Castro-Camilo, D. and Huser, R. (2019) Local likelihood estimation of complex tail dependence structures, applied to U.S. precipitation extremes. Journal of the American Statistical Association, (doi:10.1080/01621459.2019.1647842).

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

http://eprints.gla.ac.uk/191026/

Deposited on: 14 August 2019

Local likelihood estimation of complex tail dependence structures, applied to U.S. precipitation extremes

Daniela Castro-Camilo[1,2] and Raphaël Huser[1]

July 6, 2019

Abstract

To disentangle the complex non-stationary dependence structure of precipitation extremes over the entire contiguous U.S., we propose a flexible local approach based on factor copula models. Our sub-asymptotic spatial modeling framework yields non-trivial tail dependence structures, with a weakening dependence strength as events become more extreme; a feature commonly observed with precipitation data but not accounted for in classical asymptotic extreme-value models. To estimate the local extremal behavior, we fit the proposed model in small regional neighborhoods to high threshold exceedances, under the assumption of local stationarity, which allows us to gain in flexibility. By adopting a local censored likelihood approach, we make inference on a fine spatial grid, and we perform local estimation by taking advantage of distributed computing resources and the embarrassingly parallel nature of this estimation procedure. The local model is efficiently fitted at all grid points, and uncertainty is measured using a block bootstrap procedure. We carry out an extensive simulation study to show that our approach can adequately capture complex, non-stationary dependencies, in addition, our study of U.S. winter precipitation data reveals interesting differences in local tail structures over space, which has important implications on regional risk assessment of extreme precipitation events.

Keywords: factor copula model, local likelihood, non-stationarity, spatial extremes, threshold exceedances.

---

[1]Computer, Electrical and Mathematical Sciences and Engineering (CEMSE) Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia. E-mails: daniela.castro@kaust.edu.sa, raphael.huser@kaust.edu.sa

[2]School of Mathematics and Statistics, University of Glasgow, Glasgow G12 8QQ, UK.

1 Introduction

Water-related extremes such as floods and droughts can heavily impact human life, affecting our society, economic stability, and environmental sustainability. In recent years, we have witnessed an acceleration of the water cycle in some areas of the globe, including an increase in the frequency and intensity of heavy precipitation. This has been sadly illustrated by a number of unprecedented hurricane events that recently hit the Caribbean islands and the Southeastern United States (U.S.), including hurricane Harvey (August–September, 2017), and hurricane Florence (August–September, 2018). Climate change has motivated the development of stochastic models for risk assessment and uncertainty quantification of extreme weather events. In the univariate context, the generalized extreme-value distribution with time-varying parameters has been used to study the effect of climate change on global precipitation annual maxima (Westra *et al.*, 2013). More recently, Fischer and Knutti (2016) have adopted an empirical approach to study heavy rainfall intensification, validating the theory predicted by early generations of general circulation models. Similarly, Hoerling *et al.* (2016) analyzed trends in U.S. heavy precipitation data. Beyond the univariate context, several studies have focused on the spatial or spatio-temporal modeling of precipitation extremes within and across different catchments (Cooley *et al.*, 2007; Thibaud *et al.*, 2013; Huser and Davison, 2014; Opitz *et al.*, 2018), leading to extreme river discharges (Asadi *et al.*, 2015) or have focused on the risk of heavy snowfall in mountainous areas (Blanchet and Davison, 2011). In these small regions, however, the spatial dependence structure is often assumed to be stationary. By contrast, in this paper we focus on modeling the complex, non-stationary, spatial dependence structure of U.S. winter precipitation extremes.

Assessing the behavior of extreme events over space entails many challenges. First, data are measured at a finite number of locations, but extrapolation is typically required at any other location of the study region. In our case, the study region is the contiguous U.S. and
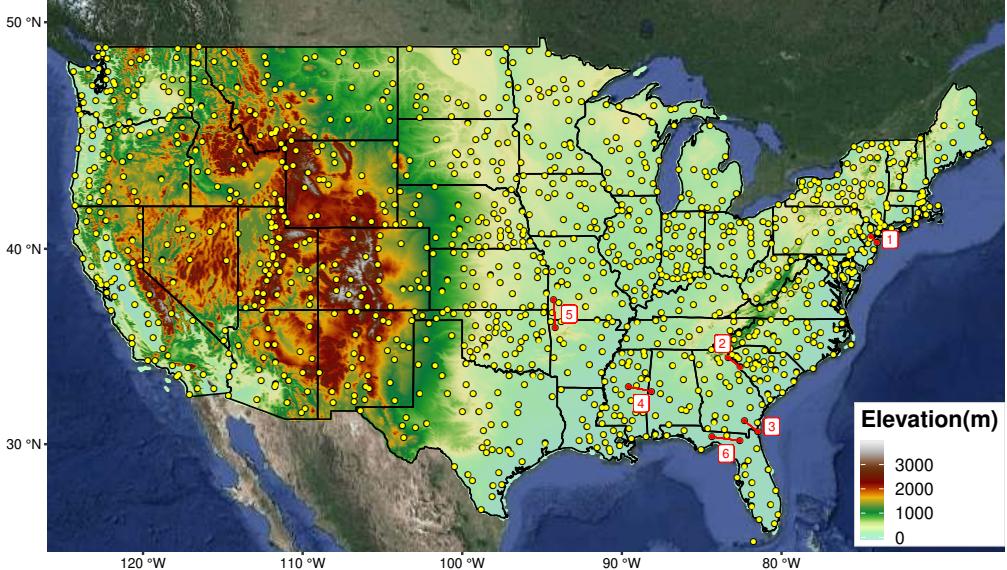
Figure 1: Topographic map of the contiguous U.S. with the 1218 weather stations (dots). Red dots connected by a red segment represent the pairs of stations chosen in Figure 2, sorted by their distances (labels).

the monitoring stations at which data are collected are displayed in Figure 1. Second, by contrast with classical geostatistics where the estimation target is usually the (conditional) mean, statistics of extremes focuses on high (or low) quantiles, and extrapolation is not only required over space, but also into the joint upper (or lower) tail of the distribution. Third, the scarcity of extremes results in uncertainty inflation, especially when very high quantiles have to be estimated, and thus it is important to develop efficient inference methods that can potentially be applied in high dimensions. Finally, a recurrent problem with large heterogeneous regions, such as the contiguous U.S., is the modeling of spatial non-stationarity. This may concern both marginal distributions and the dependence structure. As pointed out by Huser and Genton (2016), misspecification of the joint distribution of extremes may result in poor estimation of spatial risk measures. There is no universal consensus on how to handle non-stationarity; however, it is often useful and realistic to assume some weak form of local stationarity. In this work, we fit a set of locally stationary spatial models to U.S. winter precipitation extremes, which contrasts with the quite rigid fully parametric model fitted by

Huser and Genton (2016), and the non-parametric approach adopted by Castro-Camilo and de Carvalho (2017), whereby a family of bivariate extremal dependence structures at different sites are linked through the action of covariates, while neglecting spatial dependence.

The literature on spatial extremes is mostly divided into two mainstream approaches: the first approach defines extreme events as block (e.g., annual) maxima, which are typically modeled using max-stable processes (Westra and Sisson, 2011; Davison and Gholamrezaee, 2012; Huser and Genton, 2016; Davison *et al.*, 2019; Vettori *et al.*, 2019); the latter can be regarded as the functional generalization of multivariate extreme-value distributions, and they arise as limits of properly renormalized maxima of random fields. The second approach defines extreme events as high threshold exceedances, which are usually modeled using generalized Pareto processes, i.e., the threshold counterpart of max-stable processes (Ferreira *et al.*, 2014; Thibaud and Opitz, 2015; de Fondeville and Davison, 2018). Both max-stable and generalized Pareto processes are asymptotic models, in the sense that they have a limiting characterization for block maxima and threshold exceedances, respectively. However, their tail dependence structure is fairly rigid; max-stable copulas are invariant to the operation of taking componentwise maxima, while generalized Pareto copulas are invariant to thresholding at higher levels. This lack of tail flexibility has recently motivated the development of sub-asymptotic spatial models for threshold exceedances (Wadsworth and Tawn, 2012; Opitz, 2016; Huser *et al.*, 2017; Huser and Wadsworth, 2019), which, unlike asymptotic models, can capture weakening extremal dependence strength on the way to their limiting generalized Pareto process. See also Huser *et al.* (2018) and Bopp *et al.* (2019) for the case of maxima. To illustrate this, Figure 2 shows, for a range of quantiles $u \in [0.5, 0.995]$, the conditional probability

$$\chi_h(u) = \Pr\{Y_1 > F_1^{-1}(u) \mid Y_2 > F_2^{-1}(u)\}, \quad Y_j = Y(\mathbf{s}_j) \sim F_j, \quad j = 1, 2, \qquad (1.1)$$

estimated for six selected pairs of stations $\{\mathbf{s}_1, \mathbf{s}_2\} \subset \mathcal{S}$ at various distances $h = \|\mathbf{s}_1 - \mathbf{s}_2\|$,
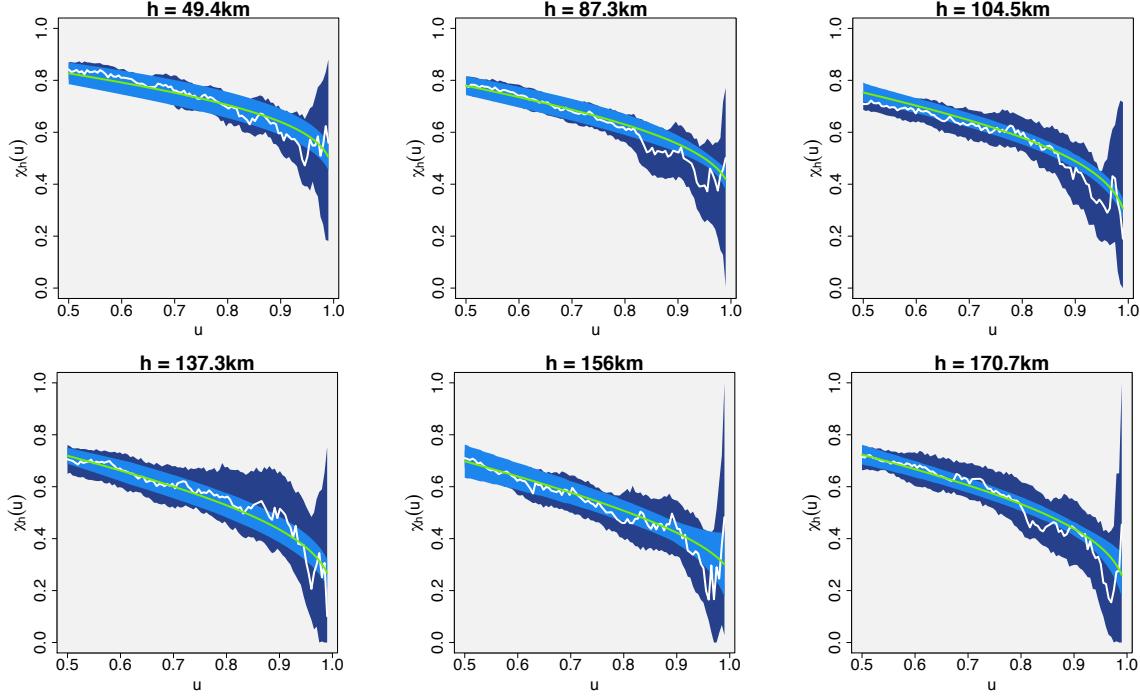
Figure 2: Empirical (white) and model-based (green) estimates for $\chi_h(u)$ as a function of the threshold $u \in [0.5, 0.995]$ for six selected pairs of stations, sorted according to their distance (labels 1–6 in Figure 1). Confidence envelopes for the empirical estimates (dark blue) are based on 300 block bootstrap samples with monthly blocks. Model-based confidence envelopes are also provided (light blue). Details on the fitted model are given in Sections 2 and 5.

where $Y(\mathbf{s})$, $\mathbf{s} \in \mathcal{S}$, denotes the 5 day-cumulative winter precipitation random field defined over the contiguous U.S. (denoted by $\mathcal{S}$), and F_j denotes the marginal distribution of $Y(\mathbf{s}_j)$ $j = 1, 2$. Whereas the function $\chi_h(u)$ in (1.1) appears to be decreasing as a function of the threshold u, a generalized Pareto process *always* assumes that $\chi_h(u) \equiv \chi_h$ is constant in u (Rootzén *et al.*, 2018) for each lag h, hence potentially leading to an overestimation of the dependence strength at higher quantiles. In Section 2, we describe a stationary factor copula model capturing non-trivial tail dependence, i.e., allowing for a positive limit $\chi_h = \lim_{u \to 1} \chi_h(u) > 0$, while at the same time having a decreasing $\chi_h(u)$ function with known limiting extreme-value dependence structure. This approach has attractive modeling features, and is computationally more convenient than the models proposed by Huser *et al.* (2017) and Huser and Wadsworth (2019). It differs fundamentally from the Laplace model of

Opitz (2016), which strongly focuses on capturing tail independence ($\chi_h = 0$) and the decay rate towards the limit, rather than tail dependence ($\chi_h > 0$), which is the type of extremal behavior suggested by our precipitation data at small spatial scales (or at least, $\chi_h(u) > 0$ remains positive for quite high thresholds u); recall Figure 2.

Inference for extreme-value models is known to be particularly cumbersome. When max-stable processes are directly or indirectly involved, the likelihood function becomes excessively prohibitive to compute in moderate to large dimensions (Castruccio *et al.*, 2016; Huser *et al.*, 2019). This has lead to the use of less efficient composite likelihood approaches (Padoan *et al.*, 2010; Westra and Sisson, 2011; Huser and Davison, 2013; Thibaud *et al.*, 2013; Huser and Davison, 2014). Alternatively, threshold approaches lead to simpler, significantly less demanding likelihoods, although their routine application to high-dimensional datasets is hampered by the need to censor observations involving non-extreme data (i.e., below a high threshold), which entails expensive multidimensional integrals; see, e.g., Wadsworth and Tawn (2014), Thibaud and Opitz (2015), Huser *et al.* (2017) and Huser and Wadsworth (2019). Other censoring schemes have been investigated by Engelke *et al.* (2015) and Opitz (2016), but they may cause some bias (Huser *et al.*, 2016). More recently, Morris *et al.* (2017) proposed to impute the censored, non-extreme, observations using a simulation-based algorithm within a Bayesian framework. In this work, although our full dataset comprises 1218 monitoring stations over the contiguous U.S., the dimensionality is considerably reduced by adopting a local estimation approach, making it possible to perform traditional censored likelihood inference in a robust and efficient way.

Our main methodological contributions can be summarized as follows: we extend the stationary exponential factor copula model by allowing the parameters to change smoothly with locations, and provide further theoretical results for it. We develop a local likelihood approach adapted to the context of extremes to fit the stationary exponential factor copula

model locally, censoring non-extreme observations. Finally, we derive simplified expression for the terms involved in the local censored likelihood function in order to compute maximum likelihood estimates in a reasonable time, while avoiding numerical instabilities.

In Section 2, we present the stationary exponential factor copula model, and its tail properties are studied. In Section 3 we discuss a censored local likelihood inference based on high threshold exceedances. In Section 4, we conduct a simulation study based on a non-stationary factor copula model with spatially-varying parameters, in order to assess the performance of our approach in various non-stationary contexts. In Section 5, we apply our proposed model to study the dependence structure of heavy precipitation over the whole contiguous U.S., and we use it to estimate return periods of spatial extreme events. Section 6 concludes with a comprehensive and critical discussion.

2 Modeling spatial extremes using factor copulas

2.1 Copula models

A copula is a multivariate probability distribution with $\text{Unif}(0, 1)$ margins. Copulas are used to describe the dependence between random variables, and may be used to link univariate marginal distributions to construct a joint distribution. Specifically, let $(X_1, \ldots, X_D)^T \in \mathbb{R}^D$ be a random vector with continuous marginals $F_j(x) = \Pr(X_j \leq x)$, $j = 1, \ldots, D$. The copula of $(X_1, \ldots, X_D)^T$ is defined through the random vector $(U_1, \ldots, U_D)^T = \{F_1(X_1), \ldots, F_D(X_D)\}^T$ (with standard uniform margins) as $C(u_1, \ldots, u_D) = \Pr(U_1 \leq u_1, \ldots, U_D \leq u_D)$. Sklar (1959) showed that each multivariate distribution $F(x_1, \ldots, x_D)$ with continuous margins $F_j(x)$ has a unique copula C, which may be expressed as

$$F(x_1, \ldots x_D) = C\{F_1(x_1), \ldots, F_D(x_D)\} \iff C(u_1, \ldots, u_D) = F\{F_1^{-1}(u_1), \ldots, F_D^{-1}(u_D)\}; \tag{2.1}$$

this implies that F can be written in terms of univariate marginal distributions chosen independently from the dependence structure between the variables. This result can be associated

with a two-step approach for inference, where margins are treated separately from the dependence structure. Based on (2.1), several copula families have been proposed and applied in practice for the modeling of environmental data, the most common one being the Gaussian copula obtained by choosing the joint distribution F to be the standard multivariate Gaussian distribution $\Phi_D(\cdot; \boldsymbol{\Sigma})$ with correlation matrix $\boldsymbol{\Sigma}$. Other more flexible elliptical copulas may be derived similarly, such as the Student-t copula, which is tail-dependent as opposed to the Gaussian copula. An alternative general family of copulas generating interesting tail dependence structures are factor copula models (Krupskii and Joe, 2015; Krupskii *et al.*, 2018), in which a random and unobserved factor affects all measurements simultaneously. In Section 2.2, we describe the construction of the stationary exponential factor copula model proposed by Krupskii *et al.* (2018), which has appealing modeling and inference properties, and we then embed it in a more general non-stationary model in Section 4.2.

2.2   The stationary exponential factor copula model

Let $Z(\mathbf{s})$, $\mathbf{s} \in \mathcal{S} \subset \mathbb{R}^2$, be a standard Gaussian process with stationary correlation function $\rho(h)$ (see Gneiting *et al.* (2006) for a review of correlation functions), and let $V \geq 0$ be an exponentially distributed random variable with rate parameter $\lambda > 0$, independent of $Z(\mathbf{s})$, and which does not depend on the spatial location $\mathbf{s} \in \mathcal{S}$. The exponential factor copula model may be expressed in continuous space through the random process

$$W(\mathbf{s}) = Z(\mathbf{s}) + V, \qquad \mathbf{s} \in \mathcal{S}. \tag{2.2}$$

The W process in (2.2) is a Gaussian *location* mixture, i.e., a standard Gaussian process with a random (exponentially distributed), constant mean. In this sense, it has similarities with the Student-t process (see e.g., Røislien and Omre, 2006), which is a Gaussian *scale* mixture (with standard deviation following a specific inverse gamma distribution). However, these models are not particular cases of each other and their dependence structures have

significant differences.

Although Model (2.2) may seem quite artificial, it is only used to generate a flexible upper tail dependence structure, which we then fit to precipitation extremes. In other words, we disregard the margins and only consider the copula associated with (2.2), which we fit to high threshold exceedances using a censored likelihood approach, reducing the contribution of small precipitation values; more details are given in Sections 3 and 5.

From (2.2), each configuration of D sites $\{\mathbf{s}_1, \ldots, \mathbf{s}_D\} \subset \mathcal{S}$ yields a D-variate copula as follows: let $Z_j = Z(\mathbf{s}_j)$ and $W_j = W(\mathbf{s}_j)$, $j = 1, \ldots, D$. The random vector $\mathbf{Z} = (Z_1, \ldots, Z_D)^T$ has a multivariate normal distribution with correlation matrix $\boldsymbol{\Sigma}_{\mathbf{Z}}$ that depends on the correlation function $\rho(h)$ and the sites' configuration, i.e., $\mathbf{Z} \sim \Phi_D(\cdot; \boldsymbol{\Sigma}_{\mathbf{Z}})$. The components of the random vector $\mathbf{W} = (W_1, \ldots, W_D)^T$ are $W_j = Z_j + V$, $j = 1, \ldots, D$, where $V \sim \mathrm{Exp}(\lambda)$ is independent of $\mathbf{Z}$; by integrating out V, the joint distribution of $\mathbf{W}$ may be expressed as

$$F_D^{\mathbf{W}}(w_1, \ldots, w_D) = \lambda \int_0^\infty \Phi_D(w_1 - v, \ldots, w_D - v; \boldsymbol{\Sigma}_{\mathbf{Z}}) \exp(-\lambda v) \mathrm{d}v, \qquad (2.3)$$

whilst its density is

$$f_D^{\mathbf{W}}(w_1, \ldots, w_D) = \lambda \int_0^\infty \phi_D(w_1 - v, \ldots, w_D - v; \boldsymbol{\Sigma}_{\mathbf{Z}}) \exp(-\lambda v) \mathrm{d}v, \qquad (2.4)$$

where $\phi_D(\cdot; \boldsymbol{\Sigma})$ is the multivariate standard normal density with correlation matrix $\boldsymbol{\Sigma}$. In Appendix A we provide simpler and computationally efficient expressions to compute (2.3), (2.4), and partial derivatives of (2.3), avoiding the integral in v. Applying (2.1), the resulting copula and its density may be written as

$$C_D^{\mathbf{W}}(u_1, \ldots, u_D) = F_D^{\mathbf{W}}(w_1, \ldots, w_D), \qquad c_D^{\mathbf{W}}(u_1, \ldots, u_D) = \frac{f_D^{\mathbf{W}}(w_1, \ldots, w_D)}{\prod_{j=1}^D f_1^{\mathbf{W}}(w_j)}, \qquad (2.5)$$

where $w_j = (F_1^{\mathbf{W}})^{-1}(u_j; \lambda)$, $j = 1, \ldots, D$, and $F_1^{\mathbf{W}}(\cdot; \lambda)$ and $f_1^{\mathbf{W}}(\cdot; \lambda)$ denote the marginal distribution and density of the W process, respectively. In particular, we can show that

$$F_1^{\mathbf{W}}(w; \lambda) = \Phi(w) - \exp(\lambda^2/2 - \lambda w)\Phi(w - \lambda), \qquad (2.6)$$

9
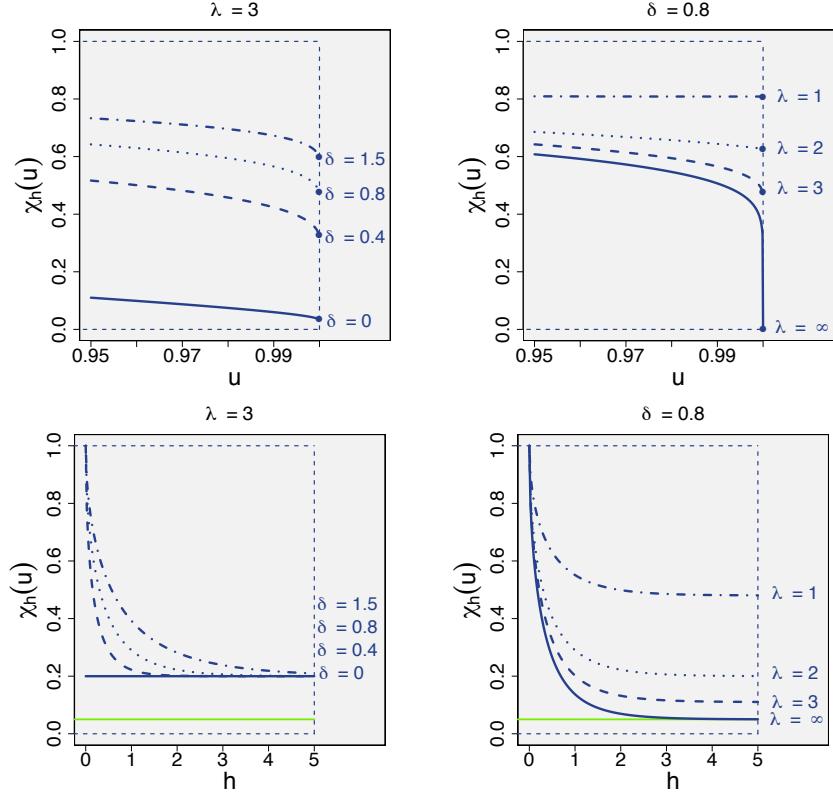
Figure 3: Conditional probability $\chi_h(u)$ as defined in (1.1) for the stationary exponential factor model with rate parameter $\lambda > 0$ and exponential correlation function $\rho(h) = \exp(-h/\delta)$, plotted with respect to the level $u \in [0.95, 1]$ at distance 0.1 (top) and with respect to the distance $h \in [0, 5]$ for fixed $u = 0.95$ (bottom). Left: fixed $\lambda = 3$ and $\delta = 0, 0.4, 0.8, 1.5$. Right: fixed $\delta = 0.8$ and $\lambda = 1, 2, 3, \infty$. The dots in the top panels correspond to the limit $\chi_h = \lim_{u \to 1} \chi_h(u)$, and the green horizontal lines in the bottom panels correspond to the complete independence case.

where $\Phi(\cdot)$ is the standard normal distribution function; see (A.3) in Appendix A.

To illustrate the tail flexibility of the stationary exponential factor copula model, Figure 3 displays the function $\chi_h(u)$ defined in (1.1) as a function of the quantile level $u \in [0.95, 1]$ and the Euclidean distance $h = \|\mathbf{s}_1 - \mathbf{s}_2\| \geq 0$, for different rate $\lambda > 0$ and range $\delta > 0$ parameters, assuming an exponential correlation function $\rho(h) = \exp(-h/\delta)$, $h \geq 0$. While the range δ controls the correlation decay with distance, the rate λ determines the overall tail dependence strength and strongly impacts the value of $\chi_h(u)$ and its limit $\chi_h = \lim_{u \to 1} \chi_h(u)$ at large distances, i.e., as $h \to \infty$. In particular, because the random factor V in (2.2) is common to all sites $\mathbf{s} \in \mathcal{S}$, spatial dependence in the W process does not vanish as $h \to \infty$. Therefore, this stationary model may be useful for (replicated) spatial data collected on a

local or regional scale, but it is generally not realistic on larger scales, such as the whole contiguous U.S. In Section 3, we develop a local estimation approach assuming that the stationary model (2.2) is only valid locally, and in Section 4.2 we discuss a flexible extension to generate non-stationary tail dependency structures. As shown by Krupskii *et al.* (2018), the distribution of the random factor characterizes the tail properties of the copula (2.5); in particular, the stationary exponential factor copula model is tail-dependent (i.e., $\chi_h > 0$) for fixed $\lambda > 0$, and its upper-tail dependence structure is governed the Hüsler and Reiss (1989) copula, which has been widely used for multivariate and spatial extremes (Davison *et al.*, 2012). Moreover, as $\lambda \to \infty$, Model (2.2) boils down to the Gaussian copula, which is tail-independent (i.e., $\chi_h = 0$). Conversely, as $\lambda \to 0$, the W process is perfectly dependent over space. Thus, the exponential factor copula model interpolates between tail independence as $\lambda \to \infty$ and perfect dependence as $\lambda \to 0$, while capturing a wide range of non-trivial tail dependence structures for $\lambda \in (0, \infty)$.

We now provide more detailed information on the sub-asymptotic behavior of Model (2.2), refining the description of its tail structure. The rate at which $\chi_h(u)$ converges to χ_h, as $u \to 1$, characterizes the flexibility of the process to capture sub-asymptotic extremal dependence. In practice, this is important, as the model will always be fitted at a finite threshold. Proposition 1 shows that the parameter λ regulates this rate of convergence; the smaller λ, the faster the convergence, and vice versa. We defer the proof to Appendix B.

Proposition 1. *Consider Model* (2.2) *with rate* $0 < \lambda < \infty$. *We have the expansion*

$$\chi_h(u) - \chi_h = -f(u) + (2 - \chi_h)\{\lambda s(u) + k(u)\}\{1 + o(1)\}, \qquad u \to 1,$$

where $f(u) \sim \frac{2\phi\{z(u)\}}{(1-u)z(u)}$, $s(u) \sim \frac{\phi\{z(u)\}}{(1-u)z(u)\{z(u)-\lambda\}}$ *and, for* $0 < \rho(h) < 1$, $k(u) \sim \frac{2\phi\{z(u)-\lambda\}}{(2-\chi_h)\{z(u)-\lambda\}}$, *with* $z(u) = (F_1^{\mathbf{W}})^{-1}(u; \lambda) \sim -\lambda^{-1}\log(1-u) + \lambda/2\{1 + o(1)\}$ *such that in the limit as* $u \to 1$,

$$\chi_h(u) - \chi_h = -f(u)\{1 + o(1)\}.$$

11

For model (2.2), we have $\chi_h = 2\{1 - \Phi(\sqrt{\gamma_h}/2)\}$, where $\gamma_h = 2\lambda^2\{1 - \rho(h)\}$; see Krupskii *et al.* (2018) and Section 1 of the Supplementary Material. The measures $\chi_h(u)$ and χ_h summarize the effects of all dependence parameters, and we use them in our simulation experiments (Section 4) and the data application (Section 5) to study the performance of our model and to assess its ability to flexibly capture different levels of extremal dependence.

Other types of tail dependence structures can be obtained using alternative distributions for the random factor V (see Krupskii *et al.*, 2018), but we here restrict ourselves to the exponential factor copula model, which yields flexible tail structures and fast inference.

3   Local likelihood inference with partial censoring

The assumption of stationarity underlying (2.2) is unrealistic over large heterogeneous regions, such as the whole contiguous U.S., but it may be the starting point for more sophisticated models. In particular, the true precipitation data generating process might be approximately stationary in small regions, providing support for non-stationary, but locally stationary, models. Here we assume that Model (2.2) provides a good approximation to the local tail dependence structure while stemming from a more complex global data generating process (e.g., such as the one used in our simulation study in Section 4.2).

Local likelihood estimation for univariate threshold exceedances was proposed by Davison and Ramesh (2000), while Anderes and Stein (2011) investigated how such an approach may be applied in the spatial context based on a single realization from a Gaussian process. Risser and Calder (2017) used a local likelihood approach to estimate the spatially-varying parameters of a non-stationary Gaussian process, modeling locally-varying anisotropies using an approach similar to the discrete mixture kernel convolution of Higdon (1998). Here we detail how to perform local likelihood estimation based on high threshold exceedances by adapting the methodology of Anderes and Stein (2011) to the joint upper tail of the

stationary copula model (2.2), under the assumption that it is valid in small regional neighborhoods. Unlike Anderes and Stein (2011), we assume that multiple replicates of the process are observed, with possibly arbitrary marginal distributions.

Since our focus is on the data's tail dependence structure, we advocate a two-step semi-parametric estimation procedure, whereby margins are first estimated at each site separately using the empirical distribution function, and the copula model is then estimated locally in a second step by maximum likelihood, censoring low (i.e., non-extreme) values to prevent them from influencing the fit. Such a two-step approach is standard in the copula literature, and has been studied in depth; see, e.g., Genest *et al.* (1995) and Joe (2014), Chapter 5.

More complex parametric approaches are also possible to estimate margins in the first step. Although marginal estimation is not our primary target, we studied the effect of the rank transformation on the estimation of the copula, by comparing it to an approach based on the generalized Pareto distribution, and the optimal scenario where marginal distributions are known exactly. We found that the use of the rank-based empirical distribution function provides a fast, robust, and reliable method, which does not largely affect the subsequent estimation of the copula (see Section 2.1 of the Supplementary Material).

The first step of our proposed estimation procedure consists of transforming the observed data non-parametrically to the uniform scale. Let $y_{1j}, \ldots, y_{Nj}$ denote N independent and identically distributed observations at the jth monitoring station, with essentially arbitrary margins. Pseudo-uniform scores may be obtained using ranks as follows:

$$u_{ij} = \frac{\mathrm{rank}(y_{ij})}{N+1}, \qquad i = 1, \ldots, N.$$

In the second step, the scores u_{ij}, $i = 1, \ldots, N$, are treated as a perfect random sample from the $\mathrm{Unif}(0, 1)$ distribution. To estimate the spatial copula structure, we then discretize the space $\mathcal{S}$ (in our case, the whole contiguous U.S.) into a fine grid $\mathcal{G} \subset \mathcal{S}$. For each grid point $\mathbf{s}_0 \in \mathcal{G}$, we assume that the local stationary copula model (2.2) with parameters

$\boldsymbol{\theta}_0 \in \boldsymbol{\Theta} \subset \mathbb{R}^p$ is valid in a small neighborhood $\mathcal{N}_{\mathbf{s}_0} \subset \mathcal{S}$ around $\mathbf{s}_0$. In what follows, these regional neighborhoods will be determined by a small number, D_0, of nearest stations from the site $\mathbf{s}_0$, thus we will write $\mathcal{N}_{\mathbf{s}_0} \equiv \mathcal{N}_{\mathbf{s}_0;D_0}$. Obviously, the choice of neighborhood is important; the stationary model (2.2) might be a poor approximation for large neighborhoods (i.e., for large D_0), whereas model fitting might be cumbersome for small neighborhoods characterized by a small number D_0 of stations. This bias-variance trade-off is tricky to deal with, and Section 5 describes one possible approach to mitigate this issue.

To estimate the local tail dependence structure, we suggest using a censored likelihood approach, which is standard in statistics for spatial extremes, though it has never been applied to Model (2.2); see, e.g., Ledford and Tawn (1996), Thibaud *et al.* (2013), Huser and Davison (2014), Wadsworth and Tawn (2014), Thibaud and Opitz (2015) and Huser *et al.* (2017). Essentially, vectors with non-extreme observations (i.e., lower than a high threshold) are partially or fully censored to prevent these low values from affecting the estimation of the extremal dependence structure. More precisely, for each grid point $\mathbf{s}_0 \in \mathcal{G}$ with associated neighborhood $\mathcal{N}_{\mathbf{s}_0;D_0}$, let $\mathbf{u}_j^0 = (u_{1j}^0, \ldots, u_{Nj}^0)^T$, $j = 1, \ldots, D_0$, denote the pseudo-uniform scores for each of the nearby stations in $\mathcal{N}_{\mathbf{s}_0;D_0}$, and let $u_j^\star$, $j = 1, \ldots, D_0$, be high marginal thresholds; in our application in Section 5, we take $u_j^\star = 0.80$ for all j. Using the notation introduced in (2.3)–(2.6), the censored local log-likelihood may be expressed as

$$
\begin{aligned}
\ell(\boldsymbol{\theta}_0 \mid \mathcal{N}_{\mathbf{s}_0;D_0}) &= \sum_{i \in \mathcal{I}_{\mathrm{nc}}} \log f_{D_0}^{\mathbf{W}}(w_{i1}^0, \ldots, w_{iD_0}^0; \boldsymbol{\theta}_0) - \sum_{i \in \mathcal{I}_{\mathrm{nc}}} \sum_{j=1}^{D_0} \log f_1^{\mathbf{W}}(w_{ij}^0; \lambda_0) \\
&\quad + N_{\mathrm{fc}} \times \log F_{D_0}^{\mathbf{W}}(w_1^\star, \ldots, w_{D_0}^\star; \boldsymbol{\theta}_0) \\
&\quad + \sum_{i \in \mathcal{I}_{\mathrm{pc}}} \log \partial_{J_i} F_{D_0}^{\mathbf{W}}(\max(w_{i1}^0, w_1^\star), \ldots, \max(w_{iD_0}^0, w_{D_0}^\star); \boldsymbol{\theta}_0) - \sum_{i \in \mathcal{I}_{\mathrm{pc}}} \sum_{j \in J_i} \log f_1^{\mathbf{W}}(w_{ij}^0; \lambda_0),
\end{aligned}
\tag{3.1}
$$

where $w_{ij}^0 = (F_1^{\mathbf{W}})^{-1}(u_{ij}^0; \lambda_0)$ and $w_j^\star = (F_1^{\mathbf{W}})^{-1}(u_j^\star; \lambda_0)$, $j = 1, \ldots, D_0$, $\mathcal{I}_{\mathrm{nc}} = \{i \in \{1, \ldots, N\} : u_{ij} > u_j^\star, \forall j = 1, \ldots, D_0\}$ is the index set of all *non-censored* observations (i.e., all vector components are extreme), $\mathcal{I}_{\mathrm{fc}} = \{i \in \{1, \ldots, N\} : u_{ij} \leq u_j^\star, \forall j = 1, \ldots, D_0\}$

14

is the index set of all *fully censored* observations (i.e., none of the vector components are extreme) with $N_{\mathrm{fc}} = |\mathcal{I}_{\mathrm{fc}}|$, $\mathcal{I}_{\mathrm{pc}} = \{1,\ldots,N\} \setminus \{\mathcal{I}_{\mathrm{nc}} \cup \mathcal{I}_{\mathrm{fc}}\}$ is the index set of all *partially censored* observations (i.e., some, but not all, vector components are extreme), $J_i = \{j \in \{1,\ldots,D_0\} : u_{ij} > u_j^\star\}$ is the index set of threshold exceedances for the ith vector of observations, and ∂_{J_i} denotes differentiation with respect to the variables indexed by the set J_i. Numerical maximization of (3.1) yields the maximum likelihood estimates $\hat{\boldsymbol{\theta}}_0$ for location $\mathbf{s}_0$. In Appendix A, we provide simple expressions for $f_{D_0}^{\mathbf{W}}$, $F_{D_0}^{\mathbf{W}}$ and $\partial_{J_i} F_{D_0}^{\mathbf{W}}$, which involve the D_0-variate Gaussian density and the multivariate Gaussian distribution in dimension D_0 and $D_0 - |J_i|$, respectively. When D_0 is large, the computation of the multivariate Gaussian distribution can significantly slow down the estimation procedure. However, thanks to our local approach, D_0 is typically quite small, and this allows to estimate the model at a reasonable computational cost. Furthermore, because the likelihood maximizations can be done independently at each grid point $\mathbf{s}_0 \in \mathcal{G}$, we can easily take advantage of distributed computing resources to perform each fit in parallel.

Following Anderes and Stein (2011), it is possible to generalize (3.1) to obtain smoother parameter estimates over space. Specifically, we can instead maximize the weighted log-likelihood function defined as

$$\ell_\omega(\boldsymbol{\theta}_0 \mid \mathcal{N}_{\mathbf{s}_0;D_0}) = \sum_{j=1}^{D_0} \omega(\|\mathbf{s}_j - \mathbf{s}_0\|) \left\{ \ell(\boldsymbol{\theta}_0|\mathcal{N}_{\mathbf{s}_0;j}) - \ell(\boldsymbol{\theta}_0|\mathcal{N}_{\mathbf{s}_0;j-1}) \right\}, \qquad (3.2)$$

where $\omega(h) \geq 0$ is a non-negative weight function, and $\{\mathcal{N}_{\mathbf{s}_0;j}; j = 0,\ldots,D_0\}$, denotes the nested sequence of subsets comprising the first j nearest neighbors of $\mathbf{s}_0$, with the convention that $\mathcal{N}_{\mathbf{s}_0;0} = \emptyset$ and $\ell(\boldsymbol{\theta}_0|\emptyset) = 0$. Choosing hard-thresholding weights with $\omega(\|\mathbf{s}_j - \mathbf{s}_0\|) = 1$ for all $\mathbf{s}_j \in \mathcal{N}_{\mathbf{s}_0;D_0}$ boils down to (3.1). However, smoother parameter estimates may be obtained by selecting soft-thresholding weights that smoothly decay to zero near the neighborhood boundaries, e.g., using the biweight function $\omega(h) = (1-(h/\tau_0)^2)_+^2$, for some bandwidth $\tau_0 > 0$. Although this estimation approach seems quite appealing, it significantly increases the

computational burden, since the weighted log-likelihood function (3.2) requires computing (3.1) in dimensions $1, \ldots, D_0$ instead of just once in dimension D_0. In Section 2.2 of the Supplementary Material, we compare hard- and soft-thesholding weights in a simulation study, and do not notice any major difference in the results. This may be explained by the fact that, unlike Anderes and Stein (2011), multiple time replicates are available to fit the model. For these reasons, and because the censored likelihood procedure is already quite intensive, we do not pursue this approach further in this paper. Subsequent simulation and real data experiments are all based on (3.1).

Note that, as mentioned above, we assume that Model (2.2) provides a good approximation to the local tail dependence structure while stemming from a more complex global data generating process. To make sure that the local stationary model (2.2) can truly come from a well-defined global stochastic process, we describe in Section 4.2 one possible way to embed the model (2.2) into a non-stationary process, which we use in our simulations.

4 Simulation study

4.1 Goals of our simulation experiments

In this section, we study the flexibility of our copula model (2.2) to locally describe complex extremal dynamics across space, and we assess the performance of our censored local estimation approach based on (3.1) at capturing such non-stationary dependence structures. We also analyze the sensitivity of the parameter estimates to the neighborhood size D_0, and the sensitivity of a profile likelihood procedure to select the smoothness parameter ν.

4.2 Data-generating scheme and simulation scenarios

We first describe how we can extend the stationary exponential factor copula model (2.2) to a non-stationary process that we use in our simulations. Further details and asymptotic results are provided in Section 1 of the Supplementary Material. Our proposed model extension,

equivalent to (2.2) for infinitesimal regions, is to consider the process

$$W(\mathbf{s}) = Z(\mathbf{s}) + \lambda_{\mathbf{s}}^{-1}E, \qquad \mathbf{s} \in \mathcal{S}, \tag{4.1}$$

where $Z(\mathbf{s})$ is a zero mean Gaussian process with non-stationary correlation function $\rho(\mathbf{s}_1, \mathbf{s}_2)$ and E is a standard exponentially distributed common factor, independent of $Z(\mathbf{s})$. The rate parameter $\lambda_{\mathbf{s}} > 0$, $\mathbf{s} \in \mathcal{S}$, is assumed to be a smooth surface, which dictates different degrees of tail dependence in distinct regions. The model in (4.1) is also tail-dependent and the limit $\chi_{12} \equiv \chi_{12}(\mathbf{s}_1, \mathbf{s}_2) = \lim_{u \to 1} \chi_{12}(u)$, with $\chi_{12}(u)$ defined analogously to (1.1) but in the non-stationary context, is $\chi_{12} = 2\{1 - \Phi(\sqrt{\gamma_{12}}/2)\}$ with $\gamma_{12} = \lambda_{\mathbf{s}_1}^2 - 2\rho(\mathbf{s}_1, \mathbf{s}_2)\lambda_{\mathbf{s}_1}\lambda_{\mathbf{s}_2} + \lambda_{\mathbf{s}_2}^2$. Although (4.1) may not realistically capture long-distance dependencies owing to the latent factor E being constant over space, its spatially-varying parameters describe the local dependence structure more flexibly; Model (4.1) is therefore useful to "think" about the results *globally*, while generating various forms of extremal dependence *locally* (or *regionally*).

In the literature, different non-stationary correlation functions $\rho(\mathbf{s}_1, \mathbf{s}_2)$ have been proposed; see, e.g., Fuentes (2001), Nychka *et al.* (2002), Stein (2005), and Paciorek and Schervish (2006). Here, we focus on a non-stationary, locally isotropic, Matérn correlation function with constant smoothness parameter $\nu > 0$, constructed through the kernel convolution approach advocated by Paciorek and Schervish (2006). Specifically, we choose

$$\rho(\mathbf{s}_1, \mathbf{s}_2) = \frac{2^{2-\nu}\delta_{\mathbf{s}_1}\delta_{\mathbf{s}_2}}{\Gamma(\nu)(\delta_{\mathbf{s}_1}^2 + \delta_{\mathbf{s}_2}^2)}\mathcal{K}_\nu\left(\frac{2\sqrt{2\nu}}{\sqrt{\delta_{\mathbf{s}_1}^2 + \delta_{\mathbf{s}_2}^2}}\|\mathbf{s}_1 - \mathbf{s}_2\|\right), \tag{4.2}$$

where $\delta_{\mathbf{s}} > 0$, $\mathbf{s} \in \mathcal{S}$, is a smoothly varying range parameter, $\Gamma(\cdot)$ is the Gamma function and $\mathcal{K}_\nu$ is the modified Bessel function of second kind of order ν. The stationary Matérn correlation function (obtained by setting $\delta_{\mathbf{s}} \equiv \delta$ for all $\mathbf{s} \in \mathcal{S}$) has become popular because of its appealing properties (Stein, 1999). In particular, a Gaussian process with Matérn correlation function is m times mean-square differentiable if and only if $\nu > m$. For $\nu = 0.5$, it boils down to the exponential correlation function, which yields continuous but non-

differentiable sample paths. As $\nu \to \infty$, sample paths are infinitely differentiable. The non-stationary correlation function defined in (4.2) is locally Matérn, and therefore it inherits these attractive properties. An extension of (4.2) allowing for varying degrees of smoothness over space has been proposed in the unpublished manuscript of Stein (2005) (see also Anderes and Stein, 2011), but in practice, estimating ν is cumbersome and conservative approaches are usually adopted.

We simulate data on a 25×25 grid in $\mathcal{S} = [1, 10]^2$ from the copula model stemming from (4.1) based on the non-stationary Matérn correlation function (4.2), generating 500 independent replicates. We choose three different levels of smoothness by fixing $\nu = 0.5, 1.5, 2.5$, and we consider three scenarios for the range $\delta_{\mathbf{s}}$ and the rate $\lambda_{\mathbf{s}}$ parameters, representing various levels of non-stationarity in the tail behavior; see Figure 4. The true parameter values and the bivariate χ_{12} measure with respect to three different reference points are shown in Figure 4 for the weakly, mildly and strongly non-stationary scenarios. In all simulations, the smoothness parameter ν is held fixed, while the rate $\lambda_{\mathbf{s}}$ and range $\delta_{\mathbf{s}}$ parameters are estimated on a 10×10 grid $\mathcal{G} \subset [1, 10]^2$ using the local estimation approach with censoring thresholds $u_j^\star = 0.95$ for all j, as described in Section 3. Thus, there are $500 \times 0.05 = 25$ exceedances per location. To compute performance metrics and to measure the uncertainty of estimated parameters, we replicated all simulation experiments 1000 times.

4.3 Simulation results

We now detail the results for the mildly non-stationary scenario (second column of Figure 4) with $\nu = 2.5$ and choosing $D_0 = 20$ nearest neighbors. Results for all simulation scenarios are summarized in Table 1. The rightmost column of Figure 4 displays mean surfaces, computed over the 1000 independent experiments, of the estimated log-rate $\log(\lambda_{\mathbf{s}})$ and range $\delta_{\mathbf{s}}$ parameters, as well as the fitted bivariate χ_{12} measure for the three different reference

18
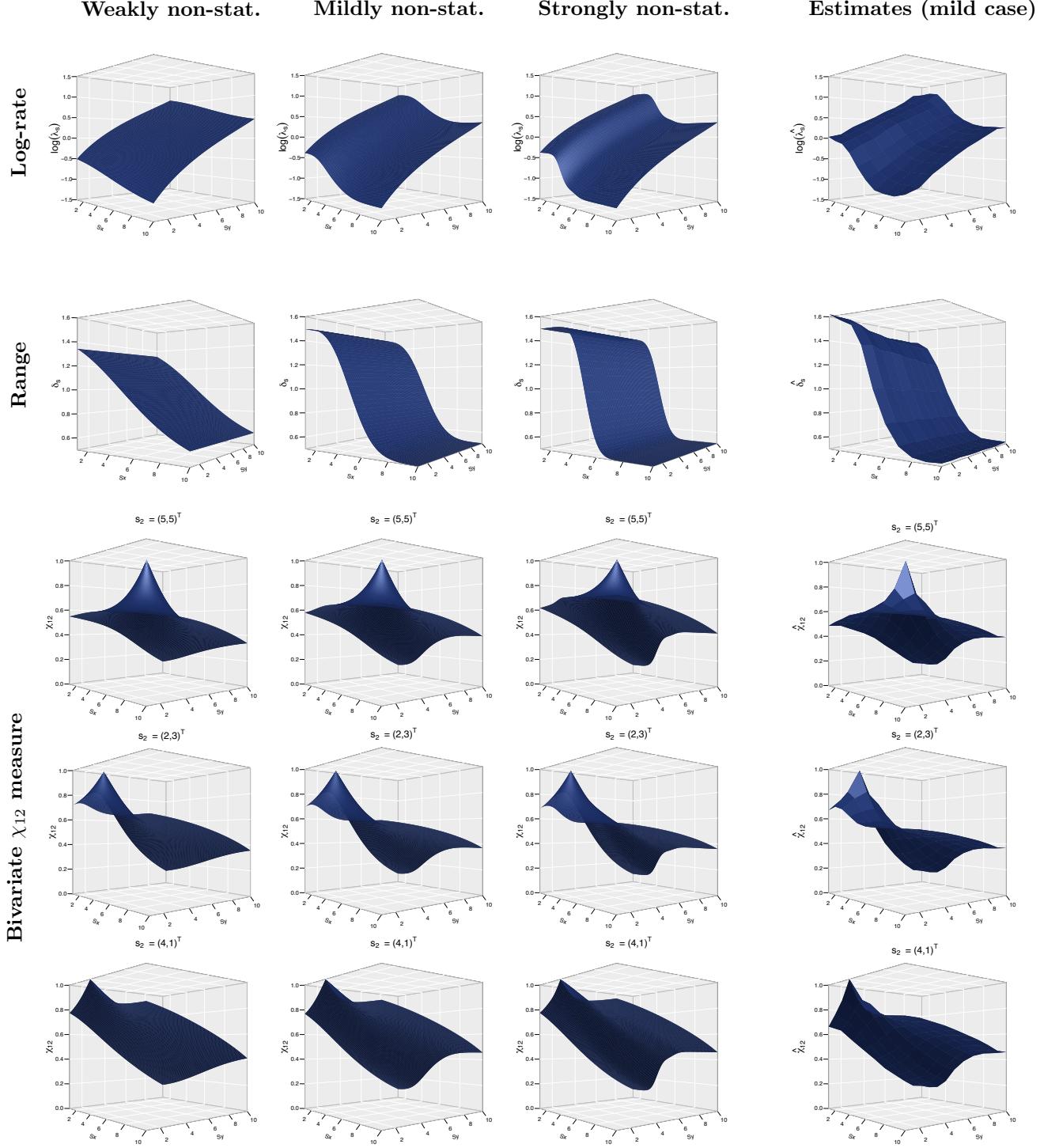
Figure 4: *Columns 1–3:* True spatially-varying log-rate $\log(\lambda_\mathbf{s})$ (1st row), range $\delta_\mathbf{s}$ (2nd row), and bivariate χ_{12} measure (3rd–5th rows) for a fixed location $\mathbf{s}_2$. Columns 1–3 correspond to different levels of non-stationarity. *Column 4:* Mean surface of estimated parameters for the mildly stationary scenario, based on 1000 simulation experiments, using the local censored likelihood approach with thresholds $u_j^\star = 0.95$ for all j, and $D_0 = 20$ nearest neighbors, as detailed in Section 3. The smoothness parameter is fixed to $\nu = 2.5$.

Table 1: RMISEs for the rate $\hat{\lambda}_{\mathbf{s}}$ and range $\hat{\delta}_{\mathbf{s}}$ parameters for each smoothness parameter $\nu = 0.5, 1.5, 2.5$, computed over 1000 replicates from the data-generating configurations discussed in Section 4.3.

Configuration	$\nu = 0.5$		$\nu = 1.5$		$\nu = 2.5$	
	$\hat{\lambda}_{\mathbf{s}}$	$\hat{\delta}_{\mathbf{s}}$	$\hat{\lambda}_{\mathbf{s}}$	$\hat{\delta}_{\mathbf{s}}$	$\hat{\lambda}_{\mathbf{s}}$	$\hat{\delta}_{\mathbf{s}}$
Weakly non-stationary	0.09	0.59	0.03	0.01	0.03	0.00
Mildly non-stationary	0.09	0.69	0.06	0.02	0.04	0.00
Strongly non-stationary	0.17	0.89	0.15	0.05	0.22	0.02

locations. By comparing the true surfaces to the mean estimates, the bias appears to be quite small overall, except perhaps at the boundaries of the study region, $\mathcal{S}$. This is a well-known drawback of local estimation approaches; because the neighborhoods are asymmetric near the boundaries, the bias is generally more severe, and this can also be noticed in our case, e.g., in regions where $\chi_{12} \approx 0$. This issue is similar to the boundary problem in kernel density estimation occurring with positively or compactly supported densities. Some strategies have been advocated to deal with it, such as taking asymmetric kernels near the boundaries, or performing local linear regression (see, e.g., Castro-Camilo *et al.*, 2018), but we do not pursue this here. Apart from this minor boundary problem, our local estimation approach succeeds in capturing the complex non-stationary dependence dynamics over space.

To assess the variability of the parameter estimates, Figure 5 displays a functional boxplot for the range parameter $\delta_{\mathbf{s}}$, projected onto the x-axis representing the only direction of variation in the true $\delta_{\mathbf{s}}$ values, and a surface boxplot for the bivariate χ_{12} measure for fixed $\mathbf{s}_2 = (2, 3)^T$. Functional and surface boxplots are the natural extensions of the classical boxplot to the case of functional data, and we refer to Sun and Genton (2011) and Genton *et al.* (2014) for their precise interpretation. As the true range parameter $\delta_{\mathbf{s}}$, $\mathbf{s} = (s_x, s_y)^T$, varies only with respect to s_x, it is easier to visualize its estimated uncertainty. As expected, $\delta_{\mathbf{s}}$ appears to be well estimated overall with higher uncertainty for larger $\delta_{\mathbf{s}}$ values. The estimated median curve follows the true curve very closely, even near the boundaries and
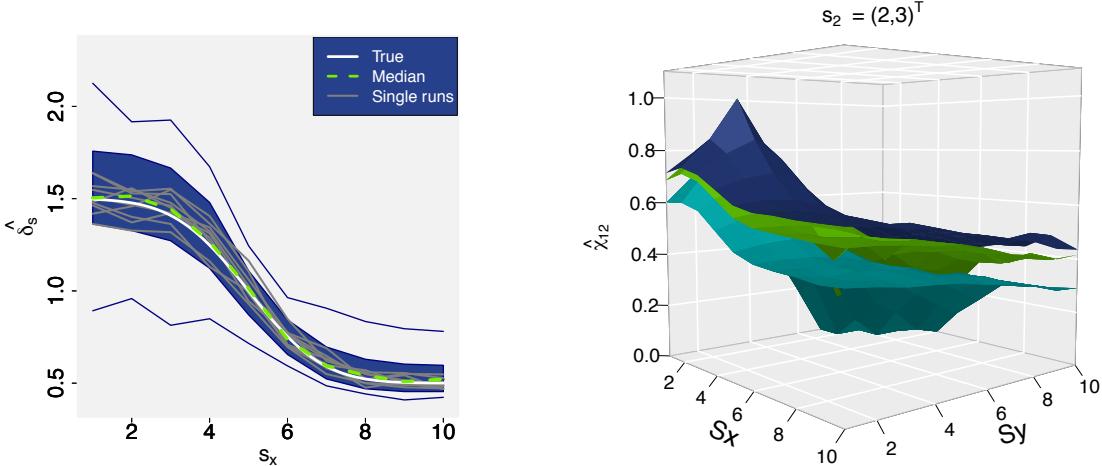
Figure 5: Functional boxplot (left) for the range $\delta_{\mathbf{s}}$, $\mathbf{s} = (s_x, s_y)^T$, plotted with respect to x-coordinate s_x, and surface boxplot (right) for the bivariate χ_{12} measure plotted as a function of location $\mathbf{s}_1$ for fixed $\mathbf{s}_2 = (2,3)^T$. In the surface boxplot, the dark and light blue surfaces represent the 1st and 3rd quartiles (i.e., the "box" of the boxplot), respectively, while the green surface is the median. The surface boxplot's "whiskers" are not displayed to improve visualization. Both panels show the results for the mildly non-stationary case with $\nu = 2.5$ and $D_0 = 20$ nearest neighbors in the local estimation approach.

around $s_x = 1$, where the true curve is the steepest, while the functional inter-quartile range is fairly narrow for all values of s_x. Moreover, the estimates for single runs, superimposed on the functional boxplot, suggest that the hard-thresholding approach produces reasonably smooth estimates, even with only about 25 threshold exceedances at each location. In our application in Section 5, we have 218 exceedances per location on average. As for the bivariate χ_{12} measure, the surface boxplot suggests that it can also be rather well estimated with relatively low uncertainty. Table 1 reports the root mean integrated squared error (RMISE) for all the data generating configurations described in Section 4.2. The results are coherent with our intuition: the estimation is more difficult for higher levels of non-stationarity and rougher random fields (i.e., with smaller ν). Overall, our simulations confirm that our local censored estimation approach provides reasonable estimates of the true range and rate parameters while capturing their dynamics over space.

To assess the performance of our local likelihood approach as a function of the neigh-

borhood size D_0, we fixed the rate parameter $\lambda_{\mathbf{s}}$ to the strongly non-stationary scenario, while considering the weakly, mildly and strongly non-stationary cases for the range parameter $\delta_{\mathbf{s}}$ (recall Figure 4). We then fit the copula model using the local likelihood approach described in Section 3 with neighborhoods $\mathcal{N}_{\mathbf{s}_0;D_0}$ defined in terms of $D_0 = 5, 10, 15, 20, 25$ nearest neighbors. Table 2 reports the RMISE for all cases. While the RMISE is quite small and fairly constant overall for the range $\delta_{\mathbf{s}}$, the rate $\lambda_{\mathbf{s}}$ seems more difficult to estimate, and it improves with lower degrees of non-stationarity for $\delta_{\mathbf{s}}$ and bigger neighborhoods (i.e., with larger D_0), even in the strongly non-stationary scenario. This suggests that the size of neighborhoods will likely be dictated by available computational resources. Unless the non-stationarity is extremely severe, it is advisable to consider large neighborhoods, as this would improve the estimation efficiency at a fairly moderate cost in bias.

Finally, using the mildly non-stationary scenario for the rate and range parameters and $\nu = 0.5$, we test an ad-hoc profile likelihood approach to select the smoothness parameter. We fit our model by setting $\nu = \nu_{\text{test}}$ with $\nu_{\text{test}} \in \{0.5, 1.0, 1.5, 2.0, 2.5\}$, and count the proportion of times that ν_{test} maximized the local log-likelihood across all grid points and the 1000 replications. Our proposed procedure is to select the "best" value for ν as the one with the largest proportion of maximized local likelihoods. Among all grid points and all replicates, the likelihood is here maximized in 37% of the cases for $\nu_{\text{test}} = 0.5$, 24% of the cases for $\nu_{\text{test}} = 1$, 21% of the cases for $\nu_{\text{test}} = 1.5$, and 18% of the cases for $\nu_{\text{test}} = 2.5$. Therefore, our profile procedure seems a reasonable method for choosing ν in a non-arbitrary way. It is important to note that this is not a method to properly *estimate* the smoothness parameter per se. Rather, the aim of our profile likelihood approach and the selection via frequencies is to select (or validate) ν in a non-arbitrary data-driven way. In general, estimation of the smoothness parameter is difficult; with a single realization in a fixed domain, not all the Matérn correlation parameters can be estimated consistently (see Zhang, 2004). Moreover,

Table 2: RMISEs for the rate $\hat{\lambda}_{\mathbf{s}}$ and range $\hat{\delta}_{\mathbf{s}}$ parameters for $\nu = 2.5$, as a function of the number D_0 of nearest neighbors used in the local estimation approach. The rate $\hat{\lambda}_{\mathbf{s}}$ was kept fixed to the strongly non-stationary case, while different degrees of non-stationarity are considered for the range $\delta_{\mathbf{s}}$.

Configuration for the range parameter $\delta_{\mathbf{s}}$	$D_0 = 5$		$D_0 = 10$		$D_0 = 15$		$D_0 = 20$		$D_0 = 25$	
	$\hat{\lambda}_{\mathbf{s}}$	$\hat{\delta}_{\mathbf{s}}$	$\hat{\lambda}_{\mathbf{s}}$	$\hat{\delta}_{\mathbf{s}}$	$\hat{\lambda}_{\mathbf{s}}$	$\hat{\delta}_{\mathbf{s}}$	$\hat{\lambda}_{\mathbf{s}}$	$\hat{\delta}_{\mathbf{s}}$	$\hat{\lambda}_{\mathbf{s}}$	$\hat{\delta}_{\mathbf{s}}$
Weakly non-stationary	1.16	0.02	0.44	0.01	0.25	0.00	0.18	0.00	0.15	0.00
Mildly non-stationary	1.33	0.03	0.61	0.01	0.44	0.01	0.34	0.01	0.30	0.01
Strongly non-stationary	1.38	0.03	0.69	0.01	0.55	0.02	0.47	0.02	0.45	0.02

when ν is estimated jointly with δ and λ, we have found that the estimated parameters become very difficult to identify.

5 Case study: U.S. winter precipitation extremes

5.1 Data description

Daily precipitation data, freely available online, were gathered from the U.S. Historical Climatological Network (USHCN). Daily precipitation data are measured in hundredths of an inch and were collected from 1900 to 2014 at the 1218 stations represented in Figure 1. To ensure data quality, we discarded all observations marked by any reliability or accuracy flag. We focus on winter data (December 21 to March 20) to remove seasonal effects, and we consider five day-cumulative precipitation, in order to capture the intensity and duration of distinct storms and to reduce the effect of temporal dependence. We conducted a preprocessing stage to check for possible temporal non-stationarity, and no evidence of non-stationarity was found (see Section 3.1 of the Supplementary Material). This procedure yields 2070 time replicates per station with 19.8% of missing data overall, which corresponds to about 2 million observations in total at all sites. The resulting dataset ranges from 0 (no rain over five days) to 3006 hundredths of an inch. The empirical 80% and 90% quantiles, plotted in Figure 6 using a logarithmic scale, reveal interesting spatial patterns that are due to marginal distributions varying smoothly over space. In order to disentangle the local

23
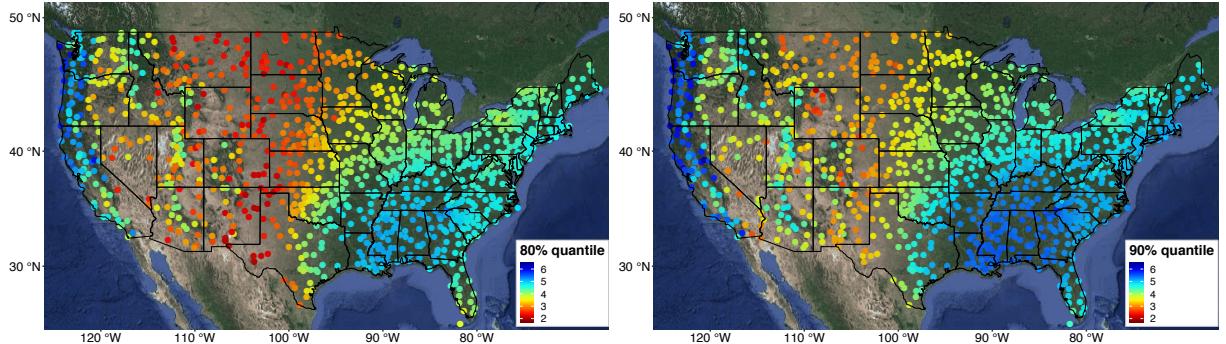
Figure 6: Empirical 80% (left) and 90% (right) quantiles of five day-cumulative winter precipitation data, observed at each of the monitoring stations, plotted on the same logarithmic color scale (units are the logarithm of hundredths of an inch).

marginal and dependence effects, we use the (two-step) local censored likelihood approach described in Section 3 and fit the exponential factor copula model; recall Sections 2 and 3.

5.2 Estimation grid and neighborhood selection

To describe the local dependence structure of precipitation extremes over the U.S., we generated a regular grid $\mathcal{G} \subset \mathcal{S}$ (using the WGS84/UTM zone 14N metric coordinate system) with 2235 grid points at an internodal distance of 60km. When plotted with respect to longitude and latitude, this results in a "distorted" grid, owing to the metric-to-degree system change.

To fit our model for extremal dependence at each grid point $\mathbf{s}_0 \in \mathcal{G}$ using the local estimation approach described in Section 3, we select a suitable number of nearest neighbors D_0, that can vary over space. Although cross-validation techniques would usually be advisable, pragmatic approaches are often adopted in practice. For example, in the time series context, Davison and Ramesh (2000) suggested selecting the local likelihood bandwidth by the naked eye, while in the spatial context, Anderes and Stein (2011) advocated a heuristic method based on a measure of spatial variation in the estimated parameters. In principle, the choice of D_0 should be such that the spatial dependence structure of threshold exceedances is approximately stationary within each selected neighborhood $\mathcal{N}_{\mathbf{s}_0;D_0}$ around $\mathbf{s}_0 \in \mathcal{G}$. Small neighborhoods (with small D_0) yield good stationary approximations, but poor statistical

24

efficiency, and vice versa, and our simulation study in Section 4 suggests that D_0 should be as large as our computational resources permit, provided the dependence structure is not overly non-stationary. In the hydrological literature, a variety of tests to assess homogeneity of marginal distributions were proposed; see for instance Lu and Stedinger (1992), Fill and Stedinger (1995), Hosking and Wallis (1993), and Hosking and Wallis (2005). Testing for stationarity of the extremal dependence structure is, however, much more complicated. Here, for simplicity, we assume that the local dependence structure of extremes is stationary whenever this is the case for margins, and we follow the recommendations of Viglione *et al.* (2007) by testing the homogeneity of the margins through a compromise between the Hosking and Wallis (1993) test and a modified Anderson–Darling test (Scholz and Stephens, 1987). Our ad-hoc neighborhood selection procedure can therefore be summarized as follows: considering an increasing nested sequence of neighborhoods, we test for marginal homogeneity until the test rejects the null hypothesis or a predefined maximum neighborhood size has been reached. In our case, we fix the maximum neighborhood to have a radius of 150km and, for computational reasons, $D_0 \leq 30$. The conditions of marginal homogeneity within a 150km radius were not satisfied at only 35 of the 2235 grid points. Although this procedure has no theoretical guarantee to be optimal, we have found that it yields reasonable estimates with our dataset.

5.3   Local likelihood inference for extreme precipitation

Following Section 3, we fitted the stationary exponential factor copula model (2.2) within small local neighborhoods, by maximizing the censored local log-likelihood (3.1) at all 2200 grid points $\mathbf{s}_0 \in \mathcal{G}$ (we discard the 35 grid points where the conditions of marginal homogeneity are not satisfied), choosing the empirical 80% quantile as a threshold to define extreme events. The left-hand panel of Figure 6 illustrates the selected threshold at each monitoring

station on the scale of the observations. As the smoothness parameter ν in (4.2) is difficult to estimate, we adopt the profile likelihood approach described in Section 4. Among all grid points, the likelihood is maximized in 58.2%, 18.6%, 11.3%, and 11.8% cases for $\nu = 0.5, 1, 1.5, 2$, respectively. To be able to easily compare rate and scale parameter estimates across space, we then choose to fix $\nu = 0.5$ at all locations, which ultimately leads to using an exponential correlation function for the underlying Gaussian process. Thanks to the embarrassingly parallel nature of our local likelihood estimation procedure, we could make an efficient use of distributed computing resources by fitting the local model at each grid point independently. On average, each set of 32 parallel fits took 3 hours. The fits were computed on a cluster with 70 nodes of 32 cores each.

To understand the dynamics of the extremal dependence strength over space, Figure 7 shows $\chi_h(u)$ at high but finite levels (specifically, $u = 0.95, 0.98$) for $h = 20, 40, 80, 160$km. Since our model is fitted locally, interpretation of $\chi_h(u)$ is possible within the selected neighborhoods, and care must be taken at locations where the radius used to fit the model is smaller than the selected distance h; see Section 3.2 in the Supplementary Material for details on the radius chosen at each grid point. As expected, tail dependence weakens at larger distances, and the patterns are quite smoothly varying over space. Extremal dependence is stronger in the Pacific West, central region (except for Colorado), mid-South, mid-West and North East. By contrast, some states in the South East (e.g., North/South Carolina and Florida) and the Rocky Mountains show weaker extremal dependence, suggesting that extreme precipitation events are more localized in these regions. To assess the variability of the estimates, we computed standard deviations for the estimates of $\chi_h(u)$ from 300 block bootstrap samples with monthly blocks. As we can see from Figure 8, estimates are relatively focused, with quite small standard deviations overall that are slightly higher with larger distances and thresholds. To complement this analysis, Section 3.3 in the Supple-
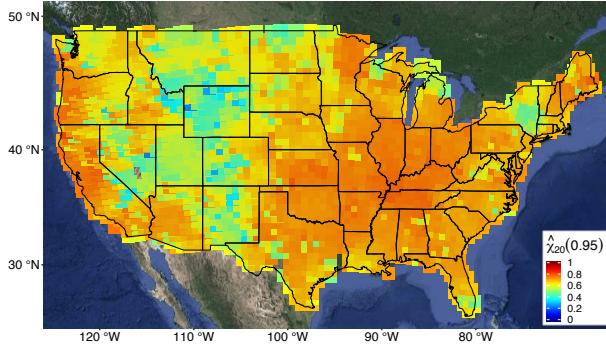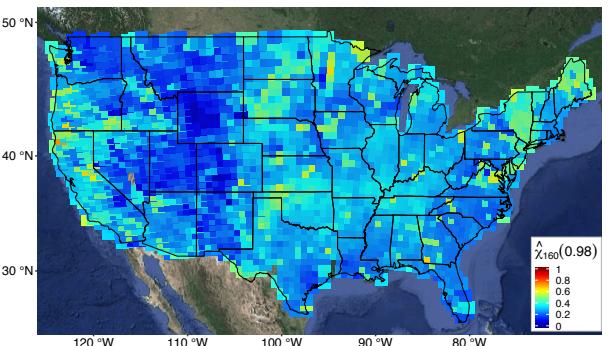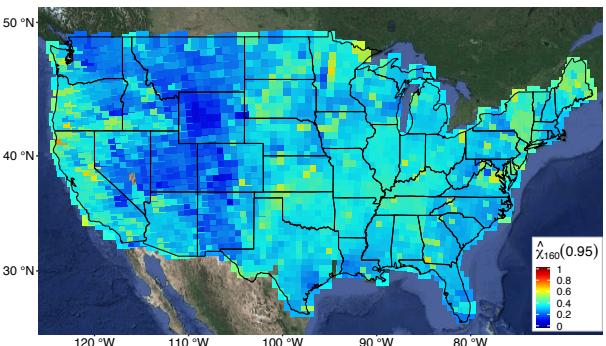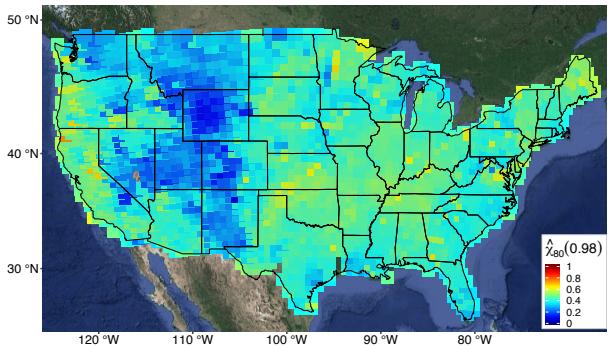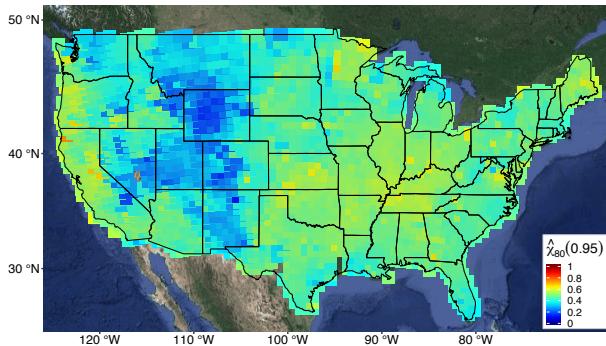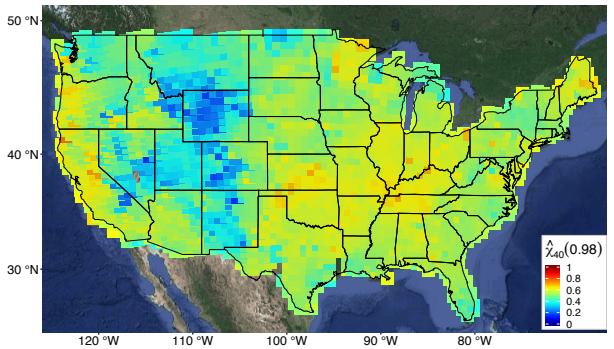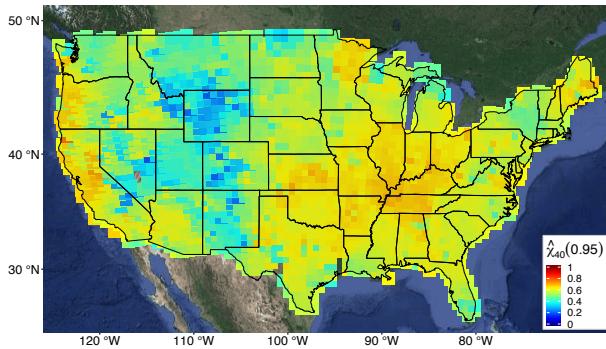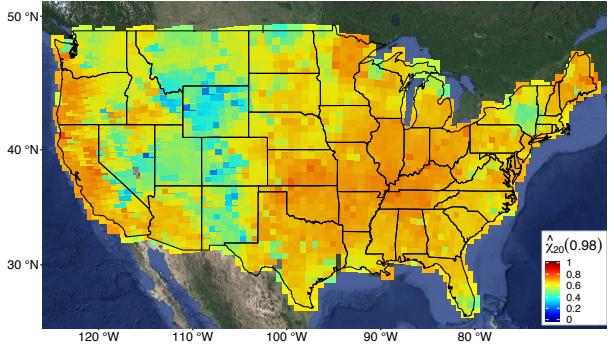
Figure 7: Estimates of $\chi_h(u)$ for $h = 20, 40, 80, 160$km (top to bottom) and $u = 0.95, 0.98$ (left to right).

mentary Material shows 95% confidence intervals for $\chi_h(u)$ at some selected locations, while Sections 3.4 and 3.5 show results for χ_h (the limiting case) and the estimated log-rate and
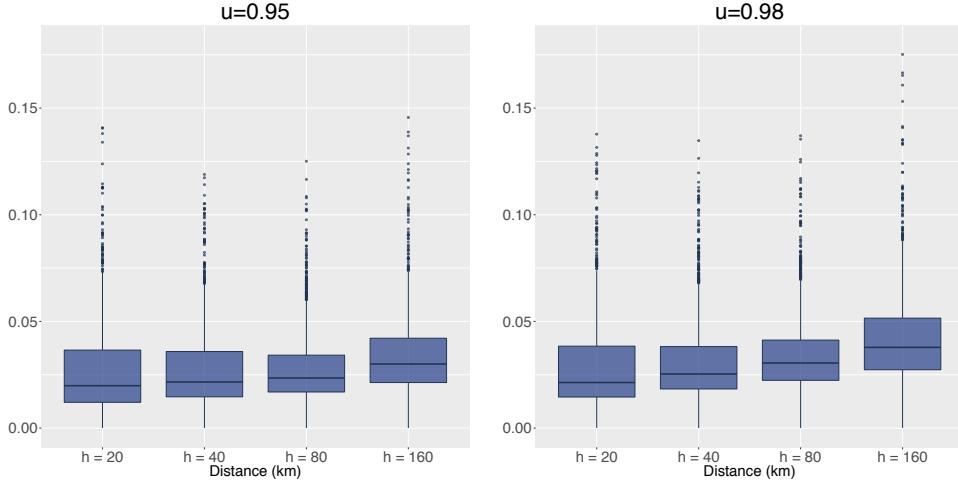
Figure 8: Summary of standard deviations for estimates of $\chi_h(u)$ for $u = 0.95$ (left) and $u = 0.98$ (right) and each distance $h = 20, 40, 80, 160$km. The standard deviations were computed from 300 block bootstrap samples with monthly blocks.

log-range parameters, respectively.

To further validate our fitted model, Figure 2 compares empirical and fitted conditional tail probabilities, as defined in (1.1). The scarcity of observations in the right tail makes the empirical estimates of $\chi_h(u)$ highly variable, which in turn yields wide empirical confidence envelopes (in dark blue). Nevertheless, the pointwise model-based estimates (in green) and the 95% confidence envelope (in light blue) suggest that our model succeeds in capturing the decaying extremal dependence between nearby sites, with relatively low uncertainty. Overall, our model has great tail flexibility, and our local estimation approach can uncover complex non-stationary patterns of extremal dependence over space, which has important implications in terms of regional risk assessment of extreme precipitation events.

Return levels and return periods are intuitive measures to quantify the risk associated with extreme events. To assess how the different spatial tail structures affect the joint occurrence of precipitation extremes, we computed the return periods (in years) associated with the events $E_{(k)} = \{Y_1^{(k)} > F_1^{-1}(u), Y_2^{(k)} > F_2^{-1}(u), Y_3^{(k)} > F_3^{-1}(u)\}$ of observing simultaneous extreme precipitation at three nearby stations within the same state k, $k = 1, \ldots, 5$,

Table 3: Pointwise estimates and 95% confidence intervals for the return periods (in years) associated with the joint probability of observing an extreme precipitation event exceeding the 94% and 99% quantiles at three locations simultaneously. The 1st column reports the state of these locations, while the 2nd column represents the average thresholds (in hundredths of an inch) for each quantile.

State	Average thresholds $u = 94\%$ / $u = 99\%$	Return period (years) Model $u = 94\%$	Empirical $u = 94\%$	Model $u = 99\%$
Louisiana	343.0 / 524.0	0.43 (0.42, 0.44)	0.46 (0.35,0.58)	1.5 (1.46, 1.55)
Mississippi	295.2 / 450.3	0.39 (0.38, 0.40)	0.51 (0.33,0.69)	1.03 (0.96,1.10)
Kentucky	267.8 / 519.4	0.41 (0.39, 0.43)	0.63 (0.45,0.82)	1.17 (1.09,1.25)
Florida	357.2 / 465.3	0.87 (0.81, 0.93)	0.92 (0.46,1.38)	3.72 (3.23, 4.21)
Tennessee	319.9 / 544.4	0.42 (0.39, 0.45)	0.75 (0.54,0.96)	1.25 (1.15,1.35)

where $u \in (0,1)$ is a high (uniform) quantile. We estimate the return periods by simulating $N = 5 \times 10^5$ replicates from our fitted copula model at each one of the 15 locations. These replicates are transformed to a common uniform scale using (2.6), and event probabilities $p_{(k)} = \Pr\{E_{(k)}\}$ are then estimated as $\hat{p}_{(k)} = N^{-1} \sum_{i=1}^{N} \mathbb{I}\{U_{i1}^{(k)} > u, U_{i2}^{(k)} > u, U_{i3}^{(k)} > u\}$, where $U_{ij}^{(k)}$ is the i-th simulated value at the j-th station ($j = 1, 2, 3$) in state k, and $\mathbb{I}(\cdot)$ is the indicator function. Return period estimates are finally obtained by taking the inverse of these estimated probabilities $\hat{p}_{(k)}$, adjusting for the time unit (years). Table 3 displays the average of the thresholds $F_1^{-1}(u)$, $F_2^{-1}(u)$ and $F_3^{-1}(u)$ used at the three stations on the scale of the data corresponding to the 94% and 99% quantiles for Louisiana, Mississippi, Kentucky, Florida and Tennessee, all of which have notoriously experienced several extreme precipitation events in the recent past. Return period estimates along with bootstrap-based 95% confidence intervals show that the selected locations in the five states are at risk of simultaneous extreme events. According to our fitted model, such spatial extreme events occur on average once every 6 to 11 months when $u = 94\%$ and once every 1 to 4 years when $u = 99\%$. Empirical estimates and confidence intervals for the return periods (provided only for $u = 94\%$ due to the high variability of empirical estimates when $u = 99\%$) show that our estimates are sensible when we take the associated uncertainty into account. These

return period estimates, of course, strongly depend on the relative positions of the selected locations. Note that the uncertainty, which accounts for both marginal and dependence estimation uncertainty, is quite low for the model-based estimates.

6 Concluding remarks

In this paper, we have developed a censored local likelihood methodology based on factor copula models for sub-asymptotic spatial extremes observed over large heterogeneous regions. Our proposed modeling approach can capture complex non-stationary patterns and is well-suited when the dependence strength weakens as events become more extreme. This contrasts with the current spatial extremes literature, which often relies on asymptotic models with a more rigid tail structure. Our extensive simulation experiments demonstrate the flexibility of our model and the efficiency of our local estimation approach based on high threshold exceedances, while providing some guidance on the selection of regional neighborhoods.

Our methodology provides a general picture of the local dependence characteristics governing precipitation extremes across the contiguous U.S.. Specifically, our fitted model reveals a diverse and complex tail dependence structure, with rich and intuitive spatial patterns. Because of the small-scale nature of the precipitation process (see, e.g., Wilby *et al.*, 1998; Katz *et al.*, 2002), our paper is focused on uncovering local dependence characteristics rather than trying to accurately capture long-distance properties. Our proposed model lacks long-distance independence, which is a common feature of most models for spatial extremes. Thus, our model must be interpreted locally, or perhaps regionally, and care is needed when extrapolating to distances higher than the ones used to fit the model. Further research should be devoted to developing models for spatial and spatio-temporal extremes that can more realistically capture long-range independence. Morris *et al.* (2017) discuss one way to achieve this based on random partitions of the study region.
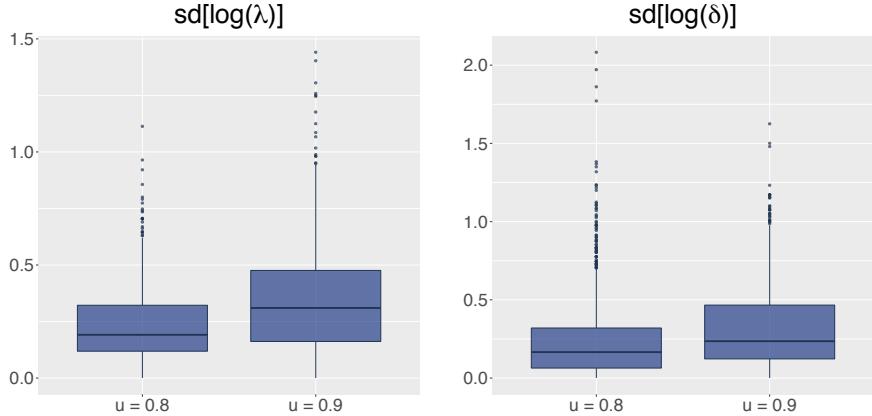
Figure 9: Boxplots of the standard deviation of the estimates of the log-rate (left) and log-range (right) parameters using the 80% and 90% quantiles.

Because our sub-asymptotic model captures weakening of dependence as events become more extreme, it is well-suited at levels less extreme than usually considered in practice. In our data application, we fitted our model at the 80%-quantile, and showed that the fit is satisfactory for a wide range of quantiles. Our approach has the additional advantage of using more data for inference, which yields lower estimation variability, as illustrated in Figure 9 which compares the fits at the 80% and 90% thresholds.

In our application, standard deviations are computed using a block bootstrap procedure, which can be computationally expensive. It would also be possible to exploit the faster-to-compute Fisher information, provided the assumption of temporal independence is satisfied.

Acknowledgements

A   Simplified expressions for the model likelihood

In this appendix, we obtain simplified expressions for the local censored likelihood function (3.1) based on Model (2.2), in order to compute the maximum likelihood estimates in a reasonable time, while avoiding numerical instabilities.

A.1   Joint density of the W process in (2.2)

From expression (2.4), we have

$$
\begin{aligned}
f_D^{\mathbf{W}}(\mathbf{w}) &= \lambda \int_0^\infty \phi_D(\mathbf{w} - v\mathbf{1}_D; \boldsymbol{\Sigma}_{\mathbf{Z}}) \exp(-\lambda v)\mathrm{d}v \\
&= \lambda \int_0^\infty (2\pi)^{-D/2}(\det \boldsymbol{\Sigma}_{\mathbf{Z}})^{-1/2} \exp\left(\frac{a_4{}^2 a_3 - a_1}{2}\right) \exp\left\{-\frac{1}{2a_3^{-1}}(v - a_4)^2\right\} \mathrm{d}v \\
&= \lambda (2\pi)^{-(D-1)/2} a_3^{-1/2}(\det \boldsymbol{\Sigma}_{\mathbf{Z}})^{-1/2} \exp\left(\frac{a_4{}^2 a_3 - a_1}{2}\right) \Phi(a_3^{1/2} a_4),
\end{aligned}
$$

where $a_1 = \mathbf{w}^T \boldsymbol{\Sigma}_{\mathbf{Z}}^{-1} \mathbf{w}$, $a_2 = \mathbf{1}_D^T \boldsymbol{\Sigma}_{\mathbf{Z}}^{-1} \mathbf{w}$, $a_3 = \mathbf{1}_D^T \boldsymbol{\Sigma}_{\mathbf{Z}}^{-1} \mathbf{1}_D$, $a_4 = (a_2 - \lambda)/a_3$, $\mathbf{w} = (w_1, \ldots, w_D)^T$, $\mathbf{1}_D = (1, \ldots, 1)^T \in \mathbb{R}^D$. The marginal density $f_1^{\mathbf{W}}(w; \lambda)$ with $D = 1$ can be easily deduced.

A.2   Joint distribution of the W process in (2.2)

Here we express the finite-dimensional distribution $F_D^{\mathbf{W}}$ as a function of conditional multivariate normal distributions. This alternative formulation allows us to use efficient routines implemented in R, substantially improving accuracy and execution time. Using integration by parts in expression (2.3), we obtain

$$
\begin{aligned}
F_D^{\mathbf{W}}(\mathbf{w}) &= \Phi_D(\mathbf{w}; \boldsymbol{\Sigma}_{\mathbf{Z}}) - \sum_{j=1}^D \int_0^\infty \Phi_{D-1}(\mathbf{w}_{-j} - v\mathbf{1}_{D-1} - \mu_{|j}; \boldsymbol{\Sigma}_{\mathbf{Z}|j})\phi(w_j - v)\exp(-\lambda v)\mathrm{d}v \\
&= \Phi_D(\mathbf{w}; \boldsymbol{\Sigma}_{\mathbf{Z}}) - \sum_{j=1}^D \exp\left(\frac{\lambda^2}{2} - \lambda w_j\right) \int_0^\infty \Phi_{D-1}\left(\mathbf{w}_{-j} - v\mathbf{1}_{D-1} - \mu_{|j}; \boldsymbol{\Sigma}_{\mathbf{Z}|j}\right)\phi(v - w_j + \lambda)\mathrm{d}v,
\end{aligned}
$$

where $\mathbf{w}_{-j} = (w_1, \ldots, w_{j-1}, w_{j+1}, \ldots, w_D)^T \in \mathbb{R}^{D-1}$ and $\mu_{|j}$ and $\boldsymbol{\Sigma}_{\mathbf{Z}|j}$ are the conditional mean and covariance matrix of $\mathbf{w}_{-j}|w_j$, respectively. Precisely,

$$\mu_{|j} = \boldsymbol{\Sigma}_{\mathbf{Z};-j,j}(w_j - v), \qquad \boldsymbol{\Sigma}_{\mathbf{Z}|j} = \boldsymbol{\Sigma}_{\mathbf{Z};-j,-j} - \boldsymbol{\Sigma}_{\mathbf{Z};-j,j}\boldsymbol{\Sigma}_{\mathbf{Z};-j,j}^T, \qquad j = 1, \ldots, D,$$

where $\boldsymbol{\Sigma}_{\mathbf{Z};-j,j}$ denotes the jth column of the matrix $\boldsymbol{\Sigma}_{\mathbf{Z}}$ with the jth row removed, etc. Now,

$$\int_0^\infty \Phi_{D-1}\left(\mathbf{w}_{-j} - v\mathbf{1}_{D-1} - \mu_{|j}; \boldsymbol{\Sigma}_{\mathbf{Z}|j}\right) \phi(v - w_j + \lambda)\mathrm{d}v$$

$$= \int_0^\infty \Phi_{D-1}\left\{\mathbf{w}_{-j} - \boldsymbol{\Sigma}_{\mathbf{Z};-j,j}w_j - v(\mathbf{1}_{D-1} - \boldsymbol{\Sigma}_{\mathbf{Z};-j,j}); \boldsymbol{\Sigma}_{\mathbf{Z}|j}\right\} \phi(v - w_j + \lambda)\mathrm{d}v$$

$$= \mathrm{Pr}\left(\mathbf{Y} \leq \boldsymbol{\Sigma}_{\mathbf{Z}|j}^{-1/2}\left\{\mathbf{w}_{-j} - \boldsymbol{\Sigma}_{\mathbf{Z};-j,j}w_j - \tilde{V}(\mathbf{1}_{D-1} - \boldsymbol{\Sigma}_{\mathbf{Z};-j,j})\right\}, \tilde{V} > 0\right)$$

$$= \mathrm{Pr}\left(\mathbf{Q}_j \leq \mathbf{w}_{-j} - \boldsymbol{\Sigma}_{\mathbf{Z};-j,j}w_j, -\tilde{V} \leq 0\right) \qquad (\text{A.1})$$

where $\mathbf{Y} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{D-1})$ follows the standard multivariate normal distribution in dimension $D-1$, and is independent of the univariate normal random variable $\tilde{V}$ with mean $w_j - \lambda$ and unit variance, and where $\mathbf{Q}_j = \boldsymbol{\Sigma}_{\mathbf{Z}|j}^{1/2}\mathbf{Y} + \tilde{V}(\mathbf{1}_{D-1} - \boldsymbol{\Sigma}_{\mathbf{Z};-j,j})$. To compute (A.1), notice that the D-dimensional vector $\mathbf{Q}_{j,0} = (\mathbf{Q}_j^T, -\tilde{V})^T$ is jointly multivariate normal, and that $\mathbf{Q}_j$ has mean $(w_j - \lambda)(\mathbf{1}_{D-1} - \boldsymbol{\Sigma}_{\mathbf{Z};-j,j})$ and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{Z}|j} + (\mathbf{1}_{D-1} - \boldsymbol{\Sigma}_{\mathbf{Z};-j,j})(\mathbf{1}_{D-1} - \boldsymbol{\Sigma}_{\mathbf{Z};-j,j})^T = \mathbf{1}_{D-1}\mathbf{1}_{D-1}^T - 2\mathbf{1}_{D-1}\boldsymbol{\Sigma}_{\mathbf{Z};-j,j}^T + \boldsymbol{\Sigma}_{\mathbf{Z};-j,-j}$. Hence, we have

$$\mathrm{Pr}\left(\mathbf{Q}_j \leq \mathbf{w}_{-j} - \boldsymbol{\Sigma}_{\mathbf{Z};-j,j}w_j, -\tilde{V} \leq 0\right) = \mathrm{Pr}\left(\mathbf{Q}_{j,0} \leq \mathbf{q}_{j,0}\right), \quad \mathbf{Q}_{j,0} = (\mathbf{Q}_j^T, -\tilde{V})^T \sim \mathcal{N}_D\left(\mu_{j,0}, \boldsymbol{\Omega}_{j,0}\right),$$

where

$$\mathbf{q}_{j,0} = \begin{pmatrix} \mathbf{w}_{-j} - \boldsymbol{\Sigma}_{\mathbf{Z};-j,j}w_j \\ 0 \end{pmatrix}, \quad \mu_{j,0} = \begin{pmatrix} (w_j - \lambda)(\mathbf{1}_{D-1} - \boldsymbol{\Sigma}_{\mathbf{Z};-j,j}) \\ \lambda - w_j \end{pmatrix},$$

$$\boldsymbol{\Omega}_{j,0} = \begin{pmatrix} \mathbf{1}_{D-1}\mathbf{1}_{D-1}^T - 2\boldsymbol{\Sigma}_{\mathbf{Z};-j,j} + \boldsymbol{\Sigma}_{\mathbf{Z};-j,-j} & \boldsymbol{\Sigma}_{\mathbf{Z};-j,j} - \mathbf{1}_{D-1} \\ \boldsymbol{\Sigma}_{\mathbf{Z};-j,j} - \mathbf{1}_{D-1} & 1 \end{pmatrix}.$$

Finally, we obtain

$$F_D^{\mathbf{W}}(\mathbf{w}) = \Phi_D(\mathbf{w}; \boldsymbol{\Sigma}_{\mathbf{Z}}) - \sum_{j=1}^D \exp\left(\frac{\lambda^2}{2} - \lambda w_j\right) \Phi_D\left(\mathbf{q}_{j,0} - \mu_{j,0}; \boldsymbol{\Omega}_{j,0}\right). \qquad (\text{A.2})$$

In particular, by setting $D = 1$, the marginal distribution may be written as

$$F_1^{\mathbf{W}}(w; \lambda) = \Phi(w) - \exp(\lambda^2/2 - \lambda w)\Phi(w - \lambda). \qquad (\text{A.3})$$

A.3 Partial derivatives of the joint distribution $F_D^{\mathbf{W}}$

We want to compute $\partial_{J_i} F_D^{\mathbf{W}}$, the partial derivative of $F_D^{\mathbf{W}}$ with respect to the variables indexed by the set $J_i \subseteq \{1, \ldots, D\}$. Without loss of generality, we here assume that $J_i = \{1, \ldots, k\}$, for some integer $1 \leq k \leq D-1$. Writing $r = D-k$, we can express the covariance matrix $\mathbf{\Sigma_Z}$ in block notation as follows:

$$\mathbf{\Sigma_Z} = \begin{pmatrix} \mathbf{\Sigma_{Z},}_k & \Sigma_{\mathbf{Z},rk}^T \\ \mathbf{\Sigma_{Z},}_{rk} & \mathbf{\Sigma_{Z},}_r \end{pmatrix}.$$

Then, writing $\mathbf{w} = (\mathbf{w}_k^T, \mathbf{w}_r^T)^T$, with $\mathbf{w}_k = (w_1, \ldots, w_k)^T$ and $\mathbf{w}_r = (w_{k+1}, \ldots, w_D)^T$, we have

$$\partial_{J_i} F_D^{\mathbf{W}}(\mathbf{w}) = \partial_{1:k}^k \lambda \int_0^\infty \Phi_D(\mathbf{w} - v\mathbf{1}_D; \mathbf{\Sigma_Z}) \exp(-\lambda v) \mathrm{d}v$$

$$= \lambda \int_0^\infty \phi_k(\mathbf{w}_k - v\mathbf{1}_k; \mathbf{\Sigma_{Z},}_k) \Phi_r(\mathbf{w}_r - v\mathbf{1}_r - \mu_{|k}; \mathbf{\Sigma_{Z|k}}) \exp(-\lambda v) \mathrm{d}v,$$

where $\mu_{|k}$ and $\mathbf{\Sigma_{Z|k}}$ are the conditional mean and covariance of $\mathbf{w}_r \mid \mathbf{w}_k$, respectively, and are obtained as $\mu_{|k} = \mathbf{\Sigma_{Z},}_{rk} \mathbf{\Sigma_{Z},}_k^{-1}(\mathbf{w}_k - v\mathbf{1}_k)$, $\mathbf{\Sigma_{Z|k}} = \mathbf{\Sigma_{Z},}_r - \mathbf{\Sigma_{Z},}_{rk} \mathbf{\Sigma_{Z},}_k^{-1} \Sigma_{\mathbf{Z},rk}^T$. Tedious but straightforward calculations yield $\phi_k(\mathbf{w}_k - v\mathbf{1}_k; \mathbf{\Sigma_{Z},}_k) \exp(-\lambda v) = C\ \phi\{b_3^{1/2}(v - b_4)\}$, where $C = (2\pi)^{-(k-1)/2} b_3^{-1/2}(\det \mathbf{\Sigma_{Z},}_k)^{-1/2} \exp\{(b_4^2 b_3 - b_1)/2\}$, with $b_1 = \mathbf{w}_k^T \mathbf{\Sigma_{Z},}_k^{-1} \mathbf{w}_k$, $b_2 = \mathbf{1}_k^T \mathbf{\Sigma_{Z},}_k^{-1} \mathbf{w}_k$, $b_3 = \mathbf{1}_k^T \mathbf{\Sigma_{Z},}_k^{-1} \mathbf{1}_k$, and $b_4 = (b_2 - \lambda)/b_3$. Therefore,

$$\partial_{J_i} F_D^{\mathbf{W}}(\mathbf{w}) = \lambda C \int_0^\infty \phi\{b_3^{1/2}(v - b_4)\} \Phi_r(\mathbf{w}_r - v\mathbf{1}_r - \mu_{|k}; \mathbf{\Sigma_{Z|k}}) \mathrm{d}v$$

$$= \lambda C \int_0^\infty \phi\{b_3^{1/2}(v - b_4)\} \Phi_r \left\{ \mathbf{w}_r - \mathbf{\Sigma_{Z},}_{rk} \mathbf{\Sigma_{Z},}_k^{-1} \mathbf{w}_k - v(\mathbf{1}_r - \mathbf{\Sigma_{Z},}_{rk} \mathbf{\Sigma_{Z},}_k^{-1} \mathbf{1}_k); \mathbf{\Sigma_{Z|k}} \right\} \mathrm{d}v$$

$$= \lambda C \operatorname{Pr} \left(\mathbf{Y} \leq \mathbf{\Sigma_{Z|k}}^{-1/2} \left\{ \mathbf{w}_r - \mathbf{\Sigma_{Z},}_{rk} \mathbf{\Sigma_{Z},}_k^{-1} \mathbf{w}_k - \tilde{V}(\mathbf{1}_r - \mathbf{\Sigma_{Z},}_{rk} \mathbf{\Sigma_{Z},}_k^{-1} \mathbf{1}_k) \right\}, \tilde{V} > 0 \right)$$

$$= \lambda C \operatorname{Pr} \left(\mathbf{Q}_r \leq \mathbf{w}_r - \mathbf{\Sigma_{Z},}_{rk} \mathbf{\Sigma_{Z},}_k^{-1} \mathbf{w}_k, -\tilde{V} \leq 0 \right), \tag{A.4}$$

where $\mathbf{Y} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_r)$ has the standard multivariate normal distribution in dimension r, and is independent of the univariate normal random variable $\tilde{V}$ with mean b_4 and variance b_3^{-1}, and where $\mathbf{Q}_r = \mathbf{\Sigma_{Z|k}}^{1/2} \mathbf{Y} + \tilde{V}(\mathbf{1}_r - \mathbf{\Sigma_{Z},}_{rk} \mathbf{\Sigma_{Z},}_k^{-1} \mathbf{1}_k)$. To compute (A.4), notice that the $(r+1)$-dimensional vector $\mathbf{Q}_{r,0} = (\mathbf{Q}_r^T, -\tilde{V})^T$ is jointly multivariate normal, and that $\mathbf{Q}_r$ has mean

$b_4(\mathbf{1}_r - \boldsymbol{\Sigma}_{\mathbf{Z},rk}\boldsymbol{\Sigma}_{\mathbf{Z},k}^{-1}\mathbf{1}_k)$ and covariance $\boldsymbol{\Sigma}_{\mathbf{Z}|k} + b_3^{-1}(\mathbf{1}_r - \boldsymbol{\Sigma}_{\mathbf{Z},rk}\boldsymbol{\Sigma}_{\mathbf{Z},k}^{-1}\mathbf{1}_k)(\mathbf{1}_r - \boldsymbol{\Sigma}_{\mathbf{Z},rk}\boldsymbol{\Sigma}_{\mathbf{Z},k}^{-1}\mathbf{1}_k)^T$. Hence,

$$\Pr\left(\mathbf{Q}_r \le \mathbf{w}_r - \boldsymbol{\Sigma}_{\mathbf{Z},rk}\boldsymbol{\Sigma}_{\mathbf{Z},k}^{-1}\mathbf{w}_k, -\tilde{V} \le 0\right) = \Pr\left(\mathbf{Q}_{r,0} \le \mathbf{q}_{r,0}\right),$$

where $\mathbf{Q}_{r,0} = (\mathbf{Q}_r^T, -\tilde{V})^T \sim \mathcal{N}_{r+1}(\mu_{r,0}, \boldsymbol{\Omega}_{r,0})$ and

$$\mathbf{q}_{r,0} = \begin{pmatrix} \mathbf{w}_r - \boldsymbol{\Sigma}_{\mathbf{Z},rk}\boldsymbol{\Sigma}_{\mathbf{Z},k}^{-1}\mathbf{w}_k \\ 0 \end{pmatrix}, \quad \mu_{r,0} = \begin{pmatrix} b_4(\mathbf{1}_r - \boldsymbol{\Sigma}_{\mathbf{Z},rk}\boldsymbol{\Sigma}_{\mathbf{Z},k}^{-1}\mathbf{1}_k) \\ -b_4 \end{pmatrix},$$

$$\boldsymbol{\Omega}_{r,0} = \begin{pmatrix} \boldsymbol{\Sigma}_{\mathbf{Z}|k} + b_3^{-1}(\mathbf{1}_r - \boldsymbol{\Sigma}_{\mathbf{Z},rk}\boldsymbol{\Sigma}_{\mathbf{Z},k}^{-1}\mathbf{1}_k)(\mathbf{1}_r - \boldsymbol{\Sigma}_{\mathbf{Z},rk}\boldsymbol{\Sigma}_{\mathbf{Z},k}^{-1}\mathbf{1}_k)^T & b_3^{-1}(\boldsymbol{\Sigma}_{\mathbf{Z},rk}\boldsymbol{\Sigma}_{\mathbf{Z},k}^{-1}\mathbf{1}_k - \mathbf{1}_r) \\ b_3^{-1}(\boldsymbol{\Sigma}_{\mathbf{Z},rk}\boldsymbol{\Sigma}_{\mathbf{Z},k}^{-1}\mathbf{1}_k - \mathbf{1}_r)^T & b_3^{-1} \end{pmatrix}.$$

Finally, we obtain

$$\partial_{J_i} F_D^{\mathbf{W}}(\mathbf{w}) = \lambda C \ \Phi_{r+1}\left(\mathbf{q}_{r,0} - \mu_{r,0}; \boldsymbol{\Omega}_{r,0}\right).$$

B Tail properties of the non-stationary exponential factor copula model

In this appendix, we provide detailed information on the sub-asymptotic joint tail behavior of the stationary exponential factor model (2.2).

Lemma 1. *In Model* (4.1), *the marginal distribution at site* $\mathbf{s} \in \mathcal{S}$, $F_{1;\mathbf{s}}^{\mathbf{W}}(w; \lambda_{\mathbf{s}})$, *satisfies*

$$F_{1;\mathbf{s}}^{\mathbf{W}}(w; \lambda_{\mathbf{s}}) = 1 - \exp(\lambda_{\mathbf{s}}^2/2 - \lambda_{\mathbf{s}}w) + \lambda_{\mathbf{s}}\phi(w)\{w(w - \lambda_{\mathbf{s}})\}^{-1} + O\{\phi(w)w^{-4}\}, \qquad w \to \infty.$$

Proof. From (A.3), we have $F_{1;\mathbf{s}}^{\mathbf{W}}(w; \lambda_{\mathbf{s}}) = \Phi(w) - \Phi(w - \lambda_{\mathbf{s}})\exp(\lambda_{\mathbf{s}}^2/2 - \lambda_{\mathbf{s}}w)$. Furthermore, thanks to a well-known expansion of the Gaussian tail, we have

$$1 - \Phi(w) = \phi(w)\{w^{-1} - w^{-3} + O(w^{-5})\}, \qquad w \to \infty. \tag{B.1}$$

Plugging (B.1) into the expression for $F_{1;\mathbf{s}}^{\mathbf{W}}(w; \lambda_{\mathbf{s}})$ yields, as $w \to \infty$,

$$
\begin{aligned}
F_{1;\mathbf{s}}^{\mathbf{W}}(w; \lambda_{\mathbf{s}}) &= 1 - \phi(w)\{w^{-1} - w^{-3} + O(w^{-5})\} \\
&\quad - \left[1 - \phi(w - \lambda_{\mathbf{s}})\{(w - \lambda_{\mathbf{s}})^{-1} - (w - \lambda_{\mathbf{s}})^{-3} + O(w^{-5})\}\right]\exp(\lambda_{\mathbf{s}}^2/2 - \lambda_{\mathbf{s}}w)
\end{aligned}
$$

$$= 1 - \exp(\lambda_{\mathbf{s}}^2/2 - \lambda_{\mathbf{s}}w) + \phi(w)\left[(w - \lambda_{\mathbf{s}})^{-1} - (w - \lambda_{\mathbf{s}})^{-3} - w^{-1} + w^{-3} + O(w^{-5})\right]$$

$$= 1 - \exp(\lambda_{\mathbf{s}}^2/2 - \lambda_{\mathbf{s}}w) + \lambda_{\mathbf{s}}\phi(w)\{w(w - \lambda_{\mathbf{s}})\}^{-1} + O\{\phi(w)w^{-4}\}.$$

$\square$

Lemma 2. *In Model* (4.1)*, the marginal quantile function at location* $\mathbf{s} \in \mathcal{S}$*,* $q_{\mathbf{s}}^{\mathbf{W}}(t; \lambda_{\mathbf{s}}) = F_{1;\mathbf{s}}^{\mathbf{W}^{-1}}(1 - t^{-1}; \lambda_{\mathbf{s}})$*, admits the expansion*

$$q_{\mathbf{s}}^{\mathbf{W}}(t; \lambda_{\mathbf{s}}) = \lambda_{\mathbf{s}}^{-1}\log t + \lambda_{\mathbf{s}}/2 - t\frac{\phi(\lambda_{\mathbf{s}}^{-1}\log t + \lambda_{\mathbf{s}}/2)}{\lambda_{\mathbf{s}}^{-2}\log^2 t - \lambda_{\mathbf{s}}^2/4}\{1 + o(1)\}, \quad t \to \infty.$$

Proof. From notational convenience, we shall write $q(t) \equiv q_{\mathbf{s}}^{\mathbf{W}}(t; \lambda_{\mathbf{s}})$. Because $1 - F_{1;\mathbf{s}}^{\mathbf{W}}\{q(t)\} = t^{-1}$ and $q(t) \to \infty$ as $t \to \infty$, we have from Lemma 1 that

$$\exp\{\lambda_{\mathbf{s}}^2/2 - \lambda_{\mathbf{s}}q(t)\} - \lambda_{\mathbf{s}}\phi\{q(t)\}\left[q(t)\{q(t) - \lambda_{\mathbf{s}}\}\right]^{-1} + O[\phi\{q(t)\}q(t)^{-4}] = t^{-1}. \qquad \text{(B.2)}$$

Noting that the leftmost term in (B.2) is dominant, this yields

$$q(t) = \lambda_{\mathbf{s}}^{-1}\log t + \lambda_{\mathbf{s}}/2 + r(t), \qquad \text{(B.3)}$$

where $r(t) \to 0$, as $t \to \infty$. Using (B.3) back into (B.2) gives

$$t^{-1}\exp\{-\lambda_{\mathbf{s}}r(t)\} - \lambda_{\mathbf{s}}\frac{\phi(\lambda_{\mathbf{s}}^{-1}\log t + \lambda_{\mathbf{s}}/2)}{\lambda_{\mathbf{s}}^{-2}\log^2 t - \lambda_{\mathbf{s}}^2/4}\{1 + o(1)\} = t^{-1}.$$

Thus, because $\exp(-x) = 1 - x\{1 + o(1)\}$, as $x \to 0$, we obtain

$$r(t) = -t\frac{\phi(\lambda_{\mathbf{s}}^{-1}\log t + \lambda_{\mathbf{s}}/2)}{\lambda_{\mathbf{s}}^{-2}\log^2 t - \lambda_{\mathbf{s}}^2/4}\{1 + o(1)\},$$

which, combined with (B.3), concludes the proof. $\square$

Proof of Proposition 1. Consider the stationary exponential factor model (2.2), and let $z(u) = F_1^{\mathbf{W}^{-1}}(u; \lambda)$ denote univariate quantiles. Thanks to (A.2), we have

$$C_2^{\mathbf{W}}(u, u) = F_2^{\mathbf{W}}\{z(u), z(u)\}$$

$$= \Phi_2\{z(u), z(u); \boldsymbol{\Sigma_Z}\} - 2\exp\left\{\lambda^2/2 - \lambda z(u)\right\} \Phi_2\left[\lambda\{1 - \rho(h)\}, z(u) - \lambda; \boldsymbol{\Omega}\right],$$

$$= \Phi_2\{z(u), z(u); \boldsymbol{\Sigma_Z}\} - 2\exp\left\{\lambda^2/2 - \lambda z(u)\right\} \Phi_2\left[\lambda\sqrt{\{1 - \rho(h)\}/2}, z(u) - \lambda; \tilde{\boldsymbol{\Omega}}\right],$$

(B.4)

where $\rho(h)$ is the underlying correlation function, and the covariance matrices $\boldsymbol{\Omega}$ and $\tilde{\boldsymbol{\Omega}}$ are

$$\boldsymbol{\Omega} = \begin{pmatrix} 2\{1 - \rho(h)\} & -\{1 - \rho(h)\} \\ -\{1 - \rho(h)\} & 1 \end{pmatrix}, \quad \tilde{\boldsymbol{\Omega}} = \begin{pmatrix} 1 & -\sqrt{\{1 - \rho(h)\}/2} \\ -\sqrt{\{1 - \rho(h)\}/2} & 1 \end{pmatrix}.$$

Now, thanks to Lemma 2, $z(u) = q_{\mathbf{s}}^{\mathbf{W}}\{(1-u)^{-1}; \lambda\} = -\lambda^{-1}\log(1-u) + \lambda/2 + s(u)$, with

$$s(u) = -\frac{\phi\{z(u)\}}{(1-u)z(u)\{z(u) - \lambda\}}\{1 + o(1)\}, \quad u \to 1, \tag{B.5}$$

such that $s(u) \to 0$, $z(u) = -\lambda^{-1}\log(1-u) + \lambda/2\,\{1 + o(1)\}$, and $z(u) \to \infty$, as $u \to 1$.

Consequently, from the definition (1.1) and (B.4)–(B.5), we can write

$$\chi_h(u) = \frac{1 - 2u + C_2^{\mathbf{W}}(u, u)}{1 - u} = 2 - \frac{1 - C_2^{\mathbf{W}}(u, u)}{1 - u} = 2 - f(u) - g(u)h(u), \tag{B.6}$$

where, using the bivariate Gaussian survivor function $\overline{\Phi}_2$, we have

$$f(u) = \frac{1 - \Phi_2\{z(u), z(u); \boldsymbol{\Sigma_Z}\}}{1 - u} = \frac{2[1 - \Phi\{z(u)\}] - \overline{\Phi}_2\{z(u), z(u); \boldsymbol{\Sigma_Z}\}}{1 - u} \tag{B.7}$$

$$= \frac{2[1 - \Phi\{z(u)\}]\{1 + o(1)\}}{1 - u} = \frac{2\phi\{z(u)\}}{(1-u)z(u)}\{1 + o(1)\} \to 0, \quad u \to 1;$$

$$g(u) = \frac{\exp\left\{\lambda^2/2 - \lambda z(u)\right\}}{1 - u} = \exp\{-\lambda s(u)\} \to 1, \quad u \to 1; \tag{B.8}$$

$$h(u) = 2\Phi_2\left[\lambda\sqrt{\{1 - \rho(h)\}/2}, z(u) - \lambda; \tilde{\boldsymbol{\Omega}}\right] \to 2\Phi\left[\lambda\sqrt{\{1 - \rho(h)\}/2}\right], \quad u \to 1. \tag{B.9}$$

The 3rd equality in (B.7) is true as bivariate Gaussian random vectors $(Z_1, Z_2)^T$ with correlation $\rho < 1$ are tail-independent, i.e., $\Pr(Z_1 > z \mid Z_2 > z) \to 0$, as $z \to \infty$ (Ledford and Tawn, 1996); the expansion in the 4th equality in (B.7) is due to (B.1); finally, $f(u) \to 0$, as $\phi\{z(u)\}/(1-u) \sim (2\pi)^{-1/2}\exp\{-\log^2(1-u)/2 - \log(1-u)\} \to 0$, as $u \to 1$.

Equations (B.6)–(B.9) imply that $\chi_h = \lim_{u \to 1} \chi_h(u) = 2 - 2\Phi[\lambda\sqrt{\{1 - \rho(h)\}/2}]$. Hence, by writing $A = \lambda\sqrt{\{1 - \rho(h)\}/2}$, we obtain as $u \to 1$

$$\chi_h(u) - \chi_h = 2\Phi(A) - f(u) - g(u)h(u)$$

$$= -f(u) + 2\Phi(A) \left[1 - \exp\{-\lambda s(u)\} \frac{\Phi_2\left\{A, z(u) - \lambda; \tilde{\boldsymbol{\Omega}}\right\}}{\Phi(A)} \right]$$

$$= -f(u) + 2\Phi(A) \left[1 - \{1 - \lambda s(u)\}\{1 - k(u)\}\{1 + o(1)\} \right]$$

$$= -f(u) + 2\Phi(A) \left\{ \lambda s(u) + k(u) \right\} \left\{ 1 + o(1) \right\}, \tag{B.10}$$

where $k(u) = \left[\Phi(A) - \Phi_2\left\{A, z(u) - \lambda; \tilde{\boldsymbol{\Omega}}\right\} \right] / \Phi(A) \to 0$ as $u \to 1$. Notice that by the l'Hospital's rule, it can be verified that for $\rho(h) > 0$, $k(u)$ satisfies the expression

$$k(u) = \frac{1 - \Phi\{z(u) - \lambda\}}{\Phi(A)}\{1 + o(1)\} = \frac{\phi\{z(u) - \lambda\}}{\Phi(A)\{z(u) - \lambda\}}\{1 + o(1)\} \to 0, \quad u \to 1. \tag{B.11}$$

Comparing the rates of convergence of $s(u)$, $f(u)$ and $k(u)$ in (B.5), (B.7) and (B.11), respectively, we deduce that $f(u)$ is dominant as $u \to 1$, and therefore from (B.10), we have

$$\chi_h(u) - \chi_h = -f(u)\{1 + o(1)\} = -\frac{2\phi\{z(u)\}}{(1 - u)z(u)}\{1 + o(1)\} = -\frac{2\phi\{-\lambda^{-1}\log(1 - u) + \lambda/2\}}{(1 - u)\{-\lambda^{-1}\log(1 - u) + \lambda/2\}}\{1 + o(1)\},$$

which concludes the proof. $\qquad\square$

References

Anderes, E. B. and Stein, M. L. (2011) Local likelihood estimation for nonstationary random fields. *Journal of Multivariate Analysis* **102**(3), 506–520.

Asadi, P., Davison, A. C. and Engelke, S. (2015) Extremes on river networks. *Annals of Applied Statistics* **9**(4), 2023–2050.

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

Bopp, G. P., Shaby, B. A. and Huser, R. (2019) A hierarchical max-infinitely divisible process for extreme areal precipitation over watersheds. *arXiv preprint 1805.06084* .

Castro-Camilo, D. and de Carvalho, M. (2017) Spectral density regression for bivariate extremes. *Stochastic environmental research and risk assessment* **31**(7), 1603–1613.

Castro-Camilo, D., de Carvalho, M. and Wadsworth, J. (2018) Time-varying extreme value dependence with application to leading european stock markets. *The Annals of Applied Statistics* **12**(1), 283–309.

Castruccio, S., Huser, R. and Genton, M. G. (2016) High-order composite likelihood inference for max-stable distributions and processes. *Journal of Computational and Graphical Statistics* **25**(4), 1212–1229.

Cooley, D. S., Naveau, P. and Nychka, D. (2007) Bayesian Spatial Modeling of Extreme Precipitation Return Levels. *Journal of American Statistical Association* **102**(479), 824–840.

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

Davison, A. C., Huser, R. and Thibaud, E. (2019) Spatial extremes. In *Handbook of Environmental and Ecological Statistics*, chapter 31, pp. 711–744. CRC Press.

Davison, A. C., Padoan, S. and Ribatet, M. (2012) Statistical Modelling of Spatial Extremes (with Discussion). *Statistical Science* **27**(2), 161–186.

Davison, A. C. and Ramesh, N. (2000) Local likelihood smoothing of sample extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **62**(1), 191–208.

de Fondeville, R. and Davison, A. C. (2018) High-dimensional peaks-over-threshold inference. *Biometrika* **105**(3), 575–592.

Engelke, S., Malinowski, A., Kabluchko, Z. and Schlather, M. (2015) Estimation of hüsler–reiss distributions and brown–resnick processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **77**(1), 239–265.

Ferreira, A., De Haan, L. *et al.* (2014) The generalized pareto process; with a view towards application and simulation. *Bernoulli* **20**(4), 1717–1737.

Fill, H. D. and Stedinger, J. R. (1995) Homogeneity tests based upon gumbel distribution and a critical appraisal of dalrymple's test. *Journal of Hydrology* **166**(1–2), 81–105.

Fischer, E. M. and Knutti, R. (2016) Observed heavy precipitation increase confirms theory and early models. *Nature Climate Change* **6**(11), 986–991.

Fuentes, M. (2001) A High Frequency Kriging Approach for Non-Stationary Environmental Processes. *Environmetrics* **12**(5), 469–483.

Genest, C., Ghoudi, K. and Rivest, L.-P. (1995) A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* **82**(3), 543–552.

Genton, M. G., Johnson, C., Potter, K., Stenchikov, G. and Sun, Y. (2014) Surface boxplots. *Stat* **3**(1), 1–11.

Gneiting, T., Genton, M. G. and Guttorp, P. (2006) Geostatistical space-time models, stationarity, separability, and full symmetry. *Monographs On Statistics and Applied Probability* **107**, 151.

Higdon, D. (1998) A process-convolution approach to modelling temperatures in the north atlantic ocean. *Environmental and Ecological Statistics* **5**(2), 173–190.

Hoerling, M., Eischeid, J., Perlwitz, J., Quan, X.-W., Wolter, K. and Cheng, L. (2016) Characterizing recent trends in U.S. heavy precipitation. *Journal of Climate* **29**(1), 2313–2332.

Hosking, J. and Wallis, J. (1993) Some statistics useful in regional frequency analysis. *Water resources research* **29**(2), 271–281.

Hosking, J. R. M. and Wallis, J. R. (2005) *Regional frequency analysis: an approach based on L-moments.* Cambridge University Press.

Huser, R. and Davison, A. C. (2013) Composite likelihood estimation for the Brown–Resnick process. *Biometrika* **100**(2), 511–518.

Huser, R. and Davison, A. C. (2014) Space–time modelling of extreme events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **76**(2), 439–461.

Huser, R., Davison, A. C. and Genton, M. G. (2016) Likelihood estimators for multivariate extremes. *Extremes* **19**(1), 79–103.

Huser, R., Dombry, C., Ribatet, M. and Genton, M. G. (2019) Full likelihood inference for max-stable data. *Stat* **8**(1), e218.

Huser, R. and Genton, M. G. (2016) Non-stationary dependence structures for spatial extremes. *Journal of agricultural, biological, and environmental statistics* **21**(3), 470–491.

Huser, R., Opitz, T. and Thibaud, E. (2017) Bridging asymptotic independence and dependence in spatial extremes using gaussian scale mixtures. *Spatial Statistics* **21**, 166–186.

Huser, R., Opitz, T. and Thibaud, E. (2018) Penultimate modeling of spatial extremes: statistical inference for max-infinitely divisible processes. *arXiv preprint 1801.02946* .

Huser, R. and Wadsworth, J. L. (2019) Modeling spatial processes with unknown extremal dependence class. *Journal of the American Statistical Association.* To appear .

Hüsler, J. and Reiss, R.-D. (1989) Maxima of normal random vectors: between independence and complete dependence. *Statistics & Probability Letters* **7**(4), 283–286.

Joe, H. (2014) *Dependence modeling with copulas.* CRC Press.

Katz, R. W., Parlange, M. B. and Naveau, P. (2002) Statistics of extremes in hydrology. *Advances in water resources* **25**(8-12), 1287–1304.

Krupskii, P., Huser, R. and Genton, M. G. (2018) Factor copula models for replicated spatial data. *Journal of the American Statistical Association* **113**(521), 467–479.

Krupskii, P. and Joe, H. (2015) Structured factor copula models: theory, inference and computation. *Journal of Multivariate Analysis* **138**, 53–73.

Ledford, A. W. and Tawn, J. A. (1996) Statistics for near independence in multivariate extreme values. *Biometrika* **83**(1), 169–187.

Lu, L.-H. and Stedinger, J. R. (1992) Sampling variance of normalized gev/pwm quantile estimators and a regional homogeneity test. *Journal of Hydrology* **138**(1-2), 223–245.

Morris, S. A., Reich, B. J., Thibaud, E. and Cooley, D. (2017) A space-time skew-t model for threshold exceedances. *Biometrics* **73**(3), 749–758.

Nychka, D., Wikle, C. K. and Royle, J. A. (2002) Multiresolution Models for Nonstationary Spatial Covariance Functions. *Statistical Modelling* **2**(4), 315–331.

Opitz, T. (2016) Modeling asymptotically independent spatial extremes based on Laplace random fields. *Spatial Statistics* **16**(1), 1–18.

Opitz, T., Huser, R., Bakka, H. and Rue, H. (2018) INLA goes extreme: Bayesian tail regression for the estimation of high spatio-temporal quantiles. *Extremes* **21**(3), 441–462.

Paciorek, C. J. and Schervish, M. J. (2006) Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics* **17**(5), 483–506.

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

Risser, M. D. and Calder, C. A. (2017) Local likelihood estimation for covariance functions with spatially-varying parameters: the convoSPAT package for R. *Journal of Statistical Software* **81**(14), 1–32.

Røislien, J. and Omre, H. (2006) T-distributed random fields: A parametric model for heavy-tailedwell-log data1. *Mathematical Geology* **38**(7), 821–849.

Rootzén, H., Segers, J. and Wadsworth, J. L. (2018) Multivariate generalized pareto distributions: Parametrizations, representations, and properties. *Journal of Multivariate Analysis* **165**, 117–131.

Scholz, F. W. and Stephens, M. A. (1987) K-sample anderson–darling tests. *Journal of the American Statistical Association* **82**(399), 918–924.

Sklar, M. (1959) Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris* **8**, 229–231.

Stein, M. L. (1999) *Interpolation of Spatial Data: Some Theory for Kriging*. First edition. New York: Springer.

Stein, M. L. (2005) Nonstationary Spatial Covariance Functions. Unpublished.

Sun, Y. and Genton, M. G. (2011) Functional boxplots. *Journal of Computational and Graphical Statistics* **20**(2), 316–334.

Thibaud, E., Mutzner, R. and Davison, A. C. (2013) Threshold modeling of extreme spatial rainfall. *Water Resources Research* **49**(8), 4633–4644.

Thibaud, E. and Opitz, T. (2015) Efficient inference and simulation for elliptical Pareto processes. *Biometrika* **102**(4), 855–870.

Vettori, S., Huser, R. and Genton, M. G. (2019) Bayesian modeling of air pollution extremes using nested multivariate max-stable processes. *Biometrics* To appear.

Viglione, A., Laio, F. and Claps, P. (2007) A comparison of homogeneity tests for regional frequency analysis. *Water Resources Research* **43**(3).

Wadsworth, J. L. and Tawn, J. A. (2012) Dependence modelling for spatial extremes. *Biometrika* **99**(2), 253–272.

Wadsworth, J. L. and Tawn, J. A. (2014) Efficient inference for spatial extreme value processes associated to log-Gaussian random functions. *Biometrika* **101**(1), 1–15.

Westra, S., Alexander, L. V. and Zwiers, F. W. (2013) Global increasing trends in annual maximum daily precipitation. *Journal of Climate* **26**(11), 3904–3918.

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

Wilby, R. L., Wigley, T., Conway, D., Jones, P., Hewitson, B., Main, J. and Wilks, D. (1998) Statistical downscaling of general circulation model output: A comparison of methods. *Water resources research* **34**(11), 2995–3008.

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