n}^d) + \sum_{k=1}^{\lambda_n/\lambda_{3n}} k^{d-1}\alpha(k\lambda_{3n}+\lambda_{2n};\lambda_{1n}^d)\Bigg) + o(K^2\lambda_n^{-2d}s_{1K}^2) \\
&= o(K^2\lambda_n^{-2d}s_{1K}^2)
\end{aligned}
$$

Thus we show that the equation (6.7) holds and it is needed only to establish the Lindeberg condition,

$$
\sum_{q=1}^{|L_{1n}|} E(S'_{1Kq})^2 I_{(|S'_{1Kq}|>\epsilon)} \longrightarrow 0, \quad \text{as } n\to\infty.
$$

Since we have

$$
\begin{aligned}
\sum_{q=1}^{|L_{1n}|}\int_{|S'_{1Kq}|>\epsilon}|S'_{1Kq}|^{2+\delta}dP &= \sum_{q=1}^{|L_{1n}|}\int_{\left|\sum_{k\in J_q} w_k\eta_k/\sigma_K\right|>\epsilon}\left|\sum_{k\in J_q} w_k\eta_k\right|^{2+\delta}\Big/\sigma_K^{2+\delta}dP \\
&\le C|L_{1n}|n^{-d\left(\frac{2+\delta}{2}\right)}\left(\frac{1}{\sigma^2}\right)^{(2+\delta)/2}\int_{\left|\sum_{k\in J_q} w_k\eta_k/\sigma_K\right|>\epsilon}\left|\sum_{k\in J_q} w_k\eta_k\right|^{2+\delta}dP \\
&\le C\left(\left[\frac{n}{\lambda_{1n}}\right]\right)^d n^{-d\left(\frac{2+\delta}{2}\right)}\int_{\left|\sum_{k\in J_q} w_k\eta_k/\sigma_K\right|>\epsilon}\left|\sum_{k\in J_q} w_k\eta_k\right|^{2+\delta}dP \\
&\le C(n^{\delta/2}\lambda_{1n})^{-d}M_k^2\int_{\left|\sum_{k\in J_q} w_k\eta_k/\sigma_K\right|>\epsilon}\left|\sum_{k\in J_q}\eta_k\right|^{2+\delta}dP \\
&\longrightarrow 0 \quad \text{as } n\to\infty,
\end{aligned}
$$

this implies that the Lindeberg condition holds. □

# Bibliography


Amemiya, T. 1985. *Advanced Econometrics.* Harvard University Press, Cambridge.

Beirlant, J., Y. Goegebeur, J. Segers and J. Teugels. 2004. *Statistics of Extremes: Theory and Applications.* New York: Wiley.

Brown, B. M. and S. I. Resnick. 1977. "Extreme values of independent stochastic processes." *Journal of Applied Probability* 14:732–739.

Coles, S. G. 1993. "Regional modelling of extreme storms via max-stable processes." *Journal of the Royal Statistical Society, Series B* 55(4):797–816.

Coles, S. G. 2001. *An Introduction to Statistical Modeling of Extreme Values.* New York: Springer Verlag.

Coles, S. G. and J. A. Tawn. 1996. "Modeling extremes of the areal rainfall process." *Journal of the Royal Statistical Society, Series B* 58:329–347.

Cooley, D., J. Cisewski, R. J. Erhardt, S. Jeon, E. Mannshardt, B. Omolo and Y. Sun. 2012. "A survey of spatial extremes: measuring spatial dependence and modeling spatial effects." *REVSTAT Statistical Journal* (accepted).

Cox, D. R. and N. Reid. 2004. "A note on pseudolikelihood constructed from marginal densities." *Biometrika* 91:729–737.

Cressie, N. A. 1993. *Statistics for Spatial Data.* New York: Wiley.

Davison, A. C. and M. M. Gholamrezaee. 2012. "Geostatistics of extremes." *Proc. R. Soc. Lond. A* 468:581–608.

de Haan, L. 1984. "A spectral representation for max-stable processes." *Annals of Probability* 12(4):1194–1204.

de Haan, L. and A. Ferreira. 2006. *Extreme Value Theory, An Introduction.* New York: Springer.

de Haan, L. and U. Stadtmüller. 1996. "Generalized regular variaton of second order." *J. Austral. Math. Soc, Ser. A* 91:381–395.

Drees, H. 1998. "On smooth statistical tail functionals." *Scandinavian Journal of Statistics* 25:187–210.

Fisher, R. A. and L. H. C. Tippett. 1928. "Limiting forms of the frequency distributions of the largest or smallest member of a sample." *Proc. Camb. Phil. Soc* 24:180–190.

Genton, M. G., Y. Ma and H. Sang. 2011. "On the likelihood function of Gaussian max-stable processes." *Biometrika* 98(2):481–488.

Gnedenko, B. V. 1943. "Sur la distribution limite du terme maximum d'une série aléatoire." *Ann. Math.* 44:423–453.

Guyon, X. 1995. *Random Fields on a Network: Modeling, Statistics, and Applications.* New York: Springer Verlag.

Kabluchko, Z., M. Schlather and L. de Haan. 2009. "Stationary max-stable fields associated to negative definite functions." *Annals of Probability* 37:2042–2065.

Lahiri, S. N. 2003. "Central Limit Theorems for weighted sums under some stochastic and fixed spatial sampling designs." *Sankhyā: The Indian Journal of Statistics* 65:356–388.

Leadbetter, M. R., G. Lindgren and H. Rootzén. 1983. *Extremes and Related Properties of Random Sequences and Series.* New York: Springer Verlag.

Lindsay, B. G. 1988. "Composite likelihood methods." *Contemporary Mathematics* 80:221–239.

Mardia, K. V. and R. J. Marshall. 1984. "Maximum likelihood estimation of models for residual covariance in spatial statistics." *Biometrika* 71(1):135–146.

Matheron, G. 1987. "Suffit-il, pour une covariance, d'être de type positif?" *Sciences de la Terre, série informatique géologique* 26:51–66.

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.

Pickands, J. 1975. "Statistical inference using extreme order statistics." *Annals of Statistics* 3:119–131.

Resnick, S. I. 1987. *Extreme Values, Regular Variation and Point Processes.* New York: Springer Verlag.

Resnick, S. I. 2007. *Heavy-tail Phenomena: Probabilistic and Statistical Modeling.* New York: Springer Verlag.

Rootzén, H. and N. Tajvidi. 2006. "Multivariate generalized Pareto distributions." *Bernoulli* 12(5):917–930.

Ruben, H. 1964. "An asymptotic expansion for the multivariate normal distribution and Mill's ratio." *Journal of Research of the National Bureau of Standards B* 68:3–11.

Schlather, M. 2002. "Models for stationary max-stable random fields." *Extremes* 5:33–44.

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

Shi, D. 1995. "Fisher information for a multivariate extreme value distribution." *Biometrika* 82(3):644–649.

Shi, D., R. L. Smith and S. G. Coles. 1992. Joint versus marginal estimation for bivariate extremes. Technical Report 2074 Department of Statistics, University of North Carolina, Chapel Hill.

Smith, R. L. 1987. "Estimating tails of probability distributions." *Annals of Statistics* 15:1174–1207.

Smith, R. L. 1989. "Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone (with discution)." *Statistical Science* 4:367–393.

Smith, R. L. 1990. "Max-stable processes and spatial extremes." Unpublished manuscript.

Smith, R. L. 2003. "Statistics of extremes, with applications in environment, insurance and finance." Chapter 1 of, *Extreme Values in Finance, Telecommunications and the Environment*, edited by B. Finkenstadt and H. Rootzen, Chapman and Hall/CRC Press, London, pp. 1-78.

Smith, R. L., J. A. Tawn and S. G. Coles. 1997. "Markov chain models for threshold exceedances." *Biometrika* 84(2):249–268.

Zhang, H. 2004. "Inconsistent estimation and asymptotically equivalent interpolations in model-based geostatistics." *Journal of the American Statistical Association* 99(465):250–261.