Spatial Statistics 8 (2014) 86–103

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta

Tapered composite likelihood for spatial max-stable models

Huiyan Sang [a,*], Marc G. Genton [b]

[a] Department of Statistics, Texas A&M University, College Station, TX 77843-3143, USA
[b] CEMSE Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia

ARTICLE INFO

Article history:
Received 6 February 2013
Accepted 18 July 2013
Available online 29 July 2013

Keywords:
Composite likelihood
Generalized extreme-value distribution
Max-stable process
Statistics of extremes
Weighted composite likelihood

ABSTRACT

Spatial extreme value analysis is useful to environmental studies, in which extreme value phenomena are of interest and meaningful spatial patterns can be discerned. Max-stable process models are able to describe such phenomena. This class of models is asymptotically justified to characterize the spatial dependence among extremes. However, likelihood inference is challenging for such models because their corresponding joint likelihood is unavailable and only bivariate or trivariate distributions are known. In this paper, we propose a tapered composite likelihood approach by utilizing lower dimensional marginal likelihoods for inference on parameters of various max-stable process models. We consider a weighting strategy based on a "taper range" to exclude distant pairs or triples. The "optimal taper range" is selected to maximize various measures of the Godambe information associated with the tapered composite likelihood function. This method substantially reduces the computational cost and improves the efficiency over equally weighted composite likelihood estimators. We illustrate its utility with simulation experiments and an analysis of rainfall data in Switzerland.

Published by Elsevier B.V.

1. Introduction

Statistical modeling of extreme values has recently drawn research attention. Many environmental problems involve extreme values observed over space, such as extreme precipitation, heavy snow, windstorms and high tides, to name a few. The primary interest in analyzing such data is to characterize recognizable and meaningful spatial patterns that are useful to understanding, predicting, and managing the risks of extreme environmental phenomena.

* Corresponding author. Tel.: +1 9798453156.
E-mail addresses: huiyan@stat.tamu.edu (H. Sang), marc.genton@kaust.edu.sa (M.G. Genton).

2211-6753/$ – see front matter. Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.spasta.2013.07.003

Recent developments of statistical models for spatial extreme values are mainly based on latent variables, on copulas and on spatial max-stable processes; see the review paper by Davison et al. (2012) and references therein. Among these models, max-stable stochastic processes have emerged as a fundamental class of models that are able to describe spatial extreme value phenomena. This is because they arise as the natural generalization of the univariate generalized extreme value (GEV) distribution in infinite dimensional continuous spaces, providing an asymptotically justified approach to modeling process extremes. Max-stable process models for spatial data were first constructed using the spectral representation of de Haan (1984). There have been several subsequent works on the construction of spatial max-stable process models (see, e.g., Smith, 1990; Schlather, 2002; Kabluchko et al., 2009; Smith and Stephenson, 2009; Davison and Gholamrezaee, 2012) and on their application (see, e.g., Coles, 1993; Buishand et al., 2008; Padoan et al., 2010).

Despite many attractive properties of max-stable process models, both classical and Bayesian inferences encounter difficulties because closed-form expressions of the corresponding joint likelihoods are typically not available except for some trivial cases. Taking advantage of the availability of low-dimensional marginal likelihoods, the composite likelihood method has been introduced for inference on the parameters of max-stable processes. Padoan et al. (2010) were the first to suggest maximum pairwise likelihood estimation for a particular class of max-stable process models, namely Gaussian extreme value processes (Smith, 1990). Genton et al. (2011) studied maximum composite likelihood estimators for the same model based on both pairs and triples. They demonstrated substantial gain in efficiency from $p = 2$ to $p = 3$ sites in $\mathbb{R}^2$ by means of a Monte Carlo simulation study. Pairwise and triplewise composite likelihoods have also been used for inference on other max-stable process models (see, e.g., Blanchet and Davison, 2011; Davison et al., 2012; Huser and Davison, 2013).

Several investigations have considered the choice of weights of the composite likelihoods of time series and spatial data to improve statistical efficiency or to reduce the computational burden associated with large datasets. In the context of Gaussian process models, Bevilacqua et al. (2012) suggested that down-weighting or excluding likelihood contributions from sites that are very far apart leads to efficiency gains over the full composite likelihood. In the context of time series, several works (see, e.g., Joe and Lee, 2009; Davis and Yau, 2011) have shown that including unnecessary pairs can cause some loss of estimation efficiency. Padoan et al. (2010) found in a simulation study that the composite likelihoods constructed only using neighboring sites can reduce the asymptotic variances of the model parameters from a Gaussian extreme value process referred to as the *Smith* model (Smith, 1990).

In this paper, we extend the results in Padoan et al. (2010) to investigate in detail the utility of a tapered composite likelihood approach to make inference on several formulations of max-stable process models, including the *Smith* model, the *Schlather* model (Schlather, 2002) and the *Brown–Resnick* model (Kabluchko et al., 2009). Two flexible and practical tapering strategies are proposed to improve the efficiency of the composite likelihood estimators. One is based on the trace of the estimated covariance matrix associated with parameter estimates, and the other is based on its determinant. The proposed tapering strategy is also useful to reducing the computational cost caused by the combinatorial explosion associated with the use of composite likelihood with large datasets. In particular, we extend our investigation beyond the tapered pairwise composite likelihood for the settings of isotropic max-stable processes to the context of anisotropic max-stable processes and triplewise composite likelihoods. Moreover, we discuss the connection between the choice of weights and the strength of the extremal dependence.

The paper is organized as follows. Section 2 reviews the theory of max-stable processes. The tapered composite likelihood approach is developed in Section 3, while Section 4 illustrates the performance of our method through a number of simulation studies. We conclude with an illustration of spatial extreme value analysis of precipitation data in Switzerland.

2. Max-stable processes

2.1. Spatial max-stable processes

Let $\{\tilde{Z}(\boldsymbol{s})^{(i)}\}$, $\boldsymbol{s} \in \mathcal{S} \subset \mathbb{R}^d$, $i = 1, \ldots, n$, be n independent replicates of a continuous spatial stochastic process, where $\mathcal{S}$ is an index set. A spatial stochastic process, $Z(\boldsymbol{s})$, is *max-stable* if there

exist continuous functions, $a_n(\boldsymbol{s}) > 0$ and $b_n(\boldsymbol{s})$, such that as $n \to \infty$,

$$\frac{\max_{i=1}^n \tilde{Z}(\boldsymbol{s})^{(i)} - b_n(\boldsymbol{s})}{a_n(\boldsymbol{s})} \stackrel{D}{=} Z(\boldsymbol{s}),$$

where $Z(\boldsymbol{s})$ is a non-degenerate random field. By the above definition, a spatial max-stable process offers a natural choice for modeling spatial extremes. The marginal distributions of the random vector $\{Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_K)\}^{\mathrm{T}}$ for any points $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_K$ belong to the class of multivariate extreme value distributions (de Haan and Resnick, 1977). In particular, the univariate marginal distributions belong to the family of GEV distributions, in which the distribution function takes the form

$$G(z; \mu, \sigma, \xi) = \begin{cases} \exp\left[-\left\{1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right\}^{-1/\xi}\right], & \xi \neq 0, \\ \exp\left\{-\exp\left(-\frac{z-\mu}{\sigma}\right)\right\}, & \xi = 0, \end{cases} \tag{1}$$

for z such that $1 + \xi(z-\mu)/\sigma > 0$. In (1), $\mu \in \mathbb{R}$ is the location parameter, $\sigma > 0$ is the scale parameter and $\xi \in \mathbb{R}$ is the shape parameter. For univariate GEV distributions, there is no loss of generality in transforming the margins to the standard Fréchet distribution, which has a distribution function $\exp(-1/z)$. Therefore, it is common to consider a *simple spatial max-stable process*, $Z(\boldsymbol{s})$, with standard Fréchet marginals.

Below, we list three specific examples of max-stable process models following two types of spectral representations of max-stable processes.

Example 1. A first max-stable process is referred to as the *Smith* model, also called the *storm profile* model (see, e.g., de Haan, 1984; Smith, 1990). Let $\{(r_i, \boldsymbol{s}_i^*),\ i \in \mathbb{N}\}$ be the points of a Poisson process on $\mathbb{R}^+ \times \mathcal{S}^*$ with intensity $r^{-2}dr\nu(d\boldsymbol{s}^*)$, where $\mathcal{S}^*$ is an arbitrary measurable set and ν is a finite measure on $\mathcal{S}^*$. Further, let f be non-negative continuous functions $\{f(\boldsymbol{s}, \boldsymbol{s}^*) : \boldsymbol{s} \in \mathcal{S} \subset \mathbb{R}^d, \boldsymbol{s}^* \in \mathcal{S}^* \subset \mathbb{R}^d\}$ such that $\int_{\mathcal{S}^*} f(\boldsymbol{s}, \boldsymbol{s}^*)\nu(d\boldsymbol{s}^*) = 1$, for all $\boldsymbol{s} \in \mathcal{S}$. Then

$$Z(\boldsymbol{s}) = \max_i\{r_i f(\boldsymbol{s}, \boldsymbol{s}_i^*)\} \tag{2}$$

is a simple max-stable process.

One common choice of $f(\boldsymbol{s}, \boldsymbol{s}_i^*)$ is the Gaussian density (Smith, 1990), i.e., $f(\boldsymbol{s}, \boldsymbol{s}_i^*) = (2\pi)^{-1}|\boldsymbol{\Sigma}|^{-1/2}$ $\exp\{-(\boldsymbol{s} - \boldsymbol{s}^*)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{s} - \boldsymbol{s}^*)/2\}$, where $\boldsymbol{\Sigma} \in \mathbb{R}^{2\times 2}$ is the covariance matrix. The square roots of the eigenvalues of $\boldsymbol{\Sigma}$ control the range of the spatial dependence. This model is called the *Gaussian extreme value process* model (Smith, 1990). Other choices include Student's t and Laplace densities (de Haan and Pereira, 2006). This construction has a nice rainfall-storms interpretation where $r_i f(\boldsymbol{s}, \boldsymbol{s}_i^*)$ is regarded as the impact at $\boldsymbol{s}$ of a storm of magnitude r_i, centered at $\boldsymbol{s}_i^*$, and of shape f. The spatial max-stable process $Z(\boldsymbol{s})$ then represents the largest impact of all storms at $\boldsymbol{s}$.

The bivariate probability distribution of the Smith model has the form:

$$-\log F(z_1, z_2) = z_1^{-1}\Phi\left\{\frac{\alpha(\boldsymbol{h})}{2} + \alpha^{-1}(\boldsymbol{h})\log\left(\frac{z_2}{z_1}\right)\right\} + z_2^{-1}\Phi\left\{\frac{\alpha(\boldsymbol{h})}{2} + \alpha^{-1}(\boldsymbol{h})\log\left(\frac{z_1}{z_2}\right)\right\},$$

where $\alpha(\boldsymbol{h}) = \sqrt{\boldsymbol{h}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{h}}$, in which $\boldsymbol{h}$ is the separation vector between the two locations, $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$, i.e., $\boldsymbol{h} = \boldsymbol{s}_1 - \boldsymbol{s}_2$.

Very recently, Genton et al. (2011) derived a closed-form expression of the joint distribution function of a Gaussian extreme value process indexed by $\mathbb{R}^d$ at $p \leq d+1$ sites, $d \geq 1$. In particular, for the spatial Gaussian max-stable process, this allows the closed-form distribution function based on triples of the Smith model (see Appendix A for details).

Schlather (2002) provided an alternative spectral representation for simple spatial max-stable processes as follows. Let $\{r_i,\ i \in \mathbb{N}\}$ be the points of a Poisson process on $\mathbb{R}^+$ with intensity dr/r^2.

Let $\{W(\boldsymbol{s}), \boldsymbol{s} \in \mathcal{S} \subset \mathbb{R}^d\}$ be a non-negative stochastic process that satisfies $E[W(\boldsymbol{s})] = 1$ for all $\boldsymbol{s} \in \mathcal{S}$ and let $\{W_i(\boldsymbol{s}),\ i \in \mathbb{N}\}$ be a collection of independent copies of this process. Then

$$Z(\boldsymbol{s}) = \max_i \{r_i W_i(\boldsymbol{s})\} \tag{3}$$

is a simple max-stable process.

Following the general representation in (3), there are several parametric formulations of $W(\boldsymbol{s})$ that allows us to construct valid simple max-stable processes. Below, we list two specific examples of max-stable process models following this representation.

Example 2. A second max-stable process is referred to as the *Schlather* model (Schlather, 2002) defined as $W_i(\boldsymbol{s}) = \max\{0, \epsilon_i(\boldsymbol{s})\}$, where $\epsilon_i(\boldsymbol{s})$ is taken to be independent copies of a stationary Gaussian process with unit variance and correlation function $\rho(\boldsymbol{h})$, scaled such that the condition $E[W_i(\boldsymbol{s})] = 1$ holds. Schlather (2002) showed that in this case the bivariate probability distribution takes the form

$$-\log F(z_1, z_2) = \frac{1}{2}\left(\frac{1}{z_1} + \frac{1}{z_2}\right)\left(1 + \left[1 - 2\frac{\{\rho(\boldsymbol{h}) + 1\}z_1 z_2}{(z_1 + z_2)^2}\right]^{1/2}\right).$$

Example 3. A third max-stable process is referred to as the *Brown–Resnick* model (see, e.g., Kabluchko et al., 2009), for which $W(\boldsymbol{s}) = \exp\{\epsilon(\boldsymbol{s}) - \gamma(\boldsymbol{s})\}$, where $\epsilon(\boldsymbol{s})$ denotes a zero mean Gaussian process with semivariogram $\gamma(\boldsymbol{h})$ and with $\epsilon(\boldsymbol{0}) = 0$ almost surely. In particular, when $\epsilon(\boldsymbol{s})$ is a Brownian process with semivariogram $\gamma(\boldsymbol{h}) = \|\boldsymbol{h}\|^2/2$, the process corresponds to the Smith Gaussian max-stable process model. The closed form for the bivariate distribution function of the Brown–Resnick process has the same expression as that of the Smith model with $\alpha(\boldsymbol{h}) = \{2\gamma(\boldsymbol{h})\}^{1/2}$. Huser and Davison (2013) derived a closed-form expression of the triplewise density function of a Brown–Resnick process.

2.2. Extremal coefficient

The extremal coefficient (see, e.g., Smith, 1990; Schlather and Tawn, 2003) is a commonly used dependence measure to summarize the amount of dependence in a multivariate extreme value distribution. Consider a pair of random variables (Z_1, Z_2) following a bivariate extreme value distribution with standard Fréchet marginal distributions. The pairwise extremal coefficient has the following connection with the bivariate extreme value distribution function:

$$P(Z_1 < z, Z_2 < z) = P(Z_1 < z)^{\eta} = \exp(-\eta/z), \tag{4}$$

$\eta \in [1, 2]$ and as η increases, the amount of dependence decreases. Here, $\eta = 1$ indicates complete dependence, whereas $\eta = 2$ indicates independence between Z_1 and Z_2. Schlather and Tawn (2003) also discussed higher-order extremal coefficients.

For the three parametric max-stable processes listed in Section 2.1, it is straightforward to derive a closed-form expression for the pairwise extremal coefficient between $Z(\boldsymbol{s}_1)$ and $Z(\boldsymbol{s}_2)$ following the bivariate distribution under each model:

- Smith model: $\eta(\boldsymbol{h}) = 2\Phi\{\alpha(\boldsymbol{h})/2\}$;
- Schlather model: $\eta(\boldsymbol{h}) = 1 + \sqrt{\{1 - \rho(\boldsymbol{h})\}/2}$;
- Brown–Resnick model: $\eta(\boldsymbol{h}) = 2\Phi\{\sqrt{\gamma(\boldsymbol{h})/2}\}$.

In particular, Davison et al. (2012) defined the notion of the *practical range*, which is analogous to the *effective range* used in conventional spatial Gaussian process models. Given a stationary isotropic correlation function, $\rho(h)$, for a Gaussian process, the effective range is defined as a distance, h, at which the correlation drops to 0.05; that is, $\rho(h) = 0.05$. For a stationary max-stable process, the practical range is defined as a pair of distances (h_-, h_+) satisfying $\eta(h_-) = 1.3$ and $\eta(h_+) = 1.7$.

3. The tapered composite likelihood inference

3.1. The weighted composite likelihood approach

Both standard Bayesian and frequentist methods are impractical due to the intractability of the full likelihood computation of the max-stable process in all but the lower dimensional cases. Composite likelihood (CL) inference is naturally considered as an appealing alternative to the full but impossible likelihood analysis.

Consider a parametric statistical model with probability density function $\{f(\mathbf{z};\boldsymbol{\theta}), \mathbf{z} \in \mathcal{Z} \subseteq \mathbb{R}^K, \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^q\}$ and a set of marginal or conditional events $\{\mathcal{A}_i : \mathcal{A}_i \subseteq \mathcal{F},\ i \in I \subseteq \mathbb{N}\}$, where $\mathcal{F}$ is some σ-algebra on $\mathcal{Z}$. The log-weighted composite likelihood (WCL) is defined by Lindsay (1988) as

$$\ell_{\mathrm{WCL}}(\boldsymbol{\theta}) = \sum_{i \in I} w_i \log f(\mathbf{z} \in \mathcal{A}_i; \boldsymbol{\theta}) \tag{5}$$

where $f(\mathbf{z} \in \mathcal{A}_i; \boldsymbol{\theta})$ is the likelihood associated with event $\mathcal{A}_i$, and $\{w_i,\ i \in I \subseteq \mathbb{N}\}$ is a set of weights. Then, $\hat{\boldsymbol{\theta}}_{\mathrm{WCL}}$ is called the maximum WCL estimator if it is the global maximum of $\ell_{\mathrm{WCL}}(\boldsymbol{\theta})$.

The definition in (5) allows for combinations of marginal and conditional densities (Cox and Reid, 2004). In this paper, we focus on the composite marginal likelihood inference, in particular, the pairwise and the triplewise case. The pairwise log-weighted composite likelihood is defined as

$$\ell_{\mathrm{WCL}}(\boldsymbol{\theta}) = \sum_{t=1}^{N} \sum_{i \neq j} w_{ij} \log f(\mathbf{z}_i^{(t)}, \mathbf{z}_j^{(t)}; \boldsymbol{\theta}),$$

where $\mathbf{z}_i^{(t)}$ is the sample of the tth replicate at the ith site, for $t = 1, \ldots, N$. Analogously, the triplewise log-weighted composite likelihood is defined as

$$\ell_{\mathrm{WCL}}(\boldsymbol{\theta}) = \sum_{t=1}^{N} \sum_{i \neq j \neq k} w_{ijk} \log f(\mathbf{z}_i^{(t)}, \mathbf{z}_j^{(t)}, \mathbf{z}_k^{(t)}; \boldsymbol{\theta}).$$

The WCL approach is known to yield estimators with sound asymptotic properties. In the case of N independent and identically distributed observations, $\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(N)}$, from the model $f(\mathbf{z};\boldsymbol{\theta})$ on $\mathbb{R}^K$ with K fixed, under the usual regularity conditions, the maximized WCL estimator, $\hat{\boldsymbol{\theta}}_{\mathrm{WCL}}$, is a consistent and unbiased parameter estimator (Lindsay, 1988; Varin and Vidoni, 2005):

$$N^{1/2}(\hat{\boldsymbol{\theta}}_{\mathrm{WCL}} - \boldsymbol{\theta}) \to N_q\{\mathbf{0}, \mathbf{G}^{-1}(\boldsymbol{\theta})\} \tag{6}$$

in distribution as $N \to +\infty$, where $\mathbf{G}(\boldsymbol{\theta}) = \mathbf{H}(\boldsymbol{\theta})\mathbf{J}^{-1}(\boldsymbol{\theta})\mathbf{H}(\boldsymbol{\theta})$ is called the Godambe information (Godambe and Heyde, 2010), $\mathbf{H}(\boldsymbol{\theta}) = E\{-\nabla^2 \ell_{\mathrm{WCL}}(\boldsymbol{\theta})\}$ and $\mathbf{J}(\boldsymbol{\theta}) = \mathrm{var}\{\nabla \ell_{\mathrm{WCL}}(\boldsymbol{\theta})\}$.

3.2. The tapered marginal composite likelihood for max-stable processes

There are many possible choices for the weights, w_i, in the WCL function defined in (5). Equal weights are by far the most popular to obtain composite likelihood estimators. One issue with equally weighted CL is that the number of pairs can be enormous when the number of observations is large, making it computationally very expensive. Moreover, several recent studies (see, e.g., Varin and Vidoni, 2005; Davis and Yau, 2011; Bevilacqua et al., 2012) have shown that further efficiency gains can be obtained by selecting appropriate unequal weights for WCL inferences. For instance, Bevilacqua et al. (2012) proposed a pairwise WCL method for the inference of a spatial Gaussian process model. They showed that setting dummy weights to exclude distant pairs for spatially correlated observations may improve the efficiency over the pairwise equally weighted CL. Varin and Vidoni (2005) investigated pairwise likelihood for generalized linear models and also suggested improved statistical efficiency by excluding pairs formed by distant observations.

Following these findings, we propose a tapered composite likelihood (TCL) approach, that aims to improve the efficiency of the estimates and to reduce the computational burden. For spatial max-stable processes, the weights of the TCL are set to depend on the distance between sites. Let γ_s denote the *taper range* in space. Let $D_2(\gamma_s)$ be a set containing all pairs of observations within γ_s units of space lag to exclude distant pairs; that is,

$$D_2(\gamma_s) = \{(i, j);\ \|\boldsymbol{s}_i - \boldsymbol{s}_j\| \leq \gamma_s\}. \tag{7}$$

Given γ_s and D_2, the pairwise weight function is defined as a thresholding/tapering function, $w_{ij} = 1$ if $(i, j) \in D_2(\gamma_s)$, $w_{ij} = 0$ otherwise. Most existing CL approaches to extreme values are based on pairs of observations (see, e.g., Padoan et al., 2010; Ribatet et al., 2012; Blanchet and Davison, 2011). For the Smith Gaussian extreme value process model and the Brown–Resnick process model, one can construct a triplewise TCL function following the trivariate distribution functions derived in Genton et al. (2011) and Huser and Davison (2013). Similarly to the pairwise case, the triplewise weight function is defined as $w_{ijk} = 1$ if $(i, j, k) \in D_3(\gamma_s)$, $w_{ijk} = 0$ otherwise, where

$$D_3(\gamma_s) = \left\{(i, j, k);\ \max_{i_1, i_2 \in (i,j,k)} \|\boldsymbol{s}_{i_1} - \boldsymbol{s}_{i_2}\| \leq \gamma_s\right\}. \tag{8}$$

The above tapering strategy clearly involves the selection of a taper range, γ_s. Different spatial dependence structures such as decay and smoothness may require different weights to achieve optimal efficiency. In particular, we note that for an anisotropic Smith model, the extremal coefficient is determined by the Mahalanobis distance, $\sqrt{(\boldsymbol{s}_1 - \boldsymbol{s}_2)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{s}_1 - \boldsymbol{s}_2)}$, controlled by $\boldsymbol{\Sigma}$ instead of the Euclidean distance. Therefore, it is plausible to exclude distant pairs or triples based on the Mahalanobis distance. However, this requires knowing the values of $\boldsymbol{\Sigma}$. One way to determine $\boldsymbol{\Sigma}$ is to estimate it from a pilot study using all pairs or triples.

3.3. Choice of the taper range

Since the maximum TCL estimator is unbiased under certain regularity conditions, various measures of the covariance matrix of estimates become natural scalar choices to compare the statistical efficiency. In this paper, we consider two commonly used measures, trace and determinant, to select the taper range. In particular, the determinant of the covariance matrix, also called the generalized variance, is often used as a measure of overall spread in a distribution (Mardia et al., 1979), while the trace of the covariance matrix, also called the total variation, equals the summation of the variances of the parameter estimates.

One way to estimate the covariance matrix of the maximum TCL estimates is to use the inverse Godambe information matrix based on the asymptotic results in (6). The Godambe information depends on γ_s in the spatial context. Below, we write $\mathbf{G}(\boldsymbol{\theta};\ \gamma_s)$. We seek the optimal taper range, γ_s^*, such that the trace or the determinant of the inverse of the Godambe information matrix is minimized; that is, $\gamma_s^* = \arg\min_{\gamma_s \in \mathbb{R}^+} \mathrm{tr}\,\{\mathbf{G}^{-1}(\boldsymbol{\theta};\ \gamma_s)\}$ or $\gamma_s^* = \arg\min_{\gamma_s \in \mathbb{R}^+} \det\{\mathbf{G}^{-1}(\boldsymbol{\theta};\ \gamma_s)\}$.

Calculations of the Godambe information matrix require the estimations of $\mathbf{H}$ and $\mathbf{J}$ in (6). For a given taper range, the Hessian matrix, $\mathbf{H}$, is usually estimated from the numerical maximization algorithm, whereas $\mathbf{J}$ is usually estimated from the variance of the observed score function. For the triplewise Smith model, we give explicit expressions of $\mathbf{J}$ in Appendix B. The evaluation of $\mathbf{J}$ requires the values of model parameters to be estimated under the assumed model.

One may also use the Jackknife method to estimate the covariance matrix of the maximum TCL estimates (Lipsitz et al., 1994). When N is not sufficiently large compared with the dimension of $\boldsymbol{\theta}$, this method may provide a more robust estimate of the covariance matrix than the asymptotic Godambe information matrix method provides. This method is also preferred in situations where the analytical derivatives of the score functions are difficult or tedious to derive. The Jackknife covariance matrix is given by $\mathrm{var}\{\hat{\boldsymbol{\theta}}_{\mathrm{TCL}}(\gamma_s)\} = \frac{N-1}{N} \sum_{t=1}^{N} \{\hat{\boldsymbol{\theta}}_{\mathrm{TCL}}^{(-t)}(\gamma_s) - \hat{\boldsymbol{\theta}}_{\mathrm{TCL}}(\gamma_s)\}\{\hat{\boldsymbol{\theta}}_{\mathrm{TCL}}^{(-t)}(\gamma_s) - \hat{\boldsymbol{\theta}}_{\mathrm{TCL}}(\gamma_s)\}^{\mathrm{T}}$, where $\hat{\boldsymbol{\theta}}_{\mathrm{TCL}}^{(-t)}(\gamma_s)$ is the TCL estimator of $\boldsymbol{\theta}$ excluding the t-th sample. Again, by minimizing the trace or the determinant of $\mathrm{var}\{\hat{\boldsymbol{\theta}}_{\mathrm{TCL}}(\gamma_s)\}$, we obtain the optimal taper range.

The taper range selection procedure can be computationally very demanding for large datasets in which likelihood functions have to be evaluated for very large numbers of pairs or triples. A computationally efficient alternative is to consider a subset of the original data. To further save computation time, instead of evaluating the TCL at every observed space lag, we can restrict the search space to a fixed set of space lags. Indeed, the simulation results presented in Section 4 confirm the utility of these two strategies to reduce the computational cost.

4. Simulation study

We design a number of simulation studies to investigate the use of the TCL method for inference on the parameters of the three max-stable process models listed in Section 2.1, including the Smith model, the Schlather model and the Brown–Resnick model.

4.1. Simulation 1: only the range parameter is unknown

To investigate the association between the optimal taper range and the spatial range parameter, we first consider only isotropic max-stable process models in which only the range parameters are unknown while the other parameters, e.g., the smoothness, are fixed. Specifically, we simulate from: an isotropic Smith Gaussian extreme value process model where $\mathbf{\Sigma}$ is a diagonal matrix ($\sigma_{12} = 0$) with $\sigma_{11}^2 = \sigma_{22}^2 = \phi^2$, i.e., $\alpha(\boldsymbol{h}) = \|\boldsymbol{h}\|/\phi$; a Schlather model with an exponential correlation function with range parameter ϕ, i.e., $\rho(\boldsymbol{h}) = \exp(-\|\boldsymbol{h}\|/\phi)$; and a Brown–Resnick model with variogram $2\gamma(\boldsymbol{h}) = \|\boldsymbol{h}\|/\phi$. For each of 100 simulation replicates, we randomly generate $K = 100$ locations uniformly in the square $[0, 100] \times [0, 100]$. We then simulate $N = 30$ max-stable process realizations under each model at the sampled K locations using the SpatialExtremes package in R (Ribatet, 2011).

We compare the TCL estimators and the equally weighted CL estimators in terms of their mean squared errors (MSE). Fig. 1 shows the relative efficiency between the TCL estimators and the equally weighted CL estimators (denoted as Complete) as a function of the range parameter ϕ for each max-stable process model. The points marked by circles in Fig. 1 show the results in which the optimal taper ranges are selected by minimizing the estimated variances of the estimates of ϕ, denoted as MSE(complete)/MSE(taper,VAR). The relative efficiencies are mostly greater than 1, suggesting that the efficiency gain by using TCL is significant. Interestingly, we observe a negative association between the relative efficiency and the spatial range for the Schlather model. But for the Smith model, our results suggest a positive association between the relative efficiency and the spatial range, indicating that parameter estimations for the Schlather model with a shorter spatial range might benefit more from the TCL inference. For the Brown–Resnick model, the relative efficiency first increases and then decreases as the spatial range increases. It is noticeable that in the case of small scale/range spatial dependence, estimations of the range parameter are relatively poor for both the equally weighted CL and the TCL. This might be due to the sampling interval being not fine enough to discover the actual extremal dependence scale. But the TCL still shows relatively better performance than the equally weighted CL in this scenario.

Ideally one may choose an optimal taper range by minimizing the squared error of the estimate of ϕ if a true parameter value is given. The points marked by crosses in Fig. 1 show the relative efficiencies using this tapering strategy for the TCL, denoted as MSE(complete)/MSE(taper,SE). Clearly this is infeasible in real data analysis. Here we use this method as a benchmark to assess the performance of the TCL by minimizing the estimated variance of the estimates. In general, the functional pattern of MSE(complete)/MSE(taper,VAR) versus the spatial range is similar to that of MSE(complete)/MSE(taper,SE). It is noticeable that MSE(taper,VAR) is larger than MSE(taper,SE) (roughly doubled) leading to less accurate parameter estimations. Although it is expected, this result suggests that there is room to improve the efficiency of the TCL estimate by considering alternative criteria for taper range selection that are closer to the squared error than the estimated variance of the estimate.

Fig. 2 shows the averaged optimal taper range as a function of the range parameter ϕ under each max-stable process model over the 100 simulations. The upper panel shows the results in which the

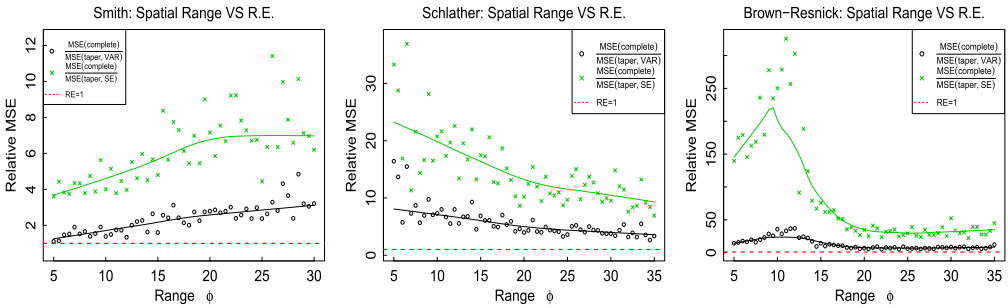

Fig. 1. Scatter plots of the relative efficiency (R.E.) between the TCL estimator and the equally weighted CL estimator as a function of the range parameter ϕ under each model overlaid with a lowess fit curve. Points marked by circles and crosses show the results in which the optimal taper ranges of the TCL estimators are selected by minimizing the variances (VAR) and the squared errors (SE) of the parameter estimate, respectively. The straight line in each scatter plot corresponds to R.E. $= 1$.

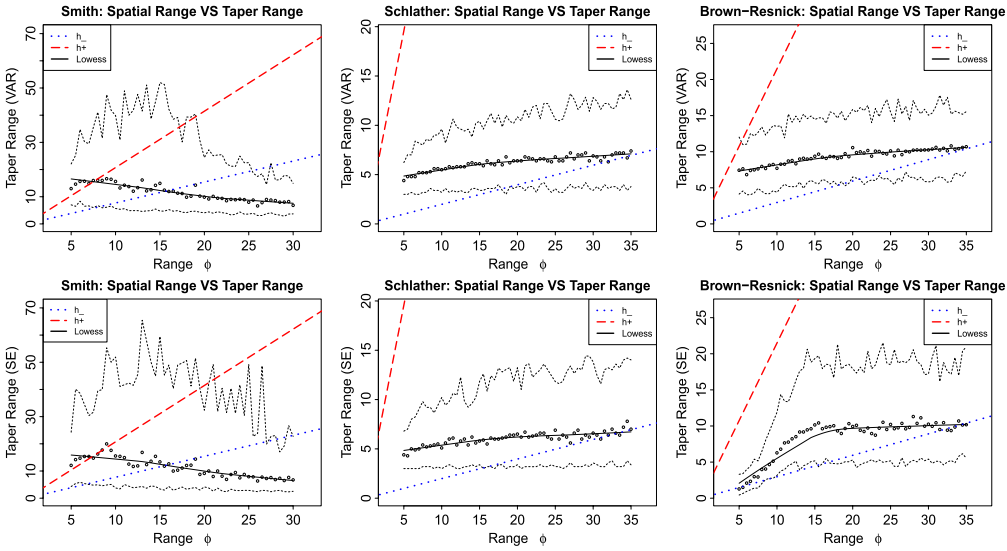

Fig. 2. Scatter plots of the median (marked by circles) and the 90% interval (in dashed curves) of the optimal taper range as a function of the range parameter ϕ overlaid with a lowess fit curve (in solid curve). The two straight lines are the practical spatial ranges h_- and h_+ as functions of the range parameter, respectively. The upper panel and lower panels show the scatter plots in which the optimal taper ranges are selected by minimizing the variances (VAR) and the squared errors (SE) of the parameter estimates under each max-stable process model, respectively.

optimal taper ranges are selected by minimizing the estimated variances of the estimates of ϕ, and the lower panel shows the results in which the optimal taper ranges are selected by minimizing the squared errors. Again, the functional patterns of the optimal taper range versus the spatial range are similar for these two tapering approaches. Specifically, for the Smith model, it seems that the longer the spatial range, the shorter the taper range is needed. In contrast, for the Schlather and the Brown–Resnick models, longer taper ranges are generally required for models with longer spatial dependence ranges. The optimal taper ranges are typically less than the h_+ for all three models. In particular, optimal taper ranges are typically greater than h_- for the Schlather model and the Brown–Resnick model. But for the Smith model, optimal taper ranges are typically less than h_- when the spatial ranges are relatively long.

We now examine the accuracy of the Godambe information-based standard error estimator. We estimate the true standard error of the TCL estimator by calculating the empirical standard error based

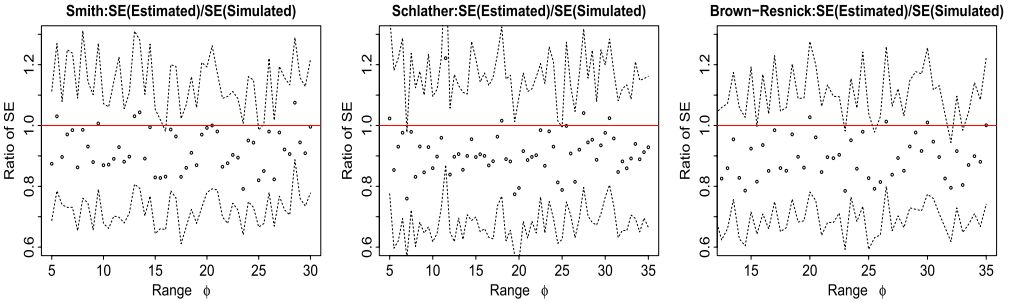

Fig. 3. Scatter plots of the mean (marked by circles) and the 90% interval (in dashed curves) of the ratio between the estimated standard error and the simulated standard error of the TCL estimator as a function of spatial range. The straight line in red corresponds to ratio of standard errors being 1.

on 100 simulated datasets. For each iteration, we calculated standard error estimates for the TCL estimator based on the Godambe information matrix by plugging in the corresponding TCL estimates from that iteration. Fig. 3 shows the ratio between the estimated standard error and the simulated standard error of the TCL estimator as a function of spatial range under each of the max-stable process models. Overall, although the mean ratios of standard errors are fairly close to 1, we do observe slight downward bias of using the Godambe information-based standard errors for all the three models. Indeed, this is not surprising, because the weighting function of the TCL is constructed by choosing an optimal taper range, which is essentially also calculated from the observed spatial extreme values. Therefore, the asymptotic distribution given in (6) may not be a valid result for the TCL estimator.

4.2. Simulation 2: all parameters are unknown

In the second simulation study, for each of the 100 simulation replicates, we fix the sampling locations at the 35 rain gauge stations in the real dataset that is analyzed in Section 5. We also experiment with the same sample size as the real data in Section 5, i.e., $N = 47$. For each max-stable process model, we obtain the equally weighted CL estimates and the two TCL estimates, in which the optimal taper range is selected by minimizing the trace (denoted as T.Trace) and the determinant (denoted as T.Det) of the associated estimated covariance matrix, respectively. We also include the results of the TCL method by minimizing the sum of the squared errors (denoted as T.SSE) of the estimates as a benchmark for comparison purposes.

For the Smith model, we consider four sets of parameters. The true parameter values are given in Table 1. The first three parameter sets correspond to middle, small and large anisotropic dependence structures. In particular, the first parameter set is chosen from the estimated covariance parameters from the real data. The last parameter setting corresponds to a strong anisotropic dependence structure which will allow us to compare the TCL estimators by excluding distant pairs based on the Euclidean distance with those based on the Mahalanobis distance. For the Smith model, the TCL estimators outperform the equally weighted CL estimators. The standard deviations of the TCL estimates are smaller than those using all pairs. The TCL also seems to reduce the biases in the estimations of the model parameters. We have observed that the longer the practical range of the Smith max-stable process, the shorter the optimal taper range. In the case of strong anisotropic dependence, the use of the Mahalanobis distance for tapering achieves smaller MSE, leading to a higher efficiency gain over the equally weighted CL, than the Euclidean distance counterpart. Moreover, it seems that the former results in fewer number of pairs selected for the TCL. We have also experimented with the triplewise TCL for the first parameter setting. Consistent with the findings in Genton et al. (2011), the triplewise equally weighted CL achieves much more accurate parameter estimations with drastically smaller standard errors than does the pairwise equally weighted CL. Efficiency is significantly further improved with the use of the triplewise TCL.

For the Schlather model, we present the results in Table 2 using a Matérn correlation function (Stein, 1999) with range parameter ϕ and smoothness parameter κ for the latent scaled Gaussian

Table 1
Smith: Results of the maximum TCL estimates of $\mathbf{\Sigma}$: mean and standard deviations (in parentheses). The column "Range" gives the median and the 90% interval (in parentheses) of the selected taper range. The median and the 90% interval (in parentheses) of the number of selected TCL objects (pairs or triples) for each sample are given in the column "CL #". The column "$\mathrm{RE}_{\sigma_{11}^2}/\mathrm{RE}_{\sigma_{12}}/\mathrm{RE}_{\sigma_{22}^2}$" gives the relative efficiencies between the TCL estimators and the equally weighted CL estimators of $(\sigma_{11}^2, \sigma_{12}, \sigma_{22}^2)$. The column "Complete" gives the results of the equally weighted pairwise CL. The columns "T.Trace", "T.Det" and "T.SSE" give the results of the pairwise TCL by minimizing the trace, the determinant and the sum of the squared errors, respectively. "T.MahD" and "T.EucD" give the results of the pairwise TCL by tapering the Mahalanobis distance and the Euclidean distance, respectively, to minimize the trace. "C.Tri" gives the results of the triplewise equally weighted CL and "T.Tri" gives the results of the triplewise TCL with tapering. h_- and h_+ are the practical distances at which the extremal coefficient is 1.3 and 1.7, respectively, under the true models.

	Taper range	CL #	σ_{11}^2	σ_{12}	σ_{22}^2	$\mathrm{RE}_{\sigma_{11}^2}/\mathrm{RE}_{\sigma_{12}}/\mathrm{RE}_{\sigma_{22}^2}$
True	$h_- = [11, 15]$, $h_+ = [30, 39]$		243	62	320	
Complete	81.78	595	245.8(36.6)	60.9(32.5)	337.6(64.7)	1.00/1.00/1.00
T.Trace	14.73(12.00, 18.06)	70(42, 91)	243.0(26.4)	58.09(21.6)	329.5(33.1)	1.94/2.19/3.79
T.Det	13.91(11.28, 16.95)	60(35, 81)	242.1(26.3)	59.0(21.3)	330.5(32.8)	1.94/2.28/3.78
T.SSE	14.00(9.23, 19.66)	62(21, 103)	242.0(22.6)	58.5(17.4)	328.2(24.0)	2.63/3.34/6.96
C.Tri	81.78	6545	243.1(0.9)	62.1(1.3)	319.6(1.3)	1.00/1.00/1.00
T.Tri	42.00	1864	243.1(0.5)	62.0(0.6)	320.1(0.7)	3.48/5.17/3.44
True	$h_- = [8, 10]$, $h_+ = [21, 28]$		121.5	31.0	160.0	
Complete	81.78	595	123.9(13.0)	32.5(13.4)	165.9(21.9)	1.00/1.00/1.00
T.Trace	15.23(13.88, 18.06)	74(60, 91)	123.5(10.3)	31.3(8.7)	161.0(14.2)	1.60/2.43/2.54
T.Det	14.73(13.17, 18.06)	70(53, 91)	123.6(10.1)	31.3(8.4)	160.8(14.6)	1.64/2.58/2.41
T.SSE	15.58(11.25, 22.66)	76(35, 137)	122.5(8.7)	31.2(7.5)	160.9(10.4)	2.29/3.26/4.74
True	$h_- = [16, 20]$, $h_+ = [42, 55]$		486	124	640	
Complete	81.78	595	496.0(87.5)	129.8(80.2)	670.3(140.2)	1.00/1.00/1.00
T.Trace	13.17(10.24, 15.98)	53(28, 77)	489.4(66.3)	121.7(50.0)	642.9(76.1)	1.76/2.58/3.55
T.Det	13.17(10.64, 17.39)	53(32, 84)	489.8(67.0)	123.8(50.1)	644.9(75.3)	1.72/2.58/3.61
T.SSE	12.71(7.64, 16.00)	49(11, 77)	489.2(58.8)	122.06(41.58)	642.5(59.2)	2.24/3.73/5.87
True	$h_- = [8, 20]$, $h_+ = [22, 53]$		640	62	121	
Complete	81.78	595	649.9(90.5)	67.1(26.8)	127.2(20.6)	1.00/1.00/1.00
T.MahD	0.88(0.67, 1.20)	68(34, 118)	644.8(71.3)	63.6(17.3)	124.57(11.63)	1.62/2.46/3.12
T.EucD	17.18(13.34, 24.23)	84(55, 155)	640.1(74.5)	63.5(18.6)	124.74(12.04)	1.49/2.15/2.90

process (see Section 2.1). Again, the first parameter set is chosen to be close to the estimated covariance parameters from the real data. The other parameter settings consist of various combinations of range and smoothness, in which the first three are chosen such that they have a common h_+. We observe significant efficiency gains using the TCL method. Consistent with the findings of the first simulation with fixed smoothness, efficiency gains are stronger and the corresponding optimal taper ranges are shorter for the Schlather models with shorter ranges of dependence or smaller smoothness parameters. It is also noticeable that the efficiency gain for the range parameter is generally larger than that for the smoothness parameter. In fact, both equally weighted CL estimates and the TCL estimates for the smoothness tend to have serious upward biases for relatively large values of smoothness. Indeed, higher orders of smoothness are typically difficult to identify from real data.

For the Brown–Resnick model, we use the variogram function $2\gamma(\boldsymbol{h}) = (\|\boldsymbol{h}\|/\phi)^\kappa$ (see Section 2.1), where ϕ is the range parameter and κ is the smoothness parameter. The parameter sets consist of five combinations of range and smoothness. Table 3 shows that the TCL leads to more accurate parameter estimates with smaller biases and standard deviations compared with the equally weighted counterparts. And we notice that the benefit of using TCL is more prominent for smoother Brown–Resnick processes. But overall the efficiency gains from using the TCL for the Brown–Resnick model are not as significant as for the Smith model and the Schlather model. The selected taper range for the Brown–Resnick model is typically the longest among all three max-stable process models. The use of the determinant of the estimated inverse Godambe information matrix for taper range selection slightly improves efficiency over the method using the trace. This also seems to be true for

Table 2
Schlather: Results of the TCL estimates for the simulated Schlather model. See Table 1 for the descriptions of the rows and the columns.

	Taper range	CL #	ϕ	κ	$\mathrm{RE}_\phi/\mathrm{RE}_\kappa$
True	$h_- = 6, h_+ = 148$		39.90	0.43	
Complete	81.78	595	43.99(18.92)	0.45(1.27e−01)	1.00/1.00
T.Trace	21.34(10.36, 54.15)	122(30, 460)	40.00(10.42)	0.53(9.80e−02)	1.99/0.84
T.Det	27.12(15.81, 40.10)	178(76, 332)	36.45(12.77)	0.48(9.79e−02)	2.13/1.33
T.SSE	24.09(12.85, 40.51)	153(51, 336)	36.92(6.66)	0.45(6.08e−02)	7.03/3.83
True	$h_- = 15, h_+ = 148$		29.50	1	
Complete	81.78	595	31.10(11.01)	1.09(3.43e−01)	1.00/1.00
T.Trace	21.93(10.36, 34.52)	127(30, 275)	25.37(8.78)	2.26(7.68e+00)	1.31/2.06e−03
T.Det	24.62(15.81, 32.58)	153(76, 250)	30.91(7.83)	2.05(8.52e+00)	1.95/1.69e−03
T.SSE	22.06(11.86, 38.24)	132(40, 317)	29.31(3.54)	1.02(9.32e−02)	9.85/13.55
True	$h_- = 20, h_+ = 148$		25.40	1.5	
Complete	81.78	595	27.15(14.22)	5.30(1.62e+01)	1.00/1.00
T.Trace	20.25(16.93, 62.52)	112(81, 525)	21.16(9.12)	6.16(1.66e+01)	2.02/0.93
T.Det	22.06(16.93, 62.09)	132(81, 520)	25.19(11.10)	5.93(1.67e+01)	1.67/0.93
T.SSE	18.07(12.85, 34.02)	91(51, 267)	24.61(4.35)	3.49(1.37e+01)	10.48/1.45
True	$h_- = 3, h_+ = 74$		19.90	0.43	
Complete	81.78	595	24.55(17.42)	0.47(1.87e−01)	1.00/1.00
T.Trace	20.80(16.93, 35.96)	117(81, 292)	18.14(6.49)	0.51(1.57e−01)	7.19/1.18
T.Det	23.00(16.93, 31.08)	142(81, 224)	19.33(6.69)	0.48(1.22e−01)	7.22/2.10
T.SSE	21.34(13.88, 29.80)	122(60, 209)	19.56(4.82)	0.45(7.40e−02)	13.96/6.13
True	$h_- = 12, h_+ = 296$		79.80	0.43	
Complete	81.78	595	85.72(30.53)	0.44(7.21e−02)	1.00/1.00
T.Trace	22.69(16.93, 57.56)	137(81, 486)	67.39(19.27)	0.48(6.22e−02)	1.84/0.82
T.Det	27.34(17.61, 42.82)	183(86, 368)	71.45(20.02)	0.47(5.79e−02)	2.05/1.06
T.SSE	27.34(15.76, 53.20)	183(76, 455)	74.33(11.31)	0.45(2.77e−02)	6.12/4.71

the other two max-stable process models. The triplewise equally weighted CL leads to slightly more efficient parameter estimations than the equally weighted pairwise CL. But the efficiency gain from using the higher-order CL is not as prominent as from the Smith model, which has a much smoother max-stable field. With the taper range fixed at the same range chosen for the tapered pairwise CL, we observe further efficiency gains by using the tapered triplewise CL.

Finally, we assess the effect of sample size on the selection of the optimal taper range and the performance of the corresponding TCL estimators. For each of the 100 simulation replicates with $K = 35$ and $N = 47$, we select the taper range by minimizing the trace and the determinant of the estimated covariance matrix of the TCL estimators based on a subset of the data. Specifically, we experiment with $m = 10$ and $m = 20$ subsamples. The selected taper range is then used to fit the complete data ($N = 47$) via the TCL approach. In Table 4, we compare this sub-sampling method for taper range selection with the TCL method in which the taper range is selected based on all samples ($m = N = 47$). In general, the optimal taper range selected based on the sub-sampling method is very close to the optimal taper range selected based on the full dataset. Consequently, the former and the latter produce comparable efficiency gains over the pairwise equally weighted CL method. As expected, the larger the sample size of the subset, the closer the efficiency gains of the two methods. For the Smith and the Brown–Resnick model, it seems that the selected taper range achieves satisfactory efficiency gains based only on $m = 10$ samples, while the Schlather model requires a larger sample size ($m = 20$) to obtain a satisfactory taper range.

5. Precipitation data

We illustrate our method using summer (June to August) maximum daily rainfall data at 35 rain gauging stations located in the north of the Alps and east of the Jura mountains in Switzerland. The

Table 3
Brown–Resnick: Results of the TCL estimates for the simulated Brown–Resnick model. See Table 1 for the descriptions of the rows and the columns.

	Taper range	CL #	ϕ	κ	$\mathrm{RE}_\phi/\mathrm{RE}_\kappa$
True	$h_- = 5.94, h_+ = 83$		30.00	0.75	
Complete	81.78	595	29.95(4.04)	0.74(8.95e−02)	1.00/1.00
T.Trace	31.49(15.08, 81.78)	232(72, 595)	27.81(3.38)	0.80(7.68e−02)	1.01/0.93
T.Det	31.84(21.43, 45.70)	239(123, 396)	28.93(3.57)	0.77(6.74e−02)	1.18/1.64
T.SSE	31.20(17.61, 81.78)	227(86, 595)	29.34(1.97)	0.75(4.85e−02)	3.79/3.43
C.Tri	81.78	6545	29.57(3.73)	0.75(8.62e−2)	1.00/1.00
T.Tri	31	654	29.43(3.56)	0.75(6.08e−2)	1.09/2.01
True	$h_- = 11.49, h_+ = 83$		38.70	1.0	
Complete	81.78	595	39.87(5.44)	0.99(9.40e−02)	1.00/1.00
T.Trace	28.64(12.83, 69.16)	204(51, 566)	36.97(4.86)	1.05(8.43e−02)	1.16/0.96
T.Det	28.64(16.93, 41.53)	204(81, 352)	38.19(4.66)	1.02(7.35e−02)	1.41/1.55
T.SSE	27.12(13.93, 81.78)	178(61, 595)	38.65(3.00)	1.00(3.96e−02)	3.43/5.65
True	$h_- = 22.25, h_+ = 83$		50.00	1.50	
Complete	81.78	595	51.15(5.85)	1.49(1.01e−01)	1.00/1.00
T.Trace	30.89(17.58, 41.21)	219(86, 346)	49.14(4.75)	1.53(7.38e−02)	1.52/1.72
T.Det	25.44(18.07, 34.15)	163(91, 270)	49.62(4.69)	1.52(7.13e−02)	1.60/1.96
T.SSE	27.62(15.81, 64.25)	188(76, 535)	49.97(2.82)	1.50(3.57e−02)	4.45/8.20
True	$h_- = 22.25, h_+ = 83$		30.00	1	
Complete	81.78	595	30.89(3.44)	0.98(9.14e−02)	1.00/1.00
T.Trace	26.79(13.93, 69.02)	176(61, 565)	28.72(2.79)	1.05(6.37e−02)	1.34/1.38
T.Det	27.48(18.01, 43.21)	186(90, 372)	29.65(2.73)	1.02(5.89e−02)	1.66/2.36
T.SSE	27.23(15.76, 81.78)	180(76, 595)	29.81(1.59)	1.00(4.33e−02)	4.90/4.66
True	$h_- = 22.25, h_+ = 83$		30.00	1.5	
Complete	81.78	595	30.90(2.61)	1.44(9.88e−02)	1.00/1.00
T.Trace	25.95(14.84, 41.19)	168(71, 346)	29.58(2.38)	1.51(7.75e−02)	1.31/2.13
T.Det	25.44(17.58, 33.14)	163(86, 255)	29.98(2.23)	1.49(6.79e−02)	1.54/2.77
T.SSE	23.00(14.40, 57.51)	142(66, 485)	29.97(1.18)	1.48(4.10e−02)	5.46/6.76

largest and smallest distances between these 35 stations are around 82 km and 4 km, respectively. The maxima rainfall data, which were used by Davison et al. (2012), were derived from the daily precipitation data from 1962 to 2008 provided by the national meteorological service, MétéoSuisse.

The major aim of our analysis is to study the dependence structures in GEV residuals by fitting various max-stable process models via the TCL method. Therefore, for simplicity we first transform the data at each location to a unit Fréchet distribution. We assume that the shape parameter is constant over space and we also assume independence across sites to obtain maximum likelihood estimators for the marginal parameters.

Next, we estimate the extreme dependence parameters under each simple max-stable process model by maximizing the equally weighted CL and the TCL. Standard deviations and covariances of the maximum CL estimates are calculated from the inverse Godambe information matrix, with the Hessian matrix, **H**, estimated from the numerical maximization algorithm, and **J** estimated from the variance of the observed score function.

Table 5 summarizes the fitted results under each max-stable process model. With the Smith model, we observe notable differences between the equally weighted pairwise CL estimates and the TCL estimates. The TCL leads to substantially smaller standard errors of the estimates. It also gives comparable estimates for h_- with those given by the Schlather model and the Brown–Resnick model. In contrast, the equally weighted CL gives much larger estimates of h_-. We note that the estimation of h_+ differs considerably among the three max-stable process models. This agrees with previous findings that the extremal correlation function may have a large impact on the estimation of h_+ (Davison et al., 2012).

Table 4
Summary of the max-stable models fitted to the simulated data. The results in row $m = \cdot$ correspond to the TCL in which the optimal taper range is selected by minimizing the determinant of the estimated covariance matrix of the TCL estimators based on a subset of data of size m. The column "R.E." reports the global relative efficiency (R.E.) between the TCL estimators and the equally weighted CL estimators.

Smith						
	Range	CL #	σ_{11}^2	σ_{12}	σ_{22}^2	R.E.
True	$(h_- = [11, 15], h_+ = [30, 39])$		243	62	320	
Complete	81.78	595	245.78(36.59)	60.89(32.48)	337.58(64.60)	1.00
$m = 10$	14.62(10.24, 26.96)	68(28, 180)	242.96(26.05)	58.83(23.36)	330.45(37.28)	2.52
$m = 20$	14.62(11.28, 18.06)	68(35, 91)	242.89(26.52)	59.13(21.43)	331.42(34.04)	2.80
$m = N = 47$	13.91(11.28, 16.95)	60(35, 81)	242.05(26.31)	58.96(21.30)	330.35(32.84)	2.94
Schlather						
	Range	CL #	ϕ	κ	R.E.	
True	$(h_- = 6, h_+ = 148)$		39.90	0.43		
Complete	81.78	595	43.99(18.92)	0.45(1.27e−01)	1.00	
$m = 10$	23.00(13.36, 50.32)	142(56, 439)	41.10(16.89)	0.45(8.41e−02)	1.31	
$m = 20$	26.47(14.82, 50.22)	173(71, 438)	38.36(12.16)	0.46(8.85e−02)	2.49	
$m = N = 47$	27.12(15.81, 40.10)	178(76, 332)	36.45(12.81)	0.48(9.79e−02)	2.13	
Brown–Resnick						
	Range	CL #	ϕ	κ	R.E.	
True	$(h_- = 6, h_+ = 83)$		30.00	0.75		
Complete	81.78	595	29.95(4.04)	0.74(8.95e−02)	1.00	
$m = 10$	31.63(19.50, 62.71)	234(102, 526)	29.45(3.59)	0.76(6.87e−02)	1.19	
$m = 20$	31.05(20.62, 57.18)	224(117, 480)	29.20(3.69)	0.77(7.21e−02)	1.10	
$m = N = 47$	31.84(21.43, 45.70)	239(123, 396)	28.93(3.57)	0.77(6.74e−02)	1.18	

The standard errors of the TCL for the Schlather and the Brown–Resnick models with fixed smoothness are smaller than those of the equally weighted counterpart, leading to efficiency gains with much fewer number of pairs included for the TCL inference. When both the range and the smoothness are unknown, the efficiency gains from using the TCL for the Schlather and the Brown–Resnick models are in general not as large as for the Smith model.

The selected taper ranges of these two models are considerably larger than those of the Smith model. The reason is that there might be some identifiability issues when jointly estimating the range and the smoothness (Zhang, 2004). Indeed, the estimates of these two parameters are typically strongly related. It is known that the determinant is a more suitable scalar measure for the covariance matrix of two highly correlated estimates. That also explains why the parameter estimates using the determinant have slightly smaller standard errors than those using the trace. We have also observed that the selected taper ranges are typically in between the empirical practical ranges of (h_-, h_+) that can be roughly estimated from the empirical extremal coefficient plot shown in Fig. 4.

Finally, we propose a model selection criterion based on the TCL to compare the three types of max-stable process models assuming that all parameters are unknown. Specifically, for a fixed taper range, γ_s, we calculate the tapered composite likelihood information criterion (TCLIC) (Takeuchi, 1976; Varin and Vidoni, 2005) for each model defined as $\text{TCLIC} = -2[\ell_{\text{TCL}}(\hat{\boldsymbol{\theta}}; \gamma_s) - \text{tr}\{\hat{\mathbf{J}}(\hat{\boldsymbol{\theta}})\hat{\mathbf{H}}(\hat{\boldsymbol{\theta}})^{-1}\}]$, where the second term is the composite log-likelihood penalty term for the effective number of parameters. Since optimal taper ranges differ across the three max-stable process models, we calculate the TCLIC scores for all three models under each taper range. Table 6 summarizes the TCLIC score for the Smith, the Schlather and the Brown–Resnick models at three different taper ranges, each of which minimizes the determinant of the inverse Godambe information under each model, respectively. We also include the results of the TCLIC in which equally weighted CL inference is used for each model. According to the TCLIC criterion, the Brown–Resnick model is superior to the other two models across different

Table 5
Maximum TCL estimates for the precipitation data. The column "R.Var" shows the ratio of the trace of the estimated covariance matrix of the estimates under each TCL method to that under the equally weighted CL method. The column "R.Det" shows the ratio of the determinant of the estimated covariance matrix of the estimates under each TCL method to that under the equally weighted CL method. Here h_- and h_+ are the practical distances at which the extremal coefficient is 1.3 and 1.7, respectively. Taper range and (h_-, h_+) are in kilometers. See Table 1 for the descriptions of other rows and columns.

Smith									
	Taper range	CL #	h_-	h_+	σ_{11}^2	σ_{12}	σ_{22}^2	R.Var	R.Det
Complete	81.78	595	12.61	33.91	267.62(4.95)	0[a]	$\sigma_{22}^2 = \sigma_{11}^2$	1.00	1.00
Taper	11.41	38	6.45	17.36	70.15(2.68)	0[a]	$\sigma_{22}^2 = \sigma_{11}^2$	3.40	3.40
Complete	81.78	595	[11.1, 14.5]	[29.9, 39.0]	242.58(4.35)	62.48(4.55)	319.93(16.48)	1.00	1.00
T.Trace	14.06	62	[6.6, 7.9]	[17.9, 21.3]	94.62(3.73)	14.70(2.37)	84.88(6.60)	4.93	30.69
T.Det	14.06	62	[6.6, 7.9]	[17.9, 21.3]	94.62(3.73)	14.70(2.37)	84.88(6.60)	4.93	30.69
Schlather									
	Taper range	CL #	h_-	h_+	ϕ	κ		R.Var	R.Det
Complete	81.78	595	6.83	134.7	34.42(5.27)	0.5[a]		1.00	1.00
Taper	13.35	55	5.49	108.2	27.67(2.82)	0.5[a]		3.49	3.49
Complete	81.78	595	5.94	147.9	39.87(16.91)	0.43(1.15e−01)		1.00	1.00
T.Trace	74.05	583	5.98	148.9	40.13(16.80)	0.43(1.12e−01)		1.01	1.05
T.Det	57.71	487	5.65	167.0	46.59(18.48)	0.39(8.69e−02)		0.84	1.15
Brown–Resnick									
	Taper range	CL #	h_-	h_+	ϕ	κ		R.Var	R.Det
Complete	81.78	595	9.12	66.0	30.73(2.02)	1[a]		1.00	1.00
Taper	13.17	53	6.10	44.1	20.54(0.99)	1[a]		4.16	4.16
Complete	81.78	595	6.30	88.2	31.81(3.67)	0.75(6.11e−02)		1.00	1.00
T.Trace	71.18	576	6.29	87.9	31.73(3.65)	0.75(5.98e−02)		1.01	1.12
T.Det	66.13	545	6.29	87.9	31.73(3.67)	0.75(5.85e−02)		1.00	1.25

[a] Denotes that the parameter was fixed.

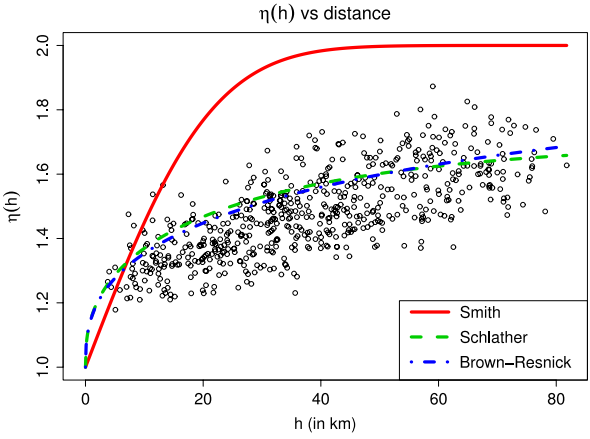

Fig. 4. Comparison between the empirical extremal coefficients (black points) and the fitted extremal coefficients using tapered pairwise composite likelihood under the Smith model (red solid line), the Schlather model (green dashed line), and the Brown–Resnick model (blue dot–dash line).

taper ranges. For example, even at a fixed taper range of $\gamma_s = 14.06$ which is optimal under the Smith model, the Brown–Resnick model still gives a smaller TCLIC, indicating a better fit. In addition, Fig. 4

Table 6
TCLIC scores of the tapered CLs with various taper ranges under the Smith, the Schlather and the Brown–Resnick models with the smallest TCLIC marked in bold.

Taper range	TCLIC (Smith)	TCLIC (Schlather)	TCLIC (Brown–Resnick)
14.06	22 453	22 237	**22 185**
57.71	187 115	184 451	**183 533**
66.13	210 402	207 512	**206 460**
81.78	229 698	226 523	**225 362**

plots the empirical extremal coefficients and the fitted extremal coefficients using tapered pairwise CL estimators under each model. The extremal coefficients seem to be well estimated under the Schlather model and the Brown–Resnick model. But the Smith model fit is obviously poorer than the other two model fits.

6. Discussion

In this paper, we have proposed a statistically efficient and computationally feasible approach to estimating the parameters of spatial max-stable process models for extreme values. The proposed method constructs a tapered composite likelihood method based on spatial pairs or triples. We propose a tapering strategy that excludes distant pairs or triples, hence substantially improving the estimation efficiency and reduces the computational burden associated with explosive combinations of equally weighted CL objects. Comprehensive simulation studies have shown that the TCL achieves significant efficiency gains over the equally weighted CL approach.

We considered only binary weights for the weighted CL inference. It is of interest to explore the use of smoothly decaying weights in future work. The proposed TCL method requires selection of a taper range to construct weights. We run a global search over a finite set of possible taper ranges to find an optimal range. It is possible to use a stochastic search algorithm that can reduce the computational cost in selecting the taper range. Alternatively, one may follow sub-sampling approaches to select the taper range. These extensions are crucial to fit max-stable process models for high-dimensional data via the TCL method.

The optimal taper range was selected by minimizing the trace/determinant of the estimated inverse Godambe information. The weighting function in the weighted CL constructed this way is essentially also calculated from the observed spatial extreme values. Therefore, whether the asymptotic distribution for the weighted CL estimator given in (6) is still valid might be questionable. One possible way to avoid such a problem is to select a taper range based on a subset of data and then to make the TCL inference based on the remaining data by fixing the selected taper range.

Acknowledgments

The research of Huiyan Sang was partially supported by NSF grant DMS-1007618. Marc G. Genton's work was partially supported by DMS-1007504 and DMS-1100492. Both authors were supported by Award Number KUS-CI-016-04, from King Abdullah University of Science and Technology (KAUST). The authors thank Dr. Anthony Davison and Dr. Mathieu Ribatet for several useful discussions regarding this work and for providing the Swiss precipitation dataset.

Appendix A. The triplewise distribution functions of the Smith model

Let $\mathbf{z} = (z_1, z_2, z_3)^{\mathrm{T}}$. Denote by $\mathbf{c}^{(j)}(\mathbf{z}) = \{[c_1^{(j)}(\mathbf{z}), c_2^{(j)}(\mathbf{z}), c_3^{(j)}(\mathbf{z})]\backslash c_j^{(j)}(\mathbf{z})\}^{\mathrm{T}}$, where $\mathbf{c}_k^{(j)}(\mathbf{z}) = (\boldsymbol{s}_j - \boldsymbol{s}_k)^{\mathrm{T}}$ $\boldsymbol{\Sigma}^{-1}(\boldsymbol{s}_j - \boldsymbol{s}_k)/2 - \log(z_j/z_k)$ $(j, k = 1, 2, 3)$. Define the 2×3 matrices $\mathbf{S} = (\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{s}_3)$, and $\boldsymbol{s}_{-j} = \mathbf{S}\backslash \boldsymbol{s}_j \in \mathbb{R}^{2\times 2}$, the matrix $\mathbf{S}$ without the column $\boldsymbol{s}_j$. Define the matrix $\boldsymbol{\Sigma}^{(j)} = (\boldsymbol{s}_j \mathbf{1}_2^{\mathrm{T}} - \boldsymbol{s}_{-j})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{s}_j \mathbf{1}_2^{\mathrm{T}} - \boldsymbol{s}_{-j})$, where $\mathbf{1}_2 = (1, 1)^{\mathrm{T}} \in \mathbb{R}^2$. When $\boldsymbol{\Sigma}^{(j)}$ is invertible, the triplewise cumulative distribution function takes the

form (Genton et al., 2011):

$$-\log F(z_1, z_2, z_3) = \sum_{j=1}^{3} \frac{1}{z_j} \Phi_2\{\mathbf{c}^{(j)}(\mathbf{z}); \boldsymbol{\Sigma}^{(j)}\} = V(\mathbf{z}), \tag{9}$$

where $\Phi_2(\cdot; \boldsymbol{\Sigma})$ denotes the cumulative distribution function of a two-dimensional Gaussian distribution with mean zero and covariance matrix $\boldsymbol{\Sigma}$.

The triplewise density function is

$$f(z_1, z_2, z_3) = F(\mathbf{z}) \left(-V_{[1]}V_{[2]}V_{[3]} + V_{[1]}V_{[2,3]} + V_{[2]}V_{[1,3]} + V_{[3]}V_{[1,2]} - V_{[1,2,3]}\right)(\mathbf{z}), \tag{10}$$

where $V_{[i_1,\, i_2,\ldots,i_k]}(\mathbf{z}) = \partial^k V(\mathbf{z})/(\partial z_{i_1} \partial z_{i_2} \cdots \partial z_{i_k})$, for $\{i_1,\ i_2, \ldots, i_k\} \subseteq \{1, 2, 3\}$. Expressions of $V_{[i_1,i_2,\ldots,i_k]}(\mathbf{z})$ are given below. Let $c^{(i)}_{-j}(\mathbf{z}) = c^{(i)}(\mathbf{z}) \setminus c^{(i)}_j(\mathbf{z})$, $\Sigma^{(i)}_{j,-j} = (\boldsymbol{s}_i - \boldsymbol{s}_j)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{s}_i - \boldsymbol{s}_{-(ij)})$, and $\Sigma^{(i)}_{-j,-j} = (\boldsymbol{s}_i - \boldsymbol{s}_{-(ij)})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{s}_i - \boldsymbol{s}_{-(ij)})$ in which $\boldsymbol{s}_{-(ij)} = \mathbf{S} \setminus (\boldsymbol{s}_i, \boldsymbol{s}_j)$. Then $V_{[i]}(\mathbf{z})$ is given by

$$V_{[i]}(\mathbf{z}) = -\frac{1}{z_i^2} \Phi_2\left\{\mathbf{c}^{(i)}(\mathbf{z}); \boldsymbol{\Sigma}^{(i)}\right\}.$$

Denote $(i_2, \ldots, i_k)$, $k \leq 3$, by $\mathcal{A}$. Then,

$$V_{[i_1,i_2,\ldots,i_k]}(\mathbf{z}) = -\frac{1}{z_{i_1} \prod_{j=1}^{k} z_{i_j}} \phi_{k-1}\left\{\mathbf{c}^{(i_1)}_{\mathcal{A}}(\mathbf{z}); \boldsymbol{\Sigma}^{(i_1)}_{\mathcal{A}}\right\} \Phi_{p-k}\left\{\mathbf{c}^{(i_1)}_{-\mathcal{A}}(\mathbf{z}) - (\boldsymbol{\Sigma}^{(i_1)}_{\mathcal{A},-\mathcal{A}})^{\mathrm{T}} (\boldsymbol{\Sigma}^{(i_1)}_{\mathcal{A},\mathcal{A}})^{-1} \mathbf{c}^{(i_1)}_{\mathcal{A}}(\mathbf{z}); \right.$$
$$\left. \boldsymbol{\Sigma}^{(i_1)}_{-\mathcal{A},-\mathcal{A}} - (\boldsymbol{\Sigma}^{(i_1)}_{\mathcal{A},-\mathcal{A}})^{\mathrm{T}} (\boldsymbol{\Sigma}^{(i_1)}_{\mathcal{A},\mathcal{A}})^{-1} \boldsymbol{\Sigma}^{(i_1)}_{\mathcal{A},-\mathcal{A}} \right\}$$

where $\mathbf{c}^{(i_1)}_{\mathcal{A}}(\mathbf{z}) = \{c^{(i_1)}_{i_2}(\mathbf{z}), \ldots, c^{(i_1)}_{i_k}(\mathbf{z})\}^{\mathrm{T}}$, $\mathbf{c}^{(i_1)}_{-\mathcal{A}}(\mathbf{z}) = \mathbf{c}^{(i_1)}(\mathbf{z}) \setminus \mathbf{c}^{(i_1)}_{\mathcal{A}}(\mathbf{z})$, $\boldsymbol{\Sigma}^{(i_1)}_{\mathcal{A},-\mathcal{A}} = (\boldsymbol{s}_i \mathbf{1}^{\mathrm{T}}_{k-1} - \boldsymbol{s}_{\mathcal{A}})^{\mathrm{T}}$ $\boldsymbol{\Sigma}^{-1} (\boldsymbol{s}_i \mathbf{1}^{\mathrm{T}}_{p-k} - \boldsymbol{s}_{-\mathcal{A}})$ and $\boldsymbol{\Sigma}^{(i_1)}_{-\mathcal{A},-\mathcal{A}} = (\boldsymbol{s}_i \mathbf{1}^{\mathrm{T}}_{p-k} - \boldsymbol{s}_{-\mathcal{A}})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{s}_i \mathbf{1}^{\mathrm{T}}_{p-k} - \boldsymbol{s}_{-\mathcal{A}})$ in which $\boldsymbol{s}_{\mathcal{A}} = (\boldsymbol{s}_{i_2}, \ldots, \boldsymbol{s}_{i_k})$ and $\boldsymbol{s}_{-\mathcal{A}} = \mathbf{S} \setminus (\boldsymbol{s}_{i_1}, \boldsymbol{s}_{\mathcal{A}})$. In particular, the third-order cross partial derivative of $V(\mathbf{z})$ is

$$V_{[1,2,3]}(\mathbf{z}) = -\frac{1}{z_i \prod_{j=1}^{3} z_j} \phi_2\left\{\mathbf{c}^{(i)}(\mathbf{z}); \boldsymbol{\Sigma}^{(i)}\right\}.$$

Appendix B. The squared score statistics of the triplewise case for the Smith model

For the triplewise TCL inference, the term $\mathbf{J}(\boldsymbol{\theta})$ of the sandwich information matrix can be estimated by

$$\frac{1}{N} \sum_{t=1}^{N} \left\{ \sum_{i \neq j \neq k} w_{ijk} \frac{\partial l_{ijk}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\} \left\{ \sum_{i \neq j \neq k} w_{ijk} \frac{\partial l_{ijk}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\}^{\mathrm{T}},$$

where $\boldsymbol{\theta} = (\sigma^2_{11}, \sigma_{12}, \sigma^2_{22})^{\mathrm{T}}$. From (10), we write the triplewise log-density as $l_{ijk}(\boldsymbol{\theta}) = -A + \log(-B + C - D)$, where $A = z_i V_{[i]} + z_j V_{[j]} + z_k V_{[k]}$, $B = V_{[i]}V_{[j]}V_{[k]}$, $C = V_{[i]}V_{[j,k]} + V_{[j]}V_{[i,k]} + V_{[k]}V_{[i,j]}$ and $D = V_{[i,j,k]}$. The first-order derivative of the triplewise log-density has the form

$$\frac{\partial l_{ijk}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -\frac{\partial A}{\partial \boldsymbol{\theta}} + \frac{-\frac{\partial B}{\partial \boldsymbol{\theta}} + \frac{\partial C}{\partial \boldsymbol{\theta}} - \frac{\partial D}{\partial \boldsymbol{\theta}}}{-B + C - D},$$

where

$$\frac{\partial A}{\partial \boldsymbol{\theta}} = z_i \frac{\partial V_{[i]}}{\partial \boldsymbol{\theta}} + z_j \frac{\partial V_{[j]}}{\partial \boldsymbol{\theta}} + z_k \frac{\partial V_{[k]}}{\partial \boldsymbol{\theta}},$$
$$\frac{\partial B}{\partial \boldsymbol{\theta}} = V_{[j]}V_{[k]} \frac{\partial V_{[i]}}{\partial \boldsymbol{\theta}} + V_{[i]}V_{[k]} \frac{\partial V_{[j]}}{\partial \boldsymbol{\theta}} + V_{[i]}V_{[j]} \frac{\partial V_{[k]}}{\partial \boldsymbol{\theta}},$$

$$\frac{\partial C}{\partial \boldsymbol{\theta}} = V_{[j,k]}\frac{\partial V_{[i]}}{\partial \boldsymbol{\theta}} + V_{[i]}\frac{\partial V_{[j,k]}}{\partial \boldsymbol{\theta}} + V_{[i,k]}\frac{\partial V_{[j]}}{\partial \boldsymbol{\theta}} + V_{[j]}\frac{\partial V_{[i,k]}}{\partial \boldsymbol{\theta}} + V_{[i,j]}\frac{\partial V_{[k]}}{\partial \boldsymbol{\theta}} + V_{[k]}\frac{\partial V_{[i,j]}}{\partial \boldsymbol{\theta}},$$
$$\frac{\partial D}{\partial \boldsymbol{\theta}} = \frac{\partial V_{[i,j,k]}}{\partial \boldsymbol{\theta}}.$$

Define $\alpha_{j,k}^{(i)}(\boldsymbol{\theta}) = \{(\boldsymbol{s}_i - \boldsymbol{s}_j)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{s}_i - \boldsymbol{s}_k)\}^{1/2}$. Let $w_j^{(i)} = c_j^{(i)}/\alpha_{j,j}^{(i)}$, $v_j^{(i)} = \alpha_{j,j}^{(i)} - w_j^{(i)}$, $\rho_{j,k}^{(i)} = (\alpha_{j,k}^{(i)})^2/(\alpha_{j,j}^{(i)}\alpha_{k,k}^{(i)})$. Write $\mathbf{w}^{(i)} = (w_j^{(i)}, w_k^{(i)})^{\mathrm{T}}$ and let $\boldsymbol{\rho}^{(i)} \in \mathbb{R}^{2\times 2}$ be the correlation matrix with correlation $\rho_{j,k}^{(i)}$. We have

$$\frac{\partial V_{[i]}}{\partial \boldsymbol{\theta}} = \frac{\partial V_{[i]}}{\partial w_j^{(i)}}\frac{\partial w_j^{(i)}}{\partial \boldsymbol{\theta}} + \frac{\partial V_{[i]}}{\partial w_k^{(i)}}\frac{\partial w_k^{(i)}}{\partial \boldsymbol{\theta}} + \frac{\partial V_{[i]}}{\partial \rho_{j,k}^{(i)}}\frac{\partial \rho_{j,k}^{(i)}}{\partial \boldsymbol{\theta}},$$

where

$$\frac{\partial V_{[i]}}{\partial w_j^{(i)}} = -\frac{1}{z_i^2}\phi(w_j^{(i)})\Phi\{w_k^{(i)} - \rho_{j,k}^{(i)}w_j^{(i)}; 1-(\rho_{j,k}^{(i)})^2\},$$
$$\frac{\partial V_{[i]}}{\partial w_k^{(i)}} = -\frac{1}{z_i^2}\phi(w_k^{(i)})\Phi\{w_j^{(i)} - \rho_{j,k}^{(i)}w_k^{(i)}; 1-(\rho_{j,k}^{(i)})^2\},$$
$$\frac{\partial V_{[i]}}{\partial \rho_{j,k}^{(i)}} = -\frac{1}{z_i^2}\phi_2(\mathbf{w}^{(i)}; \boldsymbol{\rho}^{(i)}),$$
$$\frac{\partial \rho_{j,k}^{(i)}}{\partial \boldsymbol{\theta}} = 2\frac{\alpha_{j,k}^{(i)}}{\alpha_{j,j}^{(i)}\alpha_{k,k}^{(i)}}\left(\frac{\partial \alpha_{j,k}^{(i)}}{\partial \boldsymbol{\theta}}\right)^{\mathrm{T}} - \frac{(\alpha_{j,k}^{(i)})^2}{(\alpha_{j,j}^{(i)})^2\alpha_{k,k}^{(i)}}\left(\frac{\partial \alpha_{j,j}^{(i)}}{\partial \boldsymbol{\theta}}\right)^{\mathrm{T}} - \frac{(\alpha_{j,k}^{(i)})^2}{\alpha_{j,j}^{(i)}(\alpha_{k,k}^{(i)})^2}\left(\frac{\partial \alpha_{k,k}^{(i)}}{\partial \boldsymbol{\theta}}\right)^{\mathrm{T}},$$
$$\frac{\partial w_j^{(i)}}{\partial \boldsymbol{\theta}} = \frac{v_j^{(i)}}{\alpha_{j,j}^{(i)}}\left(\frac{\partial \alpha_{j,j}^{(i)}}{\partial \boldsymbol{\theta}}\right)^{\mathrm{T}}, \qquad \frac{\partial w_k^{(i)}}{\partial \boldsymbol{\theta}} = \frac{v_k^{(i)}}{\alpha_{k,k}^{(i)}}\left(\frac{\partial \alpha_{k,k}^{(i)}}{\partial \boldsymbol{\theta}}\right)^{\mathrm{T}},$$

where the expressions for $(\partial\alpha_{j,j}^{(i)}/\partial\boldsymbol{\theta})^{\mathrm{T}}$, $(\partial\alpha_{k,k}^{(i)}/\partial\boldsymbol{\theta})^{\mathrm{T}}$, $(\partial\alpha_{j,k}^{(i)}/\partial\boldsymbol{\theta})^{\mathrm{T}}$ are given in (11).

Let $t_{ij,1} = (s_{i,1} - s_{j,1})$, $t_{ij,2} = (s_{i,2} - s_{j,2})$, $t_{ik,1} = (s_{i,1} - s_{k,1})$, $t_{ik,2} = (s_{i,2} - s_{k,2})$. We have:

$$\left(\frac{\partial \alpha_{j,k}^{(i)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^{\mathrm{T}} = \frac{1}{2\alpha_{j,k}^{(i)}|\boldsymbol{\Sigma}|^2}\left\{t_{ij,1}t_{ik,1}, (t_{ij,1}t_{ik,2} + t_{ij,2}t_{ik,1})/2, t_{ij,2}t_{ik,2}\right\} \times \begin{pmatrix} -\sigma_{22}^4 & 2\sigma_{22}^2\sigma_{12} & -\sigma_{12}^2 \\ 2\sigma_{22}^2\sigma_{12} & -2(\sigma_{11}^2\sigma_{22}^2 + \sigma_{12}^2) & 2\sigma_{11}^2\sigma_{12} \\ -\sigma_{12}^2 & 2\sigma_{11}^2\sigma_{12} & -\sigma_{11}^4 \end{pmatrix}. \tag{11}$$

Similarly

$$\frac{\partial V_{[i,j]}}{\partial \boldsymbol{\theta}} = \frac{\partial V_{[i,j]}}{\partial w_j^{(i)}}\frac{\partial w_j^{(i)}}{\partial \boldsymbol{\theta}} + \frac{\partial V_{[i,j]}}{\partial w_k^{(i)}}\frac{\partial w_k^{(i)}}{\boldsymbol{\theta}} + \frac{\partial V_{[i,j]}}{\partial \rho_{j,k}^{(i)}}\frac{\partial \rho_{j,k}^{(i)}}{\partial \boldsymbol{\theta}} - \frac{\partial V_{[i,j]}}{\alpha_{j,j}^{(i)}}\left(\frac{\partial \alpha_{j,j}^{(i)}}{\partial \boldsymbol{\theta}}\right)^{\mathrm{T}},$$

where

$$\frac{\partial V_{[i,j]}}{\partial w_j^{(i)}} = \frac{w_j^{(i)}\phi(w_j^{(i)})\Phi\{w_k^{(i)} - \rho_{j,k}^{(i)}w_j^{(i)}; 1-(\rho_{j,k}^{(i)})^2\} + \rho_{j,k}^{(i)}\phi(w^{(i)}; \rho^{(i)})}{z_i^2 z_j \alpha_{j,j}^{(i)}},$$
$$\frac{\partial V_{[i,j]}}{\partial w_k^{(i)}} = -\frac{\phi_2(w^{(i)}; \rho^{(i)})}{z_i^2 z_j \alpha_{j,j}^{(i)}},$$
$$\frac{\partial V_{[i,j]}}{\partial \rho_{j,k}^{(i)}} = \frac{(w_j^{(i)} - \rho_{j,k}^{(i)}w_k^{(i)})\phi_2(w^{(i)}; \rho^{(i)})}{z_i^2 z_j \alpha_{j,j}^{(i)}\{1-(\rho_{j,k}^{(i)})^2\}},$$

and

$$\frac{\partial V_{[i,j,k]}}{\partial \boldsymbol{\theta}} = \frac{\partial V_{[i,j,k]}}{\partial w_j^{(i)}}\frac{\partial w_j^{(i)}}{\partial \boldsymbol{\theta}} + \frac{\partial V_{[i,j,k]}}{\partial w_k^{(i)}}\frac{\partial w_k^{(i)}}{\boldsymbol{\theta}} + \frac{\partial V_{[i,j,k]}}{\partial \rho_{j,k}^{(i)}}\frac{\partial \rho_{j,k}^{(i)}}{\partial \boldsymbol{\theta}} - \frac{V_{[i,j,k]}}{\alpha_{j,j}^{(i)}}\left(\frac{\partial \alpha_{j,j}^{(i)}}{\partial \boldsymbol{\theta}}\right)^{\mathrm{T}} - \frac{V_{[i,j,k]}}{\alpha_{k,k}^{(i)}}\left(\frac{\partial \alpha_{k,k}^{(i)}}{\partial \boldsymbol{\theta}}\right)^{\mathrm{T}},$$

where

$$\frac{\partial V_{[i,j,k]}}{\partial w_j^{(i)}} = \frac{1}{z_i^2 z_j z_k \alpha_{j,j}^{(i)} \alpha_{k,k}^{(i)}} \frac{(w_j^{(i)} - \rho_{j,k}^{(i)} w_k^{(i)})\phi_2(w^{(i)}; \rho^{(i)})}{1 - (\rho_{j,k}^{(i)})^2},$$

$$\frac{\partial V_{[i,j,k]}}{\partial w_k^{(i)}} = \frac{1}{z_i^2 z_j z_k \alpha_{j,j}^{(i)} \alpha_{k,k}^{(i)}} \frac{(w_k^{(i)} - \rho_{j,k}^{(i)} w_j^{(i)})\phi_2(w^{(i)}; \rho^{(i)})}{1 - (\rho_{j,k}^{(i)})^2},$$

$$\frac{\partial V_{[i,j,k]}}{\partial \rho_{jk}^{(i)}} = -\frac{\phi_2(w^{(i)}; \rho^{(i)})}{z_i^2 z_j z_k \alpha_{j,j}^{(i)} \alpha_{k,k}^{(i)}} \left[\frac{\rho_{jk}^{(i)}}{1 - (\rho_{j,k}^{(i)})^2} + \frac{(w_k^{(i)} - \rho w_j^{(i)})(w_j^{(i)} - \rho w_k^{(i)})}{\{1 - (\rho_{j,k}^{(i)})^2\}^2}\right].$$

References

Bevilacqua, M., Gaetan, C., Mateu, J., Porcu, E., 2012. Estimating space and space–time covariance functions for large datasets: a weighted composite likelihood approach. Journal of the American Statistical Association 107 (497), 268–280.

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

Buishand, T.A., de Haan, L., Zhou, C., 2008. On spatial extremes: with application to a rainfall problem. The Annals of Applied Statistics 2 (2), 624–642.

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.

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

Davis, R.A., Yau, C.Y., 2011. Comments on pairwise likelihood in time series models. Statistica Sinica 21, 255–277.

Davison, A.C., Gholamrezaee, M.M., 2012. Geostatistics of extremes. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science 468 (2138), 581–608.

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

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

de Haan, L., Pereira, T.T., 2006. Spatial extremes: models for the stationary case. Annals of Statistics 34, 146–168.

de Haan, L., Resnick, S.I., 1977. Limit theory for multivariate sample extremes. Probability Theory and Related Fields 40 (4), 317–337.

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

Godambe, V.P., Heyde, C.C., 2010. Quasi-likelihood and optimal estimation. In: Selected Works of C.C. Heyde. pp. 386–399.

Huser, R., Davison, A.C., 2013. Composite likelihood estimation for the Brown–Resnick process. Biometrika 100, 511–518.

Joe, H., Lee, Y., 2009. On weighting of bivariate margins in pairwise likelihood. Journal of Multivariate Analysis 100 (4), 670–685.

Kabluchko, Z., Schlather, M., de Haan, L., 2009. Stationary max-stable fields associated to negative definite functions. The Annals of Probability 37 (5), 2042–2065.

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

Lipsitz, S., Dear, K., Zhao, L., 1994. Jackknife estimators of variance for parameter estimates from estimating equations with applications to clustered survival data. Biometrics 50 (3), 842–846.

Mardia, K.V., Kent, J.T., Bibby, J.M., 1979. Multivariate Analysis. Academic Press.

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

Ribatet, M., 2011. SpatialExtremes: modelling spatial extremes. R Package Version 1.8-0. URL http://CRAN.R-project.org/package=SpatialExtremes.

Ribatet, M., Cooley, D., Davison, A., 2012. Bayesian inference from composite likelihoods, with an application to spatial extremes. Statistica Sinica 22, 813–845.

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

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

Smith, R.L., 1990. Max-stable processes and spatial extremes. Unpublished Manuscript 205.

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

Stein, M., 1999. Interpolation of Spatial Data: Some Theory for Kriging. Springer-Verlag.

Takeuchi, K., 1976. Distribution of information statistics and criteria adequacy of models. Mathematical Science 153, 12–18.

Varin, C., Vidoni, P., 2005. A note on composite likelihood inference and model selection. Biometrika 92, 519–528.

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