Statistical inference for max-stable processes in space and time

Richard A. Davis

Department of Statistics, Columbia University, New York, United States.

Claudia Klüppelberg

Center for Mathematical Sciences, Technische Universität München, Garching, Germany.

Christina Steinkohl

Center for Mathematical Sciences and Institute for Advanced Study, Technische Universität München, Garching, Germany.

Abstract. Max-stable processes have proved to be useful for the statistical modelling of spatial extremes. Several families of max-stable random fields have been proposed in the literature. One such representation is based on a limit of normalised and rescaled pointwise maxima of stationary Gaussian processes that was first introduced by Kabluchko et al. (2009). This paper deals with statistical inference for max-stable space-time processes that are defined in an analogous fashion. We describe pairwise likelihood estimation, where the pairwise density of the process is used to estimate the model parameters. For regular grid observations we prove strong consistency and asymptotic normality of the parameter estimates as the joint number of spatial locations and time points tends to infinity. Furthermore, we discuss extensions to irregularly spaced locations. A simulation study shows that the proposed method works well for these models.

1. Introduction

Max-stable processes have proved to be useful in the modelling of spatial extremes. Typically, meteorological extremes like heavy rainfall or extreme wind speeds are modelled using extreme value theory. As an example consider radar rainfall measurements which are given on a grid for several time points. To analyse the extremal dependence structure in space and time one can take maxima of the rainfall measurements over several locations and over a certain time range as for instance daily maxima.

Several families of max-stable processes have been proposed in the literature including, for example, Brown and Resnick (1977), de Haan (1984), Kabluchko et al. (2009), and Schlather (2002). Recently, models for extreme values observed in a space-time setting have generated a great deal of interest. First approaches can be found in Davis and Mikosch (2008), Huser and Davison (2012), Kabluchko (2009), and Davis et al. (2011).

In this paper, we follow the approach considered in Davis et al. (2011), who extend the max-stable process introduced in Kabluchko et al. (2009) to a space-time setting. The process is constructed as the limit of normalised and rescaled maxima of independent replications of some stationary Gaussian space-time process. The underlying correlation function of the Gaussian process is assumed to belong to a parametric model, whose parameters describe smoothness of the correlation function near the origin. As pointed out in Davis et al. (2011), the resulting parametric model is very general since the condition imposed is satisfied by a broad class of correlation functions. The limit process is then given by a Brown-Resnick process with a specific extremal dependence structure resulting from the assumption on the underlying correlation function.

The main difficulty in deriving parameter estimates in such models is the fact that the finite-dimensional distribution function and, thus, the density is intractable, which precludes the use of standard maximum likelihood procedures. On the other hand, pairwise likelihood methods, where only the pairwise density is needed, can be implemented. These methods go back to Besag (1974), and there is an extensive literature available dealing with applications and properties of the estimates, see for example Cox and Reid (2004), Lindsay (1988), Varin (2007), or Varin and Vidoni (2005). Recent work concerning the application of pairwise likelihood methods to max-stable random fields can be found in Huser and Davison (2012) and Padoan et al. (2009).

We first study the asymptotic behaviour of the pairwise likelihood estimates for the Brown-Resnick process for locations which lie on a regular grid and equidistant time points. In contrast to previous studies we assume a spatial and temporal dependence structure and show consistency and asymptotic normality of the pairwise likelihood estimates for a jointly increasing number of spatial locations and time points. In addition, theorems in the literature addressing asymptotic properties for pairwise likelihood estimates often have restrictive assumptions which might not be reasonable in practical applications. For the setting considered in this paper very weak assumptions are sufficient. In particular, we only need standard assumptions on the parameter space and an identifiability condition for the pairwise density. In addition we discuss two extensions of our results to settings, where locations are irregularly spaced.

Our paper is organized as follows. In Section 2 we introduce the max-stable space-time process for which inference properties will be considered in subsequent sections. Section 3 describes pairwise likelihood estimation and the particular setting for our model. In Sections 4 and 5 we prove strong consistency and asymptotic normality, when locations lie on a regular grid and for equidistant time points. In Section 6 we discuss two possible ways of redefining the set of spatial locations, which can be irregularly spaced, for which consistency and asymptotic normality of the pairwise likelihood estimates still holds. A simulation study evaluating the performance of the estimates is presented in Section 7.

2. Description of the model

We start with the process that will be used for modelling extremes in space and time; details can be found in Davis et al. (2011). Let $\left\{Z(\boldsymbol{s},t), \boldsymbol{s}\in\mathbb{R}^d, t\in[0,\infty)\right\}$ denote a stationary space-time Gaussian process on $\mathbb{R}^d\times[0,\infty)$ with mean zero and variance one. Stationarity here means that for all $\boldsymbol{h}\in\mathbb{R}^d$ and $u\geq 0$, the process $\{Z(\boldsymbol{s}+\boldsymbol{h},t+u), \boldsymbol{s}\in\mathbb{R}^d, t\in[0,\infty)\}$ has the same finite-dimensional distributions as Z. For the correlation function

$$\rho(\boldsymbol{h},u)=\mathbb{E}\left[Z(\boldsymbol{s},t)Z(\boldsymbol{s}+\boldsymbol{h},t+u)\right],$$

where $\boldsymbol{h}\in\mathbb{R}^d$ is the spatial lag and $u\geq 0$ is the time lag, we make the following assumption that will be used throughout the paper.

Assumption 2.1. *There exist sequences of constants $s_n\to 0$, $t_n\to 0$ as $n\to\infty$, such that*

$$(\log n)(1-\rho(s_n\boldsymbol{h},t_n u))\to\delta(\boldsymbol{h},u)>0,\quad \text{as } n\to\infty.$$

Assumption 2.1 is natural in the context of stationary space-time models; the correlation function tends to one at a certain rate as the space-time lag approaches zero. The following proposition is a space-time extension of a result in Kabluchko et al. (2009), which was also considered in Davis et al. (2011). It can also be derived from Theorem 2 together with Theorem 6 in Kabluchko (2011).

PROPOSITION 2.2. *Let $\left\{Z_j(\boldsymbol{s},t), \boldsymbol{s} \in \mathbb{R}^d, t \in [0,\infty)\right\}, j = 1,\dots,n$, be independent replications of the space-time Gaussian process described above and let $\{\xi_j, j \in \mathbb{N}\}$ denote points of a Poisson random measure on $(0,\infty]$ with intensity measure $\xi^{-2}d\xi$. Suppose Assumption 2.1 is satisfied. Then the random fields $\left\{\eta_n(\boldsymbol{s},t)), \boldsymbol{s} \in \mathbb{R}^d, t \in [0,\infty)\right\}$, defined for $n \in \mathbb{N}$ by*

$$\eta_n(\boldsymbol{s},t) = \bigvee_{j=1}^{n} -\frac{1}{\log(\Phi(Z_j(s_n\boldsymbol{s}, t_n t)))}, \;\; \boldsymbol{s} \in \mathbb{R}^d, t \in [0,\infty), \tag{1}$$

converge weakly in the space of continuous functions on $\mathbb{R}^d \times [0,\infty)$ to the stationary Brown-Resnick process

$$\eta(\boldsymbol{s},t) = \bigvee_{j=1}^{\infty} \xi_j \exp\left\{W_j(\boldsymbol{s},t) - \delta(\boldsymbol{s},t)\right\}, \tag{2}$$

where the deterministic function δ is given in Assumption 2.1 and $\left\{W_j(\boldsymbol{s},t), \boldsymbol{s} \in \mathbb{R}^d, t \in [0,\infty)\right\}$, $j \in \mathbb{N}$ are independent replications of a Gaussian process with stationary increments, $W(\mathbf{0},0) = 0$, $\mathbb{E}(W(\boldsymbol{s},t)) = 0$ and covariance function, given for $\boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathbb{R}^d, t_1, t_2 \in [0,\infty)$ by

$$\mathbb{C}ov\left(W(\boldsymbol{s}_1,t_1), W(\boldsymbol{s}_2,t_2)\right) = \delta(\boldsymbol{s}_1,t_1) + \delta(\boldsymbol{s}_2,t_2) - \delta(\boldsymbol{s}_1 - \boldsymbol{s}_2, t_1 - t_2),$$

where $\boldsymbol{s}_1 - \boldsymbol{s}_2$ is defined componentwise. The bivariate distribution function of η can be expressed in closed form (based on a well-known result by Hüsler and Reiss (1989)) for $x_1, x_2 > 0$ as

$$F(x_1,x_2) = \exp\left\{-\frac{1}{x_1}\Phi\left(\frac{\log\frac{x_2}{x_1}}{2\sqrt{\delta(\boldsymbol{h},u)}} + \sqrt{\delta(\boldsymbol{h},u)}\right) - \frac{1}{x_2}\Phi\left(\frac{\log\frac{x_1}{x_2}}{2\sqrt{\delta(\boldsymbol{h},u)}} + \sqrt{\delta(\boldsymbol{h},u)}\right)\right\}, \tag{3}$$

where Φ denotes the standard normal distribution function.

Many correlation functions satisfy the following condition, which will be used throughout.

CONDITION 2.3. *The correlation function has an expansion around zero, given by*

$$\rho(\boldsymbol{h},u) = 1 - \theta_1\|\boldsymbol{h}\|^{\alpha_1} - \theta_2|u|^{\alpha_2} + O(\|\boldsymbol{h}\|^{\alpha_1} + |u|^{\alpha_2}), \quad \boldsymbol{h} \in \mathbb{R}^d, u \in \mathbb{R},$$

where $0 < \alpha_1, \alpha_2 \le 2$ and $\theta_1, \theta_2 > 0$.

REMARK 2.4. *The condition is satisfied by many space-time correlation functions including, for example, Gneiting's class of correlation functions; cf. Gneiting (2002). For a detailed analysis of Gneiting's class and further examples we refer to Davis et al. (2011), Section 4, where Condition 2.3 is proved for several classes of correlation functions. Note that Condition 2.3 is restricted to isotropic correlation functions. A natural generalization for possibly anisotropic correlation functions would be written as*

$$\rho(h_1,\dots,h_d,u) = 1 - \sum_{j=1}^{d}\theta_j|h_j|^{\alpha_j} - \theta_{d+1}|u|^{\alpha_{d+1}} + O\Big(\sum_{j=1}^{d}|h_j|^{\alpha_j} + |u|^{\alpha_{d+1}}\Big).$$

Condition 2.3 allows for an explicit expression of the limit function δ in Assumption 2.1:

$$\delta(\boldsymbol{h}, u) = \theta_1 \|\boldsymbol{h}\|^{\alpha_1} + \theta_2 |u|^{\alpha_2}, \tag{4}$$

where the scaling sequences $(s_n)_{n\in\mathbb{N}}$ and $(t_n)_{n\in\mathbb{N}}$ can be chosen as $s_n = (\log n)^{1/\alpha_1}$ and $t_n = (\log n)^{1/\alpha_2}$. The parameters $\alpha_1, \alpha_2 \in (0, 2]$ relate to the smoothness of the underlying Gaussian process in space and time, where the case $\alpha_1 = \alpha_2 = 2$ corresponds to a mean-square differentiable process. A further property of the model defined in Proposition 2.2 is a closed form expression for the *tail dependence coefficient* given by

$$\chi(\boldsymbol{h}, u) = \lim_{x\to\infty} P\left(\eta(\boldsymbol{s}_1, t_1) > F^{\leftarrow}_{\eta(\boldsymbol{s}_1,t_1)}(x) \mid \eta(\boldsymbol{s}_2, t_2) > F^{\leftarrow}_{\eta(\boldsymbol{s}_2,t_2)}(x)\right),$$

where $\boldsymbol{h} = \boldsymbol{s}_1 - \boldsymbol{s}_2$ and $u = t_1 - t_2$. As derived in Section 3 of Davis et al. (2011), we obtain

$$\chi(\boldsymbol{h}, u) = 2\left(1 - \Phi(\sqrt{\delta(\boldsymbol{h}, u)})\right) = 2\left(1 - \Phi(\sqrt{\theta_1 \|\boldsymbol{h}\|^{\alpha_1} + \theta_2 |u|^{\alpha_2}})\right). \tag{5}$$

3. Pairwise likelihood estimation

In this section, we describe the pairwise likelihood procedure for estimating the parameters of the Brown-Resnick process (2), when the underlying correlation function satisfies Condition 2.3. Composite likelihood methods have been used whenever the full likelihood is not available or intractable. We present the general definition of composite and pairwise likelihood functions for a space-time setting in Section 3.1. Afterwards, we rewrite the pairwise likelihood for regular grid observations.

3.1. *Composite likelihood estimation for the space-time setting*

Composite likelihood methods go back to Besag (1974) and Lindsay (1988) and there is now a vast literature available, from a theoretical and an applied point of view on the topic. For more information we refer to Varin (2007), who presents an overview of existing models and inference including extensive references. In the most general setting the composite log-likelihood function is given by

$$l_c(\boldsymbol{\psi}, \boldsymbol{x}) = \sum_{i=1}^{q} w_i \log f_{\boldsymbol{\psi}}(\boldsymbol{x} \in A_i),$$

where for $i = 1, \dots, p$ the sets A_i describe measurable events and the w_i are non-negative weights associated to the events. From this general form special composite likelihood functions can be derived. The *(weighted) pairwise log-likelihood function* is defined by

$$PL(\boldsymbol{\psi}; \boldsymbol{x}) = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j} \log f_{\boldsymbol{\psi}}(x_i, x_j), \tag{6}$$

where $\boldsymbol{x} = (x_1, \dots, x_n)$ is the data vector, $f_{\boldsymbol{\psi}}(x_i, x_j)$ is the density of the bivariate observations (x_i, x_j), and the $w_{i,j}$ are weights which can be used for example to reduce the number of pairs included in the estimation. The parameter estimates are obtained by maximizing (6).

As noted in Cox and Reid (2004), for dependent observations, estimates based on the composite likelihood need not be consistent or asymptotically normal. This is important for space-time applications, since all components may be highly dependent across space and time. We describe the pairwise likelihood estimation for observations from the Brown-Resnick process (2), where the underlying correlation function satisfies Condition 2.3. The resulting parameter vector is given by $\boldsymbol{\psi} = (\theta_1, \alpha_1, \theta_2, \alpha_2)$. The pairwise likelihood for a general setting with M locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_M$ and T time points $0 \leq t_1 < \cdots < t_T < \infty$ is given by

$$PL^{(M,T)}(\boldsymbol{\psi}) = \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} \sum_{k=1}^{T-1} \sum_{l=k+1}^{T} w_{i,j}^{(M)} w_{k,l}^{(T)} \log f_{\boldsymbol{\psi}}(\eta(\boldsymbol{s_i}, t_k), \eta(\boldsymbol{s}_j, t_l)), \tag{7}$$

where $w_{i,j}^{(M)} \geq 0$ and $w_{k,l}^{(T)} \geq 0$ denote spatial and temporal weights, respectively, and $f_{\boldsymbol{\psi}}$ is the bivariate density given as derivative of the distribution function in (3). Since it is expected that space-time pairs, which are far apart in space or in time, have only little influence on the dependence parameters to be estimated, we define the weights, such that in the estimation only pairs with a maximal spatio-temporal distance of (r, p) are included, i.e.,

$$w_{i,j}^{(M)} = \mathbb{1}_{\{\|\boldsymbol{s}_i - \boldsymbol{s}_j\| \leq r\}}, \qquad w_{k,l}^{(T)} = \mathbb{1}_{\{|t_k - t_l| \leq p\}}, \tag{8}$$

where $\|\cdot\|$ denotes any arbitrary norm on $\mathbb{R}^d$. The pairwise likelihood estimates are given by

$$(\hat{\theta}_1, \hat{\alpha}_1, \hat{\theta}_2, \hat{\alpha}_2) = \underset{(\theta_1, \alpha_1, \theta_2, \alpha_2)}{\arg\max} \; PL^{(M,T)}(\theta_1, \alpha_1, \theta_2, \alpha_2). \tag{9}$$

Using the definition of the weights in (8), the log-likelihood function in (7) can be rewritten as

$$PL^{(M,T)}(\boldsymbol{\psi}) = \sum_{i=1}^{M-1} \sum_{\substack{j=i+1 \\ \|\boldsymbol{s}_i - \boldsymbol{s}_j\| \leq r}}^{M} \sum_{k=1}^{T-p} \sum_{l=k+1}^{\min\{k+p,T\}} \log f_{\boldsymbol{\psi}}(\eta(\boldsymbol{s}_i, t_k), \eta(\boldsymbol{s}_j, t_l)). \tag{10}$$

3.2. *Pairwise likelihood estimation for regular grid observations*

The proof of strong consistency and asymptotic normality in Sections 4 and 5 is based on the assumption that locations lie on a regular grid and that time points are equidistant. The following condition summarizes the sampling scheme.

Condition 3.1. *We assume that the locations lie on a regular d-dimensional lattice,*

$$S_m = \left\{(i_1, \ldots, i_d),\ i_1, \ldots, i_d \in \{1, \ldots, m\}\right\}.$$

Further assume that the time points are equidistant and given by the set $\{1, \ldots, T\}$.

For later purposes, we rewrite the pairwise log-likelihood function under Condition 3.1 in the following way. Define $\mathcal{H}_r$ as the set of all vectors with non-negative integer-valued components $\boldsymbol{h}$ without the $\mathbf{0}$-vector, which point to other sites in the set of locations within distance r, i.e.,

$$\mathcal{H}_r = \mathbb{N}^d \cap B(\mathbf{0}, r) \backslash \{\mathbf{0}\},$$

where $B(\mathbf{0},r)=\{\boldsymbol{s}:\|\boldsymbol{s}\|<r\}$. Nott and Rydén (1999) call this the *design mask*. We denote by $|\mathcal{H}_r|$ the cardinality of the set $\mathcal{H}_r$. In our application, we will use design masks according to the Euclidean distance; for example with $d=2$ (cf. Figure 1),

$$\mathcal{H}_3=\{(1,0),(0,1),(1,1),(0,2),(2,0),(1,2),(2,1),(2,2),(0,3),(3,0)\}.$$

Using Condition 3.1 and the design mask, the *pairwise log-likelihood function* in (9) can be rewritten as

$$\begin{aligned}PL^{(m,T)}(\boldsymbol{\psi})&=\sum_{\boldsymbol{s}\in S_m}\sum_{t=1}^{T}\sum_{\substack{\boldsymbol{h}\in\mathcal{H}_r\\ \boldsymbol{s}+\boldsymbol{h}\in S_m}}\sum_{\substack{u=1\\ t+u\le T}}^{p}\log f_{\boldsymbol{\psi}}(\eta(\boldsymbol{s},t),\eta(\boldsymbol{s}+\boldsymbol{h},t+u))\\&=\sum_{\boldsymbol{s}\in S_m}\sum_{t=1}^{T}g_{\boldsymbol{\psi}}\left(\boldsymbol{s},t;r,p\right)-\mathcal{R}^{(m,T)}(\boldsymbol{\psi}),\end{aligned}\tag{11}$$

where

$$g_{\boldsymbol{\psi}}(\boldsymbol{s},t;r,p)=\sum_{\boldsymbol{h}\in\mathcal{H}_r}\sum_{u=1}^{p}\log f_{\boldsymbol{\psi}}(\eta(\boldsymbol{s},t),\eta(\boldsymbol{s}+\boldsymbol{h},t+u)),\tag{12}$$

and $\mathcal{R}^{(m,T)}(\boldsymbol{\psi})$ is a boundary term, given by

$$\mathcal{R}^{(m,T)}(\boldsymbol{\psi})=\sum_{\boldsymbol{s}\in S_m}\sum_{t=1}^{T}\sum_{\substack{\boldsymbol{h}\in\mathcal{H}_r\\ \boldsymbol{s}+\boldsymbol{h}\notin S_m}}\sum_{\substack{u=1\\ t+u>T}}^{p}\log f_{\boldsymbol{\psi}}(\eta(\boldsymbol{s},t),\eta(\boldsymbol{s}+\boldsymbol{h},t+u)).\tag{13}$$

Figure 1 depicts a spatial grid with length $m=6$, where the inner square is the set of observed locations S_m and the points in the outer polygon are endpoints of pairs which are in the boundary term $\mathcal{R}^{(m,T)}$. The figure visualizes the case $\mathcal{H}_2$, which is represented by the quarter circles.

4. Strong consistency of the pairwise likelihood estimates for regular grid observations

In this section we establish strong consistency for the pairwise likelihood estimates based on regular grid observations introduced in Section 3.2. For univariate time series models Davis and Yau (2011) proved strong consistency of the composite likelihood estimates in full detail. For max-stable random fields with replicates, which are independent in time, Padoan et al. (2009) showed consistency and asymptotic normality for the pairwise likelihood estimates. In contrast to previous studies, where either the spatial or the time domain increases, we show strong consistency as the space-time domain increases jointly.

4.1. Ergodic properties for max-stable processes

Stoev and Taqqu (2005) introduced extremal integrals as an analogy to sum-stable integrals. We briefly explain the notion of an extremal integral. The basis for the definition are α-Fréchet sup-measures. Given a measure space $(E,\mathcal{E},\mu)$ with $\sigma-$finite, positive measure μ, the set-indexed random process $\{M_\alpha(A),A\in\mathcal{E}\}$ is called an *independently scattered $\alpha-$Fréchet sup-measure with control measure* μ, if

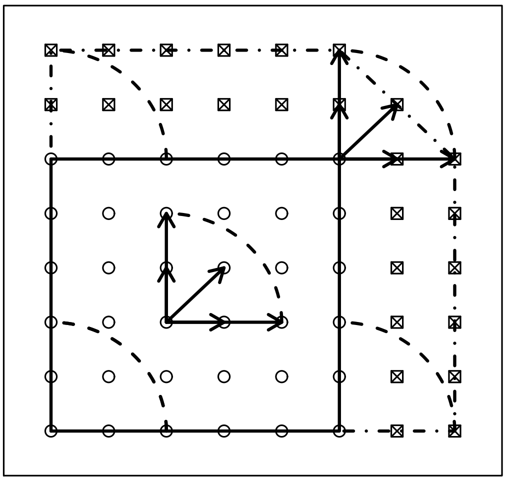


Figure 1. Visualization of the boundary term $\mathcal{R}^{(m,T)}$ for $d = 2$, $m = 6$ and any fixed time point; the set S_m of locations is the inner square and the outer polygon represents the endpoints of pairs in the boundary

(a) for disjoint $A_1, \ldots, A_n \in \mathcal{E}$, the random variables $M_\alpha(A_1), \ldots, M_\alpha(A_n)$ are independent,

(b) for $A \in \mathcal{E}$
$$P(M_\alpha(A) \le x) = \exp\left\{-\mu(A)x^{-\alpha}\right\} \mathbb{1}_{\{x>0\}},$$
i.e., $M_\alpha(A)$ is $\alpha-$Fréchet distributed with scale parameter $\mu(A)^{1/\alpha}$,

(c) for disjoint $A_j \in \mathcal{E}, j \in \mathbb{N}$, with $\bigcup_{j\in\mathbb{N}} A_j \in \mathcal{E}$,
$$M_\alpha\left(\bigcup_{j\in\mathbb{N}} A_j\right) = \bigvee_{j\in\mathbb{N}} M_\alpha(A_j).$$

For a non-negative simple function $f : E \to \mathbb{R}$, $f(x) = \sum_{j=1}^n a_j \mathbb{1}_{A_j}(x)$, where $A_1, \ldots, A_n \in \mathcal{E}$ are disjoint, the *extremal integral* $\overset{e}{\int}$ is defined by
$$\overset{e}{\int_E} f(x) M_\alpha(dx) := \bigvee_{j=1}^n a_j M_\alpha(A_j),$$
and the integral is independent of the representation of f. This definition can be extended stepwise from simple functions to nondecreasing sequences of simple functions and finally to any non-negative function $f : E \to \mathbb{R}$ satisfying $\int_E (f(x))^\alpha \mu(dx) < \infty$. Based on the extremal integral representation of max-stable processes Stoev (2008) establishes conditions under which a max-stable process is ergodic. Wang et al. (2011) extend these results to a

spatial setting. In the following, let $\tau_{(h_1,\dots,h_d,u)}$ denote the multiparameter shift-operator. In accordance with the definitions and results in Wang et al. (2011), we define ergodic and mixing space-time processes.

Definition 4.1. *Let* $\left\{\eta(\boldsymbol{s},t), \boldsymbol{s}\in\mathbb{R}^d, t\in[0,\infty)\right\}$ *be a strictly stationary space-time process. The process is called*

(a) ergodic*, if for all* $A,B\in\sigma\left\{\eta(\boldsymbol{s},t), \boldsymbol{s}\in\mathbb{R}^d, t\in[0,\infty)\right\}$

$$\lim_{m_1,\dots,m_d,T\to\infty}\frac{1}{m_1\cdots m_d T}\sum_{h_1=1}^{m_1}\cdots\sum_{h_d=1}^{m_d}\sum_{u=1}^{T}P\left(A\cap\tau_{(h_1,\dots,h_d,u)}(B)\right)=P(A)P(B), \tag{14}$$

where $m_1,\dots,m_d,T\to\infty$ *means that each individual component of* $(m_1,\dots,m_k,T)$ *tends to infinity.*

(b) mixing*, if*

$$\lim_{n\to\infty}P\left(A\cap\tau_{(s_{1,n},\dots,s_{d,n},t_n)}(B)\right)=P(A)P(B), \tag{15}$$

for all sequences $\{(s_{1,n},\dots,s_{d,n},t_n),n\in\mathbb{N}\}$ *with* $\max\{|s_{1,n}|,\dots,|s_{d,n}|,|t_n|\}\to\infty.$

Note in (14) that in contrast to the ergodic theorem in Wang et al. (2011), the number of terms in each sum is not equal, since we have an additional sum for the time component. Using Theorem 6.1.2 in Krengel (1985), we can relate the conventional definition of ergodicity to the one given above. We focus on max-stable processes with extremal integral representation

$$\eta(s_1,\dots,s_d,t)=\int_E^{e} U_{(s_1,\dots,s_d,t)}(f)dM_1, \tag{16}$$

where $U_{(s_1,\dots,s_d,t)}:L^1(\mu)\to L^1(\mu)$ given by $U_{(s_1,\dots,s_d,t)}(f)=f\circ\tau_{(s_1,\dots,s_d,t)}$ is a group of max-linear automorphisms with $U_{(0,\dots,0,0)}(f)=f$, M_1 is an independently scattered 1−Fréchet random sup-measure with control measure μ, where (E,μ) can be chosen as the standard Lebesgue space $(\mathbb{R},\lambda)$. The following result is a direct extension of the uniparameter theorem established in Stoev (2008), Theorem 3.4, and its multiparameter counterpart:

Proposition 4.2 (Wang et al. (2011), Theorem 5.3). *The max-stable process defined in* (16) *is mixing, if and only if*

$$\int_E U_{(s_{1,n},\dots,s_{d,n},t_n)}(f)\wedge U_{(0,\dots,0,0)}(f)d\mu=\int_E U_{(s_{1,n},\dots,s_{d,n},t_n)}(f)\wedge f d\mu\to 0, \tag{17}$$

for all sequences $\{(s_{1,n},\dots,s_{d,n},t_n)\}$ *with* $\max\{|s_{1,n}|,\dots,|s_{d,n}|,|t_n|\}\to\infty$ *as* $n\to\infty$.

Wang et al. (2011) showed that the ergodic theorem stated above holds for mixing max-stable processes with extremal integral representation (16) in the case of $T=m$. The extension to the multiparameter case where $T\neq m$ is a simple generalization using Theorem 6.1.2 in Krengel (1985), which is a multiparameter extension of the Ackoglus ergodic theorem. Ergodic properties of Brown-Resnick processes have been studied for the uniparameter case in Stoev and Taqqu (2005) and Wang and Stoev (2010). The Brown-Resnick process (2) has a stochastic representation

$$\left\{\int_E^{e}\exp\left\{W(\boldsymbol{s},t)-\delta(\boldsymbol{s},t)\right\}dM_1,\quad \boldsymbol{s}\in\mathbb{R}^d,t\in[0,\infty)\right\}, \tag{18}$$

where M_1 is a random 1-Fréchet sup-measure on the probability space $(\Omega, \mathcal{E}, \mathbb{P})$ on which the Gaussian process W is defined. The intensity is $\mathbb{P}$, the probability measure which defines the Gaussian process W. We summarize the results in the following proposition.

PROPOSITION 4.3. *If δ satisfies* (4)*, the Brown-Resnick process given above in* (18) *is mixing in space and time. The strong law of large numbers holds: for every measurable function $g : \mathbb{R} \to \mathbb{R}$ such that $\mathbb{E}\left[|g(\eta(\boldsymbol{s}_1, 1))|\right] < \infty$;*

$$\frac{1}{m^d T} \sum_{i_1=1}^{m} \cdots \sum_{i_d=1}^{m} \sum_{t=1}^{T} g(\eta((i_1, \ldots, i_d), t)) = \sum_{\boldsymbol{s} \in S_m} \sum_{t=1}^{T} g(\eta(\boldsymbol{s}, t))$$
$$\overset{a.s.}{\to} \mathbb{E}\left[g(\eta((1, \ldots, 1), 1))\right] = \mathbb{E}\left[\eta(\boldsymbol{s}_1, 1)\right] \quad m, T \to \infty.$$

4.2. Consistency for large m and T

In the following we show that the pairwise likelihood estimate resulting from maximizing (11) for the Brown-Resnick process (2) is strongly consistent.

THEOREM 4.4. *Assume that the correlation function ρ satisfies Condition 2.3 such that $\delta(\boldsymbol{h}, u) = \theta_1 \|\boldsymbol{h}\|^{\alpha_1} + |u|^{\alpha_2}$ with parameter vector $\boldsymbol{\psi} = (\theta_1, \alpha_1, \theta_2, \alpha_2)$. Suppose further that the true parameter vector $\boldsymbol{\psi}^* = (\theta_1^*, \alpha_1^*, \theta_2^*, \alpha_2^*)$ lies in a compact set Ψ, which does not contain $\mathbf{0}$ and which satisfies for some $c > 0$*

$$\Psi \subseteq \{\min\{\theta_1, \theta_2\} > c, \alpha_1, \alpha_2 \in (0, 2]\}. \tag{19}$$

Assume also that the identifiability condition

$$\boldsymbol{\psi} = \widetilde{\boldsymbol{\psi}} \quad \Leftrightarrow \quad f_{\boldsymbol{\psi}}(\eta(\boldsymbol{s}_1, t_1), \eta(\boldsymbol{s}_2, t_2)) = f_{\widetilde{\boldsymbol{\psi}}}(\eta(\boldsymbol{s}_1, t_1), \eta(\boldsymbol{s}_2, t_2)), \tag{20}$$

is satisfied for all $\boldsymbol{s}_1, \boldsymbol{s}_2 \in S_m, t_1, t_2 \in \{t_1, \ldots, t_T\}$. It then follows that the pairwise likelihood estimate

$$\hat{\boldsymbol{\psi}} = \underset{\boldsymbol{\psi} \in \Psi}{\arg\max}\, PL^{(m,T)}(\boldsymbol{\psi}) \tag{21}$$

for observations from the Brown-Resnick process (2) *is strongly consistent, i.e., $\hat{\boldsymbol{\psi}} \overset{a.s.}{\to} \boldsymbol{\psi}^*$ as $m, T \to \infty$.*

REMARK 4.5. *For the identifiability condition* (20) *we consider different cases according to the maximal space-time lag (r, p) included in the composite likelihood. The pairwise density depends on the spatial distance $\boldsymbol{h}$ and the time lag u only through the function $\delta(\boldsymbol{h}, u) = \theta_1 \|\boldsymbol{h}\|^{\alpha_1} + \theta_2 |u|^{\alpha_2}$. For specific combinations of (r, p) not all parameters are identifiable. Strong consistency still holds for the remaining parameters. Table 1 lists the various scenarios.*

PROOF (THEOREM 4.4). To show strong consistency of the estimates (21) we follow the method of Wald (1946). From (11) we show

$$\frac{1}{m^d T} PL^{(m,T)}(\boldsymbol{\psi}) = \frac{1}{m^d T} \Big(\sum_{\boldsymbol{s} \in S_m} \sum_{t=1}^{T} g_{\boldsymbol{\psi}}\left(\boldsymbol{s}, t; r, p\right) - \mathcal{R}^{(m,T)}(\boldsymbol{\psi}) \Big) \overset{a.s.}{\to} PL(\psi),$$

as $m, T \to \infty$, where $PL(\psi) := \mathbb{E}\left[g_{\boldsymbol{\psi}}(\boldsymbol{s}_1, 1; r, p)\right]$ and $g_{\boldsymbol{\psi}}$ and $\mathcal{R}^{(m,T)}(\boldsymbol{\psi})$ are defined in (12) and (13), respectively.

We use the following three steps.

Table 1. Identifiable parameters corresponding to different maximal space-time lags (r,p) included in the pairwise likelihood function.

Maximal spatial lag r	Maximal temporal lag p	Identifiable parameters
0	1	θ_2
0	>1	$\theta_2,\ \alpha_2$
1	0	θ_1
>1	0	$\theta_1,\ \alpha_1$
1	1	$\theta_1,\ \theta_2$
1	>1	$\theta_1,\ \theta_2,\ \alpha_2$
>1	1	$\theta_1,\ \alpha_1,\ \theta_2$
>1	>1	$\theta_1,\ \alpha_1,\ \theta_2,\ \alpha_2$

(C1) Strong law of large numbers: Uniformly on the compact parameter space Ψ,

$$\frac{1}{m^d T}\sum_{\boldsymbol{s}\in S_m}\sum_{t=1}^{T} g_{\boldsymbol{\psi}}\left(\boldsymbol{s},t;r,p\right) \xrightarrow{a.s.} PL(\boldsymbol{\psi}) = \mathbb{E}\left[g_{\boldsymbol{\psi}}(\boldsymbol{s}_1,1;r,p)\right], \quad m,T\to\infty,$$

(C2) $\frac{1}{m^d T}\mathcal{R}^{(m,T)}(\boldsymbol{\psi}) \overset{a.s.}{\to} 0, \quad m,T\to\infty$, and

(C3) the limit function $PL(\boldsymbol{\psi})$ in (C1) is uniquely maximized at the true parameter vector $\boldsymbol{\psi}^* \in \boldsymbol{\Psi}$.

We first prove (C1). For fixed $\boldsymbol{\psi}\in\Psi$ the convergence in (C1) follows immediately from Proposition 4.3 together with the fact that $g_{\boldsymbol{\psi}}$ in (12) is a measurable function of lagged versions of $\eta(\boldsymbol{s},t)$. To prove uniform convergence we have from (3) for $x_1,x_2>0$

$$\begin{aligned}
\log f_{\boldsymbol{\psi}}(x_1,x_2) &= -V(x_1,x_2) + \log(V_1(x_1,x_2)V_2(x_1,x_2) - V_{12}(x_1,x_2)),\\
V(x_1,x_2) &= \Phi(q_1)/x_1 + \Phi(q_2)/x_2,\\
V_1(x_1,x_2) &= \frac{\partial V(x_1,x_2)}{\partial x_1}, \qquad V_2(x_1,x_2) = \frac{\partial V(x_1,x_2)}{\partial x_2}, \qquad V_{12}(x_1,x_2) = \frac{\partial^2 V(x_1,x_2)}{\partial x_1 \partial x_2},
\end{aligned}$$

and

$$q_1 = q_1(x_1,x_2) = \frac{\log(x_2/x_1)}{2\sqrt{\delta(\boldsymbol{h},u)}} + \sqrt{\delta(\boldsymbol{h},u)} \quad\text{and}\quad q_2 = q_2(x_1,x_2) = \frac{\log(x_1/x_2)}{2\sqrt{\delta(\boldsymbol{h},u)}} + \sqrt{\delta(\boldsymbol{h},u)}.$$

For $x_1,x_2>0$ the log-density $\log f_{\boldsymbol{\psi}}(x_1,x_2)$ can be bounded as follows.

$$\begin{aligned}
|\log f_{\boldsymbol{\psi}}(x_1,x_2)| &= |-V(x_1,x_2) + \log(V_1(x_1,x_2)V_2(x_1,x_2) - V_{12}(x_1,x_2)|\\
&\le |\Phi(q_1)/x_1| + |\Phi(q_2)/x_2| + |V_1(x_1,x_2)V_2(x_1,x_2) - V_{12}(x_1,x_2)|\\
&\le \frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_1^2x_2^2} + \frac{1}{2\sqrt{\delta(\boldsymbol{h},u)}}\left(\frac{1}{x_1^2x_2^2} + \frac{1}{x_1^3x_2} + \frac{1}{x_1^2x_2^2} + \frac{1}{x_1x_2^3} + \frac{1}{x_1^2x_2} + \frac{1}{x_1x_2^2}\right)\\
&\quad + \frac{1}{4\delta(\boldsymbol{h},u)}\left(\frac{1}{x_1^2x_2^2} + \frac{1}{x_1^3x_2} + \frac{1}{x_1x_2^3} + \frac{1}{x_1^2x_2^2} + \left|\frac{q_1}{x_1^2x_2} + \frac{q_2}{x_1x_2^2}\right|\right),
\end{aligned}$$

where $\Phi(\cdot)\le 1$ was used. Finally note that

$$\frac{q_1}{4\delta(\boldsymbol{h},u)x_1^2x_2} = \frac{\log(x_2/x_1) + 2\delta(\boldsymbol{h},u)}{8(\delta(\boldsymbol{h},u))^{3/2}x_1^2x_2} \le \frac{1}{8(\delta(\boldsymbol{h},u))^{3/2}x_1^3} + \frac{1}{4\sqrt{\delta(\boldsymbol{h},u)}x_1x_2^2}.$$

Since the marginal distributions of the Brown-Resnick process (2) are assumed to be standard Fréchet, it follows that for every fixed location $\boldsymbol{s} \in S_m$ and fixed time point $t \in \{1, \ldots, T\}$ the random variable $1/\eta(\boldsymbol{s}, t)$ is standard exponentially distributed with all moments finite. Using Hölder's inequality, it follows that

$$\mathbb{E}\left[\left|\log f_{\boldsymbol{\psi}}(\eta(\boldsymbol{s}_1, t_1), \eta(\boldsymbol{s}_2, t_2))\right|\right] \leq K_1 + \frac{K_2}{2\sqrt{\delta(\boldsymbol{h}, u)}} + \frac{K_3}{4\delta(\boldsymbol{h}, u)} + \frac{K_4}{8(\delta(\boldsymbol{h}, u))^{3/2}},$$

where $K_1, K_2, K_3, K_4 > 0$ are finite constants. Since the parameter space Ψ is assumed to be compact and together with assumption (19), δ can be bounded away from zero, i.e.,

$$\delta(\boldsymbol{h}, u) \geq \min\{\theta_1, \theta_2\}\left(\|\boldsymbol{h}\|^{\alpha_1} + |u|^{\alpha_2}\right) > c(\|\boldsymbol{h}\|^{\alpha_1} + |u|^{\alpha_2}) > \tilde{c} > 0, \tag{22}$$

where $\tilde{c} > 0$ is independent of the parameters. Therefore,

$$\mathbb{E}\left[\left|\log f_{\boldsymbol{\psi}}(\eta(\boldsymbol{s}_1, t_1), \eta(\boldsymbol{s}_2, t_2))\right|\right] < K_1 + \frac{K_2}{2\sqrt{\tilde{c}}} + \frac{K_3}{4\tilde{c}} + \frac{K_4}{8\tilde{c}^{3/2}} =: K_5 < \infty, \tag{23}$$

where $K_5 > 0$. From (22) and (23) it follows that

$$\mathbb{E}\left[\sup_{\boldsymbol{\psi} \in \boldsymbol{\Psi}} \left|\log f_{\boldsymbol{\psi}}(\eta(\boldsymbol{s}_1, 1), \eta(\boldsymbol{s}_1 + \boldsymbol{h}, 1 + u))\right|\right] < \infty,$$

which implies $\mathbb{E}\left[\sup_{\boldsymbol{\psi} \in \boldsymbol{\Psi}} |g_{\boldsymbol{\psi}}(\boldsymbol{s}_1, 1, r, p)|\right] < \infty$. By Theorem 2.7 in Straumann (2004) uniform convergence in (C1) follows.

Turning to (C2), note from (13) that by similar arguments as above

$$\begin{aligned}
&\mathbb{E}\left[\left|\frac{1}{m^d T} \mathcal{R}^{(m,T)}(\boldsymbol{\psi})\right|\right] \\
&\leq \frac{1}{m^d T} \sum_{\boldsymbol{s} \in S_m} \sum_{\substack{\boldsymbol{h} \in \mathcal{H}_r \\ \boldsymbol{s}+\boldsymbol{h} \notin S_m}} \sum_{t=1}^{T} \sum_{\substack{u=1 \\ t+u>T}}^{p} \mathbb{E}\left[\left|\log f_{\boldsymbol{\psi}}(\eta(\boldsymbol{s}, t), \eta(\boldsymbol{s} + \boldsymbol{h}, t + u))\right|\right] \\
&\leq \frac{1}{m^d T} \sum_{\boldsymbol{s} \in S_m} \sum_{\substack{\boldsymbol{h} \in \mathcal{H}_r \\ \boldsymbol{s}+\boldsymbol{h} \notin S_m}} \sum_{t=1}^{T} \sum_{\substack{u=1 \\ t+u>T}}^{p} K_5 \leq \frac{K_5 K_6}{mT} \to 0, \quad m, T \to \infty,
\end{aligned}$$

where we used the bound derived in (23) and the fact that the number of space-time points in the boundary is of order m^{d-1} (independent of T) and, therefore, can be bounded by $K_6 m^{d-1}$ with $K_6 > 0$ a constant independent of m and T.

We denote by $\mathcal{B}_{m,T}$ the set of "boundary" points, i.e.,

$$\mathcal{B}_{m,T} = \{\boldsymbol{s} \in S_m : \boldsymbol{s} + \boldsymbol{h} \notin \mathcal{H}_r\} \times \{t \in \{1, \ldots, T\} : t + u > T\}.$$

Then,

$$\mathcal{R}^{(m,T)}(\boldsymbol{\psi}) = \sum_{\boldsymbol{h} \in \mathcal{H}_r} \sum_{u=1}^{p} \sum_{(\boldsymbol{s}, t) \in \mathcal{B}_{m,T}} \log f_{\boldsymbol{\psi}}(\eta(\boldsymbol{s}, t), \eta(\boldsymbol{s} + \boldsymbol{h}, t + u)).$$

By Proposition 4.3 and (23) it follows uniformly on Ψ, that

$$\sum_{\boldsymbol{h}\in\mathcal{H}_r}\sum_{u=1}^{p}\frac{1}{|\mathcal{B}_{m,T}|}\sum_{(\boldsymbol{s},t)\in\mathcal{B}_{m,T}}\log f_{\boldsymbol{\psi}}(\eta(\boldsymbol{s},t),\eta(\boldsymbol{s}+\boldsymbol{h},t+u))$$
$$\to\mathbb{E}\left[\sum_{\boldsymbol{h}\in\mathcal{H}_r}\sum_{u=1}^{p}\log f_{\boldsymbol{\psi}}(\eta(\boldsymbol{s}_1,1),\eta(\boldsymbol{s}_1+\boldsymbol{h},1+u))\right],\quad m,T\to\infty.$$

Therefore,

$$\frac{1}{m^dT}\mathcal{R}^{(m,T)}(\boldsymbol{\psi})\le\frac{K_6}{mT}\sum_{\boldsymbol{h}\in\mathcal{H}_r}\sum_{u=1}^{p}\frac{1}{|\mathcal{B}_{m,T}|}\sum_{(\boldsymbol{s},t)\in\mathcal{B}_{m,T}}\log f_{\boldsymbol{\psi}}(\eta(\boldsymbol{s},t),\eta(\boldsymbol{s}+\boldsymbol{h},t+u))\stackrel{a.s.}{\longrightarrow}0,$$

since $\mathbb{E}\left[|\log f_{\boldsymbol{\psi}}(\eta(\boldsymbol{s},t),\eta(\boldsymbol{s}+\boldsymbol{h},t+u))|\right]<\infty$. This proves (C2).

To prove (C3), note that by Jensen's inequality

$$\mathbb{E}\left[\log\left(\frac{f_{\boldsymbol{\psi}}(x_1,x_2)}{f_{\boldsymbol{\psi}^*}(x_1,x_2)}\right)\right]\le\log\left(\mathbb{E}\left[\frac{f_{\boldsymbol{\psi}}(x_1,x_2)}{f_{\boldsymbol{\psi}^*}(x_1,x_2)}\right]\right)=0$$

and, hence,

$$PL(\boldsymbol{\psi})\le PL(\boldsymbol{\psi}^*)$$

for all $\boldsymbol{\psi}\in\Psi$. So, $\boldsymbol{\psi}^*$ maximizes $PL(\boldsymbol{\psi})$ and is the unique optimum if and only if there is equality in Jensen's inequality. However, this is precluded by (20). □

5. Asymptotic normality of the pairwise likelihood estimates for regular grid observations

In order to prove asymptotic normality of the pairwise likelihood estimates resulting from maximizing (11) we need the following results for the pairwise log-density. The proofs can be found in Appendix A.

LEMMA 5.1. *Consider the Brown-Resnick process in* (2)*, where the underlying correlation function satisfies Condition 2.3. Further assume that all conditions from Theorem 4.4 hold.*

(1) The gradient of the bivariate log-density satisfies

$$\mathbb{E}\left[|\nabla_{\boldsymbol{\psi}}\log f_{\boldsymbol{\psi}}(\eta(\boldsymbol{s}_1,t_1),\eta(\boldsymbol{s}_2,t_2))|^3\right]<\infty$$

(2) The Hessian of the pairwise log-density satisfies

$$\mathbb{E}\left[\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\left|\nabla^2_{\boldsymbol{\psi}}\log f_{\boldsymbol{\psi}}(\eta(\boldsymbol{s}_1,t_1),\eta(\boldsymbol{s}_2,t_2))\right|\right]<\infty.$$

The absolute values of the vector in (1) and the matrix in (2) are perceived componentwise.

Assuming asymptotic normality of the pairwise score function $\nabla_{\boldsymbol{\psi}}PL^{(m,T)}(\boldsymbol{\psi})$ it is relatively routine to show that the pairwise likelihood estimates are asymptotically normal. We formulate the first result.

THEOREM 5.2. *Assume that the conditions of Theorem 4.4 hold. In addition, assume that a central limit theorem holds for the gradient of $g_{\boldsymbol{\psi}}$ defined in* (12) *in the following sense*

$$\frac{1}{m^{d/2}\sqrt{T}} \sum_{\boldsymbol{s}\in S_m} \sum_{t=1}^{T} \nabla_{\boldsymbol{\psi}} g_{\boldsymbol{\psi}^*}(\boldsymbol{s},t;r,p) \xrightarrow{d} \mathcal{N}(0,\Sigma),\ m,T\to\infty, \tag{24}$$

where $\boldsymbol{\psi}^*$ is the true parameter vector and Σ is some covariance matrix. Then it follows that the pairwise likelihood estimate in (21) *satisfies*

$$m^{d/2}\sqrt{T}(\hat{\boldsymbol{\psi}}-\boldsymbol{\psi}^*) \xrightarrow{d} \mathcal{N}(0,F^{-1}\Sigma(F^{-1})^T),\ m,T\to\infty,$$

where

$$F = \mathbb{E}\left[-\nabla^2_{\boldsymbol{\psi}} g_{\boldsymbol{\psi}^*}(\boldsymbol{s}_1,1;r,p)\right].$$

PROOF. We use a standard Taylor expansion of the pairwise score function around the true parameter vector and obtain

$$\begin{aligned}
m^{d/2}\sqrt{T}(\hat{\boldsymbol{\psi}}-\boldsymbol{\psi}^*) &= -\left(\frac{1}{m^dT}\nabla^2_{\boldsymbol{\psi}} PL^{(m,T)}(\tilde{\boldsymbol{\psi}})\right)^{-1}\left(\frac{1}{m^{d/2}\sqrt{T}}\nabla_{\boldsymbol{\psi}} PL^{(m,T)}(\boldsymbol{\psi}^*)\right)\\
&= -\left(\frac{1}{m^dT}\sum_{\boldsymbol{s}\in S_m}\sum_{t=1}^{T}\nabla^2_{\boldsymbol{\psi}} g_{\tilde{\boldsymbol{\psi}}}(\boldsymbol{s},t;r,p) - \frac{1}{m^dT}\nabla^2_{\boldsymbol{\psi}}\mathcal{R}^{(m,T)}(\tilde{\boldsymbol{\psi}})\right)^{-1}\\
&\quad\times\left(\frac{1}{m^{d/2}\sqrt{T}}\sum_{\boldsymbol{s}\in S_m}\sum_{t=1}^{T}\nabla_{\boldsymbol{\psi}} g_{\boldsymbol{\psi}^*}(\boldsymbol{s},t;r,p) - \frac{1}{m^{d/2}\sqrt{T}}\nabla_{\boldsymbol{\psi}}\mathcal{R}^{(m,T)}(\boldsymbol{\psi}^*)\right)\\
&= -(I_1-I_2)^{-1}(J_1-J_2),
\end{aligned}$$

where $\tilde{\boldsymbol{\psi}}\in[\hat{\boldsymbol{\psi}},\boldsymbol{\psi}^*]$. By (24) J_1 converges weakly to a normal distribution with mean 0 and covariance matrix Σ. By using the same arguments as in the proof of Theorem 4.4 together with (24) we have that $J_2 \overset{P}{\to} 0$. Since the underlying space-time process in the likelihood function is mixing, it follows that the process $\left\{\nabla^2_{\boldsymbol{\psi}} g_{\boldsymbol{\psi}}(\boldsymbol{s},t;r,p), \boldsymbol{s}\in\mathbb{Z}^d, t\in\mathbb{N}\right\}$ is mixing as a measurable function of mixing and lagged processes. To prove the uniform convergence we verify that

$$\mathbb{E}\left[\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\left|\nabla^2_{\boldsymbol{\psi}} g_{\boldsymbol{\psi}}(\boldsymbol{s}_1,1;r,p)\right|\right]<\infty.$$

This follows immediately from Lemma 5.1. Putting this together with the fact that $\tilde{\boldsymbol{\psi}}\in[\hat{\boldsymbol{\psi}},\boldsymbol{\psi}^*]$, and because of the strong consistency of $\hat{\boldsymbol{\psi}}$, it follows that

$$I_1 \xrightarrow{a.s.} \mathbb{E}\left[\nabla^2_{\boldsymbol{\psi}} g_{\boldsymbol{\psi}^*}(\boldsymbol{s}_1,1;r,p)\right] =: -F.$$

Using the strong law of large numbers for $\left\{\nabla^2_{\boldsymbol{\psi}}\log f_{\boldsymbol{\psi}}(\eta(\boldsymbol{s},t),\eta(\boldsymbol{s}+\boldsymbol{h},t+u)\right\}$ it follows in the same way as in the proof of Theorem 4.4 that $I_2 \overset{a.s.}{\to} 0$ as $m,T\to\infty$. Combining these results, we obtain by Slutzky's lemma

$$m^{d/2}\sqrt{T}(\hat{\boldsymbol{\psi}}-\boldsymbol{\psi}^*) \xrightarrow{d} \mathcal{N}(0,F^{-1}\Sigma(F^{-1})^T),\quad m,T\to\infty.$$

□

In the next section we provide a sufficient condition for (24).

5.1. Asymptotic normality and α-mixing

In this section we consider asymptotic normality of the parameters estimates for the Brown-Resnick process in (2). Under the assumption of α-mixing of the random field the key is to show asymptotic normality for the pairwise score function. For an increasing time domain and fixed number of locations asymptotic normality of the pairwise likelihood estimates was shown in Huser and Davison (2012). The main difference between a temporal setting and a space-time setting is the definition of the α-mixing coefficients and the resulting assumptions to obtain a central limit theorem for the score function.

We apply the central limit theorem for random fields established in Bolthausen (1982) to the pairwise score function of the pairwise likelihood in our model. In a second step we verify the α-mixing conditions for the Brown-Resnick process (2), where the underlying correlation function satisfies Condition 2.3. First, we define the α-mixing coefficients in a space-time setting as follows. Define the distances

$$d((\boldsymbol{s}_1,t_1),(\boldsymbol{s}_2,t_2)) = \max\left\{\max_{1\le i\le d}|\boldsymbol{s}_1(i)-\boldsymbol{s}_2(i)|,|t_1-t_2|\right\}, \quad \boldsymbol{s}_1,\boldsymbol{s}_2\in\mathbb{Z}^d, t_1,t_2\in\mathbb{N},$$
$$d(\Lambda_1,\Lambda_2) = \inf\left\{d((\boldsymbol{s}_1,t_1),(\boldsymbol{s}_2,t_2)),(\boldsymbol{s}_1,t_1)\in\Lambda_1,(\boldsymbol{s}_2,t_2)\in\Lambda_2\right\}, \quad \Lambda_1,\Lambda_2\subset\mathbb{Z}^d\times\mathbb{N},$$

where $\boldsymbol{s}_k = (s_k(1),\dots,s_k(d)),\ k=1,2$. Let further $\mathcal{F}_{\Lambda_i} = \sigma\left\{\eta(\boldsymbol{s},t),(\boldsymbol{s},t)\in\Lambda_i\right\}$ for $i=1,2$. The mixing coefficients are defined for $k,l\in\mathbb{N}$ and $n\in\mathbb{N}\cup\{\infty\}$ by

$$\alpha_{k,l}(n) = \sup\left\{|P(A_1\cap A_2)-P(A_1)P(A_2)|:\ A_i\in\mathcal{F}_{\Lambda_i},|\Lambda_1|\le k,|\Lambda_2|\le l,d(\Lambda_1,\Lambda_2)\ge n\right\} \tag{25}$$

and depend on the sizes and the distance of the sets Λ_1 and Λ_2.

A space-time process is called *α-mixing*, if $\alpha_{k,l}(n)\to 0$ as $n\to\infty$ for all $k,l>0$. We assume that the process $\left\{\eta(\boldsymbol{s},t),\boldsymbol{s}\in\mathbb{Z}^d,t\in\mathbb{N}\right\}$ is α-mixing with mixing coefficients defined in (25), from which it follows that the space-time process

$$\left\{\nabla_{\boldsymbol{\psi}} g_{\psi}(\boldsymbol{s},t;r,p),\boldsymbol{s}\in\mathbb{Z}^d,t\in\mathbb{N}\right\}. \tag{26}$$

is α-mixing for all $\boldsymbol{\psi}\in\Psi$. We apply Bolthausen's central limit theorem this process. By adjusting the assumptions on the α-mixing coefficients we obtain the following proposition.

PROPOSITION 5.3. *We consider the Brown-Resnick process* (2) *with* $\delta(\boldsymbol{h},u)=\theta_1\|\boldsymbol{h}\|^{\alpha_1}+\theta_2|u|^{\alpha_2}$. *Assume, that the following conditions hold:*

(1) The process $\left\{(\eta(\boldsymbol{s},t),\boldsymbol{s}\in\mathbb{Z}^d,t\in\mathbb{N}\right\}$ *is α-mixing.*

(2) The α-mixing coefficients in (25)*satisfy*
$\sum_{n=1}^{\infty} n^d\alpha_{k,l}(n)<\infty$ *for* $k+l\le 4(|\mathcal{H}_r|+1)(p+1)$ *and* $\alpha_{(|\mathcal{H}_r|+1)(p+1),\infty}(n)=o(n^{-(d+1)})$.

(3) There exists some $\beta>0$ such that

$$\mathbb{E}\left[\left|\nabla_{\boldsymbol{\psi}} g_{\boldsymbol{\psi}^*}(\boldsymbol{s},t;r,p)\right|^{2+\beta}\right]<\infty \quad \textit{and}$$
$$\sum_{n=1}^{\infty} n^d\alpha_{(|\mathcal{H}_r|+1)(p+1),(|\mathcal{H}_r|+1)(p+1)}(n)^{\beta/(2+\beta)}<\infty.$$

Then,

$$\frac{1}{m^{d/2}\sqrt{T}} \sum_{\boldsymbol{s} \in S_m} \sum_{t=1}^{T} \nabla_{\boldsymbol{\psi}} g_{\boldsymbol{\psi}^*}(\boldsymbol{s}, t; r, p) \xrightarrow{d} \mathcal{N}(0, \Sigma), \;\; m, T \to \infty,$$

where $\Sigma = \sum_{\boldsymbol{s} \in \mathbb{Z}^d} \sum_{t \in \mathbb{N}} \mathbb{C}ov\left(\nabla_{\boldsymbol{\psi}} g_{\boldsymbol{\psi}^*}(\boldsymbol{s}_1, 1; r, p), \nabla_{\boldsymbol{\psi}} g_{\boldsymbol{\psi}^*}(\boldsymbol{s}, t; r, p)\right)$.

Recent work by Dombry and Eyi-Minko (2012) deals with strong mixing properties for max-stable random fields. By using a point process representation of max-stable processes together with coupling techniques, they show that the α-mixing coefficients can be bounded by a function of the tail dependence coefficient. A direct extension to the space-time setting gives the following lemma.

Lemma 5.4 (Dombry and Eyi-Minko (2012), Corollary 2.2). *Consider a stationary max-stable space-time process* $\left\{\eta(\boldsymbol{s}, t), \boldsymbol{s} \in \mathbb{Z}^d, t \in \mathbb{N}\right\}$ *with arbitrary tail dependence coefficient* $\chi(\boldsymbol{h}, u)$*. The* α*-mixing coefficients* (25) *satisfy*

$$\alpha_{k,l}(n) \leq kl \sup_{\max\{\|\boldsymbol{h}\|, |u|\} \geq n} \chi(\boldsymbol{h}, u) \quad \text{and} \quad \alpha_{k,\infty}(n) \leq k \sum_{\max\{\|\boldsymbol{h}\|, |u|\} \geq n} \chi(\boldsymbol{h}, u).$$

In the following we show that Proposition 5.3 applies for the Brown-Resnick process (2) with tail dependence coefficient χ as in (5). By using the inequality for the normal tail probability $1 - \Phi(x) = \overline{\Phi}(x) \leq e^{-x^2/2}$ for $x > 0$ it follows that

$$\begin{aligned}
\alpha_{k,l}(n) &\leq 4kl \sup_{\max\{\|\boldsymbol{h}\|, |u|\} \geq n} \left(1 - \Phi\left(\sqrt{\delta(\boldsymbol{h}, u)}\right)\right) \leq 4kl \sup_{\max\{\|\boldsymbol{h}\|, |u|\} \geq n} \exp\left\{-\frac{\delta(\boldsymbol{h}, u)}{2}\right\} \\
&= 4kl \sup_{\max\{\|\boldsymbol{h}\|, |u|\} \geq n} \exp\left\{-\frac{1}{2}(\theta_1 \|\boldsymbol{h}\|^{\alpha_1} + \theta_2 |u|^{\alpha_2})\right\} \\
&\leq 4kl \sup_{\max\{\|\boldsymbol{h}\|, |u|\} \geq n} \exp\left\{-\frac{1}{2} \min\{\theta_1, \theta_2\} \left(\max\{\|\boldsymbol{h}\|, |u|\}\right)^{\min\{\alpha_1, \alpha_2\}}\right\}.
\end{aligned}$$

For $n \to \infty$, the right hand side tends to zero for all $k, l \geq 0$. Thus, $\left\{\eta(\boldsymbol{s}, t), \boldsymbol{s} \in \mathbb{Z}^d, t \in \mathbb{N}\right\}$ is α-mixing. Furthermore, for $k + l \leq 4(|\mathcal{H}_r| + 1)(p + 1)$ the coefficients satisfy

$$\begin{aligned}
\sum_{n=1}^{\infty} n^d \alpha_{k,l}(n) &\leq 4kl \sum_{n=1}^{\infty} n^d \sup_{\max\{\|\boldsymbol{h}\|, |u|\} \geq n} \exp\left\{-\frac{1}{2}(\theta_1 \|\boldsymbol{h}\|^{\alpha_1} + \theta_2 |u|^{\alpha_2})\right\} \\
&\leq 4kl \sum_{n=1}^{\infty} n^d \exp\left\{-\frac{1}{2} \min\{\theta_1, \theta_2\} n^{\min\{\alpha_1, \alpha_2\}}\right\} < \infty.
\end{aligned}$$

In addition,

$$n^{d+1} \alpha_{(|\mathcal{H}_r|+1)(p+1), \infty}(n) \leq n^{d+1} (|\mathcal{H}_r| + 1)(p + 1) \sum_{x \geq n} \exp\left\{-\frac{1}{2} \min\{\theta_1, \theta_2\} x^{\min\{\alpha_1, \alpha_2\}}\right\},$$

where the right hand side converges to zero as $n \to \infty$, which finally proves (2). As for (3), from Lemma 5.1 and using $\beta = 1$ we know that

$$\mathbb{E}\left[|\nabla_{\boldsymbol{\psi}} g_{\boldsymbol{\psi}^*}(\boldsymbol{s}, t; r, p)|^{(2+\beta)}\right] < \infty.$$

By the same arguments as in the proof of (2) above the second condition in (3) holds.

By combining the above results with Theorem 5.2 we obtain asymptotic normality for the parameter estimates $\hat{\boldsymbol{\psi}}$ for an increasing number of space-time locations. We summarize this result as follows.

THEOREM 5.5. *Assume that the conditions of Theorem 4.4 hold. Then,*

$$(m^d T)^{1/2}(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}^*) \xrightarrow{d} \mathcal{N}(0, F^{-1}\Sigma(F^{-1})^\intercal), \quad m, T \to \infty,$$

where

$$F = \mathbb{E}\left[-\nabla^2_{\boldsymbol{\psi}} g_{\boldsymbol{\psi}^*}(\boldsymbol{s}_1, 1; r, p)\right]$$

and

$$\Sigma = \sum_{\boldsymbol{s}\in\mathbb{Z}^d} \sum_{t\in\mathbb{N}} \mathbb{C}ov\left(\nabla_{\boldsymbol{\psi}} g_{\boldsymbol{\psi}^*}(\boldsymbol{s}_1, 1; r, p), \nabla_{\boldsymbol{\psi}} g_{\boldsymbol{\psi}^*}(\boldsymbol{s}, t; r, p)\right).$$

REMARK 5.6. *Unfortunately, we cannot provide a closed form expression for the asymptotic covariance matrix. The matrix F is the expected Hessian matrix of the pairwise log-likelihood function and an estimate is given by its empirical version*

$$\hat{F} = -\sum_{\boldsymbol{s}\in S_m} \sum_{t=1}^{T} \sum_{\boldsymbol{h}\in\mathcal{H}_r} \sum_{u=1}^{p} \nabla^2_{\boldsymbol{\psi}} \log f_{\hat{\boldsymbol{\psi}}}(\eta(\boldsymbol{s}, t), \eta(\boldsymbol{s}+\boldsymbol{h}, t+u)),$$

which can be obtained numerically from the optimization routine used to maximize the pairwise likelihood function. The calculation of Σ or estimates for Σ seems to be a difficult task. We therefore rely on resampling methods like the bootstrap or on the jackknife for obtaining estimates of the variance and confidence regions. For example a block bootstrap procedure could be applied which approximates the distribution of $\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}$. The situation here is similar to the estimation of the extremogram, where bootstrap methods have been suggested to construct asymptotically correct confidence bands (see Davis and Mikosch (2009) and Davis et al. (2012)). The justification of resampling methods is the subject of another paper.

6. Extension to irregularly spaced locations

So far we have assumed that the spatial sampling locations lie on a regular grid. In the following we discuss two settings, where the locations are irregularly spaced.

6.1. Deterministic irregularly spaced lattice

One way to extend our results to irregularly spaced locations is to invoke the ideas in Bai et al. (2012) as adapted from Jenish and Prucha (2009). Let

$$D \subset \mathbb{R}^d \times [0, \infty) \times \mathbb{R}^d \times [0, \infty)$$

denote an infinitely countable lattice such that all elements of D have distances of at least $d_0 > 0$:

$$\|(\boldsymbol{s}_1, t_1, \boldsymbol{s}_2, t_2) - (\boldsymbol{s}_3, t_3, \boldsymbol{s}_4, t_4)\| > d_0$$

for any $(\boldsymbol{s}_1, t_1, \boldsymbol{s}_2, t_2), (\boldsymbol{s}_3, t_3, \boldsymbol{s}_4, t_4) \in D$, where $\|\cdot\|$ is an arbitrary norm. Note that D describes pairs of space-time locations. Further let $\{D_n : n \in \mathbb{N}\}$ be a sequence of arbitrary finite subsets of D satisfying $|D_n| \to \infty$ as $n \to \infty$, where $|\cdot|$ denotes the cardinality. In addition the sets D_n contain only pairs of space-time locations for which $\|\boldsymbol{s}_1 - \boldsymbol{s}_2\| \leq r$, $|t_1 - t_2| \leq p$ and at least one of the lags $\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|$ and $|t_1 - t_2|$ is larger than zero. The pairwise log-likelihood function (see general definition in (10)) is now given by

$$PL^{(n)}(\boldsymbol{\psi}) = \sum_{(\boldsymbol{s}_1, t_1, \boldsymbol{s}_2, t_2) \in D_n} \log f_{\boldsymbol{\psi}}(\eta(\boldsymbol{s}_1, t_1), \eta(\boldsymbol{s}_2, t_2)).$$

Denote by $\mathcal{S} \times \mathcal{T}$ the sampling region with cardinality $|\mathcal{S} \times \mathcal{T}| = n$. To prove consistency of the pairwise likelihood estimates Theorems 2 and 3 in Jenish and Prucha (2009) are used to show that the pairwise log-likelihood function satisfies a law of large numbers. Using the same arguments as in Theorem 4.4 in Section 4.2 we can show that the estimates are consistent, i.e.,

$$\hat{\boldsymbol{\psi}} \xrightarrow{P} \boldsymbol{\psi}^*, \quad n \to \infty.$$

Compared to the conditions needed to prove Theorem 4.4 the stronger assumption that the pairwise log-density is uniformly $L_{1+\delta}$ integrable (for a definition see Section 3.1 in Bai et al. (2012)) has to be shown. For the Brown-Resnick process (2) and the assumptions in Theorem 4.4 this can be verified in a similar fashion as in the derivation of the upper bound for the log-density in the proof of Theorem 4.4, (C1).

To show asymptotic normality of the estimates, Bai et al. (2012) use Theorem 1 in Jenish and Prucha (2009) assuming eight conditions, where the first two define the setting for the space-time locations. For the Brown-Resnick process with $\delta(\boldsymbol{h}, u) = \theta_1 \|\boldsymbol{h}\|^{\alpha_1} + \theta_2 |u|^{\alpha_2}$ and together with the assumptions in Theorem 4.4 all conditions except their Assumptions (7) and (8) can be shown. For our setting, Assumptions (7) and (8) in Bai et al. (2012) are equivalent to

$$n\mathbb{V}ar(\nabla_{\boldsymbol{\psi}} PL^n(\boldsymbol{\psi})) \to \Sigma \quad \text{and} \quad \mathbb{E}\left[\nabla^2_{\boldsymbol{\psi}} PL^n(\boldsymbol{\psi})\right] \to F, \quad n \to \infty, \tag{27}$$

for all $\boldsymbol{\psi} \in \Psi$, where F and Σ are positive definite matrices. Note that by using Theorem 2 and 3 in Jenish and Prucha (2009) together with the arguments in the proof of Theorem 5.2 we can show the first part of Assumption 8 in Bai et al. (2012), i.e.

$$\sup_{\boldsymbol{\psi} \in \Psi} \left| \nabla^2_{\boldsymbol{\psi}} PL^{(n)}(\boldsymbol{\psi}) - \mathbb{E}\left[\nabla^2_{\boldsymbol{\psi}} PL^{(n)}(\boldsymbol{\psi})\right] \right| \to 0.$$

Altogether, in contrast to the regular grid case we have two additional assumptions (27), which must be checked for the sampling scheme employed.

6.2. Random locations generated by a Poisson process

In the following we assume that locations are taken at random locations. For simplicity we consider a spatial random field and no time component. We use the ideas and results in Karr (1986) and Li et al. (2008) to redefine the pairwise likelihood function and to show that the resulting estimates are asymptotically normal. Let $\{\eta(\boldsymbol{s}), \boldsymbol{s} \in \mathbb{R}^d\}$ be the max-stable random field defined analogously to (2), where δ is now given by $\delta(\boldsymbol{h}) = \theta_1 \|\boldsymbol{h}\|^{\alpha_1}$, and let N denote a Poisson random measure with mean measure $\nu\lambda(\cdot)$, where λ is Lebesgue measure, i.e., N is a stationary homogeneous Poisson process with intensity parameter ν which is assumed to be known. As before, we denote by S_m the set of possible spatial locations where the process is observed. Suppose that the set S_m is convex and compact. Following Karr (1986) we define

$$N^{(2)}(d\boldsymbol{s}_1, d\boldsymbol{s}_2) = N(d\boldsymbol{s}_1)N(d\boldsymbol{s}_2)\mathbb{1}_{\{\boldsymbol{s}_1 \neq \boldsymbol{s}_2\}}, \quad \boldsymbol{s}_1, \boldsymbol{s}_2 \in S_m.$$

The pairwise log-likelihood function is now given by

$$PL^{(m)}(\boldsymbol{\psi}) = \int\limits_{S_m}\int\limits_{S_m} w(\boldsymbol{s}_1 - \boldsymbol{s}_2)\log f_{\boldsymbol{\psi}}(\eta(\boldsymbol{s}_1), \eta(\boldsymbol{s}_2))N^{(2)}(d\boldsymbol{s}_1, d\boldsymbol{s}_2), \tag{28}$$

where w is some positive weight function. We adapt Lemma A.2 Li et al. (2008) to show that the pairwise score function satisfies a central limit theorem. The variance calculation is different from Li et al. (2008) in the sense that we investigate the pairwise score function instead of a kernel smoothed estimator of a covariance function, which requires different arguments.

Lemma 6.1. *Assume that locations are generated by a stationary homogeneous Poisson process N with intensity ν. Suppose further that the following conditions hold.*

(a) The sets S_m satisfy

$$\lambda(S_m) = O(m^d), \quad \textit{and} \quad \lambda(\partial S_m) = O(m^{d-1}),$$

where λ denotes the Lebesgue measure and ∂S_m is the boundary of S_m.

(b) The random field $\{\eta(\boldsymbol{s}), \boldsymbol{s} \in \mathbb{R}^d\}$ is α-mixing with mixing coefficients as in (25) *for which hold*

$$\sup_{k\in\mathbb{N}} \frac{1}{k^2}\alpha_{k,k}(r) = O(r^{-\epsilon}), \quad \textit{for some } \epsilon > 0.$$

(c) Let w be a positive weight function satisfying

$$\int\limits_{\mathbb{R}^d} w(\boldsymbol{u})d\boldsymbol{u} < \infty \quad \textit{and} \quad \int\limits_{\mathbb{R}^d} w(\boldsymbol{u})^2 d\boldsymbol{u} < \infty.$$

(d) The gradient of the bivariate log-density satisfies

$$\mathbb{E}\left[|\nabla_{\boldsymbol{\psi}} \log f_{\boldsymbol{\psi}^*}(\eta(\boldsymbol{s}_1), \eta(\boldsymbol{s}_2))|\right] < \infty \quad \textit{and} \quad \mathbb{E}\left[|\nabla_{\boldsymbol{\psi}} \log f_{\boldsymbol{\psi}^*}(\eta(\boldsymbol{s}_1), \eta(\boldsymbol{s}_2))|^2\right] < \infty.$$

(e) Define $S_m - S_m$ as the the set of all pairwise differences in S_m and let

$$A_{\boldsymbol{\psi}^*}(\boldsymbol{s}_1, \boldsymbol{s}_2) = \nabla_{\boldsymbol{\psi}} \log f_{\boldsymbol{\psi}^*}(\eta(\boldsymbol{s}_1), \eta(\boldsymbol{s}_2)).$$

Then,

$$\begin{aligned}
&\iiint\limits_{(S_m - S_m)^3} w(\boldsymbol{v}_1)w(\boldsymbol{v}_3 - \boldsymbol{v}_2)\mathbb{E}\left[A_{\boldsymbol{\psi}^*}(\boldsymbol{v}_1, \boldsymbol{0})A_{\boldsymbol{\psi}^*}(\boldsymbol{v}_2, \boldsymbol{v}_3)\right] \\
&\quad \times \frac{\lambda(S_m \cap (S_m + \boldsymbol{v}_1) \cap (S_m + \boldsymbol{v}_2) \cap (S_m + \boldsymbol{v}_3))}{\lambda(S_m)} d\boldsymbol{v}_1 d\boldsymbol{v}_2 d\boldsymbol{v}_3 \\
&\to \iiint\limits_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d} w(\boldsymbol{v}_1)w(\boldsymbol{v}_3 - \boldsymbol{v}_2)\mathbb{E}\left[A_{\boldsymbol{\psi}^*}(\boldsymbol{v}_1, \boldsymbol{0})A_{\boldsymbol{\psi}^*}(\boldsymbol{v}_2, \boldsymbol{v}_3)\right] d\boldsymbol{v}_1 d\boldsymbol{v}_2 d\boldsymbol{v}_3, \quad m \to \infty.
\end{aligned}$$

(f) There exists $\beta > 0$, such that

$$\sup_{m>0} \mathbb{E}\left[|\sqrt{\lambda(S_m)}\nabla_{\boldsymbol{\psi}} PL^{(m)}(\boldsymbol{\psi}^*)|^{2+\beta}\right] < C_\beta$$

for some constant $C_\beta > 0$.

Then

$$\frac{1}{\sqrt{\lambda(S_m)}}\nabla_{\boldsymbol{\psi}} PL^{(m)}(\boldsymbol{\psi}^*) \xrightarrow{d} \mathcal{N}(0,\Sigma), \quad m \to \infty,$$

where

$$\Sigma = \frac{2}{\nu^2}\int_{\mathbb{R}^d} w^2(\boldsymbol{v})\mathbb{E}\left[A^2_{\boldsymbol{\psi}^*}(\boldsymbol{v},\boldsymbol{0})\right]d\boldsymbol{v} + \frac{4}{\nu}\mathbb{V}ar\left(\int_{\mathbb{R}^d} w(\boldsymbol{u})A_{\boldsymbol{\psi}^*}(\boldsymbol{u},\boldsymbol{0})d\boldsymbol{u}\right)$$
$$+ \iiint_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d} w(\boldsymbol{v}_1)w(\boldsymbol{v}_3-\boldsymbol{v}_2)\mathbb{C}ov\left[A_{\boldsymbol{\psi}^*}(\boldsymbol{v}_1,\boldsymbol{0}),A_{\boldsymbol{\psi}^*}(\boldsymbol{v}_2,\boldsymbol{v}_3)\right]d\boldsymbol{v}_1 d\boldsymbol{v}_2 d\boldsymbol{v}_3. \tag{29}$$

Proof. We calculate the expectation and the variance of the pairwise score function $\nabla_{\boldsymbol{\psi}} PL^{(m)}(\boldsymbol{\psi}^*)$. By using standard properties of the Poisson process it follows that

$$\mathbb{E}\left[\nabla_{\boldsymbol{\psi}} PL^{(m)}(\boldsymbol{\psi}^*)\right] = \mathbb{E}\left[\iint_{S_m\times S_m} w(\boldsymbol{s}_1-\boldsymbol{s}_2)A_{\boldsymbol{\psi}^*}(\boldsymbol{s}_1,\boldsymbol{s}_2)N^{(2)}(d\boldsymbol{s}_1,d\boldsymbol{s}_2)\right]$$
$$= \nu^2 \iint_{S_m\times S_m} w(\boldsymbol{s}_1-\boldsymbol{s}_2)\mathbb{E}\left[A_{\boldsymbol{\psi}^*}(\boldsymbol{s}_1,\boldsymbol{s}_2)\right]d\boldsymbol{s}_1 d\boldsymbol{s}_2 = 0,$$

To calculate the variance note that

$$\lambda(S_m)^{-1}\mathbb{V}ar(\nu^{-2}\nabla_{\boldsymbol{\psi}} PL^{(m)}(\boldsymbol{\psi}^*))$$
$$= \lambda(S_m)^{-1}\nu^{-4}\iiiint_{S_m^4} w(\boldsymbol{s}_1-\boldsymbol{s}_2)w(\boldsymbol{s}_3-\boldsymbol{s}_4)\mathbb{E}\left[A_{\boldsymbol{\psi}^*}(\boldsymbol{s}_1,\boldsymbol{s}_2)A_{\boldsymbol{\psi}^*}(\boldsymbol{s}_3,\boldsymbol{s}_4)\right]$$
$$\times \mathbb{E}\left[N^{(2)}(d\boldsymbol{s}_1,d\boldsymbol{s}_2)N^{(2)}(d\boldsymbol{s}_3,d\boldsymbol{s}_4)\right].$$

The expectation $\mathbb{E}\left[N^{(2)}(d\boldsymbol{s}_1,d\boldsymbol{s}_2)N^{(2)}(d\boldsymbol{s}_3,d\boldsymbol{s}_4)\right]$ can be calculated by using standard properties of the Poisson process leading to seven terms as stated in Karr (1986) or Li et al. (2008). In the limit, some of these terms are the same. We calculate the three representative different parts. We denote by $\epsilon_{\boldsymbol{s}}(\cdot)$ the Dirac measure.

$$\lambda(S_m)^{-1}\nu^{-4}\iiiint_{S_m^4} w(\boldsymbol{s}_1-\boldsymbol{s}_2)w(\boldsymbol{s}_3-\boldsymbol{s}_4)\mathbb{E}\left[A_{\boldsymbol{\psi}^*}(\boldsymbol{s}_1,\boldsymbol{s}_2)A_{\boldsymbol{\psi}^*}(\boldsymbol{s}_3,\boldsymbol{s}_4)\right]\nu^2 d\boldsymbol{s}_1 d\boldsymbol{s}_2 \epsilon_{\boldsymbol{s}_1}(d\boldsymbol{s}_3)\epsilon_{\boldsymbol{s}_2}(d\boldsymbol{s}_4)$$
$$= \lambda(S_m)^{-1}\nu^{-2}\iint_{S_m^2} w^2(\boldsymbol{s}_1-\boldsymbol{s}_2)\mathbb{E}\left[A^2_{\boldsymbol{\psi}^*}(\boldsymbol{s}_1-\boldsymbol{s}_2,\boldsymbol{0})\right]d\boldsymbol{s}_1 d\boldsymbol{s}_2$$
$$= \nu^{-2}\int_{S_m-S_m} w^2(\boldsymbol{u})\mathbb{E}\left[A^2_{\boldsymbol{\psi}^*}(\boldsymbol{u},\boldsymbol{0})\right]\frac{\lambda(S_m\cap(S_m+\boldsymbol{u}))}{\lambda(S_m)}d\boldsymbol{u} = (1)$$

Since $\lambda(S_m \cap (S_m + \boldsymbol{u}))/\lambda(S_m) \to 1$ as $m \to \infty$ for every fixed $\boldsymbol{u} \in \mathbb{R}^d$ (see Lemma 3.2 in Karr (1986)), and due to the fact that

$$\int\limits_{S_m - S_m} w^2(\boldsymbol{u}) \mathbb{E}\left[A^2_{\boldsymbol{\psi}^*}(\boldsymbol{u}, \boldsymbol{0})\right] \frac{\lambda(S_m \cap (S_m + \boldsymbol{u}))}{\lambda(S_m)} d\boldsymbol{u} \le \int\limits_{S_m - S_m} w^2(\boldsymbol{u}) \mathbb{E}\left[A^2_{\boldsymbol{\psi}^*}(\boldsymbol{u}, \boldsymbol{0})\right] d\boldsymbol{u}$$
$$\to \int\limits_{\mathbb{R}^d} w^2(\boldsymbol{u}) \mathbb{E}\left[A^2_{\boldsymbol{\psi}^*}(\boldsymbol{u}, \boldsymbol{0})\right] d\boldsymbol{u} < \infty$$

it follows by dominated convergence that (1) converges to

$$\nu^{-2} \int\limits_{\mathbb{R}^d} \mathbb{V}ar\left(w(\boldsymbol{u}) A_{\boldsymbol{\psi}^*}(\boldsymbol{u}, \boldsymbol{0})\right) d\boldsymbol{u}.$$

Using similar arguments,

$$\lambda(S_m)^{-1} \nu^{-4} \iiiint\limits_{S_m^4} w(\boldsymbol{s}_1 - \boldsymbol{s}_2) w(\boldsymbol{s}_3 - \boldsymbol{s}_4) \mathbb{E}\left[A_{\boldsymbol{\psi}^*}(\boldsymbol{s}_1, \boldsymbol{s}_2) A_{\boldsymbol{\psi}^*}(\boldsymbol{s}_3, \boldsymbol{s}_4)\right] \nu^3 d\boldsymbol{s}_1 d\boldsymbol{s}_2 \epsilon_{\boldsymbol{s}_1}(d\boldsymbol{s}_3) d\boldsymbol{s}_4$$
$$= \lambda(S_m)^{-1} \nu^{-1} \iiint\limits_{S_m^3} w(\boldsymbol{s}_1 - \boldsymbol{s}_2) w(\boldsymbol{s}_1 - \boldsymbol{s}_4) \mathbb{E}\left[A_{\boldsymbol{\psi}^*}(\boldsymbol{0}, \boldsymbol{s}_2 - \boldsymbol{s}_1) A_{\boldsymbol{\psi}^*}(\boldsymbol{0}, \boldsymbol{s}_4 - \boldsymbol{s}_1)\right] d\boldsymbol{s}_1 d\boldsymbol{s}_2 d\boldsymbol{s}_4$$
$$= \nu^{-1} \iint\limits_{(S_m - S_m)^2} w(\boldsymbol{v}_1) w(\boldsymbol{v}_2) \mathbb{E}\left[A_{\boldsymbol{\psi}^*}(\boldsymbol{v}_1, \boldsymbol{0}) A_{\boldsymbol{\psi}^*}(\boldsymbol{v}_2, \boldsymbol{0})\right] \frac{\lambda(S_m \cap (S_m + \boldsymbol{v}_1) \cap (S_m + \boldsymbol{v}_2))}{\lambda(S_m)} d\boldsymbol{v}_1 d\boldsymbol{v}_2$$
$$\to \nu^{-1} \mathbb{V}ar\left(\int\limits_{\mathbb{R}^d} w(\boldsymbol{u}) A_{\boldsymbol{\psi}^*}(\boldsymbol{u}, \boldsymbol{0}) d\boldsymbol{u}\right).$$

For the last term we obtain

$$\lambda(S_m)^{-1} \nu^{-4} \iiiint\limits_{S_m^4} w(\boldsymbol{s}_1 - \boldsymbol{s}_2) w(\boldsymbol{s}_3 - \boldsymbol{s}_4) \mathbb{E}\left[A_{\boldsymbol{\psi}^*}(\boldsymbol{s}_1, \boldsymbol{s}_2) A_{\boldsymbol{\psi}^*}(\boldsymbol{s}_3, \boldsymbol{s}_4)\right] \nu^4 d\boldsymbol{s}_1 d\boldsymbol{s}_2 d\boldsymbol{s}_3 d\boldsymbol{s}_4$$
$$= \iiint\limits_{(S_m - S_m)^3} w(\boldsymbol{v}_1) w(\boldsymbol{v}_3 - \boldsymbol{v}_2) \mathbb{E}\left[A_{\boldsymbol{\psi}^*}(\boldsymbol{v}_1, \boldsymbol{0}) A_{\boldsymbol{\psi}^*}(\boldsymbol{v}_2, \boldsymbol{v}_3)\right]$$
$$\times \frac{\lambda(S_m \cap (S_m + \boldsymbol{v}_1) \cap (S_m + \boldsymbol{v}_2) \cap (S_m + \boldsymbol{v}_3))}{\lambda(S_m)} d\boldsymbol{v}_1 d\boldsymbol{v}_2 d\boldsymbol{v}_3.$$

Altogether, as $m \to \infty$,

$$\mathbb{V}ar((\lambda(S_m))^{-1/2} \nabla_{\boldsymbol{\psi}} PL^{(m)}(\boldsymbol{\psi}^*)) \to \Sigma$$
$$= 2\nu^{-2} \int\limits_{\mathbb{R}^d} \mathbb{V}ar(w(\boldsymbol{v}) A_{\boldsymbol{\psi}^*}(\boldsymbol{v}, \boldsymbol{0})) d\boldsymbol{v} + 4\nu^{-1} \mathbb{V}ar\left(\int\limits_{\mathbb{R}^d} w(\boldsymbol{u}) A_{\boldsymbol{\psi}^*}(\boldsymbol{u}, \boldsymbol{0}) d\boldsymbol{u}\right)$$
$$+ \iiint\limits_{\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d} \mathbb{C}ov\left[w(\boldsymbol{v}_1) A_{\boldsymbol{\psi}^*}(\boldsymbol{v}_1, \boldsymbol{0}), w(\boldsymbol{v}_3 - \boldsymbol{v}_2) A_{\boldsymbol{\psi}^*}(\boldsymbol{v}_2, \boldsymbol{v}_3)\right] d\boldsymbol{v}_1 d\boldsymbol{v}_2 d\boldsymbol{v}_3.$$

The central limit theorem for $\nabla_{\boldsymbol{\psi}} PL^{(m)}(\boldsymbol{\psi}^*)$ follows in exactly the same way as in Lemma A.2. together with Lemma A.4. in Li et al. (2008). □

In a second step we show that the estimates resulting by maximizing the pairwise log-likelihood function in (28) are asymptotically normal.

THEOREM 6.2. *In addition to the conditions in Lemma 6.1 assume that*

(h) $\mathbb{E}\left[|\nabla^2_{\boldsymbol{\psi}} \log f_{\boldsymbol{\psi}^*}(\eta(\boldsymbol{s}_1), \eta(\boldsymbol{s}_2))|\right] < \infty$ *and* $\mathbb{E}\left[|\nabla^2_{\boldsymbol{\psi}} \log f_{\boldsymbol{\psi}^*}(\eta(\boldsymbol{s}_1), \eta(\boldsymbol{s}_2))|^2\right] < \infty$, *and*

(i) $\sup\limits_{\boldsymbol{\psi}\in\Psi} \int\limits_{\mathbb{R}^d} w(\boldsymbol{u})\mathbb{E}\left[\nabla^2_{\boldsymbol{\psi}} \log f_{\boldsymbol{\psi}}(\eta(\boldsymbol{u}), \eta(\boldsymbol{0}))\right] d\boldsymbol{u} < \infty$, *and*

(j) further, as $m \to \infty$ *and for fixed* $\boldsymbol{\psi} \in \Psi$,

$$\iiint\limits_{(S_m - S_m)^3} w(\boldsymbol{v}_1)w(\boldsymbol{v}_3 - \boldsymbol{v}_2)\mathbb{E}\left[\nabla_{\boldsymbol{\psi}} A_{\boldsymbol{\psi}}(\boldsymbol{v}_1, \boldsymbol{0})\nabla_{\boldsymbol{\psi}} A_{\boldsymbol{\psi}}(\boldsymbol{v}_2, \boldsymbol{v}_3)\right]$$
$$\times \frac{\lambda(S_m \cap (S_m + \boldsymbol{v}_1) \cap (S_m + \boldsymbol{v}_2) \cap (S_m + \boldsymbol{v}_3))}{\lambda(S_m)} d\boldsymbol{v}_1 d\boldsymbol{v}_2 d\boldsymbol{v}_3$$
$$\to \iiint\limits_{\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d} w(\boldsymbol{v}_1)w(\boldsymbol{v}_3 - \boldsymbol{v}_2)\mathbb{E}\left[\nabla_{\boldsymbol{\psi}} A_{\boldsymbol{\psi}}(\boldsymbol{v}_1, \boldsymbol{0})\nabla_{\boldsymbol{\psi}} A_{\boldsymbol{\psi}}(\boldsymbol{v}_2, \boldsymbol{v}_3)\right] d\boldsymbol{v}_1 d\boldsymbol{v}_2 d\boldsymbol{v}_3.$$

Then, the pairwise likelihood estimate $\hat{\boldsymbol{\psi}}$ *is asymptotically normal:*

$$\sqrt{\lambda(S_m)}(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}^*) \xrightarrow{d} \mathcal{N}(0, F^{-1}\Sigma(F^{-1})^\intercal), \quad m \to \infty,$$

where Σ *is defined in* (29) *and*

$$F = \int\limits_{\mathbb{R}^d} w(\boldsymbol{u})\mathbb{E}\left[\nabla^2_{\boldsymbol{\psi}} \log f_{\boldsymbol{\psi}^*}(\eta(\boldsymbol{u}), \eta(\boldsymbol{0}))\right] d\boldsymbol{u}.$$

and

PROOF. For the second derivative of the pairwise log-likelihood function (28) we obtain for fixed $\boldsymbol{\psi} \in \Psi$

$$\mathbb{E}\left[\lambda(S_m)^{-1}\nu^{-2}\nabla^2 PL^{(m)}(\boldsymbol{\psi})\right] = \lambda(S_m)^{-1} \iint\limits_{S_m \times S_m} w(\boldsymbol{s}_1 - \boldsymbol{s}_2)\mathbb{E}\left[\nabla_{\boldsymbol{\psi}} A_{\boldsymbol{\psi}}(\boldsymbol{s}_1, \boldsymbol{s}_2)\right] d\boldsymbol{s}_1 d\boldsymbol{s}_2$$
$$= \int\limits_{(S_m - S_m)} w(\boldsymbol{u})\mathbb{E}\left[\nabla_{\boldsymbol{\psi}} A_{\boldsymbol{\psi}}(\boldsymbol{s}_1 - \boldsymbol{s}_2, \boldsymbol{0})\right] \frac{\lambda(S_m \cap (S_m + \boldsymbol{u}))}{\lambda(S_m)} d\boldsymbol{u}$$
$$\to \int\limits_{\mathbb{R}^d} w(\boldsymbol{u})\mathbb{E}\left[\nabla_{\boldsymbol{\psi}} A_{\boldsymbol{\psi}}(\boldsymbol{u}, \boldsymbol{0})\right] d\boldsymbol{u}, \quad m \to \infty.$$

Using the same argument as for the pairwise score function it follows that $\mathbb{V}ar(\lambda(S_m)^{-1}\nu^{-2}\nabla^2_{\boldsymbol{\psi}} PL^{(m)}(\boldsymbol{\psi})) \to 0$. This shows pointwise convergence of $\nabla^2_{\boldsymbol{\psi}} PL^{(m)}(\tilde{\boldsymbol{\psi}})$ to

$\int_{\mathbb{R}^d} w(\boldsymbol{u})\mathbb{E}\left[\nabla^2_{\boldsymbol{\psi}} \log f_{\boldsymbol{\psi}}(\eta(\boldsymbol{u}),\eta(\boldsymbol{0}))\right] d\boldsymbol{u}$. The uniform convergence is implied by Assumption (i). Therefore,

$$\lambda(S_m)^{-1}\nu^{-2}\nabla^2_{\boldsymbol{\psi}} PL^{(m)}(\boldsymbol{\psi}) \to F.$$

Using a Taylor expansion

$$0 = \nabla_{\boldsymbol{\psi}} PL^{(m)}(\hat{\boldsymbol{\psi}}) = \frac{1}{\lambda(S_m)\nu^2}\nabla_{\boldsymbol{\psi}} PL^{(m)}(\boldsymbol{\psi}^*) + \left(\frac{1}{\lambda(S_m)\nu^2}\nabla^2_{\boldsymbol{\psi}} PL^{(m)}(\tilde{\boldsymbol{\psi}})\right)(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}^*),$$

we obtain

$$\sqrt{\lambda(S_m)}(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}^*) = -\left(\frac{1}{\lambda(S_m)\nu^2}\nabla^2_{\boldsymbol{\psi}} PL^{(m)}(\tilde{\boldsymbol{\psi}})\right)^{-1}\left(\frac{1}{\sqrt{\lambda(S_m)}\nu^2}\nabla_{\boldsymbol{\psi}} PL^{(m)}(\boldsymbol{\psi}^*)\right).$$

Together with the central limit theorem for the pairwise score function (Lemma 6.1) it follows that

$$\sqrt{\lambda(S_m)}(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}^*) \xrightarrow{d} \mathcal{N}(0,\Sigma).$$

□

REMARK 6.3. *First note that the rate of convergence here is* $\sqrt{\lambda(S_m)} = O(m^{d/2})$ *(see Assumption (a)) which is the same as for regular grids. For the max-stable random field in* (2) *satisfying Condition 2.3, Assumption (d) was shown in Lemma 5.1. Assumptions (f) and (h) can be shown in the same way as Lemma 5.1. The condition (b) on the* α*-mixing coefficients is easily verified using Lemma 5.4, from which follows that*

$$\frac{1}{k^2}\alpha_{k,k}(r) \le \exp\{-\theta_1 r_1^{\alpha}/2\} \le C r_1^{-\alpha},$$

where $C > 0$ *is some constant.*

7. Simulation study

We illustrate the small sample behaviour of the pairwise likelihood estimation for spatial dimension $d = 2$ in a simulation experiment. The setup for this study is:

(a) The spatial locations consisted of a 10×10 grid

$$S_{10} = \left\{\boldsymbol{s}_{(i_1,i_2)} = (i_1,i_2), i_1, i_2 \in \{1,\dots,10\}\right\}.$$

The time points are chosen equidistantly, $1 < \cdots < T = 100$.

(b) One hundred independent Gaussian space-time processes $Z_j(s_n\boldsymbol{s}, t_n t), j = 1,\dots,100$ were generated using the R-package `RandomFields` with covariance function $\rho(s_n\boldsymbol{h}, t_n u)$. We use the following correlation function for the underlying Gaussian random field.

$$\rho(\boldsymbol{h},u) = (1 + \theta_1\|\boldsymbol{h}\|^{\alpha_1} + \theta_2|u|^{\alpha_2})^{-3/2}.$$

Assumption 2.1 is fulfilled and the limit function δ is given by

$$\lim_{n\to\infty}(\log n)(1-\rho(s_n\boldsymbol{h}, t_n u)) = \delta(\boldsymbol{h},u) = \frac{3}{2}\theta_1\|\boldsymbol{h}\|^{\alpha_1} + \frac{3}{2}\theta_2|u|^{\alpha_2},$$

where $s_n = (\log n)^{1/\alpha_1}$ and $t_n = (\log n)^{1/\alpha_2}$.

(c) The simulated processes were transformed to standard Fréchet margins using the transformation $-1/\log(\Phi(Z_j(\boldsymbol{s},t)))$ for $\boldsymbol{s} \in S_{10}$ and $t \in \{1,\dots,T\}$.

(d) The pointwise maximum of the transformed Gaussian random fields was computed and rescaled by $1/1000$ to obtain an approximation of a max-stable random field, i.e.,

$$\eta(\boldsymbol{s},t) = \frac{1}{1000} \bigvee_{j=1}^{1000} -\frac{1}{\log\left(\Phi(Z_j(s_n \boldsymbol{s}, t_n t))\right)}, \ \boldsymbol{s} \in S_{10}, \ t \in \{1,\dots,T\}.$$

(e) The parameters $\theta_1, \alpha_1, \theta_2$ and α_2 for different combinations of maximal space-time lags (r,p) were estimated by maximizing (11). The program is adjusted such that it takes care of identifiability issues, when some of the parameters are not identifiable, cf. Remark 4.5.

(f) Steps (a)-(e) are repeated 100 times.

Note first, that we only get an approximation of a Brown-Resnick process since we cannot choose $n = \infty$. There is not doubt that the simulation of the marginal distribution is accurate. Before estimating the parameters we checked the bivariate extremal dependence structure by estimating the tail dependence coefficient separately in time and space. A comparison with simulations from the bivariate limit distribution function showed that the realizations are appropriate. Figures 2 and 3 show the resulting estimates as a function of (r,p), where the true parameter set is given by $\boldsymbol{\psi}^* = (\theta_1^*, \alpha_1^*, \theta_2^*, \alpha_2^*) = (0.06, 1, 0.04, 1)$. Figure 2 shows boxplots of the resulting estimates for the spatial parameters θ_1 and α_1. The horizontal axis shows the different maximal space-time lags included in the pairwise likelihood function from (11). We also show qq-plots against a normal distribution for all parameters and different combinations of r and p in Figure 4. In addition to the graphical output we calculate the root mean square error (RMSE) and the mean absolute error (MAE) to see how the choice of (r,p) influences the estimation.

We make the following observations. As already pointed out by Davis and Yau (2011) and Huser and Davison (2012), there might be a loss in efficiency if too many pairs are included in the estimation. This can be explained by the fact that pairs get more and more independent as the space-time lag increases. Adding more and more pairs to the pairwise log-likelihood function can introduce some noise which decreases the efficiency. This is evident in Figure 3 for the temporal parameter α_2, where the estimates vary more around the mean as more pairs are included in the estimation.

An interesting observation for our model is that using a maximal spatial lag of $\mathbf{0}$ or a maximal temporal lag 0, respectively, leads to very good results. For the spatial parameters, the space-time lags which lead to the lowest RMSE and MAE are $(2,0)$ for θ_1 and $(2,0)$ (RMSE) or $(3,0)$ (MAE) for α_2 (see Table 2), i.e., we use all pairs within a spatial distance of 2 or 3 at the same time point. Basically, this suggests that we could also estimate the spatial parameters based on each individual random field for fixed time points and then take the mean over all estimates in time. The same holds for the time parameters θ_2 and α_2, where the best results in the sense of the lowest RMSE and MAE are obtained for the space-time lags $(0,3)$, i.e., if we use all pairwise densities corresponding to the space-time pairs $(\boldsymbol{s},t_1)$ and $(\boldsymbol{s},t_2)$, where $|t_2 - t_1| \leq 3$ (see Table 3). The reason for this observation is that the parameters of the underlying space-time correlation function get "separated" in the extremal setting in the sense that for example a spatial lag equal to zero does not affect the temporal parameters θ_1 and α_1 and vice versa.

Table 2. RMSE and MAE based on 100 simulations for the spatial estimates θ_1 and α_1 for different combinations of maximal space-time lags (r, p).

θ_1	(1,0)	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(2,0)	(2,1)	(2,2)
RMSE	0.0099	0.0118	0.0121	0.0122	0.0123	0.0124	0.0103	0.0104	0.0105
MAE	0.0071	0.0090	0.0092	0.0093	0.0094	0.0095	0.0080	0.0081	0.0081
	(2,3)	(2,4)	(2,5)	(3,0)	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)
RMSE	0.0104	0.0104	0.0104	0.0106	0.0107	0.0108	0.0108	0.0107	0.0108
MAE	0.0081	0.0081	0.0081	0.0082	0.0083	0.0083	0.0084	0.0083	0.0084
α_1	(2,0)	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(3,0)	(3,1)	(3,2)
RMSE	0.1338	0.1398	0.1530	0.1492	0.1543	0.1569	0.1351	0.1409	0.1579
MAE	0.1078	0.1124	0.1154	0.1137	0.1233	0.1252	0.1050	0.1106	0.1127
	(3,3)	(3,4)	(3,5)	(4,0)	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)
RMSE	0.1596	0.1639	0.1649	0.1423	0.1483	0.1614	0.1673	0.1735	0.1751
MAE	0.1228	0.1291	0.1297	0.1120	0.1176	0.1114	0.1276	0.1372	0.1385

Table 3. RMSE and MAE based on 100 simulations for the spatial estimates θ_2 and α_2 for different combinations of maximal space-time lags (r, p).

$\hat{\theta}_2$	(0,1)	(0,2)	(0,3)	(0,4)	(0,5)	(1,1)	(1,2)	(1,3)	(1,4)
RMSE	0.0182	0.0182	0.0182	0.0182	0.0182	0.0184	0.0183	0.0183	0.0183
MAE	0.0171	0.0171	0.0171	0.0171	0.0171	0.0173	0.0172	0.0171	0.0171
	(1,5)	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(3,1)	(3,2)	(3,3)
RMSE	0.0183	0.0187	0.0186	0.0185	0.0185	0.0185	0.0188	0.0188	0.0186
MAE	0.0171	0.0175	0.0174	0.0173	0.0174	0.0173	0.0176	0.0176	0.0174
$\hat{\alpha}_2$	(0,2)	(0,3)	(0,4)	(0,5)	(1,2)	(1,3)	(1,4)	(1,5)	(2,2)
RMSE	0.1317	0.1269	0.1280	0.1289	0.1442	0.1401	0.1426	0.1438	0.1463
MAE	0.1008	0.0989	0.1015	0.1035	0.1086	0.1079	0.1139	0.1147	0.1179
	(2,3)	(2,4)	(2,5)	(3,2)	(3,3)	(3,4)	(3,5)	(4,2)	(4,3)
RMSE	0.1532	0.1580	0.1619	0.1473	0.1531	0.1589	0.1642	0.1549	0.1607
MAE	0.1242	0.1275	0.1294	0.1169	0.1223	0.1273	0.1317	0.1233	0.1284

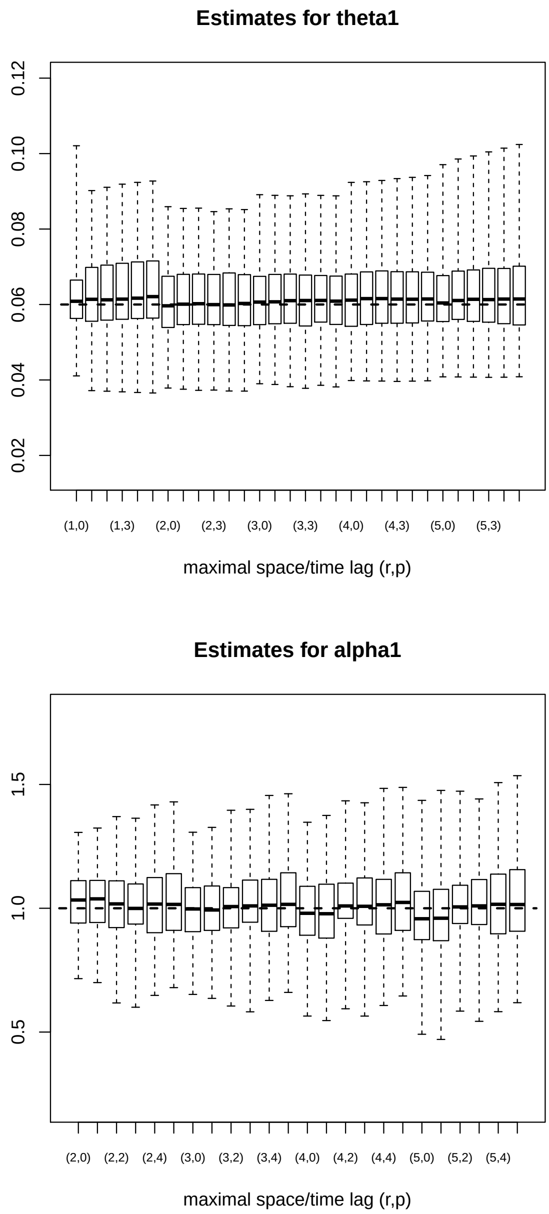
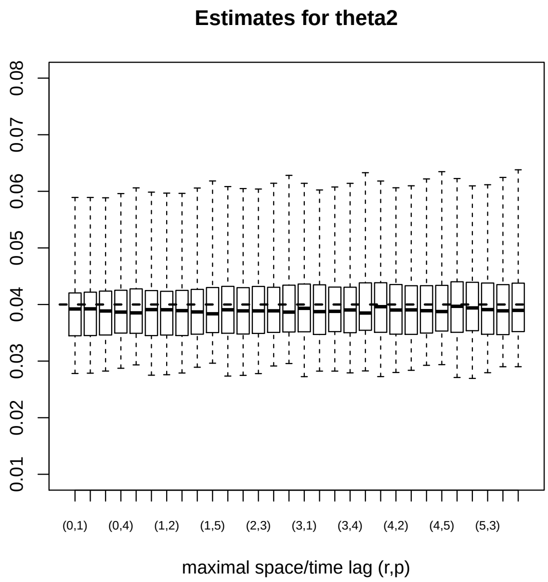


Figure 2. Estimates for θ_1 and α_1 (spatial parameters) as a function of maximal space-time lags (r, p). Each boxplot represents the estimates for 100 simulations. The dashed line represents the true value.



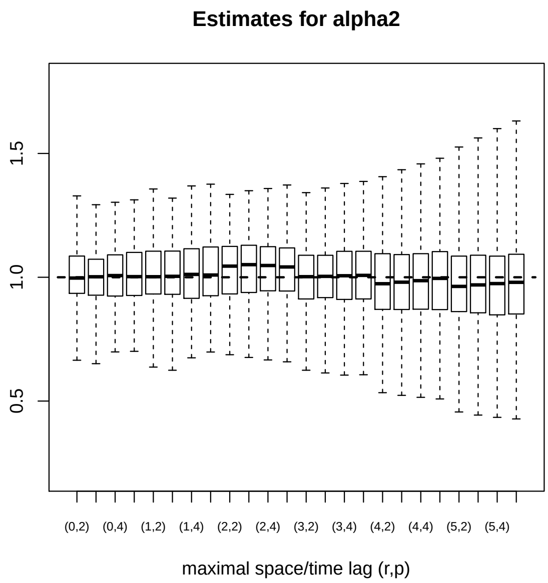


Figure 3. Estimates for θ_2 and α_2 (spatial parameters) as a function of maximal space-time lags (r, p). Each boxplot represents the estimates for 100 simulations. The dashed line represents the true value.

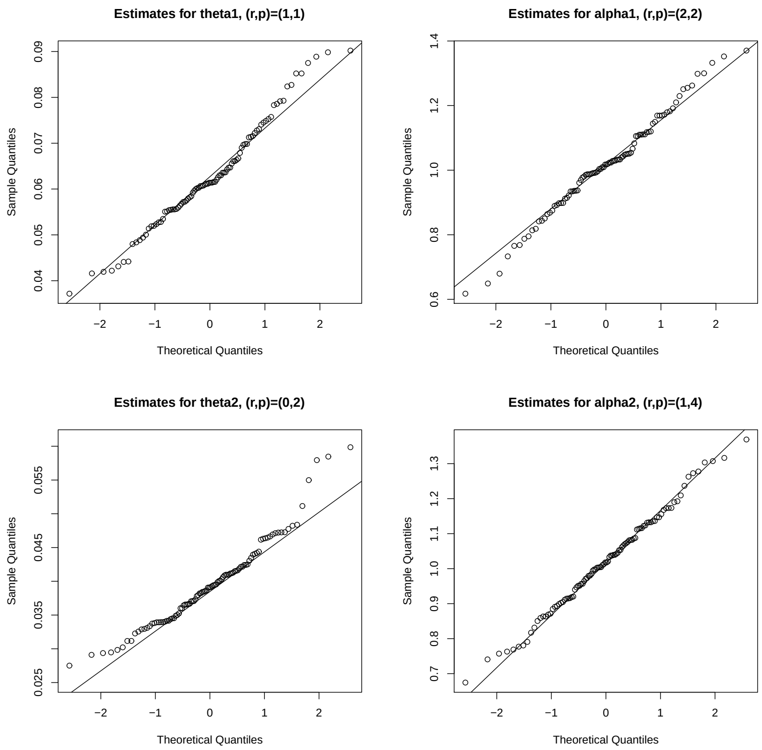


Figure 4. QQ-plots for estimates against normal distribution, where for each parameter we chose a random combination of r and p.

Acknowledgments
All authors gratefully acknowledge the support by the TUM Institute for Advanced Study (TUM-IAS). The third author additionally likes to thank the International Graduate School of Science and Engineering (IGSSE) of the Technische Universität München for their support. The research of Richard A. Davis was also supported in part by the National Science Foundation grant DMS-1107031.

References

Bai, Y., X.-K. Song, and T. Raghunathan (2012). Joint composite estimating functions in spatiotemporal models. *Journal of the Royal Statistical Society B 75*(5).

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society B 34*(2), 192–236.

Bolthausen, E. (1982). On the central limit theorem for stationary mixing random fields. *The Annals of Probability 10*(4), 1047–1050.

Brown, B. and S. Resnick (1977). Extreme values of independent stochastic processes. *Journal of Applied Probability 14*(4), 732–739.

Cox, D. and N. Reid (2004). Miscellanea: A note on pseudolikelihood constructed from marginal densities. *Biometrika 91*(3), 729 – 737.

Davis, R., C. Klüppelberg, and C. Steinkohl (2011). Max-stable processes for extremes of processes observed in space and time. *Preprint*. arXiv:1107.4464v1 [stat.ME].

Davis, R. and T. Mikosch (2008). Extreme value theory for space-time processes with heavy-tailed distributions. *Stochastic Processes and their Applications 118*, 560–584.

Davis, R. and T. Mikosch (2009). The extremogram: A correlogram for extreme events. *Bernoulli 15*(4), 977–1009.

Davis, R., T. Mikosch, and I. Cribben (2012). Towards estimating extremal series dependence via the bootstrapped extremogram. *Journal of Econometrics 17*(1), 142–152.

Davis, R. and C. Yau (2011). Comments on pairwise likelihood in time series models. *Statistica Sinica 21*, 255–277.

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

Dombry, C. and F. Eyi-Minko (2012). Strong mixing properties of max-infinitely divisible random fields. arXiv:1201.4645v1.

Gneiting, T. (2002). Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association 95*, 590–600.

Huser, R. and A. Davison (2012). Space-time modelling for extremes. *Preprint*. arXiv:1201.3245v1 [stat.ME].

Hüsler, J. and R.-D. Reiss (1989). Maxima of normal random vectors: between independence and complete dependence. *Statistics and Probability Letters 7*, 283–286.

Jenish, N. and I. Prucha (2009). Central limit theorems and uniform laws of large numbers for arrays of random fields. *Journal of Econometrics 150*(1), 86–98.

Kabluchko, Z. (2009). Extremes of space-time Gaussian processes. *Stochastic Processes and their Applications 119*, 3962 – 3980.

Kabluchko, Z. (2011). Extremes of independent Gaussian processes. *Extremes 14*(3), 285–310.

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

Karr, A. F. (1986). Inference for stationary random fields given Poisson samples. *Advances in Applied Probability 18*(2), 406 – 422.

Krengel, U. (1985). *Ergodic Theorems.* de Gruyter, Berlin.

Li, B., M. Genton, and M. Sherman (2008). On the asymptotic joint distribution of sample space-time covariance estimators. *Bernoulli 14*(1), 208–248.

Lindsay, B. (1988). Composite likelihood methods. *Contemporary Mathematics 80*, 221–239.

Nott, D. J. and T. Rydén (1999). Pairwise likelihood methods for inference in image models. *Biometrika 86*(3), 661–676.

Padoan, S., M. Ribatet, and S. Sisson (2009). Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association (Theory and Methods) 105*(489), 263–277.

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

Stoev, S. (2008). On the ergodicity and mixing of max-stable processes. *Stochastic Processes and their Applications 18*(9), 1679–1705.

Stoev, S. and M. Taqqu (2005). Extremal stochastic integrals: a parallel between max-stable processes and α-stable processes. *Extremes 8*, 237–266.

Straumann, D. (2004). *Estimation in Conditionally Heteroscedastic Time Series Models.* Lecture Notes in Statistics, Springer, Berlin.

Varin, C. (2007). On composite marginal likelihoods. *Advances in Statistical Analysis 92*, 1–28.

Varin, C. and P. Vidoni (2005). Pairwise likelihood inference and model selection. *Biometrika 92*(3), 519–528.

Wald, A. (1946). Note on the consistency of the maximum likelihood estimate. *The Annals of Mathematical Statistics 29*, 595 – 601.

Wang, Y., P. Roy, and S. Stoev (2011). Ergodic properties of sum- and max-stable stationary random fields via null and positive group actions. *Technical Report*. arXiv:0911.0610v2[math.PR].

Wang, Y. and S. Stoev (2010). On the structure and representations of max-stable processes. *Advances in Applied Probability 42*(3), 855–877.

A. Proof of Lemma 5.1

In the following, we use the same abbreviations as in the proof of Theorem 4.4. The gradient of the bivariate log-density with respect to the parameter vector $\boldsymbol{\psi}$ is given by

$$\nabla_{\boldsymbol{\psi}} \log f(x_1, x_2) = \frac{\partial \log f(x_1, x_2)}{\partial \delta} \nabla_{\boldsymbol{\psi}} \delta.$$

Assume in the following that all parameters $\theta_1, \alpha_1, \theta_2$ and α_2 are identifiable. Since all partial derivatives

$$\frac{\partial \delta}{\partial \theta_1} = \|\boldsymbol{h}\|^{\alpha_1}, \; \frac{\partial \delta}{\partial \theta_2} = |u|^{\alpha_2}, \; \frac{\partial \delta}{\partial \alpha_1} \theta_1 \alpha_1 \|\boldsymbol{h}\|^{\alpha_1 - 1}, \; \frac{\partial \delta}{\partial \alpha_2} = \theta_2 \alpha_2 |u|^{\alpha_2 - 1},$$

as well as all second order partial derivatives can be bounded from below and above for $0 < \min\{\|\boldsymbol{h}\|, |u|\}, \max\{\|\boldsymbol{h}\|, |u|\} < \infty$ using assumption (19) and, independently of the parameters θ_1, θ_2, α_1 and α_2, it suffices to show that

$$\mathbb{E}_{\boldsymbol{\psi}^*}\left[\left|\frac{\partial \log f_{\boldsymbol{\psi}}(\eta(\boldsymbol{s}_1, t_1), \eta(\boldsymbol{s}_2, t_2))}{\partial \delta}\right|^3\right] < \infty$$

and

$$\mathbb{E}_{\boldsymbol{\psi}^*}\left[\sup_{\boldsymbol{\psi} \in \boldsymbol{\Psi}}\left|\frac{\partial^2 \log f_{\boldsymbol{\psi}}(\eta(\boldsymbol{s}_1, t_1), \eta(\boldsymbol{s}_2, t_2))}{\partial \delta}\right|\right] < \infty.$$

Since δ can be bounded away from zero using assumption (19), we can treat δ as a constant. For simplification we drop the argument in the following equalities. Define

$$V_1 = \frac{\partial V}{\partial x_1}, \; V_2 = \frac{\partial V}{\partial x_2}, \quad \text{and } V_{12} = \frac{\partial^2 V}{\partial x_1 x_2}.$$

The partial derivative of the bivariate log-density with respect to δ has the following form

$$\frac{\partial \log f_{\boldsymbol{\psi}}}{\partial \delta} = -\frac{\partial V}{\partial \delta} + (V_1 V_2 - V_{12})^{-1}\left(\frac{\partial V_1}{\partial \delta} V_2 + V_1 \frac{\partial V_2}{\partial \delta} - \frac{\partial V_{12}}{\partial \delta}\right).$$

We identify stepwise the "critical" terms, where "critical" means higher order terms of functions of x_1 and x_2. To give an idea on how to handle the components in the derivatives, we describe one such step. Note that $(V_1 V_2 - V_{12})^{-1}$ can be written as

$$(V_1 V_2 - V_{12})^{-1} = \frac{x_1 x_2}{g_1\left(\frac{1}{x_1}, \frac{1}{x_2}, \frac{1}{x_1 x_2}, \frac{1}{x_1^2}, \frac{1}{x_2^2}\right)},$$

where g_1 describes the sum of the components together with additional multiplicative factors. By using

$$\frac{\partial \Phi(q_{\boldsymbol{\psi}}^{(1)})}{\partial \delta} = \frac{q_{\boldsymbol{\psi}}^{(1)}}{2\delta} \varphi(q_{\boldsymbol{\psi}}^{(1)}) \quad \text{and} \quad \frac{\partial \varphi(q_{\boldsymbol{\psi}}^{(1)})}{\partial \delta} = -\frac{(q_{\boldsymbol{\psi}}^{(1)})^2}{2\delta} \varphi(q_{\boldsymbol{\psi}}^{(1)}),$$

where $q_{\boldsymbol{\psi}}^{(1)} = \log(x_2/x_1)/(2\sqrt{\delta}) + \sqrt{\delta}$, we have

$$\frac{\partial V_1}{\partial \delta} V_2 = g_2\left(\frac{1}{x_1^2, x_2^2}, \frac{q_1}{x_1^2 x_2^2}, \frac{q_1^2}{x_1^2 x_2^2}, \frac{1}{x_1^3 x_2}, \frac{q_1}{x_1^3 x_2}, \frac{q_1^2}{x_1^3 x_2}, \frac{1}{x_1 x_2^3}, \frac{q_1)^2}{x_1 x_2^3}\right),$$

where g_2 is a linear function of the components. By combining the two representations above, we obtain that all terms in
$(V_1V_2 - V_{12})^{-1}(\partial V_1/\partial\delta)V_2$ are of the form

$$\frac{|\log x_1|^{k_1}|\log x_2|^{k_2}}{x_1^{k_3}x_2^{k_4}}, \quad k_1,k_2,k_3,k_4 \geq 0. \tag{30}$$

The second derivative of the bivariate log-density with respect to δ is given by

$$\begin{aligned}\frac{\partial^2 \log f_{\boldsymbol{\psi}}}{(\partial\delta)^2} =& -\frac{\partial^2 V}{(\partial\delta)^2} - (V_1V_2 - V_{12})^{-2}\left(\frac{\partial V_1}{\partial\delta}V_2 + V_1\frac{\partial V_2}{\partial\delta} - \frac{\partial V_{12}}{\partial\delta}\right)^2 \\ &+ (V_1V_2 - V_{12})^{-1}\left(\frac{\partial^2 V_1}{(\partial\delta)^2}V_2 + 2\frac{\partial V_1}{\partial\delta}\frac{\partial V_2}{\partial\delta} + V_1\frac{\partial^2 V_2}{(\partial\delta)^2} - \frac{\partial^2 V_{12}}{(\partial\delta)^2}\right)\end{aligned}$$

Stepwise calculation of the single components shows that all terms are also of form (30). This implies that for both statements it suffices to show that for all $k_1,k_2,k_3,k_4 \geq 0$

$$\mathbb{E}\left[\frac{(\log\eta(\boldsymbol{s},t))^{k_1}(\log\eta(\boldsymbol{s},t))^{k_2}}{|\eta(\boldsymbol{s},t)|^{k_3}|\eta(\boldsymbol{s},t)|^{k_4}}\right] < \infty.$$

Since $\eta(\boldsymbol{s},t)$ is standard Fréchet $\log(\eta(\boldsymbol{s},t))$ is standard Gumbel and $1/\eta(\boldsymbol{s},t)$ is standard exponential. Using Hölder's inequality, we obtain

$$\begin{aligned}&\mathbb{E}\left[\frac{|\log(\eta(\boldsymbol{s},t))|^{k_1}|\log(\eta(\boldsymbol{s},t))|^{k_2}}{|\eta(\boldsymbol{s},t)|^{k_3}|\eta(\boldsymbol{s},t)|^{k_4}}\right] \\ &< \left(\mathbb{E}\left[|\log(\eta(\boldsymbol{s},t))|^{4k_1}\right]\mathbb{E}\left[|\log(\eta(\boldsymbol{s},t))|^{4k_2}\right]\right)^{1/2}\left(\mathbb{E}\left[\left|\frac{1}{\eta(\boldsymbol{s},t)}\right|^{4k_3}\right]\mathbb{E}\left[\left|\frac{1}{\eta(\boldsymbol{s},t)}\right|^{4k_4}\right]\right)^{1/2} < \infty,\end{aligned}$$

since all moments of the exponential and the Gumbel distributions are finite.