Anisotropic Brown-Resnick space-time processes: estimation and model assessment

Sven Buhl*   Claudia Klüppelberg*

June 7, 2016

Abstract

Spatially isotropic max-stable processes have been used to model extreme spatial or space-time observations. One prominent model is the Brown-Resnick process, which has been successfully fitted to time series, spatial data and space-time data. This paper extends the process to possibly anisotropic spatial structures. For regular grid observations we prove strong consistency and asymptotic normality of pairwise maximum likelihood estimates for fixed and increasing spatial domain, when the number of observations in time tends to infinity. We also present a statistical test for isotropy versus anisotropy. We apply our test to precipitation data in Florida, and present some diagnostic tools for model assessment. Finally, we present a method to predict conditional probability fields and apply it to the data.

1   Introduction

Max-stable processes, such as the Brown-Resnick process, have been successfully fitted to time series, spatial and recently to space-time data. Methods for inference include pairwise likelihood based on the bivariate density of the models (cf. Padoan et al. [27]), censored likelihood (cf. Wadsworth and Tawn [32]) or threshold-based approaches (cf. Engelke et al. [15]). In Davis et al. [8] a spatially isotropic Brown-Resnick space-time process is suggested and applied to precipitation data. Pairwise maximum likelihood estimates are shown to be strongly consistent and asymptotically normal, provided the domain of observations increases jointly in space and time. Their approach is restricted to isotropic spatial dependence.

In the present paper we generalise the Brown-Resnick model to allow anisotropy in space. The new model allows for different extremal behaviour along orthogonal spatial directions. Anisotropy is

---

*Center for Mathematical Sciences, Technische Universität München, 85748 Garching, Boltzmannstrasse 3, Germany, email: sven.buhl@tum.de , cklu@tum.de, url: http://www.statistics.ma.tum.de

often observed on Earth, for example in Middle Europe with its westerly winds or near the equator where trade winds involve predominant easterlies. All dependence parameters are summarised in the semivariogram of an underlying Gaussian space-time process. This semivariogram then defines the dependence structure of the max-stable process and, as a consequence, the tail dependence coefficient between two process values evaluated at two location and two time points.

Furthermore, since in real world applications, observations are often recorded over a large number of time points, but only at a comparably small number of spatial locations, we consider both a fixed and increasing spatial domain in combination with an increasing temporal domain. For both settings, fixed and increasing spatial domain, we prove strong consistency and asymptotic normality of the pairwise maximum likelihood estimates in the anisotropic model based on regular grid observations. This requires in particular to prove space-time and temporal mixing conditions in both settings for the anisotropic model.

We also provide tests for isotropy versus anisotropy again in both settings, which are designed for the new model. The asymptotic normality of the parameter estimates determines in principle the rejection areas of the test. However, the covariance matrices of the normal limit laws are not available in closed form. We formulate a subsampling procedure in the terminology of the Brown-Resnick space-time process and prove its convergence for fixed and increasing spatial domain.

We conclude with an analysis of space-time block maxima of radar rainfall measurements in Florida. Firstly, we present a simple procedure to test whether they originate from a max-stable process. As this cannot be rejected, we fit the Brown-Resnick space-time model to the data, using pairwise maximum likelihood estimation. Subsequently we apply the new isotropy test. Both the estimation and the test are based on the setting of a fixed spatial domain and increasing time series. In particular, since the Brown-Resnick space-time process satisfies the strong mixing conditions for increasing spatial and time domain as well as for fixed spatial and increasing time domain, the estimation and test procedure are independent of the specific setting: it works in both settings in exactly the same way, taking the different asymptotic covariance matrices into account. Finally, we assess the goodness of fit of the estimated model by a simulation diagnostics based on a large number of i.i.d. simulated anisotropic Brown-Resnick space-time processes. As a result, there is no statistical significance that the anisotropic Brown-Resnick space-time process with the fitted parameters should be rejected.

Our paper is organised as follows. In Section 2 we present the Brown-Resnick space-time model, which allows for anisotropic effects in space, and various dependence measures, including the parameterised dependence function. In Section 3 we compute the pairwise maximum likelihood estimates for the new model and prove their strong consistency and asymptotic normality for both settings, fixed and increasing spatial domain. Section 4 presents hypothesis tests for spatial isotropy and derives rejection areas based on a subsampling procedure. A data analysis is performed in Section 5 with focus on model assessment. The isotropy test rejects spatial isotropy for these data in favour of our new anisotropic model. Based on two other test procedures, we conclude that the anisotropic Brown-Resnick space-time process with the given dependence parameters is an appropriate model for the block-maxima data. We conclude by predicting conditional probability fields, which give the probability of a high value (for example of the amount of precipitation) at some space-time location given a high value at some other location.

2 Spatially anisotropic Brown-Resnick processes

Throughout the paper we consider a *stationary Brown-Resnick space-time process* with representation

$$\eta(\boldsymbol{s},t) = \bigvee_{j=1}^{\infty} \left\{ \xi_j \, e^{W_j(\boldsymbol{s},t)-\delta(\boldsymbol{s},t)} \right\}, \quad (\boldsymbol{s},t) \in \mathbb{R}^d \times [0,\infty), \tag{2.1}$$

where $\{\xi_j : j \in \mathbb{N}\}$ are points of a Poisson process on $[0,\infty)$ with intensity $\xi^{-2}d\xi$, the *dependence function* δ is nonnegative and conditionally negative definite and $\{W_j(\boldsymbol{s},t) : \boldsymbol{s} \in \mathbb{R}^d, t \in [0,\infty)\}$ are independent replicates of a Gaussian process
$\{W(\boldsymbol{s},t) : \boldsymbol{s} \in \mathbb{R}^d, t \in [0,\infty)\}$ with stationary increments, $W(\boldsymbol{0},0) = 0$, $\mathbb{E}[W(\boldsymbol{s},t)] = 0$ and covariance function

$$\mathbb{C}ov[W(\boldsymbol{s}^{(1)},t^{(1)}), W(\boldsymbol{s}^{(2)},t^{(2)})] = \delta(\boldsymbol{s}^{(1)},t^{(1)}) + \delta(\boldsymbol{s}^{(2)},t^{(2)}) - \delta(\boldsymbol{s}^{(1)} - \boldsymbol{s}^{(2)}, t^{(1)} - t^{(2)}).$$

Representation (2.1) goes back to de Haan [10] and Giné et al. [18]. Brown-Resnick processes have been studied by Brown and Resnick [5] in a time series context, as a spatial model by Kabluchko et al. [23], and in a space-time setting by Davis et al. [7] and Huser and Davison [19]. The univariate margins of the process η follow standard Fréchet distributions.

There are various quantities to describe the dependence in (2.1):

- In geostatistics, the dependence function δ is termed the *semivariogram* of the process $\{W(\boldsymbol{s},t)\}$: For $(\boldsymbol{s}^{(1)},t^{(1)}), (\boldsymbol{s}^{(2)},t^{(2)}) \in \mathbb{R}^d \times [0,\infty)$, it holds that

$$\mathbb{V}ar[W(\boldsymbol{s}^{(1)},t^{(1)}) - W(\boldsymbol{s}^{(2)},t^{(2)})] = 2\delta(\boldsymbol{s}^{(1)} - \boldsymbol{s}^{(2)}, t^{(1)} - t^{(2)}).$$

- For $\boldsymbol{h} \in \mathbb{R}^d$ and $u \in \mathbb{R}$, the *tail dependence coefficient* $\chi(\boldsymbol{h},u)$ is given by (cf. Kabluchko et al. [23], Remark 25 or Davis et al. [7], Section 3)

$$\chi(\boldsymbol{h},u) := \lim_{y \to \infty} \mathbb{P}\big(\eta(\boldsymbol{s}^{(1)},t^{(1)}) > y \mid \eta(\boldsymbol{s}^{(2)},t^{(2)}) > y\big) = 2\left(1 - \Phi\left(\sqrt{\frac{\delta(\boldsymbol{h},u)}{2}}\right)\right), \tag{2.2}$$

 where $\boldsymbol{h} = \boldsymbol{s}^{(1)} - \boldsymbol{s}^{(2)}$, $u = t^{(1)} - t^{(2)}$, and Φ denotes the standard normal distribution function.

- For $D = \{(\boldsymbol{s}^{(1)},t^{(1)}), \ldots, (\boldsymbol{s}^{(|D|)},t^{(|D|)})\}$ and $\boldsymbol{y} = (y_1, \ldots, y_{|D|}) > \boldsymbol{0}$ the finite-dimensional margins are given by

$$\mathbb{P}(\eta(\boldsymbol{s}^{(1)},t^{(1)}) \le y_1, \eta(\boldsymbol{s}^{(2)},t^{(2)}) \le y_2, \ldots, \eta(\boldsymbol{s}^{(|D|)},t^{(|D|)}) \le y_{|D|}) = e^{-V_D(\boldsymbol{y})}. \tag{2.3}$$

 Here V_D denotes the *exponent measure*, which is homogeneous of order -1.

- The *extremal coefficient* ξ_D for any finite set $D \subset \mathbb{R}^d \times [0,\infty)$ is defined through

$$\mathbb{P}(\eta(\boldsymbol{s}^{(1)},t^{(1)}) \le y, \eta(\boldsymbol{s}^{(2)},t^{(2)}) \le y, \ldots, \eta(\boldsymbol{s}^{(|D|)},t^{(|D|)}) \le y) = e^{-\xi_D/y}, \quad y > 0;$$

 i.e., $\xi_D = V_D(1, \ldots, 1)$. If $|D| = 2$, then (cf. Beirlant et al. [1], Section 9.5.1)

$$\chi(\boldsymbol{s}^{(1)} - \boldsymbol{s}^{(2)}, t^{(1)} - t^{(2)}) = 2 - \xi_D.$$

3

In this paper we assume the dependence function δ to be given for spatial lag $\boldsymbol{h}$ and time lag u by

$$\delta(\boldsymbol{h}, u) = \sum_{j=1}^{d} C_j |h_j|^{\alpha_j} + C_{d+1} |u|^{\alpha_{d+1}}, \quad (\boldsymbol{h}, u) = (h_1, \ldots, h_d, u) \in \mathbb{R}^{d+1}, \tag{2.4}$$

with parameters $C_j > 0$ and $\alpha_j \in (0, 2]$ for $j = 1, \ldots, d+1$.

Model (2.4) allows for different rates of decay of extreme dependence in different directions. This particularly holds along the axes of a d-dimensional spatial grid, but also for other directions. For example in the case $d = 2$, the decreases of dependence along the directions $(1, 2)$ and $(2, 1)$ differ. Model (2.4) can be generalised by a simple rotation to a setting, where not necessarily the axes, but other principal orthogonal directions play the major role. The rotation angle then needs to be estimated together with the other model parameters. A similar approach has been applied to introduce geometric or zonal anisotropy into a spatial isotropic model (see e.g. Blanchet and Davison [2], Section 4.2, or Engelke et al. [15], Section 5.2). For a justification of model (2.4) see Buhl [6], Sections 4.1 and 4.2. There it is shown that Brown-Resnick processes with this dependence function arise as limits of appropriately rescaled maxima of Gaussian processes with a large variety of correlation functions.

3   Pairwise maximum likelihood estimation

We extend the pairwise maximum likelihood procedure described in Davis et al. [8] for spatially isotropic space-time Brown-Resnick processes to the anisotropic case. We focus on the difference introduced by the spatial anisotropy and refer to the corresponding formulas in Davis et al. [8], where also a short introduction to composite likelihood estimation and further references can be found.

The pairwise likelihood function uses the bivariate distribution function of $(\eta(\boldsymbol{s}, t), \eta(\boldsymbol{s} + \boldsymbol{h}, t + u)) \overset{d}{=} (\eta(\boldsymbol{0}, 0), \eta(\boldsymbol{h}, u))$ (equal in distribution by stationarity) for $\boldsymbol{h} \in \mathbb{R}^d$ and $u \in \mathbb{R}$, which is given as

$$G(y_1, y_2) = \exp\{-V(y_1, y_2)\}, \quad y_1, y_2 > 0, \tag{3.1}$$

where the exponent measure $V = V_D$ for $D = \{(\boldsymbol{s}^{(1)}, t^{(1)}), (\boldsymbol{s}^{(2)}, t^{(2)})\}$ has the representation

$$V(y_1, y_2)$$
$$= \frac{1}{y_1} \Phi\left(\frac{\log(y_2/y_1)}{\sqrt{2\delta(\boldsymbol{h}, u)}} + \sqrt{\frac{\delta(\boldsymbol{h}, u)}{2}} \right) + \frac{1}{y_2} \Phi\left(\frac{\log(y_1/y_2)}{\sqrt{2\delta(\boldsymbol{h}, u)}} + \sqrt{\frac{\delta(\boldsymbol{h}, u)}{2}} \right), \tag{3.2}$$

which is a particular form of Eq. (2.7) in Hüsler and Reiss [20]. The dependence function δ is given by (2.4). For a derivation of (3.2) see for instance Oesting [26], Satz und Definition 2.4.

From this we can calculate the pairwise density $g(y_1, y_2) = g_{\boldsymbol{\theta}}(y_1, y_2)$ of G by differentiation. The parameter vector $\boldsymbol{\theta} = (C_1, \ldots, C_{d+1}, \alpha_1, \ldots, \alpha_{d+1})$ lies in the parameter space

$$\Theta := \{(C_1, \ldots, C_{d+1}, \alpha_1, \ldots, \alpha_{d+1}) : C_j \in (0, \infty), \alpha_j \in (0, 2], j = 1, \ldots, d+1\}.$$

We focus on data on a regular spatial grid and at equidistant time points. More precisely, we assume that the spatial observations lie on a regular d-dimensional lattice,

$$\mathcal{S}_M = \{\boldsymbol{s} = (s_1, \ldots, s_d) \in \{1, \ldots, M\}^d\}$$

4

for $M \in \mathbb{N}$, and that the time points are given by the set $\mathcal{T}_T = \{1, \ldots, T\}$ for $T \in \mathbb{N}$.

For the computation of the pairwise likelihood it is common not to include observations on all available space-time pairs, but only on those that lie within some prespecified spatio-temporal distance. This is motivated by the fact that pairs which lie sufficiently far apart in a space-time sense have little influence on the dependence parameters, see Nott and Rydén [25], Section 2.1. To express this notationally, we take inspiration from that paper and use a design mask adapted to the anisotropic setting,

$$\mathcal{H}_{\boldsymbol{r}} := \left\{ \boldsymbol{h} = (h_1, \ldots, h_d) \in \mathbb{N}_0^d : \boldsymbol{h} \leq \boldsymbol{r} \right\}, \quad \boldsymbol{r} = (r_1, \ldots, r_d) \in \mathbb{N}_0^d. \tag{3.3}$$

We are now ready to define the pairwise log-likelihood function and the resulting estimate.

Definition 1 (Pairwise likelihood estimate). *The pairwise log-likelihood function based on space-time pairs, whose maximum spatial lag is $\boldsymbol{r} \in \mathbb{N}_0^d$ and maximum time lag is $p \in \mathbb{N}_0$, such that $(\boldsymbol{r}, p) \neq (\boldsymbol{0}, 0)$, is defined as*

$$PL^{(M,T)}(\boldsymbol{\theta}) := \sum_{\boldsymbol{s} \in \mathcal{S}_M} \sum_{t=1}^{T} \sum_{\substack{\boldsymbol{h} \in \mathcal{H}_{\boldsymbol{r}} \\ \boldsymbol{s}+\boldsymbol{h} \in \mathcal{S}_M}} \sum_{\substack{u=0 \\ t+u \leq T}}^{p} \mathbb{1}_{\{(\boldsymbol{h},u) \neq (\boldsymbol{0},0)\}} \log \left\{ g_{\boldsymbol{\theta}} \left(\eta(\boldsymbol{s}, t), \eta(\boldsymbol{s}+\boldsymbol{h}, t+u) \right) \right\}$$

$$= \sum_{\boldsymbol{s} \in \mathcal{S}_M} \sum_{t=1}^{T} q_{\boldsymbol{\theta}}(\boldsymbol{s}, t; \boldsymbol{r}, p) - \mathcal{R}^{(M,T)}(\boldsymbol{\theta}), \quad \boldsymbol{\theta} \in \Theta, \tag{3.4}$$

where

$$q_{\boldsymbol{\theta}}(\boldsymbol{s}, t; \boldsymbol{r}, p) := \sum_{\boldsymbol{h} \in \mathcal{H}_{\boldsymbol{r}}} \sum_{u=0}^{p} \mathbb{1}_{\{(\boldsymbol{h},u) \neq (\boldsymbol{0},0)\}} \log \left\{ g_{\boldsymbol{\theta}} \left(\eta(\boldsymbol{s}, t), \eta(\boldsymbol{s}+\boldsymbol{h}, t+u) \right) \right\} \tag{3.5}$$

and

$$\mathcal{R}^{(M,T)}(\boldsymbol{\theta}) := \sum_{\boldsymbol{s} \in \mathcal{S}_M} \sum_{t=1}^{T} \sum_{\boldsymbol{h} \in \mathcal{H}_{\boldsymbol{r}}} \sum_{u=0}^{p} \mathbb{1}_{\{\boldsymbol{s}+\boldsymbol{h} \notin \mathcal{S}_M \text{ or } t+u > T\}} \log \left\{ g_{\boldsymbol{\theta}} \left(\eta(\boldsymbol{s}, t), \eta(\boldsymbol{s}+\boldsymbol{h}, t+u) \right) \right\}$$

$$= \sum_{\boldsymbol{h} \in \mathcal{H}_{\boldsymbol{r}}} \sum_{u=0}^{p} \sum_{(\boldsymbol{s},t) \in \mathcal{G}_{M,T}(\boldsymbol{h},u)} \log \left\{ g_{\boldsymbol{\theta}} \left(\eta(\boldsymbol{s}, t), \eta(\boldsymbol{s}+\boldsymbol{h}, t+u) \right) \right\}, \tag{3.6}$$

with

$$\mathcal{G}_{M,T}(\boldsymbol{h}, u) := \left\{ (\boldsymbol{s}, t) \in \mathcal{S}_M \times \mathcal{T}_T : \boldsymbol{s} + \boldsymbol{h} \notin \mathcal{S}_M \text{ or } t + u > T \right\}. \tag{3.7}$$

for $(\boldsymbol{h}, u) \in \mathbb{N}^{d+1}$. The pairwise maximum likelihood estimate (PMLE) is given by

$$\widehat{\boldsymbol{\theta}} = \operatorname*{argmax}_{\boldsymbol{\theta} \in \Theta} PL^{(M,T)}(\boldsymbol{\theta}). \tag{3.8}$$

We derive the asymptotic properties of the PMLE for two scenarios. The first one is based on regularly spaced observations with an increasing spatio-temporal domain. For this scenario we follow the proofs in Davis et al. [8] and show that the properties of strong consistency and asymptotic normality also hold if the dependence structure δ allows for spatially anisotropic effects as in (2.4). In the second scenario, the observations are taken from a fixed spatial domain and an increasing temporal domain.

5

3.1 Increasing spatio-temporal domain

Lemma 1. *For $(\boldsymbol{h}, u) \in \mathcal{H}_{\boldsymbol{r}} \times \{0, \ldots, p\}$, it holds that*

$$|\mathcal{G}_{M,T}(\boldsymbol{h}, u)| \le K_2(M^{d-1}T + M^d),$$

where K_2 is a constant independent of M and T.

Proof. The number of space-time points within the space-time observation area, from which some grid point outside the observation area is within a lag $(\boldsymbol{h}, u) \in \mathcal{H}_{\boldsymbol{r}} \times \{1, \ldots, p\}$, is bounded by $M^{d-1}T \sum_{j=1}^{d} r_j + M^d p$. Thus we obtain

$$|\mathcal{G}_{M,T}(\boldsymbol{h}, u)| \le M^{d-1}T \sum_{j=1}^{d} r_j + M^d p \le K_2(M^{d-1}T + M^d),$$

where $K_2 := \max\left\{ \sum_{j=1}^{d} r_j, p \right\}$ is a constant independent of M and T. □

Theorem 1 (Strong consistency for large M and T). *Let $\{\eta(\boldsymbol{s}, t) : \boldsymbol{s} \in \mathbb{R}^d, t \in [0, \infty)\}$ be a Brown-Resnick process as in (2.1) with dependence structure*

$$\delta(\boldsymbol{h}, u) = \sum_{j=1}^{d} C_j |h_j|^{\alpha_j} + C_{d+1}|u|^{\alpha_{d+1}}, \quad (\boldsymbol{h}, u) \in \mathbb{R}^{d+1},$$

where $0 < \alpha_j \le 2$ and $C_j > 0$ for $j = 1, \ldots, d+1$. Denote the parameter vector by

$$\boldsymbol{\theta} = (C_1, \ldots, C_{d+1}, \alpha_1, \ldots, \alpha_{d+1}).$$

Assume that the true parameter vector $\boldsymbol{\theta}^\star$ lies in a compact set

$$\Theta^\star \subset \{(C_1, \ldots, C_{d+1}, \alpha_1, \ldots, \alpha_{d+1}) : C_j \in (0, \infty), \alpha_j \in (0, 2], \ j = 1, \ldots, d+1\}. \tag{3.9}$$

Suppose that the following identifiability condition holds for all $(\boldsymbol{s}, t) \in \mathcal{S}_M \times \mathcal{T}_T$:

$$\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}} \quad \Leftrightarrow \tag{3.10}$$
$$g_{\boldsymbol{\theta}}\left(\eta(\boldsymbol{s}, t), \eta(\boldsymbol{s}+\boldsymbol{h}, t+u)\right) = g_{\tilde{\boldsymbol{\theta}}}\left(\eta(\boldsymbol{s}, t), \eta(\boldsymbol{s}+\boldsymbol{h}, t+u)\right), \quad \boldsymbol{h} \in \mathcal{H}_{\boldsymbol{r}}, \quad 0 \le u \le p.$$

Then, the PMLE

$$\widehat{\boldsymbol{\theta}}^{(M,T)} = \underset{\boldsymbol{\theta} \in \Theta^\star}{\arg\max}\, PL^{(M,T)}(\boldsymbol{\theta})$$

is strongly consistent:

$$\widehat{\boldsymbol{\theta}}^{(M,T)} \overset{a.s.}{\to} \boldsymbol{\theta}^\star \text{ as } M, T \to \infty.$$

Proof. The proof uses the method of Wald [33]. One aim is to show that for some chosen maximum space-time lag $(\boldsymbol{r}, p) \in \mathbb{N}_0^{d+1} \setminus \{\boldsymbol{0}\}$ and $\boldsymbol{\theta} \in \Theta^\star$,

$$\frac{1}{M^d T} PL^{(M,T)}(\boldsymbol{\theta})$$
$$= \frac{1}{M^d T}\left(\sum_{\boldsymbol{s} \in \mathcal{S}_M} \sum_{t=1}^{T} q_{\boldsymbol{\theta}}(\boldsymbol{s}, t; \boldsymbol{r}, p) - \mathcal{R}^{(M,T)}(\boldsymbol{\theta}) \right) \overset{a.s.}{\to} \mathrm{PL}(\boldsymbol{\theta}) := \mathbb{E}[q_{\boldsymbol{\theta}}(\boldsymbol{1}, 1; \boldsymbol{r}, p)]$$

as $M, T \to \infty$. This is done by verifying the following two limit results: Uniformly on $\Theta^\star$,

(A) $\dfrac{1}{M^d T} \sum\limits_{\boldsymbol{s} \in \mathcal{S}_M} \sum\limits_{t=1}^{T} q_{\boldsymbol{\theta}}(\boldsymbol{s}, t; \boldsymbol{r}, p) \overset{\text{a.s.}}{\to} \text{PL}(\boldsymbol{\theta})$ as $M, T \to \infty$,

(B) $\dfrac{1}{M^d T} \mathcal{R}^{(M,T)}(\boldsymbol{\theta}) \overset{\text{a.s.}}{\to} 0$ as $M, T \to \infty$.

Furthermore, we need to show:

(C) The limit function $\text{PL}(\boldsymbol{\theta})$ is uniquely maximized at the true parameter vector $\boldsymbol{\theta}^\star \in \Theta^\star$.

We show (A). The almost sure convergence holds because $q_{\boldsymbol{\theta}}(\cdot)$ is a measurable function of lagged versions of $\eta(\boldsymbol{s}, t)$ for $\boldsymbol{s} \in \mathcal{S}_M$, $t \in \mathcal{T}_T$. Proposition 3 of Davis et al. [8] implies a strong law of large numbers. What remains to show is that the convergence is uniform on the compact parameter space $\Theta^\star$. This can be done by carefully following the lines of the proof of Theorem 1 of Davis et al. [8], adapting it to the spatially anisotropic setting. For details we refer to Buhl [6], Theorem 4.4. We find that there is a positive finite constant K_1, independent of $\boldsymbol{\theta}, M$ and T, such that

$$\mathbb{E}\big[|\log g_{\boldsymbol{\theta}}\big(\eta(\boldsymbol{s}^{(1)}, t^{(1)}), \eta(\boldsymbol{s}^{(2)}, t^{(2)})\big)|\big] < K_1, \quad (\boldsymbol{s}^{(1)}, t^{(1)}), (\boldsymbol{s}^{(2)}, t^{(2)}) \in \mathbb{N}^{d+1}, \tag{3.11}$$

and that $\mathbb{E}\big[\sup_{\boldsymbol{\theta} \in \Theta^\star} |q_{\boldsymbol{\theta}}(\boldsymbol{1}, 1; \boldsymbol{r}, p)|\big] < \infty$. Theorem 2.7 of Straumann [31] implies that the convergence is uniform.

Next we show (B). Using Proposition 3 of Davis et al. [8] and (3.11) we have that, uniformly on $\Theta^\star$,

$$\sum_{\boldsymbol{h} \in \mathcal{H}_{\boldsymbol{r}}} \sum_{u=0}^{p} \frac{1}{|\mathcal{G}_{M,T}(\boldsymbol{h}, u)|} \sum_{(\boldsymbol{s}, t) \in \mathcal{G}_{M,T}(\boldsymbol{h}, u)} \log\left\{ g_{\boldsymbol{\theta}}\left(\eta(\boldsymbol{s}, t), \eta(\boldsymbol{s} + \boldsymbol{h}, t + u)\right)\right\}$$
$$\overset{\text{a.s.}}{\to} \mathbb{E}\Big[\sum_{\boldsymbol{h} \in \mathcal{H}_{\boldsymbol{r}}} \sum_{u=0}^{p} \log\left\{ g_{\boldsymbol{\theta}}\left(\eta(\boldsymbol{1}, 1), \eta(\boldsymbol{1} + \boldsymbol{h}, 1 + u)\right)\right\}\Big] \text{ as } M, T \to \infty.$$

By Lemma 1 and (3.11) it follows that, uniformly on $\Theta^\star$,

$$\frac{1}{M^d T}|\mathcal{R}^{(M,T)}(\boldsymbol{\theta})|$$
$$\leq K_2\Big(\frac{1}{M} + \frac{1}{T}\Big)\Big| \sum_{\boldsymbol{h} \in \mathcal{H}_{\boldsymbol{r}}} \sum_{u=0}^{p} \frac{1}{|\mathcal{G}_{M,T}(\boldsymbol{h}, u)|} \sum_{(\boldsymbol{s}, t) \in \mathcal{G}_{M,T}(\boldsymbol{h}, u)} \log\left\{ g_{\boldsymbol{\theta}}\left(\eta(\boldsymbol{s}, t), \eta(\boldsymbol{s} + \boldsymbol{h}, t + u)\right)\right\}\Big|$$
$$\overset{\text{a.s.}}{\to} 0 \quad \text{as } M, T \to \infty,$$

Finally, we prove (C). Let $\boldsymbol{\theta} \neq \boldsymbol{\theta}^\star$. For $\boldsymbol{s} \in \mathcal{S}_M$ and $t \in \mathcal{T}_T$, Jensen's inequality yields

$$\mathbb{E}\left[\log\left\{\frac{g_{\boldsymbol{\theta}}\left(\eta(\boldsymbol{s}, t), \eta(\boldsymbol{s} + \boldsymbol{h}, t + u)\right)}{g_{\boldsymbol{\theta}^\star}\left(\eta(\boldsymbol{s}, t), \eta(\boldsymbol{s} + \boldsymbol{h}, t + u)\right)}\right\}\right] \leq \log\left\{\mathbb{E}\left[\frac{g_{\boldsymbol{\theta}}\left(\eta(\boldsymbol{s}, t), \eta(\boldsymbol{s} + \boldsymbol{h}, t + u)\right)}{g_{\boldsymbol{\theta}^\star}\left(\eta(\boldsymbol{s}, t), \eta(\boldsymbol{s} + \boldsymbol{h}, t + u)\right)}\right]\right\}$$
$$= \log\left\{\int\limits_{(0,\infty)^2} \frac{g_{\boldsymbol{\theta}}(y_1, y_2)}{g_{\boldsymbol{\theta}^\star}(y_1, y_2)} g_{\boldsymbol{\theta}^\star}(y_1, y_2)\, \mathrm{d}(y_1, y_2)\right\}$$
$$= \log\left\{\int\limits_{(0,\infty)^2} g_{\boldsymbol{\theta}}(y_1, y_2)\, \mathrm{d}(y_1, y_2)\right\} = 0,$$

and it directly follows from (3.5) that $\text{PL}(\boldsymbol{\theta}) \leq \text{PL}(\boldsymbol{\theta}^\star)$. As $\boldsymbol{\theta} \neq \boldsymbol{\theta}^\star$, the identifiability condition (3.10) yields (C). $\qquad\square$

r_1	r_2	p	identifiable parameters
1	0	0	C_1
1	1	0	C_1, C_2
1	1	1	C_1, C_2, C_3
> 1	0	0	C_1, α_1
> 1	> 1	> 1	$C_1, \alpha_1, C_2, \alpha_2, C_3, \alpha_3$

Table 1: Identifiable parameters for model (2.4) with $d = 2$ for some examples of maximum space-time lags (r_1, r_2, p).

Remark 1. There are combinations of maximum space-time lags that lead to non-identifiable parameters, see Table 1. However, Theorem 1 still applies to all identifiable parameters (cf. Davis et al. [8], Remark 2).

Next we prove asymptotic normality of the PMLE defined in (3.8). As in the proof of Theorem 1 we follow the lines of proof of Davis et al. [8], Section 5, adapting the arguments to the anisotropic setting. We start with some basic results needed throughout the remainder of the section.

Lemma 2. *Assume that all conditions of Theorem 1 are satisfied. Then for $\boldsymbol{s}^{(1)}, \boldsymbol{s}^{(2)} \in \mathbb{R}^d$ and $t^{(1)}, t^{(2)} \in [0, \infty)$, the following assertions hold componentwise:*

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

$$\mathbb{E}\left[\left|\left|\nabla_{\boldsymbol{\theta}} \log g_{\boldsymbol{\theta}}(\eta(\boldsymbol{s}^{(1)}, t^{(1)}), \eta(\boldsymbol{s}^{(2)}, t^{(2)}))\right|\right|^3\right] < \infty, \quad \boldsymbol{\theta} \in \Theta^\star.$$

(2) The Hessian matrix of the bivariate log-density satisfies

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta} \in \Theta^\star} \left|\nabla^2_{\boldsymbol{\theta}} \log g_{\boldsymbol{\theta}}(\eta(\boldsymbol{s}^{(1)}, t^{(1)}), \eta(\boldsymbol{s}^{(2)}, t^{(2)}))\right|\right] < \infty.$$

Proof. Assume identifiability of all parameters C_j, α_j for $j = 1, \ldots, d+1$. For $y_1, y_2 \in (0, \infty)$ and for $(\boldsymbol{h}, u) \in \mathbb{R}^{d+1} \setminus \{\boldsymbol{0}\}$ lengthy but simple calculations of derivatives of (3.1) yield

$$\nabla_{\boldsymbol{\theta}} \log g_{\boldsymbol{\theta}}(y_1, y_2) = \frac{\partial \log g_{\boldsymbol{\theta}}(y_1, y_2)}{\partial \delta(\boldsymbol{h}, u)} \nabla_{\boldsymbol{\theta}} \delta(\boldsymbol{h}, u),$$

$$\frac{\partial \delta(\boldsymbol{h}, u)}{\partial C_j} = |h_j|^{\alpha_j}, \quad \frac{\partial \delta(\boldsymbol{h}, u)}{\partial \alpha_j} = C_j |h_j|^{\alpha_j} \log |h_j|, \quad j = 1, \ldots d,$$

and

$$\frac{\partial \delta(\boldsymbol{h}, u)}{\partial C_{d+1}} = |u|^{\alpha_{d+1}}, \quad \frac{\partial \delta(\boldsymbol{h}, u)}{\partial \alpha_{d+1}} = C_{d+1} |u|^{\alpha_{d+1}} \log |u|.$$

By compactness of the parameter space, as required in (3.9), we can bound those first partial derivatives as well as the second order partial derivatives from above and below. So it remains to show that for $\boldsymbol{s}^{(1)}, \boldsymbol{s}^{(2)} \in S$ and $t^{(1)}, t^{(2)} \in T$,

$$\mathbb{E}_{\boldsymbol{\theta}^\star}\left[\left|\frac{\partial \log\{g_{\boldsymbol{\theta}}(\eta(\boldsymbol{s}^{(1)}, t^{(1)}), \eta(\boldsymbol{s}^{(2)}, t^{(2)}))\}}{\partial \delta(\boldsymbol{h}, u)}\right|^3\right] < \infty$$

and

$$\mathbb{E}_{\boldsymbol{\theta}^\star}\left[\sup_{\boldsymbol{\theta} \in \Theta^\star} \left|\frac{\partial^2 \log\{g_{\boldsymbol{\theta}}(\eta(\boldsymbol{s}^{(1)}, t^{(1)}), \eta(\boldsymbol{s}^{(2)}, t^{(2)}))\}}{\partial^2 \delta(\boldsymbol{h}, u)}\right|\right] < \infty,$$

where the function $\delta(\boldsymbol{h}, u)$ can be treated as a constant since it is bounded away from 0 by (3.9). Hence, for the rest of the proof we refer to that of Davis et al. [8], Lemma 1. $\square$

For a central limit theorem we need certain mixing properties for a space-time setting (cf. Davis et al. [8], Section 5.1 and Huser and Davison [19], Section 3.2).

Definition 2 (Mixing coefficients and α-mixing). *Let $\{\eta(\boldsymbol{s}, t) : \boldsymbol{s} \in \mathbb{Z}^d, t \in \mathbb{N}\}$ be a space-time process. Let d be some metric induced by a norm on $\mathbb{R}^{d+1}$. For $\Lambda_1, \Lambda_2 \subset \mathbb{Z}^d \times \mathbb{N}$ let*

$$d(\Lambda_1, \Lambda_2) := \inf\{d((\boldsymbol{s}^{(1)}, t^{(1)}), (\boldsymbol{s}^{(2)}, t^{(2)})) : (\boldsymbol{s}^{(1)}, t^{(1)}) \in \Lambda_1, (\boldsymbol{s}^{(2)}, t^{(2)}) \in \Lambda_2\}.$$

(1) For $k, \ell, n \geq 0$ the mixing coefficients *are defined as*

$$\alpha_{k,\ell}(n) := \sup\{|\mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_1)\mathbb{P}(A_2)| :$$
$$A_1 \in \mathcal{F}_{\Lambda_1}, A_2 \in \mathcal{F}_{\Lambda_2}, |\Lambda_1| \leq k, |\Lambda_2| \leq \ell, d(\Lambda_1, \Lambda_2) \geq n\}, \tag{3.12}$$

where $\mathcal{F}_{\Lambda_i} = \sigma(\eta(\boldsymbol{s}, t) : (\boldsymbol{s}, t) \in \Lambda_i)$ for $i = 1, 2$.

(2) $\{\eta(\boldsymbol{s}, t) : \boldsymbol{s} \in \mathbb{Z}^d, t \in \mathbb{N}\}$ is called α-mixing if for all $k, \ell > 0$,

$$\alpha_{k,\ell}(n) \to 0, \quad n \to \infty.$$

Recall from Eq. (2.2) that for $(\boldsymbol{h}, u) \in \mathbb{R}^{d+1}$ with δ as in (2.4) the tail dependence coefficient of the Brown-Resnick process is given by

$$\chi(\boldsymbol{h}, u) = 2\left(1 - \Phi\left(\sqrt{\frac{1}{2}\left[C_1|h_1|^{\alpha_1} + \cdots + C_d|h_d|^{\alpha_d} + C_{d+1}|u|^{\alpha_{d+1}}\right]}\right)\right).$$

Corollary 2.2 of Dombry and Eyi-Minko [12] links the α-mixing coefficients with the tail dependence coefficients, and we will use this for the next result.

Proposition 1. *Let $\{\eta(\boldsymbol{s}, t) : \boldsymbol{s} \in \mathbb{R}^d, t \in [0, \infty)\}$ be the Brown-Resnick process (2.1) with dependence function δ given by (2.4). Then the process $\{\eta(\boldsymbol{s}, t) : \boldsymbol{s} \in \mathbb{Z}^d, t \in \mathbb{N}\}$ is α-mixing, where the mixing coefficients in (3.12) satisfy for $\mathcal{H}_{\boldsymbol{r}}$ as in (3.3)*

(1) $\sum_{n=1}^{\infty} n^d \alpha_{k,\ell}(n) < \infty$ for $k + l \leq 4(|\mathcal{H}_{\boldsymbol{r}}| + 1)(p + 1)$,

(2) $\alpha_{(|\mathcal{H}_{\boldsymbol{r}}|+1)(p+1),\infty}(n) = o(n^{-(d+1)})$ as $n \to \infty$,

(3) $\sum_{n=1}^{\infty} n^d \alpha_{(|\mathcal{H}_{\boldsymbol{r}}|+1)(p+1),(|\mathcal{H}_{\boldsymbol{r}}|+1)(p+1)}(n)^{\frac{1}{3}} < \infty$.

Proof. Note that for $(\boldsymbol{h}, u) \in \mathbb{R}^{d+1}$, by the equivalence of norms, for some positive constant L,

$$d((\boldsymbol{h}, u), (\boldsymbol{0}, 0)) \leq \frac{1}{L} \max\{|h_1|, \ldots, |h_d|, |u|\}$$

Therefore, for $n \in \mathbb{N}$, presuming $d((\boldsymbol{h}, u), (\boldsymbol{0}, 0)) \geq n$ results in $\max\{|h_1|, \ldots, |h_d|, |u|\} \geq Ln$, so that by Corollary 2.2 and Eq. (3) of Dombry and Eyi-Minko [12] we get

$$\alpha_{k,\ell}(n) \leq 2k\ell \sup_{d((\boldsymbol{h},u),(\boldsymbol{0},0)) \geq n} \chi(\boldsymbol{h}, u) \leq 2k\ell \sup_{\max\{|h_1|,\ldots,|h_d|,|u|\} \geq Ln} \chi(\boldsymbol{h}, u), \tag{3.13}$$

$$\alpha_{k,\infty}(n) \leq 2k \sum_{d((\boldsymbol{h},u),(\boldsymbol{0},0)) \geq n} \chi(\boldsymbol{h}, u) \leq 2k \sum_{\max\{|h_1|,\ldots,|h_d|,|u|\} \geq Ln} \chi(\boldsymbol{h}, u). \tag{3.14}$$

9

In the following we use the notation $\|(\boldsymbol{h}, u)\|_\infty := \max\{|h_1|, \ldots, |h_d|, |u|\}$ for $(\boldsymbol{h}, u) \in \mathbb{Z}^d \times \mathbb{N}$. Using $1 - \Phi(x) \le \exp\{-\frac{1}{2}x^2\}$ for $x > 0$ and Eq. (2.2) and (3.13), we find for all $k, \ell \ge 0$,

$$
\begin{aligned}
\alpha_{k,\ell}(n) &\le 4k\ell \sup_{\|(\boldsymbol{h},u)\|_\infty \ge Ln} \left(1 - \Phi\left(\sqrt{\frac{\delta(\boldsymbol{h}, u)}{2}}\right)\right) \\
&\le 4k\ell \sup_{\|(\boldsymbol{h},u)\|_\infty \ge Ln} \exp\left\{-\frac{\delta(\boldsymbol{h}, u)}{4}\right\} \\
&= 4k\ell \sup_{\|(\boldsymbol{h},u)\|_\infty \ge Ln} \exp\left\{-\frac{1}{4}\left[C_1|h_1|^{\alpha_1} + \ldots + C_d|h_d|^{\alpha_d} + C_{d+1}|u|^{\alpha_{d+1}}\right]\right\} \\
&\le 4k\ell \sup_{\|(\boldsymbol{h},u)\|_\infty \ge Ln} \exp\left\{-\frac{1}{4}\min\{|C_1|, \ldots, |C_{d+1}|\}\|(\boldsymbol{h}, u)\|_\infty^{\min\{\alpha_1, \ldots, \alpha_{d+1}\}}\right\} \\
&\le 4k\ell \exp\left\{-\frac{1}{4}\min\{|C_1|, \ldots, |C_{d+1}|\}(Ln)^{\min\{\alpha_1, \ldots, \alpha_{d+1}\}}\right\} \qquad\qquad (3.15)
\end{aligned}
$$

$\to 0$ as $n \to \infty$.

This implies α-mixing.

By similar arguments we obtain by (3.14) for all $k \ge 0$,

$$
\begin{aligned}
&\alpha_{k,\infty}(n) \le \\
&4k \sum_{\|(\boldsymbol{h},u)\|_\infty \ge Ln} \exp\left\{-\frac{1}{4}\min\{|C_1|, \ldots, |C_{d+1}|\}\|(\boldsymbol{h}, u)\|_\infty^{\min\{\alpha_1, \ldots, \alpha_{d+1}\}}\right\}. \qquad (3.16)
\end{aligned}
$$

We use the above bounds to prove assertions (1)-(3).

(1) For $k + \ell \le 4(|\mathcal{H}_{\boldsymbol{r}}| + 1)(p + 1)$ we have by (3.15),

$$
\sum_{n=1}^\infty n^d \alpha_{k,\ell}(n) \le 4k\ell \sum_{n=1}^\infty n^d \exp\left\{-\frac{1}{4}\min\{|C_1|, \ldots, |C_{d+1}|\}(Ln)^{\min\{\alpha_1, \ldots, \alpha_{d+1}\}}\right\} < \infty.
$$

(2) First note that the number of grid points $(\boldsymbol{h}, u) \in \mathbb{R}^{d+1}$ with $\|(\boldsymbol{h}, u)\|_\infty = i$ for $i \in \mathbb{N}$ equals $(i+1)^{d+1} - i^{d+1}$, and is therefore of order $\mathcal{O}(i^d)$. We use (3.16) and a more precise estimate than in part (1) to obtain for sufficiently large n

$$
\begin{aligned}
&n^{d+1} \alpha_{(|\mathcal{H}_{\boldsymbol{r}}|+1)(p+1),\infty}(n) \\
&\le 4n^{d+1}(|\mathcal{H}_{\boldsymbol{r}}| + 1)(p + 1) \\
&\qquad \sum_{\|(\boldsymbol{h},u)\|_\infty \ge Ln} \exp\left\{-\frac{1}{4}\min\{|C_1|, \ldots, |C_{d+1}|\}\|(\boldsymbol{h}, u)\|_\infty^{\min\{\alpha_1, \ldots, \alpha_{d+1}\}}\right\} \\
&\le K_3 n^{d+1}(|\mathcal{H}_{\boldsymbol{r}}| + 1)(p + 1) \sum_{i=\lfloor Ln \rfloor}^\infty i^d \exp\left\{-\frac{1}{4}\min\{C_1, \ldots, C_{d+1}\}i^{\min\{\alpha_1, \ldots, \alpha_{d+1}\}}\right\}
\end{aligned}
$$

$\to 0$ as $n \to \infty$,

where K_3 is a positive constant. Convergence to 0 follows using the integral test for power series convergence and Lemma 4, Eq. (A.1).

(3) We find, using again (3.15),

$$
\sum_{n=1}^\infty n^d \alpha_{(|\mathcal{H}_{\boldsymbol{r}}|+1)(p+1),(|\mathcal{H}_{\boldsymbol{r}}|+1)(p+1)}(n)^{\frac{1}{3}}
$$

10

$$\leq \left(4\left[\,(|\mathcal{H}_{\boldsymbol{r}}|+1)(p+1)\right]^{2}\right)^{\frac{1}{3}}$$

$$\cdot \sum_{n=1}^{\infty} n^{d} \exp\left\{-\frac{1}{12}\min\{C_{1},\ldots,C_{d+1}\}(Ln)^{\min\{\alpha_{1},\ldots,\alpha_{d+1}\}}\right\}$$

$$< \infty$$

as in (1). $\qquad\qquad\square$

Because of Lemma 2 and Proposition 1 the following central limit theorem of Bolthausen [3] holds.

Corollary 1. *Consider the process* $\{\nabla_{\boldsymbol{\theta}} q_{\boldsymbol{\theta}^{\star}}(\boldsymbol{s},t;\boldsymbol{r},p) : \boldsymbol{s} \in \mathbb{Z}^{d}, t \in \mathbb{N}\}$. *Then*

$$\frac{1}{M^{\frac{d}{2}}\sqrt{T}}\sum_{\boldsymbol{s}\in\mathcal{S}_{M}}\sum_{t=1}^{T}\nabla_{\boldsymbol{\theta}}q_{\boldsymbol{\theta}^{\star}}(\boldsymbol{s},t;\boldsymbol{r},p) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0},\Sigma_{1}) \ \text{as } M,T\to\infty,$$

where

$$\Sigma_{1} := \sum_{s_{1}=-\infty}^{\infty}\cdots\sum_{s_{d}=-\infty}^{\infty}\sum_{t=1}^{\infty}\mathrm{Cov}\left[\nabla_{\boldsymbol{\theta}}q_{\boldsymbol{\theta}^{\star}}(\mathbf{1},1;\boldsymbol{r},p),\nabla_{\boldsymbol{\theta}}q_{\boldsymbol{\theta}^{\star}}(s_{1},\ldots,s_{d},t;\boldsymbol{r},p)\right]. \qquad (3.17)$$

Now we formulate the main result of this section.

Theorem 2 (Asymptotic normality for large M and T)**.** *Assume the same conditions as in Theorem 1. Then*

$$\sqrt{M^{d}T}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^{\star}) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0},\tilde{\Sigma}_{1}) \ \text{as } M,T\to\infty, \qquad (3.18)$$

where $\tilde{\Sigma}_{1} := F_{1}^{-1}\Sigma_{1}(F_{1}^{-1})^{\top}$ *with* Σ_{1} *given in* (3.17) *and*

$$F_{1} := \mathbb{E}\left[-\nabla_{\boldsymbol{\theta}}^{2}q_{\boldsymbol{\theta}^{\star}}(\mathbf{1},1;\boldsymbol{r},p)\right].$$

Proof. A Taylor expansion of the score function $\nabla_{\boldsymbol{\theta}}PL^{(M,T)}(\boldsymbol{\theta})$ around the true parameter vector $\boldsymbol{\theta}^{\star}$ yields for some $\tilde{\boldsymbol{\theta}} \in [\widehat{\boldsymbol{\theta}},\boldsymbol{\theta}^{\star}]$:

$$\mathbf{0} = \nabla_{\boldsymbol{\theta}}PL^{(M,T)}(\widehat{\boldsymbol{\theta}}) = \nabla_{\boldsymbol{\theta}}PL^{(M,T)}(\boldsymbol{\theta}^{\star}) + \nabla_{\boldsymbol{\theta}}^{2}PL^{(M,T)}(\tilde{\boldsymbol{\theta}})(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^{\star}).$$

Therefore,

$$M^{\frac{d}{2}}\sqrt{T}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^{\star}) = -\left(\frac{1}{M^{d}T}\nabla_{\boldsymbol{\theta}}^{2}PL^{(M,T)}(\tilde{\boldsymbol{\theta}})\right)^{-1}\left(\frac{1}{M^{\frac{d}{2}}\sqrt{T}}\nabla_{\boldsymbol{\theta}}PL^{(M,T)}(\boldsymbol{\theta}^{\star})\right)$$

$$= -\left(\frac{1}{M^{d}T}\sum_{\boldsymbol{s}\in\mathcal{S}_{M}}\sum_{t=1}^{T}\nabla_{\boldsymbol{\theta}}^{2}q_{\tilde{\boldsymbol{\theta}}}(\boldsymbol{s},t;\boldsymbol{r},p) - \frac{1}{M^{d}T}\nabla_{\boldsymbol{\theta}}^{2}\mathcal{R}^{(M,T)}(\tilde{\boldsymbol{\theta}})\right)^{-1}$$

$$\left(\frac{1}{M^{\frac{d}{2}}\sqrt{T}}\sum_{\boldsymbol{s}\in\mathcal{S}_{M}}\sum_{t=1}^{T}\nabla_{\boldsymbol{\theta}}q_{\boldsymbol{\theta}^{\star}}(\boldsymbol{s},t;\boldsymbol{r},p) - \frac{1}{M^{\frac{d}{2}}\sqrt{T}}\nabla_{\boldsymbol{\theta}}\mathcal{R}^{(M,T)}(\boldsymbol{\theta}^{\star})\right)$$

$$=: -(I_{1}-I_{2})^{-1}(J_{1}-J_{2}).$$

Note the following:

- Corollary 1 implies that $J_{1} \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0},\Sigma_{1})$ as $M,T\to\infty$.

- Using representation (3.6) of the boundary term $\mathcal{R}^{(M,T)}(\cdot)$ and Lemma 1, we find

$$\|J_2\| = \frac{1}{M^{\frac{d}{2}}\sqrt{T}}\left\|\sum_{\boldsymbol{h}\in\mathcal{H}_{\boldsymbol{r}}}\sum_{u=0}^{p}\sum_{(\boldsymbol{s},t)\in\mathcal{G}_{M,T}(\boldsymbol{h},u)}\nabla_{\boldsymbol{\theta}}\log\{g_{\boldsymbol{\theta}^\star}(\eta(\boldsymbol{s},t),\eta(\boldsymbol{s}+\boldsymbol{h},t+u))\}\right\|$$

$$\leq \sqrt{K_2}\frac{\sqrt{M^{d-1}T+M^d}}{M^{\frac{d}{2}}\sqrt{T}}$$

$$\left\|\sum_{\boldsymbol{h}\in\mathcal{H}_{\boldsymbol{r}}}\sum_{u=0}^{p}\frac{1}{\sqrt{|\mathcal{G}_{M,T}(\boldsymbol{h},u)|}}\sum_{(\boldsymbol{s},t)\in\mathcal{G}_{M,T}(\boldsymbol{h},u)}\nabla_{\boldsymbol{\theta}}\log\{g_{\boldsymbol{\theta}^\star}(\eta(\boldsymbol{s},t),\eta(\boldsymbol{s}+\boldsymbol{h},t+u))\}\right\|$$

$$\leq \sqrt{K_2}(\frac{1}{\sqrt{M}}+\frac{1}{\sqrt{T}})$$

$$\left\|\sum_{\boldsymbol{h}\in\mathcal{H}_{\boldsymbol{r}}}\sum_{u=0}^{p}\frac{1}{\sqrt{|\mathcal{G}_{M,T}(\boldsymbol{h},u)|}}\sum_{(\boldsymbol{s},t)\in\mathcal{G}_{M,T}(\boldsymbol{h},u)}\nabla_{\boldsymbol{\theta}}\log\{g_{\boldsymbol{\theta}^\star}(\eta(\boldsymbol{s},t),\eta(\boldsymbol{s}+\boldsymbol{h},t+u))\}\right\|$$

In the same way as done in Corollary 1 for the process $\{\nabla_{\boldsymbol{\theta}}q_{\boldsymbol{\theta}^\star}(\boldsymbol{s},t;\boldsymbol{r},p):\boldsymbol{s}\in\mathbb{Z}^d,t\in\mathbb{N}\}$, we can apply Bolthausen's central limit theorem to the processes $\{\nabla_{\boldsymbol{\theta}}\log\{g_{\boldsymbol{\theta}^\star}(\eta(\boldsymbol{s},t),\eta(\boldsymbol{s}+\boldsymbol{h},t+u))\}:$ $\boldsymbol{s}\in\mathbb{Z}^d,t\in\mathbb{N}\}$ for $\boldsymbol{h}\in\mathcal{H}_{\boldsymbol{r}}$, $u\in\{0,\ldots,p\}$. We conclude that

$$\sum_{\boldsymbol{h}\in\mathcal{H}_{\boldsymbol{r}}}\sum_{u=0}^{p}\frac{1}{\sqrt{|\mathcal{G}_{M,T}(\boldsymbol{h},u)|}}\sum_{(\boldsymbol{s},t)\in\mathcal{G}_{M,T}(\boldsymbol{h},u)}\nabla_{\boldsymbol{\theta}}\log\{g_{\boldsymbol{\theta}^\star}(\eta(\boldsymbol{s},t),\eta(\boldsymbol{s}+\boldsymbol{h},t+u))\}$$

converges weakly to a normal distribution as $M,T\to\infty$, and it follows that $J_2\xrightarrow{P}\boldsymbol{0}$ as $M,T\to\infty$.

- As $\{\eta(\boldsymbol{s},t):\boldsymbol{s}\in\mathbb{Z}^d,t\in\mathbb{N}\}$ is α-mixing, the process

$$\{\nabla_{\boldsymbol{\theta}}^2 q_{\boldsymbol{\theta}}(\boldsymbol{s},t;\boldsymbol{r},p):\boldsymbol{s}\in\mathbb{Z}^d,t\in\mathbb{N}\}$$

is α-mixing as a set of measurable functions of mixing lagged processes. Furthermore, as $\tilde{\boldsymbol{\theta}}\in[\widehat{\boldsymbol{\theta}},\boldsymbol{\theta}^\star]$ and $\widehat{\boldsymbol{\theta}}$ is strongly consistent, we have that $I_1\xrightarrow{\text{a.s.}}-F_1$ as $M,T\to\infty$. The convergence is uniform on $\Theta^\star$ by Lemma 2 which implies that

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\Theta^\star}\left|\nabla_{\boldsymbol{\theta}}^2 q_{\boldsymbol{\theta}}(\boldsymbol{1},1;\boldsymbol{r},p)\right|\right]<\infty.$$

- Concerning I_2, the law of large numbers applied to

$$\left\{\nabla_{\boldsymbol{\theta}}^2\log\{g_{\boldsymbol{\theta}}(\eta(\boldsymbol{s},t),\eta(\boldsymbol{s}+\boldsymbol{h},t+u))\}:\boldsymbol{s}\in\mathbb{Z}^d,t\in\mathbb{N}\right\}$$

results in the fact that, in the same way as in part (B) of the proof of Theorem 1, $I_2\xrightarrow{\text{a.s.}}\boldsymbol{0}$ as $M,T\to\infty$.

Finally, summarising these results, Slutzky's Lemma yields (3.18). $\qquad\square$

3.2 Fixed spatial domain and increasing temporal domain

As before we compute the PMLE based on observations on the area $\mathcal{S}_M \times \mathcal{T}_T$, but now we consider M fixed, whereas T tends to infinity.

We define the temporal α-mixing coefficients (cf. Ibragimov and Linnik [21], Definition 17.2.1 or Bradley [4], Definition 1.6).

Definition 3 (Temporal mixing coefficients and temporal α-mixing). *Let $\{\eta(\boldsymbol{s}, t) : \boldsymbol{s} \in \mathcal{S}_M, t \in \mathbb{N}\}$ be a space-time process. Consider the metric $d(\cdot)$ of Definition 2.*

(1) Let $\mathcal{T}^{(1)}, \mathcal{T}^{(2)} \subset \mathbb{N}$. For $n \geq 0$ the temporal α-mixing coefficients are defined as

$$\alpha(n) := \sup\{|P(A_1 \cap A_2) - P(A_1)P(A_2)| :$$
$$A_1 \in \mathcal{F}_{\mathcal{S}_M \times \mathcal{T}^{(1)}}, A_2 \in \mathcal{F}_{\mathcal{S}_M \times \mathcal{T}^{(2)}}, d(\mathcal{S}_M \times \mathcal{T}^{(1)}, \mathcal{S}_M \times \mathcal{T}^{(2)}) \geq n\}, \qquad (3.19)$$

where $\mathcal{F}_{\mathcal{S}_M \times \mathcal{T}^{(i)}} = \sigma(\eta(\boldsymbol{s}, t) : (\boldsymbol{s}, t) \in \mathcal{S}_M \times \mathcal{T}^{(i)})$ for $i = 1, 2$.

(2) $\{\eta(\boldsymbol{s}, t) : \boldsymbol{s} \in \mathcal{S}_M, t \in \mathbb{N}\}$ is called temporally α-mixing, if

$$\alpha(n) \to 0, \quad n \to \infty. \qquad (3.20)$$

Proposition 2. *Let $\{\eta(\boldsymbol{s}, t) : \boldsymbol{s} \in \mathbb{R}^d, t \in [0, \infty)\}$ be the Brown-Resnick process (2.1) with dependence function δ given by (2.4). Then the process $\{\eta(\boldsymbol{s}, t) : \boldsymbol{s} \in \mathcal{S}_M, t \in \mathbb{N}\}$ is temporally α-mixing, where the mixing coefficients (3.19) satisfy*

$$\sum_{n=1}^{\infty} |\alpha(n)|^{\frac{1}{3}} < \infty. \qquad (3.21)$$

Proof. We use Eq. (3) and Corollary 2.2 of Dombry and Eyi-Minko [12] and (2.2) to obtain for $n \in \mathbb{N}$

$$\alpha(n)$$
$$\leq 2 \sup_{d(\mathcal{S}_M \times \mathcal{T}^{(1)}, \mathcal{S}_M \times \mathcal{T}^{(2)}) \geq n} \sum_{\substack{(\boldsymbol{s}^{(1)}, t^{(1)}) \\ \in \mathcal{S}_M \times \mathcal{T}^{(1)}}} \sum_{\substack{(\boldsymbol{s}^{(2)}, t^{(2)}) \\ \in \mathcal{S}_M \times \mathcal{T}^{(2)}}} \chi(\boldsymbol{s}^{(1)} - \boldsymbol{s}^{(2)}, t^{(1)} - t^{(2)})$$

$$= 4 \sup_{d(\mathcal{S}_M \times \mathcal{T}^{(1)}, \mathcal{S}_M \times \mathcal{T}^{(2)}) \geq n} \sum_{\substack{(\boldsymbol{s}^{(1)}, t^{(1)}) \\ \in \mathcal{S}_M \times \mathcal{T}^{(1)}}} \sum_{\substack{(\boldsymbol{s}^{(2)}, t^{(2)}) \\ \in \mathcal{S}_M \times \mathcal{T}^{(2)}}}$$

$$\left(1 - \Phi\left(\sqrt{\frac{1}{2}\left[C_1 |s_1^{(1)} - s_1^{(2)}|^{\alpha_1} + \cdots + C_d |s_d^{(1)} - s_d^{(2)}|^{\alpha_d} + C_{d+1}|t^{(1)} - t^{(2)}|^{\alpha_{d+1}} \right]} \right) \right)$$

$$\leq 4M^{2d} \sup_{d(\mathcal{S}_M \times \mathcal{T}^{(1)}, \mathcal{S}_M \times \mathcal{T}^{(2)}) \geq n} \sum_{\substack{(t^{(1)}, t^{(2)}) \\ \in \mathcal{T}^{(1)} \times \mathcal{T}^{(2)}}} \left(1 - \Phi\left(\sqrt{\frac{1}{2}\left[C_{d+1}|t^{(1)} - t^{(2)}|^{\alpha_{d+1}} \right]} \right) \right)$$

$$\leq 4M^{2d} \sup_{d(\mathcal{S}_M \times \mathcal{T}^{(1)}, \mathcal{S}_M \times \mathcal{T}^{(2)}) \geq n} \sum_{\substack{(t^{(1)}, t^{(2)}) \\ \in \mathcal{T}^{(1)} \times \mathcal{T}^{(2)}}} \exp\left\{ -\frac{1}{4} C_{d+1}|t^{(1)} - t^{(2)}|^{\alpha_{d+1}} \right\},$$

where the last inequality follows from $1 - \Phi(x) \leq \exp\{-\frac{1}{2}x^2\}$ for $x > 0$. We bound $\alpha(n)$ for large n further by

$$\alpha(n) \leq 4M^{2d} \sum_{t^{(1)} \in \{-\infty, \ldots, 0\}} \sum_{t^{(2)} \in \{n, \ldots, \infty\}} \exp\left\{ -\frac{1}{4} C_{d+1}|t^{(1)} - t^{(2)}|^{\alpha_{d+1}} \right\}.$$

In the double sum a temporal lag $u = |t^{(1)} - t^{(2)}| \geq n$ appears exactly $u - (n-1)$ times. This yields

$$\alpha(n) \leq 4M^{2d} \sum_{u=n}^{\infty} (u - (n-1)) \exp\left\{ -\frac{1}{4} C_{d+1} u^{\alpha_{d+1}} \right\}$$

$$\leq 4M^{2d} \sum_{u=n}^{\infty} u \exp\left\{ -\frac{1}{4} C_{d+1} u^{\alpha_{d+1}} \right\}.$$

Convergence of the series (3.21) now follows by the integral test and Lemma 4. $\qquad\square$

In the following we show that strong consistency of the PMLE also holds, if the spatial domain remains fixed.

Theorem 3 (Strong consistency for fixed M and large T)**.** *Assume the same conditions as in Theorem 1 restricted to the fixed space $\mathcal{S}_M$. Then the PMLE*

$$\widehat{\boldsymbol{\theta}}^{(M,T)} = \arg\max_{\boldsymbol{\theta} \in \Theta^\star} PL^{(M,T)}(\boldsymbol{\theta})$$

is strongly consistent, that is,

$$\widehat{\boldsymbol{\theta}}^{(M,T)} \overset{\text{a.s.}}{\to} \boldsymbol{\theta}^\star \ \text{as } T \to \infty.$$

Proof. For $\boldsymbol{\theta} \in \Theta^\star$ and $t \in \mathbb{N}$, set

$$q_{\boldsymbol{\theta}}^M(t; \boldsymbol{r}, p) := \sum_{\boldsymbol{s} \in \mathcal{S}_M} \sum_{\substack{\boldsymbol{h} \in \mathcal{H}_{\boldsymbol{r}} \\ \boldsymbol{s}+\boldsymbol{h} \in \mathcal{S}_M}} \sum_{\substack{u=0 \\ t+u \leq T}}^{p} \mathbb{1}_{\{(\boldsymbol{h},u) \neq (\boldsymbol{0},0)\}} \log\left\{ g_{\boldsymbol{\theta}}\left(\eta(\boldsymbol{s},t), \eta(\boldsymbol{s}+\boldsymbol{h}, t+u) \right) \right\}.$$

Then

$$PL^{(M,T)}(\boldsymbol{\theta}) = \sum_{t=1}^{T} q_{\boldsymbol{\theta}}^M(t; \boldsymbol{r}, p).$$

Following carefully the lines of the proof of Theorem 1, the following conditions hold for fixed spatial domain:

(A) $\dfrac{1}{T} \sum\limits_{t=1}^{T} q_{\boldsymbol{\theta}}^M(t; \boldsymbol{r}, p) \overset{\text{a.s.}}{\to} PL^M(\boldsymbol{\theta}) := \mathbb{E}[(q_{\boldsymbol{\theta}}^M(1; \boldsymbol{r}, p)]$ as $T \to \infty$ uniformly on the compact parameter space $\Theta^\star$. The main argument is that $q_{\boldsymbol{\theta}}^M(\cdot)$ is a function of temporally mixing lagged processes, then we apply again Theorem 2.7 of Straumann [31].

(B) The limit function $PL^M(\boldsymbol{\theta})$ is uniquely maximised at the true parameter vector $\boldsymbol{\theta}^\star \in \Theta^\star$.

$\qquad\square$

Now we formulate the main result of this section.

Theorem 4 (Asymptotic normality for fixed M and large T)**.** *Assume the same conditions as in Theorem 1 restricted to the fixed space $\mathcal{S}_M$. Then*

$$\sqrt{T}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star) \overset{\mathcal{D}}{\to} \mathcal{N}(\boldsymbol{0}, \tilde{\Sigma}_2) \ \text{as } T \to \infty, \tag{3.22}$$

where $\tilde{\Sigma}_2 := F_2^{-1} \Sigma_2 (F_2^{-1})^\top$ with

$$F_2 := \mathbb{E}[-\nabla_{\boldsymbol{\theta}}^2 q_{\boldsymbol{\theta}^\star}^M(1; \boldsymbol{r}, p)]$$

and

$$\Sigma_2 := \mathbb{V}\mathrm{ar}[\nabla_{\boldsymbol{\theta}} q^M_{\boldsymbol{\theta}^\star}(1; \boldsymbol{r}, p)] + 2 \sum_{t=2}^{\infty} \mathbb{C}\mathrm{ov}[\nabla_{\boldsymbol{\theta}} q^M_{\boldsymbol{\theta}^\star}(1; \boldsymbol{r}, p), \nabla_{\boldsymbol{\theta}} q^M_{\boldsymbol{\theta}^\star}(t; \boldsymbol{r}, p)].$$

Proof. By its definition as a function of lagged temporally mixing processes, $(\nabla_{\boldsymbol{\theta}} q^M_{\boldsymbol{\theta}^\star}(t; \boldsymbol{r}, p))_{t \in \mathbb{N}}$ is also temporally α-mixing with coefficients $\alpha'(n) = \alpha(n-p)$. Furthermore,

$$\mathbb{E}\left[\nabla_{\boldsymbol{\theta}} \log\{g_{\boldsymbol{\theta}^\star}(\eta(\mathbf{0}, 0), \eta(\boldsymbol{h}, u))\}\right] = 0, \quad (\boldsymbol{h}, u) \in \mathbb{N}_0^{d+1},$$

because Lemma 2 implies regularity conditions of the pairwise log-likelihood (3.4) allowing to interchange differentiation and integration. Now note that Lemma 2 and Proposition 2 imply that

- $\mathbb{E}[|\nabla_{\boldsymbol{\theta}} q^M_{\boldsymbol{\theta}^\star}(t; \boldsymbol{r}, p)|^3] < \infty$ for $t \in \mathbb{N}$ and every maximum spatial lag $\boldsymbol{r}$ and time lag p, and that

- $\sum_{n=1}^{\infty} |\alpha'(n)|^{\frac{1}{3}} < \infty$.

Therefore, the conditions of Theorem 18.5.3 of Ibragimov and Linnik [21] (see also Bradley [4], Theorem 10.7) are satisfied and we conclude that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \nabla_{\boldsymbol{\theta}} q^M_{\boldsymbol{\theta}^\star}(t; \boldsymbol{r}, p) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \Sigma_2) \text{ as } T \to \infty. \tag{3.23}$$

Taylor expansion of the score function $\nabla_{\boldsymbol{\theta}} PL^{(M,T)}(\boldsymbol{\theta})$ around the true parameter vector $\boldsymbol{\theta}^\star$ yields for some $\tilde{\boldsymbol{\theta}} \in [\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}^\star]$:

$$\mathbf{0} = \nabla_{\boldsymbol{\theta}} PL^{(M,T)}(\widehat{\boldsymbol{\theta}}) = \nabla_{\boldsymbol{\theta}} PL^{(M,T)}(\boldsymbol{\theta}^\star) + \nabla_{\boldsymbol{\theta}}^2 PL^{(M,T)}(\tilde{\boldsymbol{\theta}})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star).$$

Therefore,

$$\sqrt{T}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star) = -\left(\frac{1}{T} \nabla_{\boldsymbol{\theta}}^2 PL^{(M,T)}(\tilde{\boldsymbol{\theta}})\right)^{-1} \left(\frac{1}{\sqrt{T}} \nabla_{\boldsymbol{\theta}} PL^{(M,T)}(\boldsymbol{\theta}^\star)\right)$$

$$= -\left(\frac{1}{T} \sum_{t=1}^{T} \nabla_{\boldsymbol{\theta}}^2 q^M_{\tilde{\boldsymbol{\theta}}}(t; \boldsymbol{r}, p)\right)^{-1} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \nabla_{\boldsymbol{\theta}} q^M_{\boldsymbol{\theta}^\star}(t; \boldsymbol{r}, p)\right) =: -I^{-1} J.$$

Note the following:

- (3.23) implies that $J \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \Sigma_2)$ as $T \to \infty$.

- Uniform convergence holds because of Lemma 2 which implies that componentwise

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta} \in \Theta^\star} |\nabla_{\boldsymbol{\theta}}^2 q^M_{\boldsymbol{\theta}}(1; \boldsymbol{r}, p)|\right] < \infty.$$

  By temporal α-mixing, since $\tilde{\boldsymbol{\theta}} \in [\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}^\star]$, and $\widehat{\boldsymbol{\theta}}$ is strongly consistent, we have $I \xrightarrow{\text{a.s.}} -F_2$ as $T \to \infty$.

Finally, summarising those results, Slutzky's Lemma yields (3.22). $\qquad\square$

Throughout this section we have proved asymptotic properties of the parameter estimates of model (2.4) by classical results for ML estimators in combination with a spatio-temporal central limit theorem. Such results can also be applied to other models like geometrically anisotropic models, provided the required rates for α-mixing hold.

4 Test for spatial isotropy

We use the results of Section 3 to formulate statistical tests for spatial isotropy versus anisotropy based on the model (2.4),

$$\delta(\boldsymbol{h}, u) = \sum_{j=1}^{d} C_j |h_j|^{\alpha_j} + C_{d+1} |u|^{\alpha_{d+1}},$$

for spatial lags $(\boldsymbol{h}, u) = (h_1, \ldots, h_d, u) \in \mathbb{R}^{d+1}$. We derive the necessary results for $d = 2$. Generalisations to higher dimensions are possible, but notationally much more involved. Again we consider the two cases of an increasing and fixed spatial domain.

Due to the structure of model (2.4) a test for isotropy versus anisotropy is a test of

$$H_0 : \{C_1 = C_2 \text{ and } \alpha_1 = \alpha_2\} \quad \text{versus} \quad H_1 : \{C_1 \neq C_2 \text{ or } \alpha_1 \neq \alpha_2\}. \tag{4.1}$$

4.1 Increasing spatial domain

From Theorem 2 we know that, under suitable regularity conditions, the PMLE

$$\widehat{\boldsymbol{\theta}} = (\widehat{C}_1, \widehat{C}_2, \widehat{C}_3, \widehat{\alpha}_1, \widehat{\alpha}_2, \widehat{\alpha}_3)$$

is asymptotically normal; more precisely, for M^2 spatial observations on a regular grid and for T equidistant time points we have

$$M\sqrt{T} \begin{pmatrix} \widehat{C}_1 - C_1 \\ \widehat{C}_2 - C_2 \\ \widehat{C}_3 - C_3 \\ \widehat{\alpha}_1 - \alpha_1 \\ \widehat{\alpha}_2 - \alpha_2 \\ \widehat{\alpha}_3 - \alpha_3 \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}(\boldsymbol{0}, \tilde{\Sigma}_1) \text{ as } M, T \to \infty, \tag{4.2}$$

where $\tilde{\Sigma}_1 \in \mathbb{R}^{6 \times 6}$ is the asymptotic covariance matrix given in Theorem 2.

Our test is based on the spatial parameters only. Moreover, we test the two equalities in H_0 separately and use Bonferroni's inequality to solve the multiple test problem.

Lemma 3. *Assume the conditions of Theorem 2. Setting $A_1 := (-1, 1, 0, 0, 0, 0)$ and $A_2 := (0, 0, 0, -1, 1, 0)$, we have that, as $M, T \to \infty$,*

$$M\sqrt{T}((\widehat{C}_2 - \widehat{C}_1) - (C_2 - C_1)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, A_1 \tilde{\Sigma}_1 A_1^\top), \tag{4.3}$$

$$M\sqrt{T}((\widehat{\alpha}_2 - \widehat{\alpha}_1) - (\alpha_2 - \alpha_1)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, A_2 \tilde{\Sigma}_1 A_2^\top). \tag{4.4}$$

Proof. We obtain the left hand side of (4.3) and (4.4) by multiplying A_1 and A_2 to (4.2), respectively. This yields the limits on the right hand side by the continuous mapping theorem. $\qquad \square$

We define

$$\theta_C := (C_2 - C_1), \widehat{\theta}_C := (\widehat{C}_2 - \widehat{C}_1), \quad \theta_\alpha := (\alpha_2 - \alpha_1), \widehat{\theta}_\alpha := (\widehat{\alpha}_2 - \widehat{\alpha}_1).$$

16

Then the multiple test problem (4.1) becomes

$$H_{0,1} : \{\theta_C = 0\} \quad \text{versus} \quad H_{1,1} : \{\theta_C \neq 0\} \tag{4.5}$$

$$H_{0,2} : \{\theta_\alpha = 0\} \quad \text{versus} \quad H_{1,2} : \{\theta_\alpha \neq 0\}. \tag{4.6}$$

Since the variances in (4.3) and (4.4) are not known explicitly, we find the rejection areas of the two tests by subsampling as suggested in Politis et al. [29], Chapter 5. Their main Assumption 5.3.1, the existence of a weak limit law of the estimates, is satisfied by Lemma 3.

We formulate the subsampling procedure in the terminology of the space-time process $\{\eta(\boldsymbol{s}, t) : \boldsymbol{s} \in \mathcal{S}_M, t \in \mathcal{T}_T\}$. We choose space-time block lengths $\boldsymbol{b} = (b_1, b_2, b_3) \geq (1, 1, 1)$ and the degree of overlap $\boldsymbol{e} = (e_1, e_2, e_3) \leq (M, M, T)$. The blocks are indexed by $\boldsymbol{i} = (i_1, i_2, i_3) \in \mathbb{N}^3$ with $i_j \leq q_j$ for $q_j := \lfloor \frac{M - b_j}{e_j} \rfloor + 1$, $j = 1, 2$ and $q_3 := \lfloor \frac{T - b_j}{e_j} \rfloor + 1$. This results in a total number of $q = q_1 q_2 q_3$ blocks, which we summarise in the set

$$\begin{aligned} E_{\boldsymbol{i},\boldsymbol{b},\boldsymbol{e}} \quad = \quad & \{(s_1, s_2, t) \in \mathcal{S}_M \times \mathcal{T}_T : (i_j - 1)e_j + 1 \leq s_j \leq (i_j - 1)e_j + b_j, j = 1, 2, \\ & (i_3 - 1)e_3 + 1 \leq t \leq (i_3 - 1)e_3 + b_3\}. \end{aligned}$$

Now we estimate θ_C and θ_α based on all observations in a block, hence getting q different estimates, which we denote by $\widehat{\theta}_{C,\boldsymbol{b},\boldsymbol{i}}$ and $\widehat{\theta}_{\alpha,\boldsymbol{b},\boldsymbol{i}}$.

In order to find rejection areas for the isotropy test, we will use Lemma 3, and take care of the unknown variance in the normal limit by a subsampling result.

Theorem 5. *Denote by $\tau_{M,T} := M\sqrt{T}$ and $\tau_{\boldsymbol{b}} = \sqrt{b_1 b_2 b_3}$ the square roots of the number of observations in total and in each block, respectively. Assume that the conditions of Theorem 2 hold and, as $M, T \to \infty$,*

(i) $b_i \to \infty$ for $i = 1, 2, 3$, such that $b_i = o(M)$ for $i = 1, 2$, and $b_3 = o(T)$ (hence, $\tau_{\boldsymbol{b}}/\tau_{M,T} \to 0$),

(ii) $\boldsymbol{e}$ does not depend on M or T.

In the following $\widehat{\theta}$ stands for either $\widehat{\theta}_C$ or $\widehat{\theta}_\alpha$. Define the empirical distribution function

$$L_{\boldsymbol{b},\widehat{\theta}}(x) := \frac{1}{q} \sum_{i_1=1}^{q_1} \sum_{i_2=1}^{q_2} \sum_{i_3=1}^{q_3} \mathbf{1}_{\{\tau_{\boldsymbol{b}}|\widehat{\theta}_{\boldsymbol{b},\boldsymbol{i}} - \widehat{\theta}| \leq x\}}, \quad x \in \mathbb{R}, \tag{4.7}$$

and the empirical quantile function

$$c_{\boldsymbol{b},\widehat{\theta}}(1 - \beta) := \inf\left\{x \in \mathbb{R} : L_{\boldsymbol{b},\widehat{\theta}}(x) \geq 1 - \beta\right\}, \quad \beta \in (0, 1). \tag{4.8}$$

Then the following statements hold for $M, T \to \infty$:

(1) Denote by $\Phi_\sigma(\cdot)$ the distribution function of a mean 0 normal random variable Z with variance

$$\sigma^2 = \begin{cases} A_1 \tilde{\Sigma}_1 A_1^\top, & \text{in case of } \widehat{\theta}_C, \\ A_2 \tilde{\Sigma}_1 A_2^\top, & \text{in case of } \widehat{\theta}_\alpha, \end{cases}$$

and recall that $2\Phi_\sigma(\cdot) - 1$ is the distribution function of $|Z|$. Then

$$L_{\boldsymbol{b},\widehat{\theta}}(x) \quad \overset{P}{\to} \quad 2\Phi_\sigma(x) - 1, \quad x \in \mathbb{R}.$$

(2) Set $J_{\widehat{\theta}}(x) := \mathbb{P}(\tau_{M,T}|\widehat{\theta} - \theta| \leq x)$ for $x \in \mathbb{R}$, then

$$\sup_{x \in \mathbb{R}} \left| L_{\boldsymbol{b}, \widehat{\theta}}(x) - J_{\widehat{\theta}}(x) \right| \quad \overset{P}{\to} \quad 0.$$

(3) For $\beta \in (0, 1)$,

$$\mathbb{P}\left(\tau_{M,T}|\widehat{\theta} - \theta| \leq c_{\boldsymbol{b}, \widehat{\theta}}(1 - \beta)\right) \to 1 - \beta. \tag{4.9}$$

Proof. We apply Corollary 5.3.1 of Politis et al. [29]. Their main Assumption 5.3.1; i.e., the existence of a continuous limit distribution, is satisfied by Lemma 3. Assumptions (i)-(ii) are also presumed by Politis et al. [29]. The required condition on the α-mixing coefficients is satisfied similarly as in the proof of Proposition 1 by Lemma 4 and the result holds. □

From (4.9), we find rejection areas for the test statistics $\tau_{M,T}\widehat{\theta}$ at confidence level $\beta \in (0, 1)$ as (recall that $\widehat{\theta}$ stands for either $\widehat{\theta}_C$ or $\widehat{\theta}_\alpha$)

$$\mathrm{Rej}_{\widehat{\theta}}^{(M,T)} := (-\infty, -c_{\boldsymbol{b}, \widehat{\theta}}(1 - \beta)) \cup (c_{\boldsymbol{b}, \widehat{\theta}}(1 - \beta), \infty) = [-c_{\boldsymbol{b}, \widehat{\theta}}(1 - \beta), c_{\boldsymbol{b}, \widehat{\theta}}(1 - \beta)]^c.$$

Bonferroni's inequality

$$\mathbb{P}(\text{reject } H_{0,1} \text{ or } H_{0,2}) \leq \mathbb{P}(\text{reject } H_{0,1}) + \mathbb{P}(\text{reject } H_{0,2}) \leq 2\beta,$$

applies and solves the multiple test problem.

4.2 Fixed spatial domain

First note that an analogue of Lemma 3 holds with rate $\sqrt{T}$ instead of $M\sqrt{T}$ and with the asymptotic covariance matrix $\tilde{\Sigma}_2$ as given in Theorem 4.

The subsampling statement corresponding to Theorem 5 then reads as follows.

Theorem 6. *Denote by $\tau_T := \sqrt{T}$ and $\tau_{b_3} = \sqrt{b_3}$ the square roots of the number of time points of observations in total and in each block, respectively. Assume that the conditions of Theorem 4 are satisfied and that Lemma 3 holds for $T \to \infty$ with rate $\sqrt{T}$ instead of $M\sqrt{T}$ and with the asymptotic covariance matrix $\tilde{\Sigma}_2$ as given in Theorem 4. Assume further that as $T \to \infty$,*

(i) $b_3 \to \infty$ such that $b_3 = o(T)$ (hence, $\tau_{b_3}/\tau_T \to 0$),

(ii) $\boldsymbol{e}$ does not depend on T,

(iii) $b_1, b_2 \to M$.

Let $\boldsymbol{b} = (b_1, b_2, b_3)$, $\tau_{\boldsymbol{b}} = \sqrt{b_1 b_2 b_3}$ and $\tau_{M,T} = M\sqrt{T}$. With $\tilde{\Sigma}_1$ as in Theorem 5 replaced by $M^2\tilde{\Sigma}_2$, conclusions (a), (b), and (c) of Theorem 5 remain true as T tends to infinity.

Proof. We apply Corollary 5.3.2 of Politis et al. [29]. The required temporal mixing condition is satisfied similarly as in the proof of Proposition 2 by Lemma 4. □

Remark 2. We can in practice apply the same procedure of subsampling as in Section 4.1. This is justified by the fact that $\tau_{b_3}/\tau_T \to 0$ implies that $\tau_{\boldsymbol{b}}/\tau_{M,T} \to 0$ as $T \to \infty$ under conditions (i)-(iii) of Theorem 6. In particular, the rejection area for $\tau_T\widehat{\theta}$ (where again $\widehat{\theta}$ stands for either $\widehat{\theta}_C$ or $\widehat{\theta}_\alpha$) is found as

$$\mathrm{Rej}_{\widehat{\theta}}^{(T)} := \frac{1}{M}\mathrm{Rej}_{\widehat{\theta}}^{(M,T)}.$$
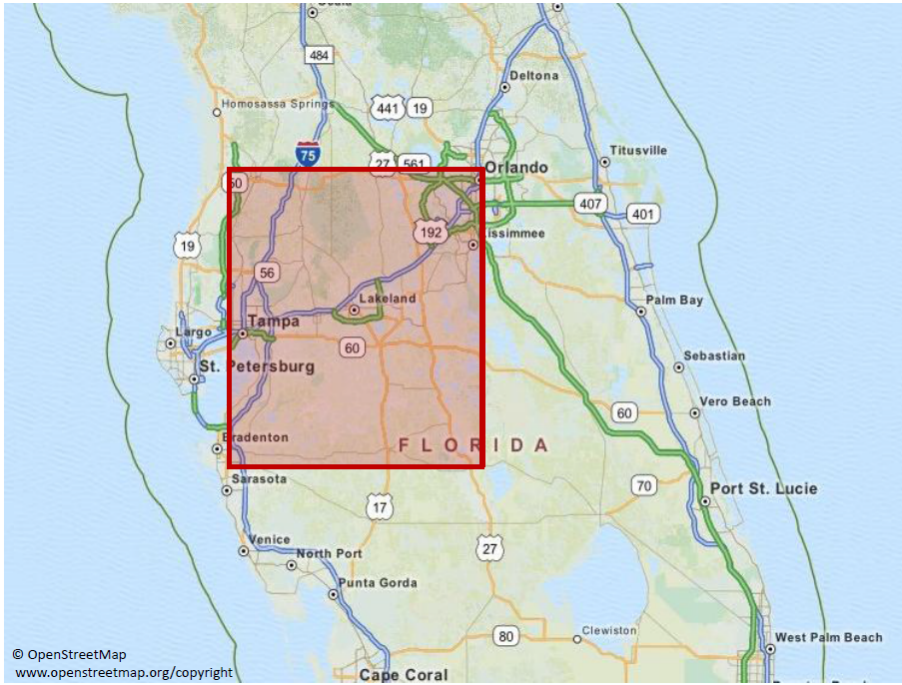
Figure 1: Rainfall observation area in Florida

5 Data analysis

We fit the Brown-Resnick space-time process (2.1) with dependence structure given by the model (2.4) to radar rainfall data, which were provided by the Southwest Florida Water Management District (SWFWMD). The data used for the analysis are rainfall measurements on a square of 120km×120km in Florida (see Figure 1) over the years 1999-2004. The raw data consist of measurements in inches on a regular grid in space every two kilometres and every 15 minutes. Since there exist wet seasons and dry seasons with almost no rain we consider only the wet season June-September. Moreover, the area is basically flat with predominant easterly winds due to its closeness to the equator and, therefore, existing trade winds. Hence, (2.4) with parameters that possibly differ along both spatial axes fits well without introducing a rotation matrix.

5.1 Data transformation and marginal modelling

We carry out a block-maxima method in space and time as follows: We calculate cumulated hourly rainfall by adding up four consecutive measurements. Then we take block-maxima over 24 consecutive hours and over 10km×10km areas; i.e., the daily maxima over 25 locations, resulting in a 12×12 grid in space for all 6×122 days of the wet seasons giving a time series of dimension 12×12 and of length 732. Taking smaller areas than 10km×10km squares or a higher temporal resolution (e.g. 12-hour-maxima) results in observations that are not max-stable and the max-stability test described in Section 5.2 would reject.

By removing possible seasonal effects, we transform the data to stationarity. We obtain the observations

$$\{\tilde{\eta}((s_1, s_2), t) : s_1, s_2 = 1, \ldots, 12, t = 1, \ldots, 732\}. \tag{5.1}$$

19

Taking daily maxima removes for every location most of the dependence in the time series. This implies that marginal parameter estimates found by maximum likelihood estimation are consistent and asymptotically normal.

To give some details: for each fixed location (s_1, s_2), we fit a univariate *generalised extreme value distribution* (cf. Embrechts et al. [14], Definition 3.4.1) to the associated time series. The estimated shape parameters are all sufficiently close to 0 to motivate a Gumbel distribution as appropriate model. We therefore fit a Gumbel distribution $\Lambda_{\mu,\sigma}(x) = \exp\{-e^{-\frac{x-\mu}{\sigma}}\}$ with parameters $\mu = \mu(s_1, s_2) \in \mathbb{R}$ and $\sigma = \sigma(s_1, s_2) > 0$ and obtain estimates $\widehat{\mu} = \widehat{\mu}(s_1, s_2)$ and $\widehat{\sigma} = \widehat{\sigma}(s_1, s_2)$.

Depending on different statistical questions and methods, we transform (5.1) either to standard Gumbel or standard Fréchet margins. In the first case we set

$$\eta_1((s_1, s_2), t) := \frac{\tilde{\eta}((s_1, s_2), t) - \widehat{\mu}}{\widehat{\sigma}}, \quad t = 1, \ldots, 732, \tag{5.2}$$

and in the latter case, with $\Lambda_{\widehat{\mu},\widehat{\sigma}}$ denoting the Gumbel distribution with estimated parameters,

$$\eta_2((s_1, s_2), t) := -\frac{1}{\log\left\{\Lambda_{\widehat{\mu},\widehat{\sigma}}(\tilde{\eta}((s_1, s_2), t))\right\}}, \quad t = 1, \ldots, 732. \tag{5.3}$$

We assess the goodness of the marginal fits by qq-plots of the observations (5.2) versus the standard Gumbel quantiles for every spatial location. Figure 2 depicts the qq-plots at four exemplary spatial locations $(1, 1)$, $(6, 8)$, $(9, 4)$ and $(11, 10)$. [1] Confidence bounds are based on the Kolmogorov-Smirnov statistic (cf. Doksum and Sievers [11], Theorem 1 and Remark 1). All graphs show a reasonably good fit.

In the following data analysis we regard (5.3) as realisations of the space-time Brown-Resnick process (2.1) with dependence structure δ as in (2.4):

$$\delta(h_1, h_2, u) = C_1|h_1|^{\alpha_1} + C_2|h_2|^{\alpha_2} + C_3|u|^{\alpha_3}, \tag{5.4}$$

with $h_1 = s_1^{(1)} - s_1^{(2)}$, $h_2 = s_2^{(1)} - s_2^{(2)}$, $u = t^{(1)} - t^{(2)}$, for two spatial locations $\boldsymbol{s}^{(1)} = (s_1^{(1)}, s_2^{(1)})$ and $\boldsymbol{s}^{(2)} = (s_1^{(2)}, s_2^{(2)})$ and two time points $t^{(1)}$ and $t^{(2)}$.

5.2 Testing for max-stability in the data

We first want to check if the block-maxima data originate from a max-stable process. A diagnostic tool is based on a multivariate Gumbel model (cf. Gabda et al. [16]), and we explain first the method in general. We assume a space-time model of a general spatial dimension $d \in \mathbb{N}$. As before, we denote the regular grid of space-time observations by

$$\mathcal{S}_M \times \mathcal{T}_T = \{1, \ldots, M\}^d \times \{1, \ldots, T\}.$$

We define a hypothesis test based on the standard Gumbel transformed space-time observations (5.2) by

$$H_0 : \{\eta_1(\boldsymbol{s}, t) : (\boldsymbol{s}, t) \in \mathbb{R}^d \times [0, \infty)\} \text{ is max-stable.} \tag{5.5}$$

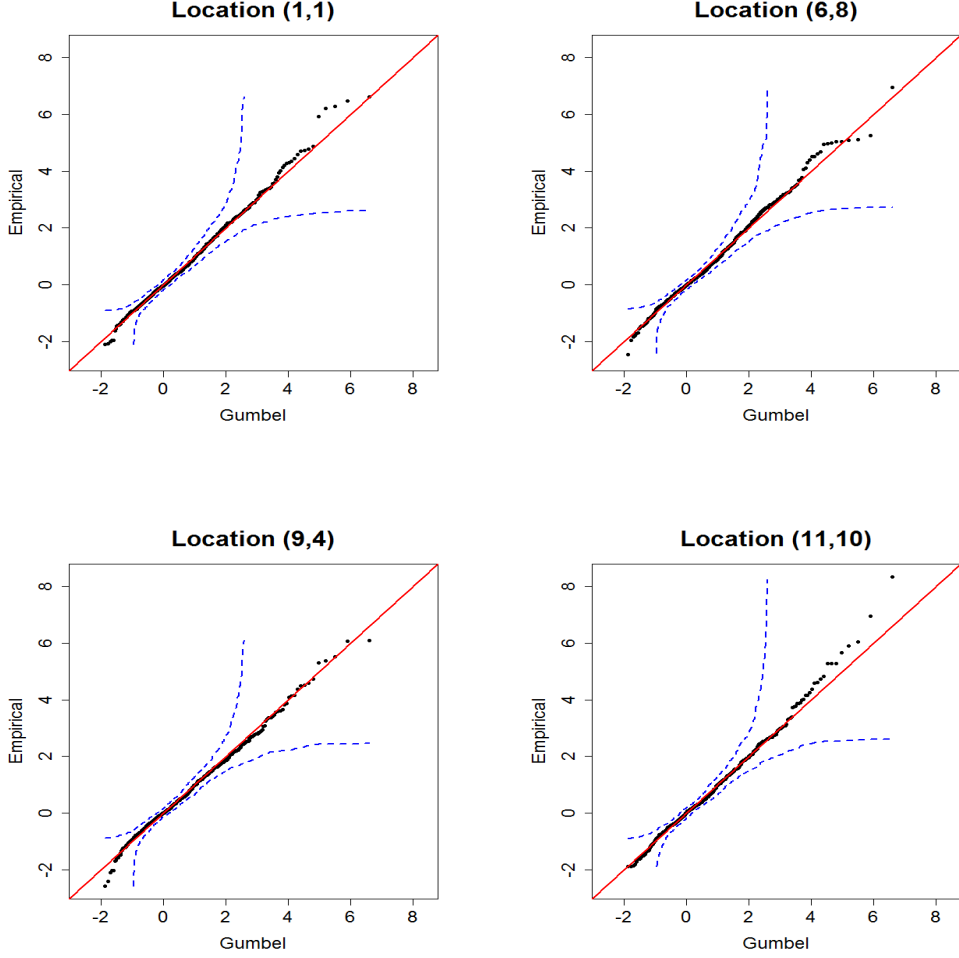[1] We use the `R`-package `extRemes` (Gilleland and Katz [17]).

Figure 2: qq-plots of the Gumbel transformed time series values versus the standard Gumbel distribution for four locations: (1,1) (top left), (6,8) (top right), (9,4) (bottom left) and (11,10) (bottom right). Dashed blue lines mark 95% confidence bounds. Solid red lines correspond to no deviation.

Under H_0 all finite-dimensional margins are max-stable; particularly, for every $D \subseteq \mathcal{S}_M \times \mathcal{T}_T$, the multivariate distribution function of $\{\eta_1(\boldsymbol{s}, t) : (\boldsymbol{s}, t) \in D\}$ is given by

$$G_D(y_1, \ldots, y_{|D|}) = \exp\{-V_D(\mathrm{e}^{y_1}, \ldots, \mathrm{e}^{y_{|D|}})\}, \quad (y_1, \ldots, y_{|D|}) \in \mathbb{R}^{|D|},$$

where V_D is the exponent measure from (2.3). Since V_D is homogeneous of order -1, the random variable

$$\eta_D := \max\{\eta_1(\boldsymbol{s}, t) : (\boldsymbol{s}, t) \in D\}$$

has univariate Gumbel distribution function

$$\mathbb{P}(\eta_D \leq y) = G_D(y, \ldots, y) = \exp\{-\mathrm{e}^{-y} V_D(1, \ldots, 1)\} = \mathrm{e}^{-\mathrm{e}^{-(y - \mu_D)}}, \quad y \in \mathbb{R}; \qquad (5.6)$$

i.e., $\mu_D := \log V_D(1, \ldots, 1)$ is the location parameter and, since $1 \leq V_D(1, \ldots, 1) \leq |D|$, we have $0 \leq \mu_D \leq \log|D|$. These considerations can be used to construct a graphical test for max-stability: First, choose different subsets D with the same fixed cardinality. Then extract several independent
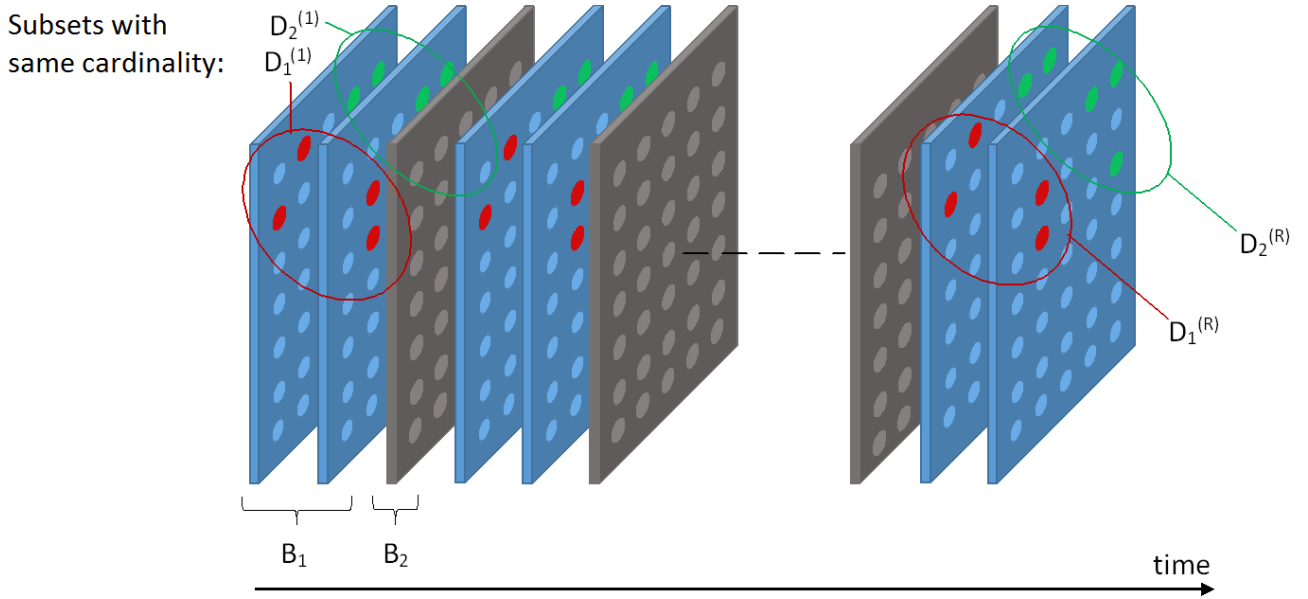
Figure 3: R independent realisations of η_D for different subsets D of the space-time observation area.

realisations of the random variables η_D from the data and test by means of a qq-plot, if they follow a Gumbel distribution.

We apply this test to the standardized Gumbel transformed data (5.2). As indicated above, taking daily maxima removes for every location most of the dependence in the time series. For this test we want to take every precaution to make sure that we work indeed with independent data. Preliminary tests show that spatial observations, which are a small number of B_2 days apart (to be specified below), show only very little time-dependence.

Consequently, we define time blocks of size B_1 of spatial observations, which are in turn separated by time blocks of size B_2 as

$$\mathcal{S}_M \times \mathcal{T}^{(i)} = \{1, \ldots, M\}^2 \times \{(i-1)(B_1 + B_2) + t : \quad t = 1, \ldots, B_1\}, \tag{5.7}$$

for $i = 1, \ldots, R = \lfloor \frac{T}{B_1 + B_2} \rfloor$. The numbers B_1 and B_2 need to be chosen in such a way that the blocks can be considered as independent. This results in R independent time blocks of length B_1 of spatial data and thus in R independent realisations of η_D for every $D \subseteq \mathcal{S}_M \times \{1, \ldots, B_1\}$. The procedure is illustrated in Figure 3.

We use these i.i.d. realisations to estimate μ_D for every D by maximum likelihood estimation restricted to $[0, \log |D|]$. Since the MLE of the location parameter of a Gumbel distribution is not unbiased (cf. Johnson et al. [22], Section 9.6), we perform a bias correction.

For the diagnostic we take $K \in \mathbb{N}$ and consider subsets D with cardinality $|D| = K$. As the total number $\binom{B_1 M^2}{K}$ of those subsets is in most cases intractably large, we randomly choose $m := \min\{R, \binom{B_1 M^2}{K}\}$ subsets and obtain in total $N = m \cdot R$ subsets, which we denote by $D_j^{(i)}$ for $j = 1, \ldots, m$ and $i = 1, \ldots, R$. For every $j = 1, \ldots, m$ we estimate μ_{D_j} by MLE based on the i.i.d. random variables $\eta_{D_j}^{(i)} := \eta_{D_j^{(i)}}$, $i = 1, \ldots, R$. Then we perform qq-plots of

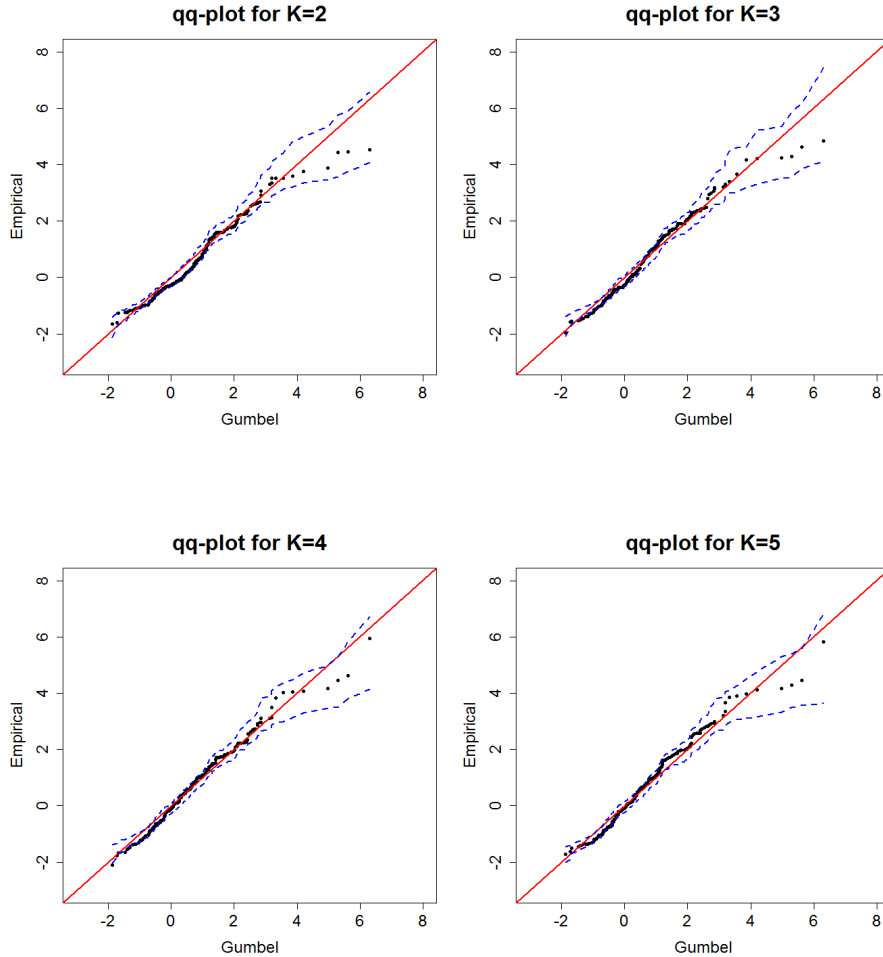$$\eta_{D_1}^{(1)} - \mu_{D_1}, \ldots, \eta_{D_m}^{(m)} - \mu_{D_m}$$

22

Figure 4: qq-plots of theoretical standard Gumbel quantiles versus the empirical quantiles (black dots). The latter correspond to the empirical distribution of maxima taken over groups of cardinality K. Dashed blue lines mark 95% pointwise confidence bounds obtained by block bootstrap. Solid red lines correspond to no deviation.

versus the standard Gumbel distribution. As a measure of variability of the estimates, non-parametric block bootstrap methods (cf. Politis and Romano [28], Section 3.2) are applied to obtain 95% pointwise confidence bounds. Using bootstrap methods, we preserve the dependence between different subsets D in the confidence intervals. Under H_0, the bisecting line should lie within these confidence bounds.

The Florida daily rainfall maxima show only little temporal dependence beyond one day. Hence we choose $B_1 = 2$ and $B_2 = 1$, which yields $R = \lfloor \frac{732}{3} \rfloor = 244$ mutually independent time blocks of spatial data. We perform the described procedure for $K = 2, 3, 4, 5$, which entails $m = R = 244$. Thus we obtain a total number of $N = 244^2 = 59\,536$ subsets. The power of this diagnostic test increases with K (cf. Gabda et al. [16]) as it gets less likely to include sets of space-time points that are K-wise independent. Figure 4 shows the results for the different choices of K. The solid red bisecting lines lie inside the confidence bounds. Hence, there is no statistically significant evidence of the space-time process generating the data not to be max-stable.

5.3 Pairwise maximum likelihood estimation

We apply the pairwise maximum likelihood estimation to the standard Fréchet transformed data (5.3). The parameters to estimate are those of the function δ in (5.4); i.e., $C_1, C_2, C_3 \in (0, \infty)$ and $\alpha_1, \alpha_2, \alpha_3 \in (0, 2]$.

In the definition of the pairwise log-likelihood function (3.4), the maximum spatial and temporal lags are specified by the numbers r_1, r_2 and p, respectively. Immediately by model (5.4) for δ, the parameters of the three different dimensions (space and time) are separated in the extremal setting. This has also been noticed in Davis et al. [8], where a simulation study in Section 7 for the isotropic model shows that estimating the spatial and temporal parameter pairs individually leads to very good results in terms of root-mean-square error and mean absolute error. Hence, for example for parameter estimates for C_1 and α_1, we can set the maximum lags corresponding to the remaining parameters equal to 0 (i.e., we set $r_2 = p = 0$). This means that we basically fit univariate models to the respective spatial and temporal parts of the dependence function (5.4). Hence, this separation simplifies the statistical estimation. However, proving asymptotic properties of the pairwise likelihood estimator in the special case of a univariate model would for instance still involve showing the required mixing conditions and thus not remove much of the complexity.

Furthermore, we know that we should not include too many lags in space or time into the likelihood, since independence effects can introduce a bias in the estimates, see for example Nott and Rydén [25], Section 2.1, or Huser and Davison [19], Section 4. On the other hand, an empirical analysis showed that extremal spatial dependence of the Florida daily rainfall maxima ranges up to lag 4 and extremal temporal dependence does not last more than one or two days, cf. Figure 7.2.6 in Steinkohl [30]. Hence, we perform the PMLE for maximum spatial and temporal lags up to 4 and 2, respectively, thus also assuring identifiability of all parameters according to Table 1. The results are summarised in Table 2. Setting r_1, r_2 or p equal to 1 results in non-identifiability of the corresponding parameters α_1, α_2 or α_3, respectively; cf. Table 1. Therefore, they are not shown in Table 2.

The combination of a rather large estimate for $\widehat{C}_3$ and a rather small estimate for $\widehat{\alpha}_3$ indicates that there is only little extremal temporal dependence, see Steinkohl [30], Section 7.2. Asymptotic 95%-confidence intervals are based on asymptotic normality of the parameter estimates and estimated using subsampling methods (cf. Section 4).

5.4 Isotropic versus anisotropic model

Using the results of Section 4, we want to apply the test (4.1) for spatial isotropy to the hypothesis

$$H_0 : \{C_1 = C_2 \text{ and } \alpha_1 = \alpha_2\} \quad \text{versus} \quad H_1 : \{C_1 \neq C_2 \text{ or } \alpha_1 \neq \alpha_2\}.$$

For the block maxima of the precipitation data we have $d = 2$, $M = 12$ and $T = 732$. This corresponds to the situation of a fixed spatial domain with $\tau_T = \sqrt{732}$.

We use the spatial PMLEs based on maximum lags 2-4, which can be read off from Table 2. We obtain the rejection areas from Theorem 6. We choose $b_1 = b_2 = 5$, thus ensuring that the full range of spatial dependence is contained in the subsamples and simultaneously achieving that their number is large. Concerning the number of time points in each subsample, we take $b_3 = 600$. Here we choose a large number to ensure that Theorem 6, where $T \to \infty$, is applicable. This results in $\tau_{b_3} = \sqrt{b_3} = \sqrt{600}$.

max. lags	$\widehat{C}_i$	$\widehat{\alpha}_i$
(2,0,0)	0.6287	0.9437
	$[0.5928, 0.6646]$	$[0.9065, 0.9808]$
(3,0,0)	0.6358	0.8599
	$[0.5989, 0.6728]$	$[0.8189, 0.9009]$
(4,0,0)	0.6438	0.8107
	$[0.6051, 0.6825]$	$[0.7690, 0.8525]$
(0,2,0)	0.7271	0.9517
	$[0.6492, 0.8050]$	$[0.8715, 1.0320]$
(0,3,0)	0.7370	0.8521
	$[0.6586, 0.8154]$	$[0.7737, 0.9305]$
(0,4,0)	0.7476	0.7931
	$[0.6677, 0.8275]$	$[0.7039, 0.8822]$
(0,0,2)	4.8378	0.1981
	$[4.4282, 5.2474]$	$[0.0177, 0.3784]$

Table 2: Estimates of the parameter pairs (C_1, α_1), (C_2, α_2) and (C_3, α_3) for different maximum spatial and temporal lags. Intervals below the point estimates are asymptotic 95%-confidence bounds based on subsampling.

In order to obtain a large number of subsamples, we further choose $e_1 = e_2 = e_3 = 1$ as the degree of overlap.

max. lag	τ_T	$\widehat{C}_2 - \widehat{C}_1$	$\tau_T(\widehat{C}_2 - \widehat{C}_1)$	$\text{Rej}_{\widehat{\theta}_C}^{(T)}$	97.5%-CI for $C_2 - C_1$	Reject $C_1 = C_2$
2	27.055	0.098	2.651	$[-2.400, 2.400]^c$	$[0.010, 0.187]$	yes
3	27.055	0.101	2.738	$[-2.392, 2.392]^c$	$[0.013, 0.190]$	yes
4	27.055	0.104	2.808	$[-2.393, 2.393]^c$	$[0.015, 0.192]$	yes

Table 3: Test results for parameters C_1 and C_2. All values are rounded to three positions after decimal point.

max. lag	τ_T	$\widehat{\alpha}_2 - \widehat{\alpha}_1$	$\tau_T(\widehat{\alpha}_2 - \widehat{\alpha}_1)$	$\text{Rej}_{\widehat{\theta}_\alpha}^{(T)}$	97.5%-CI for $(\alpha_2 - \alpha_1)$	Reject $\alpha_1 = \alpha_2$
2	27.055	0.008	0.216	$[-2.162, 2.162]^c$	$[-0.072, 0.088]$	no
3	27.055	-0.008	-0.216	$[-2.130, 2.130]^c$	$[-0.087, 0.071]$	no
4	27.055	-0.018	-0.477	$[-2.342, 2.342]^c$	$[-0.104, 0.069]$	no

Table 4: Test results for parameters α_1 and α_2. All values are rounded to three positions after decimal point.

Tables 3 and 4 present the results of the two tests at individual confidence levels $\beta = 2.5\%$ giving a test for (4.1) at a confidence level $2\beta = 5\%$ by Bonferroni's inequality. The differences $(\widehat{C}_2 - \widehat{C}_1)$ and $(\widehat{\alpha}_2 - \widehat{\alpha}_1)$ can be obtained from Table 2.

Since we can reject the individual hypothesis that $C_1 = C_2$ at a confidence level of 2.5%, we can reject the overall hypothesis H_0 of (4.1) at a confidence level of 5% and conclude that our data

originate from a spatially anisotropic max-stable Brown-Resnick process. Further note the interesting fact that, although the asymptotic confidence interval for the difference $C_2 - C_1$ does not include 0, the individual intervals for C_1 and C_2 overlap, see Table 2. This is due to the fact that the individual confidence bounds are estimated independently of each other, whereas the estimated bounds for the difference reflect how far the parameter estimates lie apart in one fixed particular (sub)sample.

5.5 Model check

Finally, having fitted the Brown-Resnick space-time model (2.1) to the precipitation data, we want to assess the quality of the fit. We take inspiration from Section 5.2 of Davison et al. [9] and compare maxima taken over subsets of the space-time precipitation data with simulated counterparts.

Similarly as in Section 5.2, we consider subsets of the observations on a regular grid for L spatial locations and for time points $1, \ldots, B_1$,

$$D = \{(s_1^{(\ell)}, s_2^{(\ell)}, 1), \ldots, (s_1^{(\ell)}, s_2^{(\ell)}, B_1) : \ell = 1, \ldots, L\}.$$

We follow the procedure as in (5.7) to extract R independent realisations of $\{\eta_1(\boldsymbol{s}, t) : (\boldsymbol{s}, t) \in D\}$ from the standard Gumbel transformed space-time observations (5.2). This yields in turn R independent realisations of $\eta_D = \max\{\eta_1(\boldsymbol{s}, t) : (\boldsymbol{s}, t) \in D\}$, which we summarise in the ordered vector $\eta_{\text{data}} := (\eta_D^{(1)}, \ldots, \eta_D^{(R)})$. Now we simulate a corresponding vector, denoted by $\widehat{\eta}_{\text{sim}} := (\widehat{\eta}_D^{(1)}, \ldots, \widehat{\eta}_D^{(R)})$. To this end we need reliable Monte Carlo values as elements of $\widehat{\eta}_{\text{sim}}$. We obtain them by simulating empirical order statistics as follows. We simulate $m \cdot R$ independent copies of the Brown-Resnick space-time process on D with dependence structure δ as in (2.4) with the PMLEs from Table 2, where we take the estimates based on maximum lag 4 (for the spatial parameters) and 2 (for the temporal parameters), which are the maximum lags, where dependence is still present. We transform the univariate margins to standard Gumbel. This results in corresponding $m \cdot R$ independent simulations of η_D and we consider them as m blocks of size R. We order the R values in each block and define $\widehat{\eta}_D^{(i)}$ as the mean of all simulated ith order statistics for $i = 1, \ldots, R$, which gives $\widehat{\eta}_{\text{sim}} := (\widehat{\eta}_D^{(1)}, \ldots, \widehat{\eta}_D^{(R)})$.

The vectors η_{data} and $\widehat{\eta}_{\text{sim}}$ are compared by qq-plots. If the fit is good, the points in the plots lie approximately on the bisecting line. Pointwise 95%-confidence bands are determined by the 2.5% and the 97.5% quantiles of the simulated order statistics. As in Section 5.3, we choose $B_1 = 2$. The number of simulations is $N = m \cdot R = 100 \cdot 244 = 24400$. Figure 5 presents the results for four exemplary groups of locations. The plots reveal a good model fit.

We carried out the simulations using the exact method recently suggested in Dombry et al. [13], Sections 3.3 and 5.2. For an overview and comparison of different simulation methods for Brown-Resnick processes we refer to Leber [24].

5.6 Application: conditional probability fields

Based on the fitted model, we want to answer questions like: Given there is extreme rain at some space-time reference point $(s_1^\star, s_2^\star, t^\star) \in \{1, \ldots, 12\}^2 \times \{1, \ldots, 732\}$, what is the estimated probability of extreme rain at some prediction space-time point (s_1^p, s_2^p, t^p)? In other words, we want to estimate the probabilities

$$\mathbb{P}\left(\tilde{\eta}((s_1^p, s_2^p), t^p) > z \mid \tilde{\eta}((s_1^\star, s_2^\star), t^\star) > z^\star\right), \tag{5.8}$$

GOF qq-plot for L=2 **GOF qq-plot for L=3**

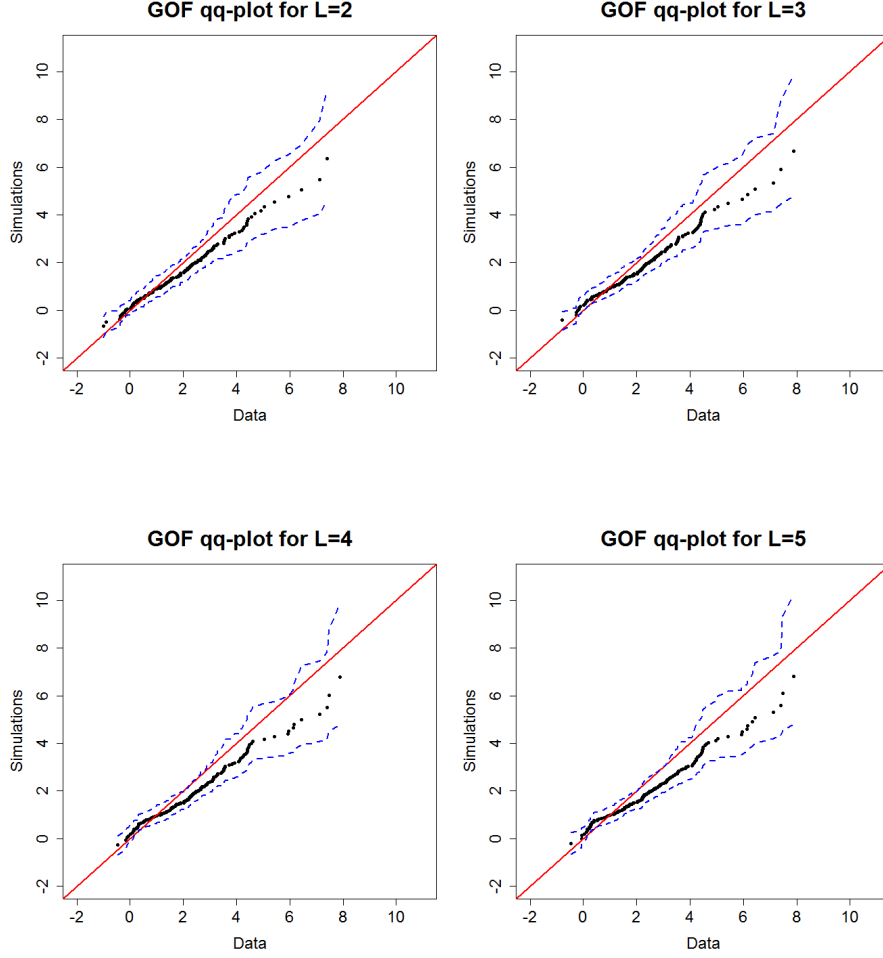**GOF qq-plot for L=4** **GOF qq-plot for L=5**

Figure 5: Goodness of fit qq-plots for different spatial locations and different L. Top left: $L = 2$: (1,1) and (1,2). Top right: $L = 3$: (1,1), (1,2) and (3,1). Bottom left: $L = 4$: (1,1), (1,2), (3,1) and (3,2). Bottom right: $L = 5$: (1,1), (1,2), (3,1), (3,2) and (2,1). PMLEs underlying the simulations are based on maximum spatial and temporal lags 4 and 2, respectively. Dashed blue lines mark 95% pointwise confidence bounds. Solid red lines correspond to no deviation.

where $\{\tilde{\eta}((s_1, s_2), t) : s_1, s_2 = 1, \ldots, 12, t = 1, \ldots, 732\}$ are the stationary observations (5.1) and z and $z^\star$ are prediction and reference rainfall levels, respectively. Denote by $\Lambda_{\mu,\sigma}$ the Gumbel distribution with location and scale parameters μ and σ (cf. Section 5.1) and set $\widehat{\mu}^p := \widehat{\mu}(s_1^p, s_2^p)$, $\widehat{\sigma}^p := \widehat{\sigma}(s_1^p, s_2^p)$, $\widehat{\mu}^\star := \widehat{\mu}(s_1^\star, s_2^\star)$ and $\widehat{\sigma}^\star := \widehat{\sigma}(s_1^\star, s_2^\star)$, which are the marginal Gumbel parameter estimates. Simple computations show that (5.8) can be estimated by

$$
\frac{1}{1 - \Lambda_{\widehat{\mu}^\star, \widehat{\sigma}^\star}(z^\star)} \Big(1 - \Lambda_{\widehat{\mu}^\star, \widehat{\sigma}^\star}(z^\star) - \Lambda_{\widehat{\mu}^p, \widehat{\sigma}^p}(z)
$$
$$
+ \exp \Big\{ - \widehat{V}_D \Big(- \frac{1}{\log\{\Lambda_{\widehat{\mu}^p, \widehat{\sigma}^p}(z)\}}, - \frac{1}{\log\{\Lambda_{\widehat{\mu}^\star, \widehat{\sigma}^\star}(z^\star)\}} \Big) \Big\} \Big),
$$

where $\widehat{V}_D$ is the estimate of the exponent measure (3.2) obtained by plugging in the PMLEs of the parameters of the dependence function δ. Figure 6 shows four predicted conditional probability fields for the reference points $(1, 1, 1)$, $(5, 6, 1)$, $(8, 10, 1)$ and $(10, 7, 1)$ and for high empirical rainfall levels
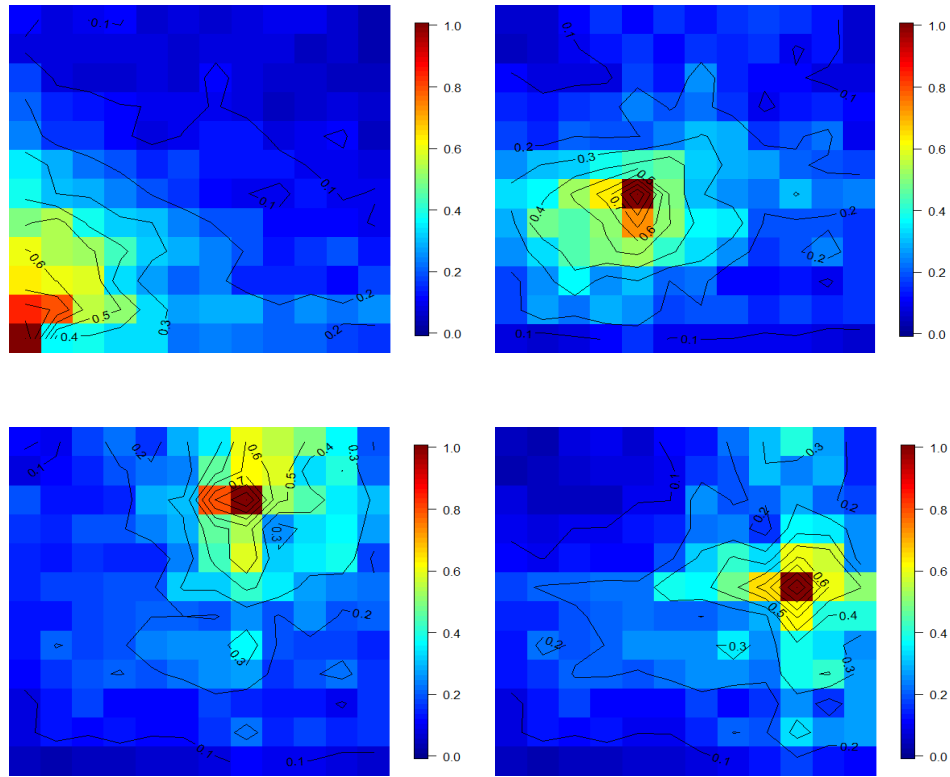
Figure 6: Predicted conditional probability fields based on daily maxima for reference space-time points (1,1,1), (5,6,1), (8,10,1) and (10,7,1) and rainfall levels $z = z^\star = 2.5$ (clockwise from the top left to the bottom right).

$z = z^\star = 2.5$. Because of the little temporal dependence in the daily maxima, we only consider equal time points for spatial predictions.

Acknowledgements

References

[1] J. Beirlant, Y. Goegebeur, J. Segers, and J. Teugels. *Statistics of Extremes, Theory and Applications.* Wiley, Chichester, 2004.

[2] J. Blanchet and A. Davison. Spatial modeling of extreme snow depth. *Ann. Appl. Stat.*, 5(3): 1699–1724, 2011.

[3] E. Bolthausen. On the central limit theorem for stationary mixing random fields. *Ann. Probab.*, 10(4):1047–1050, 1982.

[4] R. Bradley. *Introduction to Strong Mixing Conditions*, volume I. Kendrick Press, Heber City, Utah, 2007.

[5] B. Brown and S. Resnick. Extreme values of independent stochastic processes. *J. Appl. Probab.*, 14(4):732–739, 1977.

[6] S. Buhl. *Modelling and Estimation of Extremes in Space and Time.* Master's thesis, Technische Universität München, 2013. Available under `https://mediatum.ub.tum.de/doc/1145694/1145694.pdf`.

[7] R. Davis, C. Klüppelberg, and C. Steinkohl. Max-stable processes for modeling extremes observed in space and time. *J. Korean Stat. Soc.*, 42(3):399–414, 2013.

[8] R. Davis, C. Klüppelberg, and C. Steinkohl. Statistical inference for max-stable processes in space and time. *J. Roy. Stat. Soc. B*, 75(5):791–819, 2013.

[9] A. Davison, S. Padoan, and M. Ribatet. Statistical modelling of spatial extremes. *Stat. Sci.*, 27 (2):161–186, 2012.

[10] L. de Haan. A spectral representation for max-stable processes. *Ann. Probab.*, 12(4):1194–1204, 1984.

[11] K. Doksum and G. Sievers. Plotting with confidence: graphical comparisons of two populations. *Biometrika*, 63(3):421–434, 1976.

[12] C. Dombry and F. Eyi-Minko. Strong mixing properties of max-infinitely divisible random fields. *Stoch. Proc. Appl.*, 122(11):3790–3811, 2012.

[13] C. Dombry, S. Engelke, and M. Oesting. Exact simulation of max-stable processes. *Biometrika*, 103, 2016. To appear.

[14] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events.* Springer, Berlin, 1997.

[15] S. Engelke, A. Malinowski, Z. Kabluchko, and M. Schlather. Estimation of Hüsler-Reiss distributions and Brown-Resnick processes. *J. Roy. Stat. Soc. B*, 77(1):239–265, 2015.

[16] D. Gabda, R. Towe, J. Wadsworth, and J. Tawn. Discussion of "Statistical Modelling of Spatial Extremes" by A.C. Davison, S.A. Padoan and M.Ribatet. *Stat. Sci.*, 27(2):189–192, 2012.

[17] E. Gilleland and R. W. Katz. New software to analyze how extremes change over time. *Eos*, 92 (2):13–14, 2011.

[18] E. Giné, M. G. Hahn, and P. Vatan. Max-infinitely divisible and max-stable sample continuous processes. *Probab. Theory Rel.*, 87:139–165, 1990.

[19] R. Huser and A. Davison. Space-time modelling of extreme events. *J. Roy. Stat. Soc. B*, 76(2): 439–461, 2014.

[20] J. Hüsler and R.-D. Reiss. Maxima of normal random vectors: between independence and complete dependence. *Stat. Probabil. Lett.*, 7(4):283–286, 1989.

[21] I. Ibragimov and Y. Linnik. *Independent and Stationary Sequences of Random Variables*. Groningen: Wolters-Noordhoff, 1971.

[22] N. Johnson, S. Kotz, and N. Balakrishnan. *Continuous Univariate Distributions - Volume 2*. Wiley, New York, 2nd edition, 1995.

[23] Z. Kabluchko, M. Schlather, and L. de Haan. Stationary max-stable fields associated to negative definite functions. *Ann. Probab.*, 37(5):2042–2065, 2009.

[24] D. Leber. *Comparison of Simulation Methods of Brown-Resnick processes*. Master's thesis, Technische Universität München, 2015. Available under `https://mediatum.ub.tum.de/doc/1286600/1286600.pdf`.

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

[26] M. Oesting. *Simulationsverfahren für Brown-Resnick-Prozesse*. Diploma thesis, Georg-August-Universität zu Göttingen, 2009. Available under `http://arxiv.org/abs/0911.4389v2`.

[27] S. Padoan, M. Ribatet, and S. Sisson. Likelihood-based inference for max-stable processes. *J. Am. Stat. Assoc.*, 105(489):263–277, 2010.

[28] D. N. Politis and J. P. Romano. Nonparametric resampling for homogeneous strong mixing random fields. *J. Multivariate Anal.*, 47:301–328, 1993.

[29] D. N. Politis, J. P. Romano, and M. Wolf. *Subsampling*. Springer, New York, 1999.

[30] C. Steinkohl. *Statistical Modelling of Extremes in Space and Time using Max-Stable Processes*. Dissertation, Technische Universität München, 2013. Available under `https://mediatum.ub.tum.de/doc/1120541/1120541.pdf`.

[31] D. Straumann. *Estimation in Conditionally Heteroscedastic Time Series Models*. Lecture Notes in Statistics, No. 182, Springer, Berlin, 2004.

[32] J. Wadsworth and J. Tawn. Efficient inference for spatial extreme value processes associated to log-gaussian random functions. *Biometrika*, 101(1):1–15, 2014.

[33] A. Wald. Note on the consistency of the maximum likelihood estimate. *Ann. Math. Stat.*, 20(4): 595–601, 1949.

A   An auxiliary lemma

Lemma 4. *The following two bounds hold true for $r \geq 1$, $\alpha \in (0,2]$ and $C > 0$:*

$$\int_y^\infty u^r \mathrm{e}^{-Cu^\alpha} \, \mathrm{d}u \sim \frac{1}{C\alpha} y^{r-\alpha+1} \mathrm{e}^{-Cy^\alpha}, \quad y \to \infty, \tag{A.1}$$

$$\int_1^\infty \left(\int_y^\infty u^r \mathrm{e}^{-Cu^\alpha} \, \mathrm{d}u \right)^{\frac{1}{3}} \mathrm{d}y < \infty. \tag{A.2}$$

Proof. First note that integrals of the form $\int_0^\infty u^r \mathrm{e}^{-Cu^\alpha} \, \mathrm{d}u$ are finite for every $r > -1$, $\alpha \in (0,2]$, and $C > 0$, since they are transformations of the gamma function $\Gamma(x) = \int_0^\infty t^{x-1} \mathrm{e}^{-t} \, \mathrm{d}t$, which exists for positive x. We prove (A.1) by an application of l'Hôpital's rule:

$$\lim_{y \to \infty} \frac{\int_y^\infty u^r \mathrm{e}^{-Cu^\alpha} \, \mathrm{d}u}{\frac{1}{C\alpha} y^{r-\alpha+1} \mathrm{e}^{-Cy^\alpha}} = \lim_{y \to \infty} \frac{-y^r \mathrm{e}^{-Cy^\alpha}}{\left(-y^r + \frac{r-\alpha+1}{C\alpha} y^{r-\alpha} \right) \mathrm{e}^{-Cy^\alpha}} = \lim_{y \to \infty} \frac{y^r}{y^r \left(1 - \frac{r-\alpha+1}{C\alpha} y^{-\alpha} \right)} = 1.$$

In order to prove (A.2) first note that it follows from (A.1) that for every $\epsilon > 0$ there exists $y_0 = y_0(\epsilon)$ such that for all $y \geq y_0$,

$$\left(\int_y^\infty u^r \mathrm{e}^{-Cu^\alpha} \, \mathrm{d}u \right)^{\frac{1}{3}} \leq (1+\epsilon) \left(\frac{1}{C\alpha} \right)^{\frac{1}{3}} y^{\frac{r-\alpha+1}{3}} \mathrm{e}^{-\frac{C}{3} y^\alpha}. \tag{A.3}$$

Now we split the double integral of (A.2) up into

$$\int_1^{y_0} \left(\int_y^\infty u^r \mathrm{e}^{-Cu^\alpha} \, \mathrm{d}u \right)^{\frac{1}{3}} \mathrm{d}y + \int_{y_0}^\infty \left(\int_y^\infty u^r \mathrm{e}^{-Cu^\alpha} \, \mathrm{d}u \right)^{\frac{1}{3}} \mathrm{d}y =: I_1 + I_2.$$

For I_1 we obtain

$$I_1 \leq \int_1^{y_0} \left(\int_y^{y_0} u^r \mathrm{e}^{-Cu^\alpha} \, \mathrm{d}u \right)^{\frac{1}{3}} \mathrm{d}y + \int_1^{y_0} \left(\int_{y_0}^\infty u^r \mathrm{e}^{-Cu^\alpha} \, \mathrm{d}u \right)^{\frac{1}{3}} \mathrm{d}y =: I_1^{(1)} + I_1^{(2)}.$$

$I_1^{(1)}$ is obviously finite, and to bound $I_1^{(2)}$ we use (A.3), which yields

$$I_1^{(2)} \leq (y_0 - 1)(1+\epsilon) \left(\frac{1}{C\alpha} \right)^{\frac{1}{3}} y_0^{\frac{r-\alpha+1}{3}} \mathrm{e}^{-\frac{C}{3} y_0^\alpha} < \infty.$$

Concerning I_2, note that

$$I_2 \leq (1+\epsilon) \left(\frac{1}{C\alpha} \right)^{\frac{1}{3}} \int_{y_0}^\infty y^{\frac{r-\alpha+1}{3}} \mathrm{e}^{-\frac{C}{3} y^\alpha} \, \mathrm{d}y,$$

which is finite by finiteness of the gamma function. $\qquad \square$