Taylor & Francis
Taylor & Francis Group

Check for updates

Hierarchical Space-Time Modeling of Asymptotically Independent Exceedances With an Application to Precipitation Data

Jean-Noël Bacro[a], Carlo Gaetan[b], Thomas Opitz[c], and Gwladys Toulemonde[d]

[a]IMAG, Université de Montpellier, CNRS, Montpellier, France; [b]DAIS, Università Ca' Foscari di Venezia, Venice, Italy; [c]BioSP, INRA, Avignon, France; [d]IMAG, Univ Montpellier, CNRS, INRIA, Montpellier, France

ABSTRACT

The statistical modeling of space-time extremes in environmental applications is key to understanding complex dependence structures in original event data and to generating realistic scenarios for impact models. In this context of high-dimensional data, we propose a novel hierarchical model for high threshold exceedances defined over continuous space and time by embedding a space-time Gamma process convolution for the rate of an exponential variable, leading to asymptotic independence in space and time. Its physically motivated anisotropic dependence structure is based on geometric objects moving through space-time according to a velocity vector. We demonstrate that inference based on weighted pairwise likelihood is fast and accurate. The usefulness of our model is illustrated by an application to hourly precipitation data from a study region in Southern France, where it clearly improves on an alternative censored Gaussian space-time random field model. While classical limit models based on threshold-stability fail to appropriately capture relatively fast joint tail decay rates between asymptotic dependence and classical independence, strong empirical evidence from our application and other recent case studies motivates the use of more realistic asymptotic independence models such as ours. Supplementary materials for this article, including a standardized description of the materials available for reproducing the work, are available as an online supplement.

1. Introduction

The French Mediterranean area is subject to heavy rainfall events occurring mainly in the fall season. Intense precipitation often leads to flash floods, which can be defined as a sudden strong rise of the water level. Flash floods often cause fatalities and important material damage. In the literature, such intense rainfalls are often called flood-risk rainfall (Carreau and Bouvier 2016); characterizing their spatio-temporal dependencies is key to understanding these processes. In this article, we consider a large dataset of hourly precipitation measurements from a study region in Southern France. We tackle the challenge of proposing a physically interpretable statistical space-time model for high threshold exceedances, which aims to capture the complex dependence and time dynamics of the data process.

Fueled by important environmental applications during the last decade, the statistical modeling of spatial extremes has undergone a fast evolution. A shift from maxima-based modeling to approaches using threshold exceedances can be observed over recent years, whose reasons lie in the capability of thresholding techniques to exploit more information from the data and to explicitly model the original event data. A first overview of approaches to modeling maxima is due to Davison, Padoan, and Ribatet (2012). A number of hierarchical models based on latent Gaussian processes (Casson and Coles 1999; Gaetan and Grigoletto 2007; Cooley, Nychka, and Naveau 2007;

Sang and Gelfand 2009) have been proposed, but they may be criticized for relying on the rather rigid Gaussian dependence with very weak dependence in the tail, while the lack of closed-form marginal distributions makes interpretation difficult and frequentist inference cumbersome. Max-stable random fields (Smith 1990; Schlather 2002; Kabluchko, Schlather, and de Haan 2009; Davison and Gholamrezaee 2012; Reich and Shaby 2012; Opitz 2013) are the natural limit models for maxima data and have spawned a very rich literature, from which the model of Reich and Shaby (2012) stands out for its hierarchical construction simplifying high-dimensional multivariate calculations and Bayesian inference. Generalized Pareto (GP) processes (Ferreira and de Haan 2014; Opitz, Bacro, and Ribereau 2015; Thibaud and Opitz 2015) are the equivalent limit models for threshold exceedances. However, the asymptotic dependence stability in these limiting processes for maxima and threshold exceedances has a tendency to be overly restrictive when asymptotic dependence strength decreases at high levels and may vanish ultimately in the case of asymptotic independence. The results from the empirical spatio-temporal exploration of our French rainfall data in Section 6.2 are strongly in favor of asymptotic independence, which appears to be characteristic for many environmental datasets (Davison, Huser, and Thibaud 2013; Thibaud, Mutzner, and Davison 2013; Tawn et al. 2018) and may arise from physical laws such as the conservation of mass. This

has motivated the development of more flexible dependence models, such as max-mixtures of max-stable and asymptotically independent processes (Wadsworth and Tawn 2012; Bacro, Gaetan, and Toulemonde 2016) or max-infinitely divisible constructions (Huser, Opitz, and Thibaud 2018) for maxima data, or Gaussian scale mixture processes (Opitz 2016; Huser, Opitz, and Thibaud 2017) for threshold exceedances, capable to accommodate asymptotic dependence, asymptotic independence and Gaussian dependence with a smooth transition. Other flexible spatial constructions involve marginally transformed Gaussian processes (Huser and Wadsworth 2019). Such threshold models can be viewed as part of the wider class of copula models (see Bortot, Coles, and Tawn 2000; Davison, Huser, and Thibaud 2013, for other examples) typically combining univariate limit distributions with dependence structures that should ideally be flexible and relatively easy to handle in practice.

Statistical inference is then often carried out assuming temporal independence in measurements typically observed at spatial sites at regularly spaced time intervals. However, developing flexible space-time modeling for extremes is crucial for characterizing the temporal persistence of extreme events spanning several time steps; such models are important for short-term prediction in applications such as the forecasting of wind power and atmospheric pollution, and for extreme event scenario generators providing inputs to impact models, for instance in hydrology and agriculture. Early spatio-temporal models for rainfall were proposed in the 1980s (Rodriguez-Iturbe, Cox, and Isham 1987; Cox and Isham 1988, and the references therein) and exploit the idea that storm events give rise to a cluster of rain cells, which are represented as cylinders in space-time. Currently, only few statistical space-time models for extremes are available. Davis and Mikosch (2008) considered extremal properties of heavy-tailed moving average processes where coefficients and the white-noise process depend on space and time, but their work was not focused on practical modeling. Sang and Gelfand (2009) proposed a hierarchical procedure for maxima data but limited to latent Gauss–Markov random fields. Davis, Klüppelberg, and Steinkohl (2013a, 2013b) extended the widely used class of Brown–Resnick max-stable processes to the space-time framework and propose pairwise likelihood inference. Spatial max-stable processes with random set elements have been proposed by Schlather (2002) and Davison and Gholamrezaee (2012), and Huser and Davison (2014) have fitted a space-time version to threshold exceedances of hourly rainfall data through pairwise censored likelihood. Huser and Davison (2014) modeled storms as discs of random radius moving at a random velocity for a random duration, leading to randomly centered space-time cylinders; our models developed in the following rely on similar geometric representations. A Bayesian approach based on spatial skew-t random fields with a random set element and temporal autoregression was proposed by Morris et al. (2017). The aforementioned space-time models may capture asymptotic dependence or exact independence at small distances but are unsuitable for dealing with residual dependence in asymptotic independence. In this article, we propose a novel approach to space-time modeling of asymptotically independent data to avoid the tendency of max-stable-like models to potentially strongly overestimate joint extreme risks. In a similar context,

Nieto-Barajas and Huerta (2017) recently proposed a spatio-temporal Pareto model for heavy-tailed data on spatial lattices, generalizing the temporal latent process model of Bortot and Gaetan (2014) to space-time.

Our model provides a hierarchical formulation for modeling spatio-temporal exceedances over high thresholds. It is defined over a continuous space-time domain and allows for a physical interpretation of extreme events spreading over space and time. Strong motivation also comes from Bortot and Gaetan (2014) by developing a generalization of their latent temporal process. Alternatively, our latent process could be viewed as a space-time version of the temporal trawl processes introduced by Barndorff-Nielsen et al. (2014) and exploited for extreme values by Noven, Veraart, and Gandy (2015), with spatial extensions recently proposed by Opitz (2017). Our approach is based on representing a GP distribution as a Gamma mixture of an exponential distribution, enabling us to keep easily tractable marginal distributions which remain coherent with univariate extreme value theory. We use a kernel convolution of a space-time Gamma random process (Wolpert and Ickstadt 1998a) based on influence zones defined as cylinders with an ellipsoidal basis to generate anisotropic spatio-temporal dependence in exceedances. We then develop statistical inference based on pairwise likelihood.

The article is structured as follows. Our hierarchical model with a detailed description of its two stages and marginal transformations is developed in Section 2. Space-time Gamma random fields are presented in Section 2.1 where we also propose the construction and formulas for the space-time objects used for kernel convolution. Section 3 characterizes tail dependence behavior in our new model yielding asymptotic independence in space and time. Statistical inference of model parameters is addressed in Section 4 based on a pairwise log-likelihood for the observed censored excesses. We show good estimation performance through a simulation study presented in Section 5 involving two scenarios of different complexity. In Section 6, we focus on the dataset and explore in detail how our fitted space-time model captures spatio-temporal extremal dependence in hourly precipitation. Since a natural choice of a reference model for asymptotically independent data is to use threshold-censored space-time Gaussian processes, we show the good relative performance of our model by comparing it to such alternatives. A discussion of our modeling approach with some potential future extensions closes the article in Section 7.

2. A Hierarchical Model for Spatio-Temporal Exceedances

When dealing with exceedances of a random variable X above a high threshold u, univariate extreme value theory suggests using the limit distribution of GP type. The GP cumulative distribution function (cdf) is defined for any $y > 0$ by

$$\mathrm{GP}(y; \sigma, \xi) = 1 - \left(1 + \xi \frac{y}{\sigma}\right)_+^{-(1/\xi)}, \qquad (1)$$

where $(a)_+ = \max(0, a)$, ξ is a shape parameter and σ a positive scale parameter. The sign of ξ characterizes the maximum domain of attraction of the distribution of X: $\xi > 0$ corresponds

to the Fréchet domain of attraction while $\xi = 0$ and $\xi < 0$ correspond to the Gumbel and Weibull ones, respectively.

When $\xi > 0$, the GP distribution can be expressed as a Gamma mixture for the rate of the exponential distribution (Reiss and Thomas 2007, p. 157), that is,

$$V|\Lambda \sim \text{Exp}(\Lambda), \quad \Lambda \sim \text{Gamma}(1/\xi, \sigma/\xi)$$
$$\Rightarrow \quad V \sim \text{GP}(\,\cdot\,; \sigma, \xi), \tag{2}$$

where $\text{Exp}(b)$ refers to the Exponential distribution with rate $b > 0$ and $\text{Gamma}(a, b)$ to the Gamma distribution with shape $a > 0$ and rate $b > 0$. Based on this hierarchical structure, we will here develop a stationary space-time construction for modeling exceedances over a high threshold, which possesses marginal GP distributions for the strictly positive excess above the threshold.

2.1. First Stage: Generic Hierarchical Space-Time Structure

We consider a stationary space-time random field $Z = \{Z(x), x \in \mathcal{X}\}$ with $x = (s, t)$ and $\mathcal{X} = \mathbb{R}^2 \times \mathbb{R}^+$, such that s indicates spatial location and t time. Without loss of generality, we assume that the margins $Z(x)$ belong to the Fréchet domain of attraction with positive shape parameter ξ. To infer the tail behavior of $\{Z(x)\}$, we focus on values exceeding a fixed high threshold u, and we consider the exceedances over u,

$$Y(x) = (Z(x) - u) \cdot \mathbf{1}_{(u,\infty)}(Z(x)). \tag{3}$$

Standard results from extreme value theory (de Haan and Ferreira 2006) establish the *GP* distribution with $\xi > 0$ in (1) as the limit of suitably renormalized positive threshold exceedances in (3), such that it represents a natural model for the values $Y(x) > 0$. Following Bortot and Gaetan (2014), we use the representation of the GP distribution as a Gamma mixture of an exponential distribution to formulate a two-stage model that induces spatio-temporal dependence arising in both the exceedance indicators $\mathbf{1}_{(u,\infty)}(Z(x))$ and the positive excesses $Z(x) - u > 0$ by integrating space-time dependence in a latent Gamma component. A key feature of our model is that it naturally links exceedance probability to the size of the excess and therefore provides a joint space-time structure of the zero part and the positive part in the zero-inflated distribution of $Y(x)$.

In the first stage, we condition on a latent space-time random field $\{\Lambda(x)\}$ with marginal distributions $\Lambda(x) \sim \text{Gamma}(\alpha, \beta)$ and assume that

$$Y(x) \mid [\Lambda(x), Y(x) > 0] \sim \text{Exp}(\Lambda(x)), \tag{4a}$$
$$\Pr(Y(x) > 0 \mid \Lambda(x)) = e^{-\kappa \Lambda(x)}, \tag{4b}$$

where $\kappa > 0$ is a parameter controlling the rate of upcrossings of the threshold. The resulting marginal distribution of $Y(x)$ conditionally on $Z(x) > u$ corresponds to the GP distribution, and the unconditional marginal cdf of $Y(x)$ is

$$F(y; \sigma, \xi) = \begin{cases} p & \text{for } y = 0, \\ p + (1-p)\text{GP}(y; \xi, \sigma) & \text{for } y > 0, \end{cases} \tag{5}$$

with shape parameter $\xi = 1/\alpha$, scale parameter $\sigma = (\kappa + \beta)/\alpha$, and with $1 - p$ the probability of an exceedance over u, that is,

$\Pr(Z(x) > u) = \Pr(Y(x) > 0) = 1 - p$. The probability of exceeding u,

$$\Pr(Z(x) > u) = \text{E}\left(\Pr(Y(x) > 0|\Lambda(x))\right) = \text{E}\left(e^{-\kappa \Lambda(x)}\right)$$
$$= \left(\frac{\beta}{\kappa + \beta}\right)^\alpha \tag{6}$$

depends on κ and corresponds to the Laplace transform of $\Lambda(x)$ evaluated at κ. The constraint $\xi > 0$ is not restrictive for dealing with precipitation in the French Mediterranean area, which is known to be heavy-tailed. For general modeling purposes, we can relax this assumption by following Bortot and Gaetan (2016): we consider a marginal transformation within the class of GP distributions for threshold exceedances, for which we suppose that $\alpha = 1$ and $\beta = 1$ for identifiability. By transforming $Y(x)$ through the probability integral transform

$$g(y) = \text{GP}^{-1}(\text{GP}(y; 1, 1 + \kappa); \sigma^*, \xi^*) \tag{7}$$
$$= (\sigma^*/\xi^*)\left\{\left(1 + \frac{y}{\kappa + 1}\right)^{\xi^*} - 1\right\} \tag{8}$$

with parameters $\xi^* \in \mathbb{R}$ and $\sigma^* > 0$ to be estimated, we get a marginally transformed random field $Y^*(x) = g(Y(x))$ which satisfies $Y^*(x) \sim \text{GP}(\,\cdot\,; \xi^*, \sigma^*)$, conditionally on $Y^*(x) > 0$. Notice that it is straightforward to develop extensions with nonstationary marginal excess distributions by injecting response surfaces $\sigma^\star(x)$ and $\xi^\star(x)$ into (8). Moreover, nonstationarity could be introduced into the latent Gamma model (4) in different ways. If $\kappa = \kappa(x)$ depends on x or other covariate information, exceedance probabilities become nonstationary. If Gamma parameters $\alpha = \alpha(x)$ and $\beta = \beta(x)$ depend on covariates, then the GP margins in $Y(x)$ become nonstationary. Finally, one could combine the two previous nonstationary extensions.

2.2. Second Stage: Space-Time Dependence With Gamma Random Fields

Spatio-temporal dependence is introduced by means of a space-time stationary random field $\{\Lambda(x), x \in \mathcal{X}\}$ with $\text{Gamma}(\alpha, \beta)$ marginal distributions. In principle, we could use an arbitrarily wide range of models with any kind of space-time dependence structure, for instance by marginally transforming a space-time Gaussian random field using the copula idea (Joe 1997). However, we here aim to propose a construction where Gamma marginal distributions arise naturally without applying rather artificial marginal transformations. Inspired by the Gamma process convolutions of Wolpert and Ickstadt (1998a), we develop a space-time Gamma convolution process with Gamma marginal distributions. The kernel shape in our construction allows for a straightforward interpretation of the dependence structure, and it offers a physical interpretation of real phenomena such as mass and particle transport. Moreover, we obtain simple analytical formulas for the bivariate distributions, which facilitates statistical inference, interpretation and the characterization of joint tail properties.

We fix $\mathcal{X} = \mathbb{R}^3$ and consider $A \in \mathcal{B}_b(\mathcal{X})$, a subset of $\mathcal{X}$ belonging to the σ-field $\mathcal{B}_b(\mathcal{X})$ restricted to bounded sets of $\mathcal{X}$. A Gamma random field $\Gamma(dx)$ (Ferguson 1973) is a nonnegative random measure defined on $\mathcal{X}$ characterized by a base measure $\alpha(dx)$ and a rate parameter β such that
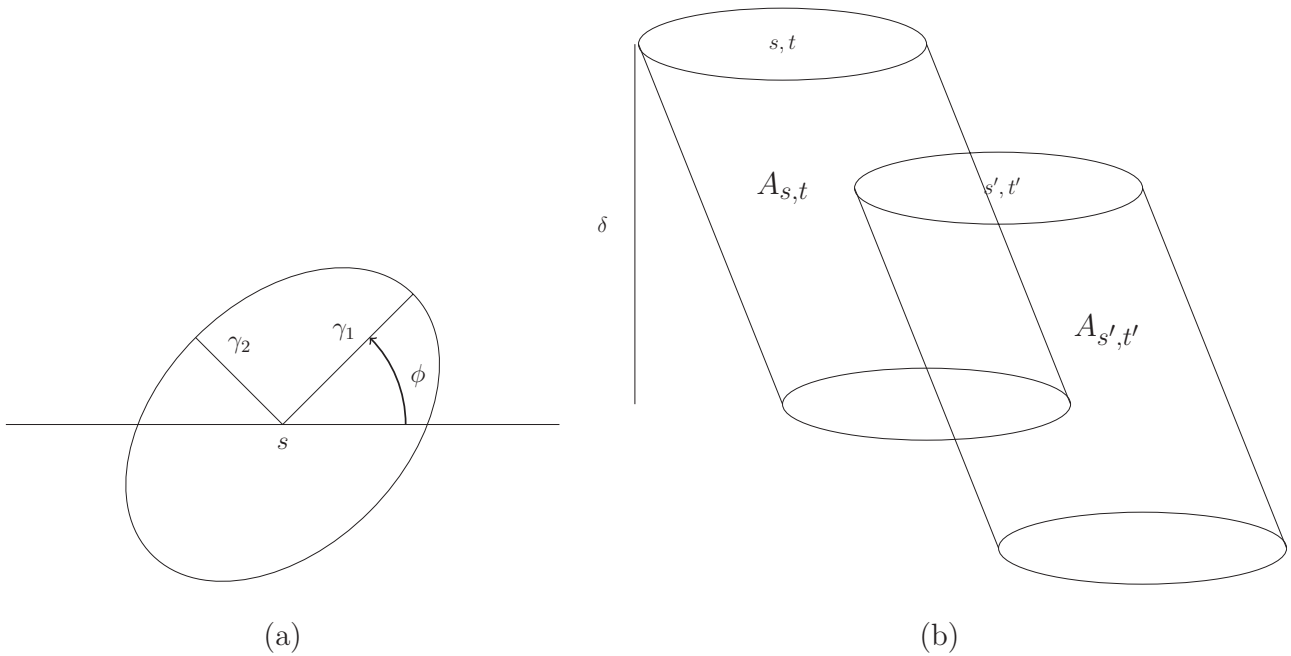
Figure 1. Space-time kernels. Left display: a spatial ellipse $E(s, \gamma_1, \gamma_2, \phi)$ centered at s. Right display: an intersection of two slated elliptical cylinders $A_{s,t}$ and $A_{s',t'}$ with duration δ.

1. $\Gamma(A) := \int_A \Gamma(dx) \sim \text{Gamma}(\alpha(A), \beta)$, with $\alpha(A) := \int_A \alpha(dx)$;
2. for any $A_1, A_2 \in \mathcal{B}_b(\mathcal{X})$ such that $A_1 \cap A_2 = \emptyset$, $\Gamma(A_1)$ and $\Gamma(A_2)$ are independent random variables.

The calculation of important formulas in this article requires the Laplace exponent of the random measure given as

$$\mathcal{L}(\phi) := -\log \text{E}\left(\exp\left\{-\int \phi(x)\Gamma(dx)\right\}\right)$$
$$= \int_{\mathcal{X}} \log\left\{1 + \frac{\phi(x)}{\beta}\right\} \alpha(dx),$$

where ϕ is any positive measurable function; in our case, it will represent the kernel function (see Appendix A). We propose to model $\{\Lambda(x), x \in \mathcal{X}\}$ as a convolution using a three-dimensional indicator kernel $K(x, x')$ with an indicator set of finite volume used to convolve the Gamma random field $\Gamma(dx)$ (Wolpert and Ickstadt 1998a), that is, $\Lambda(x) = \int K(x, x')\Gamma(dx')$. The shape of the kernel can be very general (although nonindicator kernels usually do not lead to Gamma marginal distributions), and particular choices may lead to nonstationary random fields, or to stationary random fields with given dependence properties such as full symmetry, separability or independence beyond some spatial distance or temporal lag. To limit model complexity and computational burden to a reasonable amount, we use the indicator kernel $K(x, x') = \mathbf{1}_A(x - x')$, for $A \in \mathcal{B}_b(\mathcal{X})$, where A is given as a slated elliptical cylinder, defining a three-dimensional set A_x that moves through $\mathcal{X}$ according to some velocity vector. More precisely, let $E(s, \gamma_1, \gamma_2, \phi)$ be an ellipse centered at $s = (s_1, s_2) \in \mathbb{R}^2$ (see Figure 1(a)), whose axes are rotated counterclockwise by the angle ϕ with respect to the coordinate axes, whose semi-axes' lengths in the rotated coordinate system are γ_1 and γ_2, respectively. A physical interpretation is that the ellipse describes the spatial influence zone of a storm centered at s. For the temporal dynamics, we

assume that the ellipses (storms) $E(s, \gamma_1, \gamma_2, \phi)$ move through space with a velocity $\omega = (\omega_1, \omega_2) \in \mathbb{R}^2$ for a duration $\delta > 0$. The volume of the intersection of two slated elliptical cylinders (see Figure 1(b)) is given by

$$V(s, t, s', t') = (\delta - |t - t'|)_+ \times \nu_2(E(s, \gamma_1, \gamma_2, \phi) \cap E(\tilde{s}, , \gamma_1, \gamma_2, \phi)),$$

where $\tilde{s} = (\tilde{s}_1, \tilde{s}_2)$ with $\tilde{s}_i = s'_i - |t' - t| \times \omega_i$, $i = 1, 2$, where $\nu_d(\cdot)$ is the Lebesgue measure on $\mathbb{R}^d$.

For two fixed locations, the strength of dependence in the random field $\Lambda(x)$ is an increasing monotone function of the intersection volume; other choices of A are possible, provided that we are able to calculate efficiently the volume of the intersection. To efficiently calculate the ellipse intersection area, we use an approach for finding the overlap area between two ellipses, which does not rely on proxy curves; see Hughes and Chraibi (2012).[1]

In the sequel, we consider the measure

$$\alpha(B) = \alpha \nu_d(B)/\nu_d(A), \quad B \in \mathcal{B}_b(\mathcal{X}). \tag{9}$$

It follows that $\Lambda(x) \sim \text{Gamma}(\alpha, \beta)$, as required for model (4). Exploiting the formulas of Appendix A, the univariate Laplace transform of $\Lambda(x)$ is

$$\text{LP}_x^{(1)}(\nu) := \text{E}\left(e^{-\nu\Lambda(x)}\right) = \left(\frac{\beta}{\nu + \beta}\right)^\alpha, \tag{10}$$

and the bivariate Laplace transform of $\Lambda(x)$ and $\Lambda(x')$ is

$$\text{LP}_{x,x'}^{(2)}(\nu_1, \nu_2) := \text{E}\left(e^{-\nu_1\Lambda(x) - \nu_2\Lambda(x')}\right)$$
$$= \left(\frac{\beta}{\nu_1 + \beta}\right)^{\alpha(A_x \setminus A_{x'})} \left(\frac{\beta}{\nu_1 + \nu_2 + \beta}\right)^{\alpha(A_x \cap A_{x'})}$$
$$\times \left(\frac{\beta}{\nu_2 + \beta}\right)^{\alpha(A_{x'} \setminus A_x)}. \tag{11}$$

---

[1]The code is open source and can be downloaded from *http://github.com/ chraibi/EEOver*.

This model for $\Lambda(x)$ is stationary, but nonstationarity in Gamma marginal distributions and/or dependence can be generated by using nonstationary indicator sets A_x whose size and shape depends on x. More general sets A_x with finite Lebesgue volume $\nu_3(A_x)$ could be used for constructing $\Lambda(x) = \Gamma(A_x)$. In all cases, the intersecting volume $\nu_3(A_{x_1} \cap A_{x_2})$ tends to zero if $\|x_2 - x_1\| \to \infty$, which establishes the property of α-mixing over space and time for the processes $\Lambda(x)$ and $Y(x)$. This property is paramount to ensure consistency and asymptotic normality in the pairwise likelihood estimation that we consider in the following (see Huser and Davison 2014).

3. Joint Tail Behavior of the Hierarchical Process

Extremal dependence in a bivariate random vector (Z_1, Z_2) can be explored based on the tail behavior of the conditional distribution $\Pr(Z_1 > F_1^\leftarrow(q)|Z_2 > F_2^\leftarrow(q))$ as q tends to 1, where $F_i^\leftarrow$, $i = 1, 2$ denotes the generalized inverse distribution functions of Z_i (Sibuya 1960; Coles, Heffernan, and Tawn 1999). The random vector (Z_1, Z_2) is said to be asymptotically dependent if a positive limit χ, referred to as the tail correlation coefficient, arises

$$\chi(q) := \frac{\Pr(Z_1 > F_1^\leftarrow(q), Z_2 > F_2^\leftarrow(q))}{\Pr(Z_2 > F_2^\leftarrow(q))} \to \chi > 0,$$
$$q \to 1^-.$$

The case $\chi = 0$ characterizes asymptotic independence.

To obtain a finer characterization of the joint tail decay rate under asymptotic independence, faster than the marginal tail decay rate, Coles, Heffernan, and Tawn (1999) introduced the $\overline{\chi}$ index defined through the limit relation

$$\overline{\chi}(q) := \frac{2 \log \Pr(Z_2 > F_2^\leftarrow(q))}{\log \Pr(Z_1 > F_1^\leftarrow(q), Z_2 > F_2^\leftarrow(q))}$$
$$- 1 \to \bar{\chi} \in (-1, 1], \qquad q \to 1^-.$$

Larger values of $|\bar{\chi}|$ correspond to stronger dependence. We now show that $\{Z(x), x \in \mathcal{X}\}$ is an asymptotic independent process, that is, for all pairs $(x, x') \in \mathcal{X}^2$ with $x \neq x'$ the bivariate random vectors $(Z(x), Z(x'))$ are asymptotically independent.

Owing to the stationarity of the process, it is easy to show that for any $(x, x') \in \mathcal{X}^2$, $x \neq x'$ and for values ν exceeding a threshold $u \geq 0$, we get

$$\Pr(Z(x) > \nu) = \mathrm{LP}_x^{(1)}(\nu - u + \kappa)$$
$$= \left(1 + \frac{\nu - u + \kappa}{\beta}\right)^{-\alpha(A_x)}$$

and

$$\Pr(Z(x) > \nu, Z(x') > \nu) = \mathrm{LP}_{x,x'}^{(2)}(\nu - u + \kappa, \nu - u + \kappa)$$
$$= \left(1 + \frac{\nu - u + \kappa}{\beta}\right)^{-\alpha(A_x \backslash A_{x'})}$$
$$\times \left(1 + \frac{2\nu - 2u + 2\kappa}{\beta}\right)^{-\alpha(A_x \cap A_{x'})}$$
$$\times \left(1 + \frac{\nu - u + \kappa}{\beta}\right)^{-\alpha(A_{x'} \backslash A_x)}.$$

To simplify notations, we set $c_0 := \alpha(A_x)$, $c_1 := \alpha(A_x \backslash A_{x'})$, $c_2 := \alpha(A_x \cap A_{x'})$, $c_3 := \alpha(A_{x'} \backslash A_x)$, such that $c_1 = c_3 = c_0 - c_2 \geq 0$ and $c_1 + 2c_2 + c_3 = 2c_0$. For $c_2 = 0$ characterizing disjoint indicator sets A_x and $A_{x'}$, it is clear that $Z(x)$ and $Z(x')$ are independent. Now, assume $u = 0$ without loss of generality and $x \neq x'$; then,

$$\chi_{x,x'}(\nu) := \frac{\Pr(Z(x) > \nu, Z(x') > \nu)}{\Pr(Z(x') > \nu)}$$
$$= \left(1 + \frac{2\nu + 2\kappa}{\beta}\right)^{-c_2} \left(1 + \frac{\nu + \kappa}{\beta}\right)^{-c_1 - c_3 + c_0}$$
$$= \left(1 + \frac{2\nu + 2\kappa}{\beta}\right)^{-c_2} \left(1 + \frac{\nu + \kappa}{\beta}\right)^{2c_2 - c_0}$$
$$\sim 2^{-c_2} \left(\frac{\nu}{\beta}\right)^{c_2 - c_0}, \qquad \text{for large } \nu.$$

Since $c_2 < c_0$, we obtain

$$\chi_{x,x'} = 0.$$

We conclude that Z is an asymptotic independent process.

To characterize the faster joint tail decay, we calculate

$$\bar{\chi}_{xx'}(\nu)$$
$$:= \frac{2 \log \Pr(Z(x) > \nu)}{\log \Pr(Z(x) > \nu, Z(x') > \nu)} - 1$$
$$= \frac{-2c_0 \log\left(1 + (\nu + \kappa)/\beta\right)}{-c_1 \log(1 + (\nu + k)/\beta) - c_2 \log\left(1 + 2(\nu + k)/\beta\right)} - 1$$
$$\qquad\qquad -c_3 \log(1 + (\nu + k)/\beta)$$
$$= \frac{2c_0}{c_1 + c_2 \dfrac{\log\left(1 + 2(\nu + k)/\beta\right)}{\log\left(1 + (\nu + k)/\beta\right)} + c_3} - 1.$$

Taking the limit for $\nu \to \infty$ yields

$$\bar{\chi}_{x,x'} = \frac{2c_0}{c_1 + c_2 + c_3} - 1 = \frac{c_2}{2c_0 - c_2},$$

which describes the ratio between the intersecting volume of A_x and $A_{x'}$ and the volume of the union of these two sets. The value of $\bar{\chi}$ confirms the asymptotic independence of the process Z. A larger intersecting volume between A_x and $A_{x'}$ corresponds to stronger dependence.

4. Composite Likelihood Inference

To infer the tail behavior of the observed data process $\{Z(x)\}$, without loss of generality assumed to have GP marginal distributions with shape parameter α, we focus on values exceeding a fixed high threshold u. We let θ denote the vector of unknown parameters. For simplicity, we assume that we have observed the excess values $Y(s_i, t)$ for a factorial design of S locations s_i, $i = 1, \ldots, S$ and T times $t = 1, \ldots, T$.

To exploit the tractability of intersecting volumes of two kernel sets, we focus on pairwise likelihood for efficient inference in our high-dimensional space-time set-up. The

pairwise (weighted) log-likelihood adds up the contributions $f(Y(s_i, t), Y(s_j, t + k); \theta)$ of the censored observations $Y(s_i, t), Y(s_j, t + k)$ and can be written

$$\text{pl}(\theta) = \sum_{t=1}^{T} \text{pl}_t(\theta) = \sum_{t=1}^{T} \sum_{k=0}^{\Delta_T} \sum_{i=1}^{S} \sum_{j=1}^{S} \{1 - 1_{\{i \geq j, k=0\}}\}$$
$$\times \log f(Y(s_i, t), Y(s_j, t + k); \theta) w_{s_i, s_j}, \quad (12)$$

where w_{s_i, s_j} is a weight defined on $[0, \infty)$ (Bevilacqua et al. 2012; Davis, Klüppelberg, and Steinkohl 2013b; Huser and Davison 2014). We opt for a cut-off weight with $w_{s_i, s_j} = 1$ if $||s_i - s_j|| \leq \Delta_S$ and 0 otherwise, which bypasses an explosion of the number of likelihood terms and shifts focus to relatively short-range distances where dependence matters most. This also avoids that the pairwise likelihood value (and therefore parameter estimation) is dominated by a large number of intermediate-range distances where dependence has already decayed to (almost) nil.

The contributions $f(Y(x), Y(x'); \theta)$ are given by

$f(y_1, y_2; \theta)$

$$= \begin{cases} \frac{\partial^2}{\partial v_1 \partial v_2} \text{LP}_{x,x'}^{(2)}(v_1, v_2) J(y_1) J(y_2) & y_1 > 0, y_2 > 0 \\[2mm] \left(-\frac{\partial}{\partial v_1} \text{LP}^{(1)}(v_1) + \frac{\partial}{\partial v_1} \text{LP}_{x,x'}^{(2)}(v_1, v_2)\right) J(y_1) & y_1 > 0, y_2 = 0 \\[2mm] \left(-\frac{\partial}{\partial v_2} \text{LP}^{(1)}(v_2) + \frac{\partial}{\partial v_2} \text{LP}_{x,x'}^{(2)}(v_1, v_2)\right) J(y_2) & y_1 = 0, y_2 > 0 \\[2mm] 1 - 2\text{LP}^{(1)}(v_1) + \text{LP}_{x,x'}^{(2)}(v_1, v_2) & y_1 = 0, y_2 = 0 \end{cases}$$

with $v_i = (\kappa + 1)\left(1 + \xi^* y_i / \sigma^*\right)^{1/\xi^*} - 1$ and $J(y_i) = \frac{\kappa+1}{\sigma^*}\left(1 + \frac{\xi^* y_i}{\sigma^*}\right)^{1/\xi^*-1}$, $i = 1, 2$. We provide analytical expressions for $\text{LP}^{(1)}$ and $\text{LP}_{x,x'}^{(2)}$ in Appendix B.

Since the space-time random field $\{\Lambda(x)\}$ is temporally α-mixing, the maximum pairwise likelihood estimator $\hat{\theta}$ can be shown to be asymptotically normal for large T under mild additional regularity conditions; see Theorem 1 of Huser and Davison (2014). The asymptotic variance is given by the inverse of the Godambe information matrix $\mathcal{G}(\theta) = \mathcal{H}(\theta)[\mathcal{J}(\theta)]^{-1}\mathcal{H}(\theta)$. Therefore, standard error evaluation requires consistent estimation of the matrices $\mathcal{H}(\theta) = E(-\nabla^2 \text{pl}(\theta))$ and $\mathcal{J}(\theta) = \text{var}(\nabla \text{pl}(\theta))$. We estimate $\mathcal{H}(\theta)$ with $\hat{\mathcal{H}} = -\nabla^2 \text{pl}(\hat{\theta})$ and $\mathcal{J}(\theta)$ through a subsampling technique (Carlstein 1986), implemented as follows. We define B overlapping blocks $D_b \subset \{1, \ldots, T\}$, $b = 1, \ldots, B$, containing d_b observations; we write pl_{D_b} for the pairwise likelihood (12) evaluated over the block D_b. The estimate of $\mathcal{J}(\theta)$ is

$$\hat{\mathcal{J}} = \frac{T}{B} \sum_{b=1}^{B} \frac{1}{d_b} \nabla pl_{D_b}(\hat{\theta}) \nabla pl_{D_b}(\hat{\theta})'.$$

The estimates $\hat{\mathcal{H}}$ and $\hat{\mathcal{J}}$ allow us to calculate the composite likelihood information criterion (Varin and Vidoni 2005)

$$\text{CLIC} = -\text{pl}(\hat{\theta}) + \text{tr}\{\hat{\mathcal{H}}^{-1}\hat{\mathcal{J}}\}$$

with lower values of CLIC indicating a better fit. Similar to Davison and Gholamrezaee (2012), we improve the interpretability of CLIC values through rescaling CLIC*=c CLIC by a positive constant c chosen to give a pairwise log-likelihood value $\text{pl}(\theta)$ comparable to the log-likelihood under independence.

5. Simulation Study

We assess the performance of the pairwise composite likelihood estimator through a small simulation study. For each replication, we consider $S = 30$ randomly chosen sites on $[0, 1] \times [0, 1]$ observed at time points $t = 1, \ldots, T = 2000$. The realizations of the Gamma random field are simulated by adapting the algorithm of Wolpert and Ickstadt (1998b). In the simulations, we fix parameters $\xi = 1$, $\sigma = 10$ and an exceedance probability of $1 - p = 0.2$. We focus on estimating dependence parameters while treating the margins as known. For estimation, we fix the site-dependent threshold u to an empirical quantile of order greater than p. Here, we fix $p = 0.9$ corresponding to $\kappa = 9$.

Two scenarios with different model complexity are considered, involving different specifications of the cylinder (see Table 1). scenario A uses a circle-based cylinder without velocity, while scenario B comes with a slated ellipse-based cylinder, yielding nonnull velocity. Technically, the model in scenario A is over-parameterized since the rotation parameter ϕ cannot change the volume of the cylinder.

Model parameters are estimated on 100 data replications using the composite likelihood approach developed in Section 4. We have considered a larger number of replications for some parameter combinations, but in general the number of 100 replications is enough to satisfactorily illustrate the estimation efficiency. The evaluation of $pl(\theta)$ depends on the choice of Δ_S and Δ_T, where greater values increase the computational cost. Results in the literature indicate that using as much as computationally possible or all of the pairs will not necessarily lead to an improvement in estimation owing to potential issues with estimation variance (see, e.g., Huser and Davison 2014). We have considered different values of Δ_T and have identified $\Delta_T = 15$ as a good compromise for the estimation quality. The parameter Δ_S has been set to 1 which is large enough with respect to the spatial domain limits. Main results are illustrated in the boxplots in Figures 2 and 3.

When the cylinder is circle-based, that is, $\gamma_1 = \gamma_2$, and without velocity (scenario A), the orientation parameter ϕ can take any value. In the simulation experiment we estimate all parameters without constraints, such that the optimization algorithm gives also an estimate of ϕ. It is reassuring to see in the boxplots of Figure 2 that the other parameters are still well estimated.

Results are fairly good for the scenario B where the velocity is nonnull. The estimates of the velocity present slightly higher variability, and the estimation of ω_2 appears slightly biased. On the other hand, the duration δ and the lengths of the semi-axes of the ellipse (γ_1 and γ_2) are still well estimated. The angle ϕ is well defined in scenario B, but it is still estimated with relatively high variability. This may seem as disappointing at first glance, but it may be due to the only moderate difference in the length of

Table 1. Design of the two simulation scenarios.

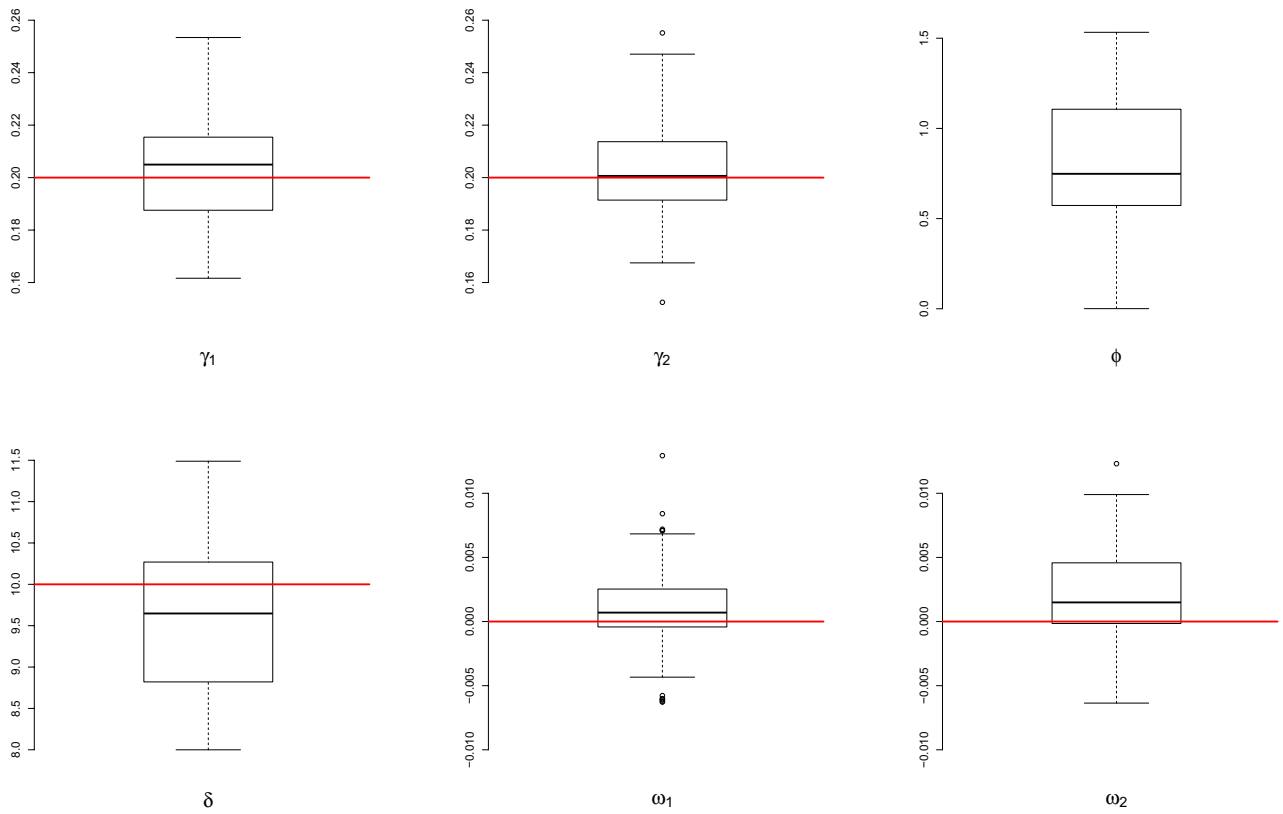| Scenario	Parameters					
	γ_1	γ_2	ϕ	δ	ω_1	ω_2
A	0.2	0.2	–	10	0.00	0.00
B	0.2	0.3	$\pi/4$	5	0.05	0.10

Figure 2. Summary of parameter estimates for scenario A of the simulation study: boxplots of parameter estimates for 100 simulated datasets.
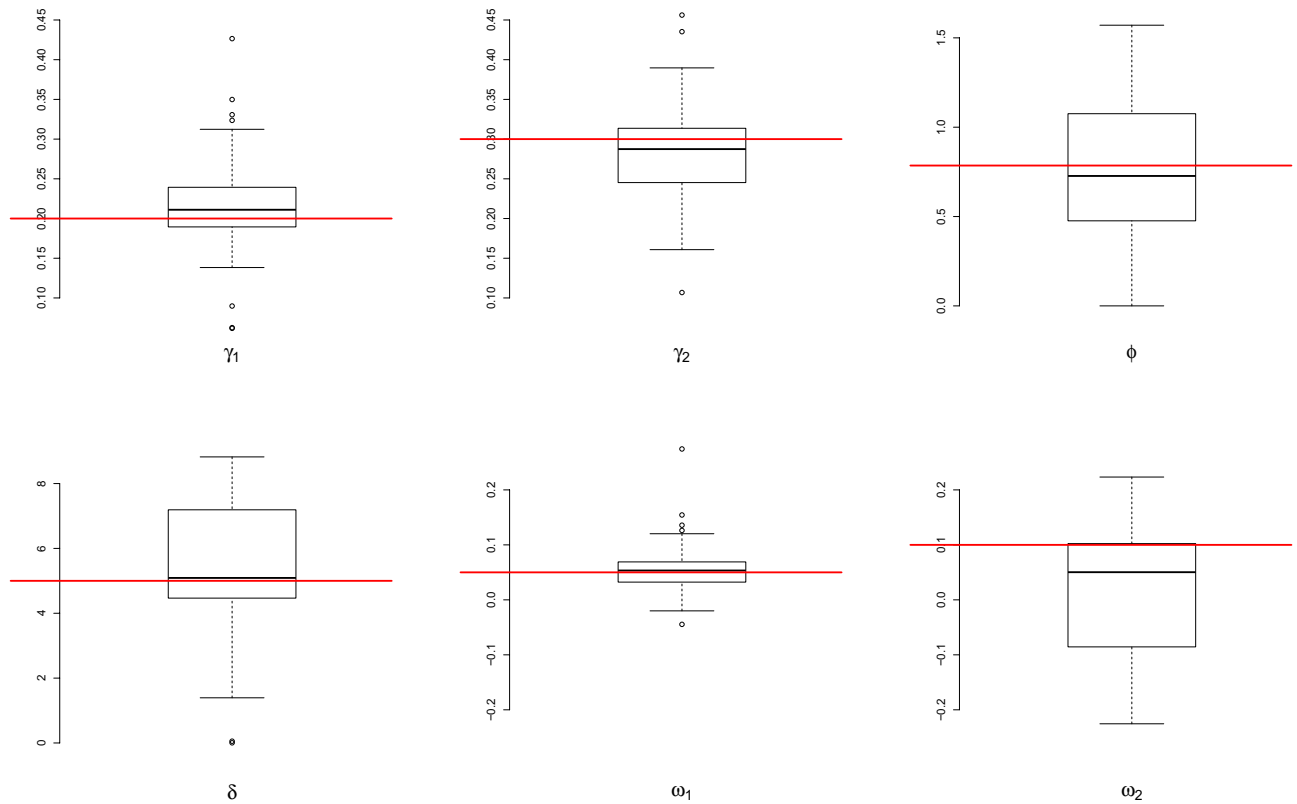

Figure 3. Summary of parameter estimates for scenario B of the simulation study: boxplots of parameter estimates for 100 simulated datasets.

the semi-axes. To check this conjecture, we consider a modified scenario B where the second semi-axis is modified from $\gamma_2 = 0.3$ to $\gamma_2 = 0.5$ and other parameters remain unchanged. As illustrated by the boxplots in Figure 4, estimation of ϕ clearly improves when the shape of the ellipse departs more strongly from a circular shape.
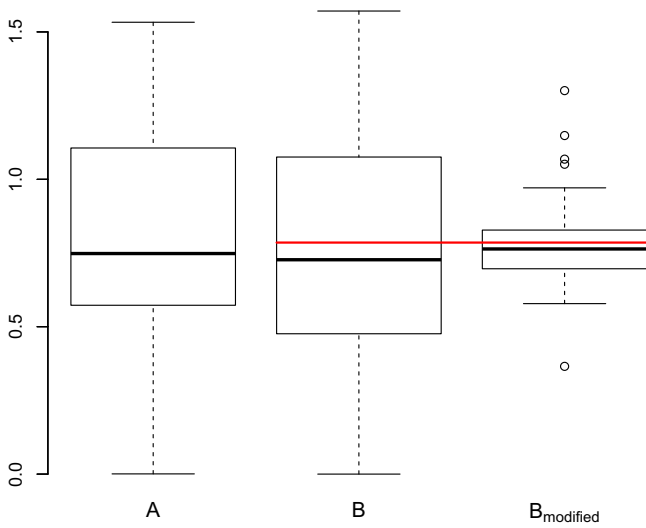
Figure 4. Parameter estimates of the simulation study: boxplots of ϕ estimates according to scenario A, scenario B, and a modified scenario B with $\gamma_2 = 0.5$.

Even with only a relatively small number of spatial sites and time steps, the simulation study shows that the pairwise composite likelihood approach leads to reliable estimates of model parameters that are well identifiable. We underline that results are consistently good whatever the complexity of the scenario.

6. Space-Time Modeling of Hourly Precipitation Data in Southern France

6.1. Data

We apply our hierarchical model to precipitation extremes observed over a study region in the South of France. Extreme rainfall events usually occur during fall season. They are mainly due to southern winds driving warm and moist air from the Mediterranean sea toward the relatively cold mountainous areas of the Cevennes and the Alps, leading to a situation which often provokes severe thunderstorms. The data were provided by Météo France (*https://publitheque.meteo.fr*). Our dataset is part of a query containing hourly observations at 213 rainfall stations for years 1993 to 2014. To avoid modeling complex seasonal trends, we keep only data from the September to November months, resulting in observations over 54,542 hr. For model fitting, we consider a subsample of 50 meteorological stations with elevations ranging from 2 to 1418 m, for which the observation series contain less than 70% of missing values over the full period. The spatial design of the stations is illustrated in Figure 5.

6.2. Exploratory Analysis

We fit the univariate model (5) for each station by fixing a threshold u that corresponds to the empirical 99% quantile. We use such a rather high probability value since we have many observations, and there is a substantial number of zero values such that a high quantile is needed to get into the tail region of the positive values. Figure 5 clearly shows that spatial nonstationarity arises in the marginal distributions.

Figure 6 displays the results of a bootstrap procedure in which we calculate estimates of $\chi(q)$ and $\bar{\chi}(q)$ for probabilities $q = 0.99, 0.995$ for pairs $Z(s, t)$, $Z(s, t + h)$ with only temporal lag, and for pairs $Z(s, t)$, $Z(s', t)$ with only spatial lag. The curves for spatial lags are the result of a smoothing procedure. Confidence bands are based on 200 bootstrap samples, drawn by the stationary bootstrap (Politis and Romano 1994). Our procedure samples temporal blocks of observations and the block length follows a geometric distribution with an average of 20 days. These plots support the assumption of asymptotic independence at all positive distances and at all positive temporal lags. Moreover, the strength of tail dependence as measured by the subasymptotic tail correlation value $\chi(q)$ strongly decreases when considering exceedances over increasingly high thresholds, which provides another clear sign of continuously decreasing and ultimately vanishing dependence strength. On the other hand, the values of the subasymptotic dependence measure $\bar{\chi}(q)$ (well adapted to asymptotic independence) decrease with increasing spatial distances or temporal lags, but they tend to stabilize at a nonzero value. This behavior indicates the presence of residual tail dependence that vanishes only asymptotically.

6.3. Modeling Spatio-Temporal Dependence

While the preceding exploratory analysis has shown that marginal distributions are not stationary, our model detailed in Section 2 requires a specific type of common marginal distributions. It would indeed be possible to extend the model to accommodate nonstationary patterns (an example can be found in Bortot and Gaetan (2016)) and to jointly estimate marginal and dependence parameters. However, our focus here is to illustrate that our modeling strategy is capable to capture complex stationary spatio-temporal dependence patterns at large values, which would render joint estimation of margins and dependence highly intricate. Therefore, we fit a GP distribution separately to each site with thresholds chosen as the empirical 99% quantile. With respect to positive precipitation, this quantile globally corresponds to a probability of 0.91, with a minimum of 0.86 and maximum of 0.95 over the 50 sites. Next, we use the estimated parameters $\hat{\xi}$ and $\hat{\sigma}$ to transform the raw exceedances $Y(x)$ observed at site x to exceedances $\tilde{Y}(x)$ with cdf (5) such that $\xi = 1$ and $\sigma = \kappa + 1$, that is,

$$\tilde{Y}(x) = (\kappa + 1) \left\{ \left(1 + \frac{\hat{\xi}\, Y(x)}{\hat{\sigma}} \right)^{1/\hat{\xi}} - 1 \right\}.$$

Since κ must satisfy $\Pr(\tilde{Y}(x) > 0) = (\kappa + 1)^{-1} = 0.01$, see Equation (6), we get $\kappa = 99$.

We fit our hierarchical models to the censored pretransformed data $\tilde{Y}(x)$ by numerically maximizing the pairwise likelihood. We set the spatial cut-off distance to $\Delta_S = 110$ km, which retains about 60% of the pairs of meteorological stations, and we choose the temporal cut-off as $\Delta_T = 10$ hr. The resulting number of pairs of observations is approximately 4.6×10^9, taking into account missing values. The full pairwise likelihood counts around 1.7×10^{11} pairs, which shows that we have attained a huge reduction. Pairwise likelihood maximization is coded in C, and it runs in parallel using the
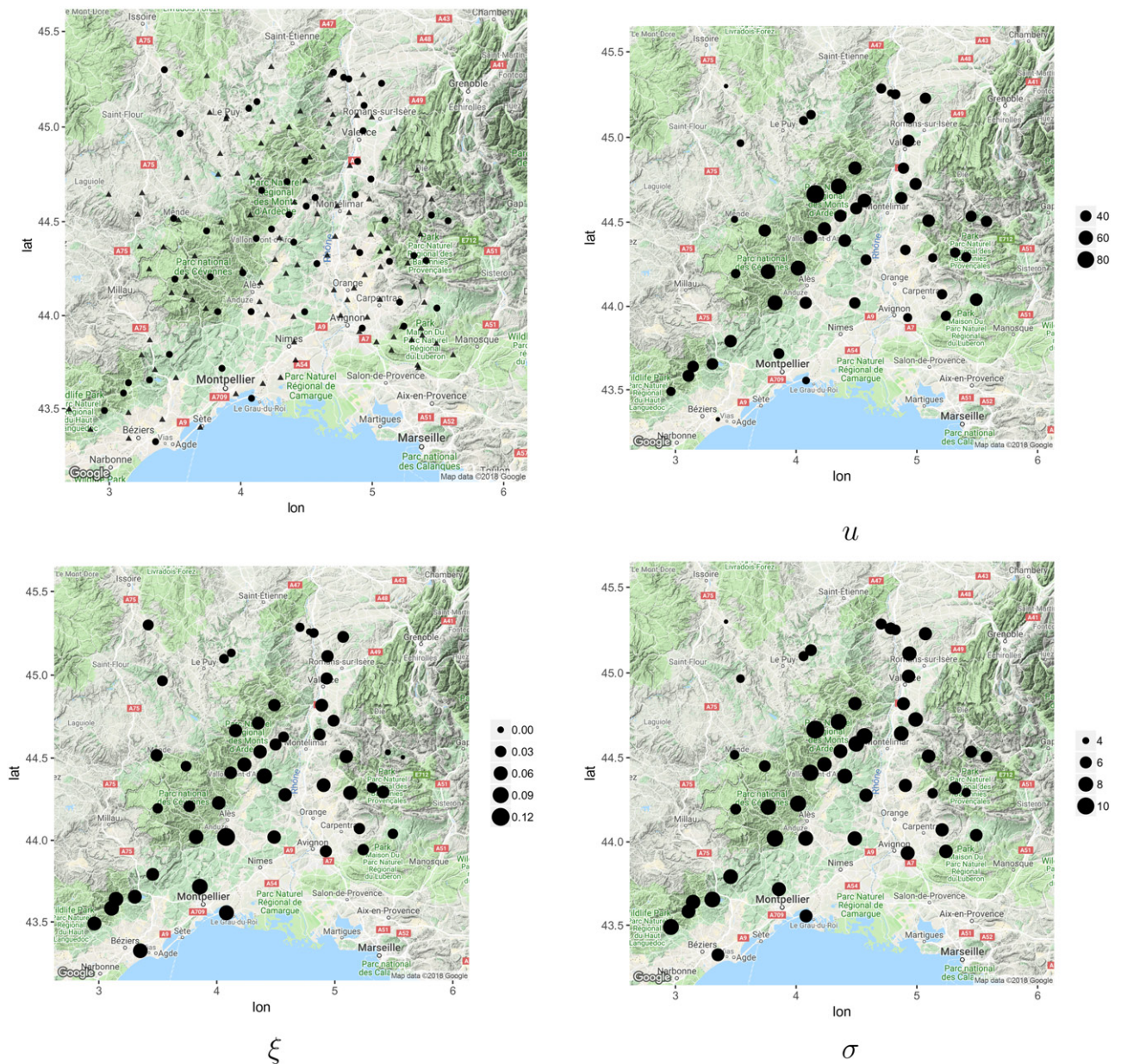
Figure 5. Precipitation data of Southern France. Top left display: topographic map showing the meteorological stations selected for our case study. Dots correspond to the stations used for fitting. In the other displays, their diameter is proportional to empirical 99% *quantiles u(s)* (top right plot) and to estimates of the GPD parameters $\xi(s)$ (bottom left plot) and $\sigma(s)$ (bottom right plot).

R library `parallel`. All calculations were carried out on a 2.6 GHz machine with 32 cores and 52 GB of memory. One evaluation of the composite likelihood requires approximately 18 seconds. For calculating standard errors and CLIC* values, we use the previously described subsampling technique based on temporal windows by considering $B = 500$ overlapping blocks, each corresponding to 1000 consecutive hours, that is, $d_b = 50 \times 1000$.

We consider two settings for the hierarchical model, with (G1) and without velocity (G2). Then, we compare these two models to three variants of a censored Gaussian space-time copula model labeled C1, C2, and C3 (Bortot, Coles, and Tawn 2000; Renard and Lang 2007; Davison, Huser, and Thibaud 2013) pertaining to the class of asymptotic independent processes. The

fits of the censored Gaussian space-time copula models match a censored Gaussian random field with transformed threshold exceedances; that is, we transform original data to standard Gaussian margins $G(x) = \Phi^{\leftarrow}(\text{GP}(\tilde{Y}(x)))$ (with Φ the standard Gaussian cdf), and we suppose that $\{G(x), x \in \mathcal{X}\}$ is a Gaussian space-time random field with space-time correlation function $\rho(x_1, x_2; \theta)$.

We denote by $\rho_e(a) = \exp(-a)$ and by $\rho_s(a) = (1 - 1.5a + 0.5a^3)1_{[0,1]}(a)$, $a \geq 0$, the exponential and spherical correlation models with scale 1, respectively. We introduce the scaled Mahalanobis distance between spatial locations s_1 and s_2, written

$$a(s_1, s_2; \tau) = \{(s_1 - s_2)'\Omega(\tau)^{-1}(s_1 - s_2)\}^{1/2},$$
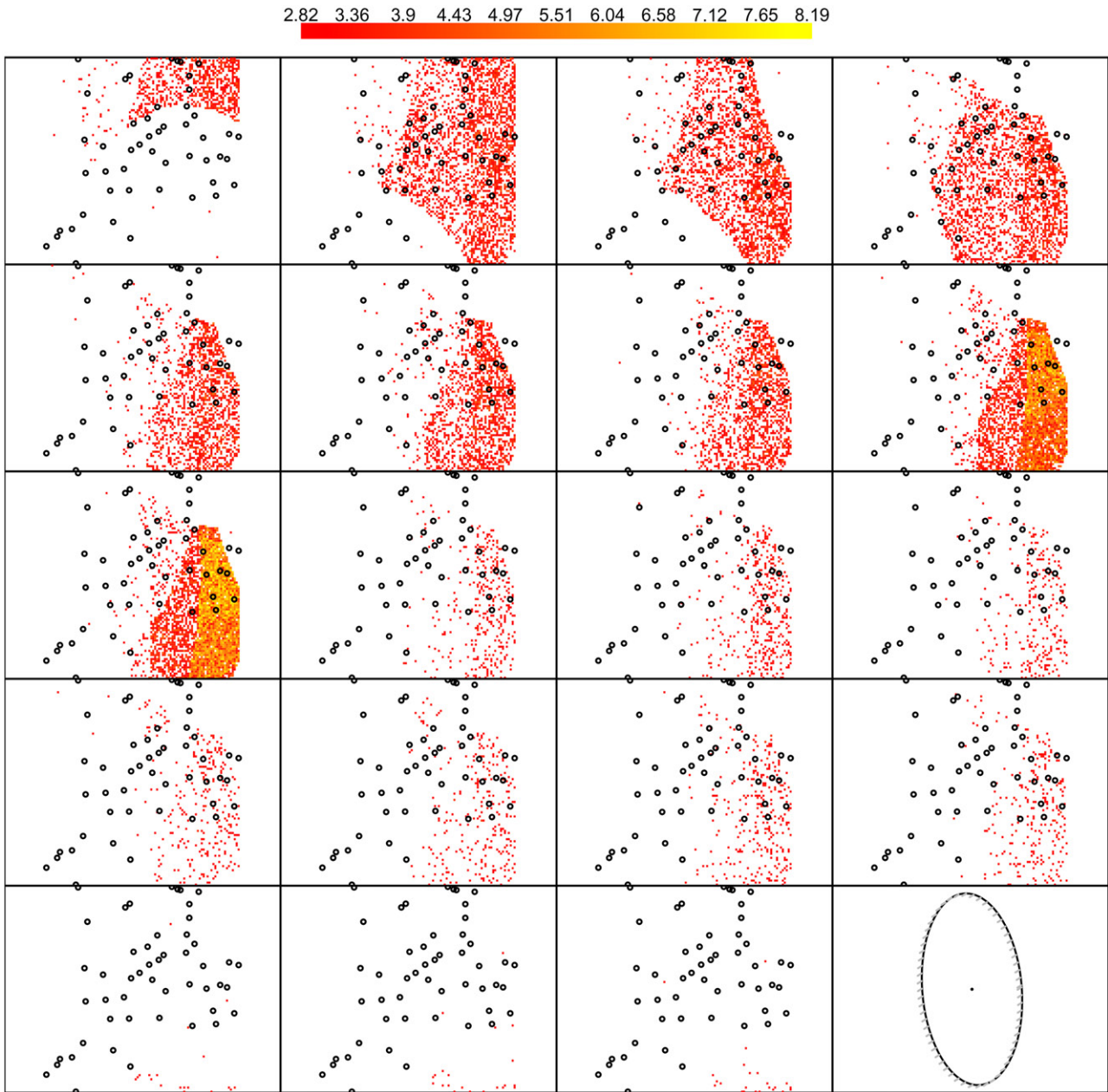
Figure 6. A simulation example showing exceedances of the 0.95-quantile for the model G1 fitted to precipitation data. Dots correspond to the stations used for fitting. The evolution over time during 19 hr is presented row-wise starting from the top left. The bottom right display illustrates the estimated ellipses, centered at the barycenter of the locations, and the movement induced by the velocity vector.

where

$$\Omega(\tau) = \begin{pmatrix} \cos(\tau_1) & -\sin(\tau_1) \\ \sin(\tau_1) & \cos(\tau_1) \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \tau_2^{-1} \end{pmatrix}$$
$$\times \begin{pmatrix} \cos(\tau_1) & \sin(\tau_1) \\ -\sin(\tau_1) & \cos(\tau_1) \end{pmatrix}.$$

The Mahalanobis distance defines elliptical isocontours. Here, $\tau_1 \in [0, \pi)$ is the angle with respect to the West-East direction, and $\tau_2 > 0$ is the length ratio of the two principal axes. We choose three specifications of the space-time correlation function:

C1. Space-time separable model:

$$\rho(x_1, x_2; \theta) = \rho_e(a(s_1, s_2; \tau)/\psi_S)\, \rho_e(|t_1 - t_2|/\psi_T), \quad (13)$$

with $\theta = (\tau_1, \tau_2, \psi_S, \psi_T)$. We assume anisotropic spatial correlation in analogy to models G1 and G2. The model is isotropic for $\tau_2 = 1$.

C2. Frozen field model 1 (see Christakos 2017, for a comprehensive account):

$$\rho(x_1, x_2; \theta) = \rho_e(a(s_1 - vt_1, s_2 - vt_2; \tau)/\psi), \quad (14)$$

where $\theta = (\tau_1, \tau_2, \psi, v')$ and $v \in \mathbb{R}^2$ is a velocity vector.

C3. Frozen field model 2 with compact support:

$$\rho(x_1, x_2; \theta) = \rho_s(a(s_1 - vt_1, s_2 - vt_2; \tau)/\psi). \quad (15)$$

In this model, two observations separated by Mahalonobis distance $a(s_1 - vt_1, s_2 - vt_2; \tau)$ greater than ψ will be independent.

Table 2. Estimates, standard errors (in italic), and CLIC* values of fitted models.

Model	Parameters						
	γ_1	γ_2	ϕ	δ	ω_1	ω_2	CLIC*
G1	165.062	318.823	0.085	20.184	0.723	0.446	404480.8
	23.459	*19.811*	*0.026*	*0.948*	*0.195*	*0.009*	
G2	175.817	294.323	0.041	20.036	0	0	404488.1
	11.879	*25.291*	*0.064*	*1.039*	*–*	*–*	
	τ_1	τ_2	ψ_S	ψ_T			CLIC*
C1	0.057	2.568	137.692	10.128			404626.2
	0.060	*0.309*	*7.615*	*0.523*			
	τ_1	τ_2	ψ_S		ν_1	ν_2	CLIC*
C2	1.034	2.025	108.755		6.672	16.358	404750.3
	0.040	*0.318*	*7.299*		*0.908*	*1.502*	
C3	0.481	5.125	174.980		6.614	10.406	405020.4
	0.005	*0.262*	*6.955*		*0.095*	*0.226*	

NOTES: Parameter units are kilometers for ϕ_S, γ_1, and γ_2, radians for ϕ and τ_1, hours for δ and ϕ_T, and kilometers per hour for $\omega_1, \omega_2, \nu_1$, and ν_2.

Evaluation of the full likelihood of the models C1, C2, and C3 requires numerical operations such as matrix inversion, matrix determinants, and high-dimensional Gaussian cdfs (Genz and Bretz 2009), which are computationally intractable in our case. Therefore, we opt again for a pairwise likelihood approach, which also simplifies model selection through the CLIC*.

Estimation results are summarized in Table 2. The CLIC* in the last column shows a preference for our hierarchical models with the best value for model G1, followed closely by G2. Estimated durations vary only slightly between G1 and G2. Estimates of ϕ differ more strongly, but one has to take into account that estimates of both semi-axis are very close. Moreover, estimates of γ_1 and γ_2 are similar for G1 and G2, which suggests coherent results for the two models and allows reliable physical interpretation of estimated parameter values. Regarding the results for model G1, we observe that the estimated parameters γ_1 and γ_2 characterize an ellipse covering a large part of the study region, which indicates relatively strong dependence even between sites that are far separated in space.

The estimate of ϕ underlines the low inclination of the ellipse, while $\gamma_2 \approx 2\gamma_1$, which leads to an ellongated shape of the ellipse. It corresponds well to the orientation of the mountain ridges in the considered region.

The estimate of δ, which may be interpreted as the average duration of extreme events, corresponds well to empirical measures of the actual durations of extreme events in the study region. The orientation of the reliefs seems to play an important role for the estimated velocity characterized by the values of ω_1 and ω_2, with ω_1 being considerably larger than ω_2. For visual illustration, Figure 7 shows a simulation of model G1 where the velocity effect in precipitation intensities becomes apparent. This simulation shows heavy precipitation arriving from the north, predominantly spreading over the eastern slopes of a mountain range in the study region, and then becoming more intense and finally gradually evacuating toward the south.

Among the Gaussian copula models, the preference goes to the separable model C1.

To underpin the good fit of our models through visual diagnostics, Figure 8 shows estimated probabilities $\Pr(Z(s, t) >$

$q | Z(s', t') > q)$ along different directions and at different temporal lags $|t - t'|$. These plots suggest that the behavior of models G1 and G2 is very close; there is no strong preference for one model over the other. The ranking of the copula models based on the CLIC* is also confirmed by the visual diagnostics. For contemporaneous observations with time lag 0, the models C1, C2, and C3 have comparable performance in capturing spatial dependence. However, for lags of 1 hr, models C2 and C3 represent the space-time interaction not satisfactorily.

Finally, we gain deeper insight into the joint tail structure of the fitted models by calculating empirical estimates $\hat{p}_i(h)$ of the multivariate conditional probability

$$\chi^*_{s_i;h}(q) := \Pr(Z(s_j, t) > q, s_j \in \partial s_i | Z(s_i, t - h) > q),$$

where ∂s_i is the set of the four nearest neighbors of site s_i, $i = 1, \ldots, 50$. We compare these values with precise Monte Carlo estimates $\tilde{p}_i^{(j)}(h)$, $j = 1, \ldots, 200$, based on a parametric bootstrap procedure using 200 simulations of the models G1, G2, and C1 with the leading CLIC* values. We compute site-specific root mean-squared errors (RMSE)

$$\text{RMSE}_i(h) = \left\{ \frac{\sum_{j=1}^{200} (\tilde{p}_i^{(j)}(h) - \hat{p}_i(h))^2}{200} \right\}^{1/2},$$

as well as the resulting total RMSE, $\text{RMSE}(h) = \sum_{i=1}^{50} \text{RMSE}_i(h)$, as an overall measure of goodness of fit. Table 3 reports such values for fitted models using contemporaneous observations or lags of 1 or 2 hr ($h = 0, 1, 2$) between the reference site and its neighbors. If we consider the quantile $q_{0.99}$ used as a threshold for fitting models, our hierarchical models present the best fit in terms of RMSE only for lagged values. However, models G1 and G2 extrapolate better for larger values of the threshold such as $q_{0.995}$.

7. Conclusions

We have proposed a novel space-time model for threshold exceedances of data with asymptotically vanishing dependence strength. In the spirit of the hierarchical modeling paradigm with latent layers to capture complex dependence and time dynamics, it is based on a latent Gamma convolution process with nonseparable space-time indicator kernels, and therefore amenable to physical interpretation. This framework leads to marginal and joint distributions that are available in closed form and are easy to handle in the extreme value context. The assumption of conditional independence as in our model is practical since it avoids the need to calculate cumulative distribution functions in large dimensions, although difficulty remains in evaluating the volume of the intersections of more than two cylinders and in calculating partial derivates for full likelihood formula. We can draw an interesting parallel to the max-stable Reich–Shaby process $Z_{\text{RS}}(x)$ (Reich and Shaby 2012), which is one of the more easily tractable spatial max-stable models and has a related construction. Indeed, the inverted process $1/Z_{\text{RS}}(x)$ can be represented as the embedding of a dependent latent convolution process (based on positive α-stable variables) for the rate of an exponential distribution. Conditional independence models cannot accurately capture
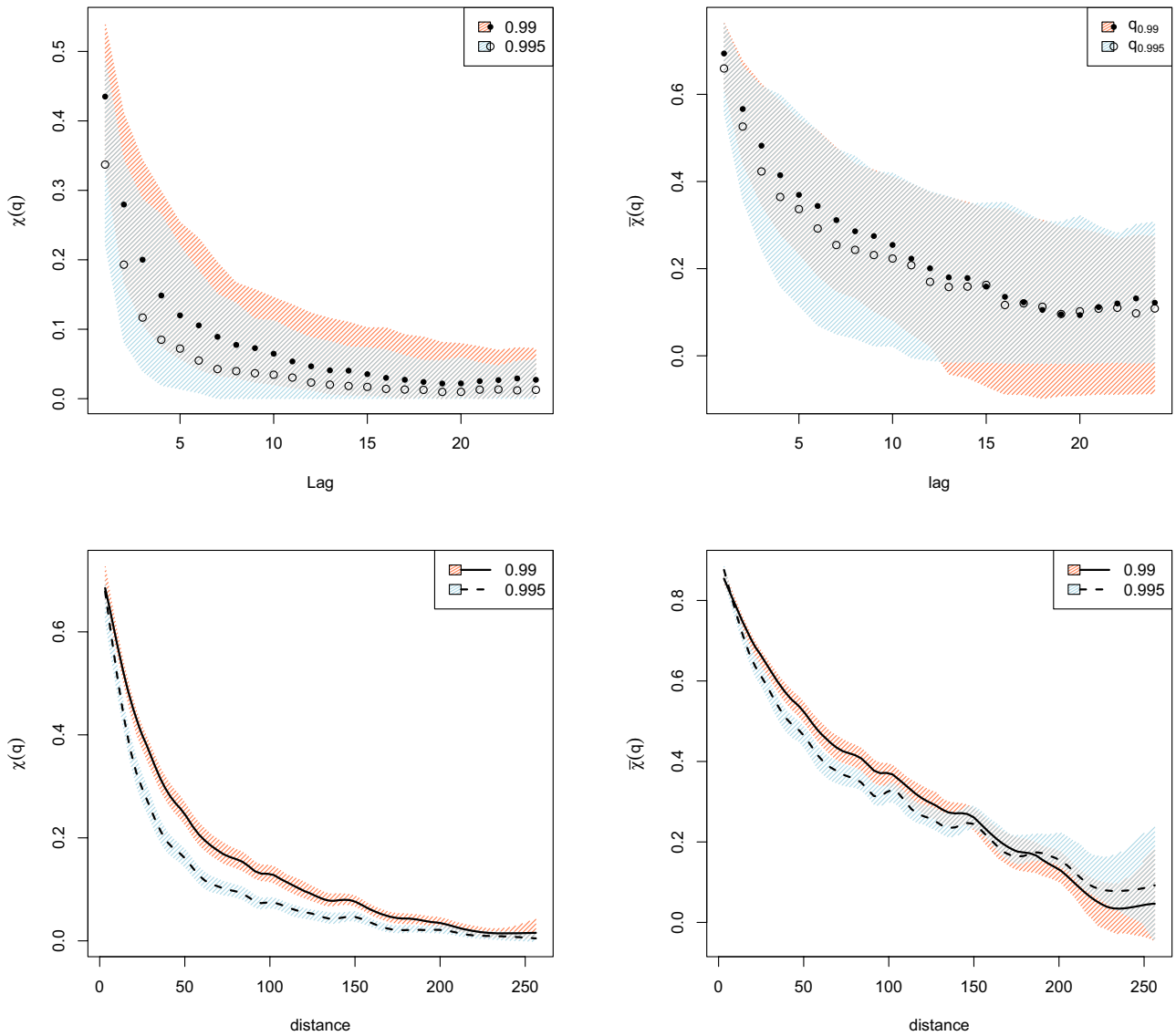
Figure 7. Empirical estimates of $\chi_X(q)$ (left panels) and $\bar{\chi}_X(q)$ (right panels) coefficients for the precipitation data. The filled region represents an approximate 95% confidence region based on a stationary bootstrap procedure.

the smoothness of the data-generating process. Nevertheless, the α-parameter in our model of the Gamma noise in (9) partially controls the smoothness of the latent Gamma field $\Lambda(s)$, with smaller values yielding more rugged surfaces.

In cases where data present asymptotic dependence, our asymptotically independent model may substantially underestimate the probability of jointly observing very high values over several space-time points. Asymptotic dependence in our construction (4) is equivalent to lower tail dependence in $\Lambda(x)$. There is no natural choice for introducing such dependence behavior, but a promising idea is to use what we label *Beta scaling*: given a temporal process $B(t)$ independent of $\Lambda(s, t)$ with $\text{Beta}(\tilde{\alpha}, \alpha)$ distributed margins, $0 < \tilde{\alpha} < \alpha$, we could replace $\Lambda(s, t)$ in our construction by the process $\tilde{\Lambda}(s, t) = B(t)\Lambda(s, t)$ possessing margins following the $\Gamma(\tilde{\alpha}, \beta)$ distribution. This construction has asymptotic dependence over space, and it will be asymptotically dependent over time if $B(t)$ has lower tail dependence. Follow-up work will explore

theoretical properties and practical implementation of such extensions.

We have developed pairwise likelihood inference for our models, which scales well with high-dimensional datasets. We point out that handling observations over irregular time steps and missing data is straightforward with our model thanks to its definition over continuous time. While we think that MCMC-based Bayesian estimation of the relatively high number of parameters may be out of reach principally due to the very high dimension of the set of latent Gamma variables in the model's current formulation, we are confident that future efforts to tackle the conditional simulation of such space-time processes based on MCMC simulation with fixed parameters could be successful; that is, by using frequentist estimation of parameters, space-time prediction requires to iteratively update only the latent Gamma field through MCMC, but not parameters.

The application of our novel model to a high-dimensional real precipitation dataset from southern France was motivated
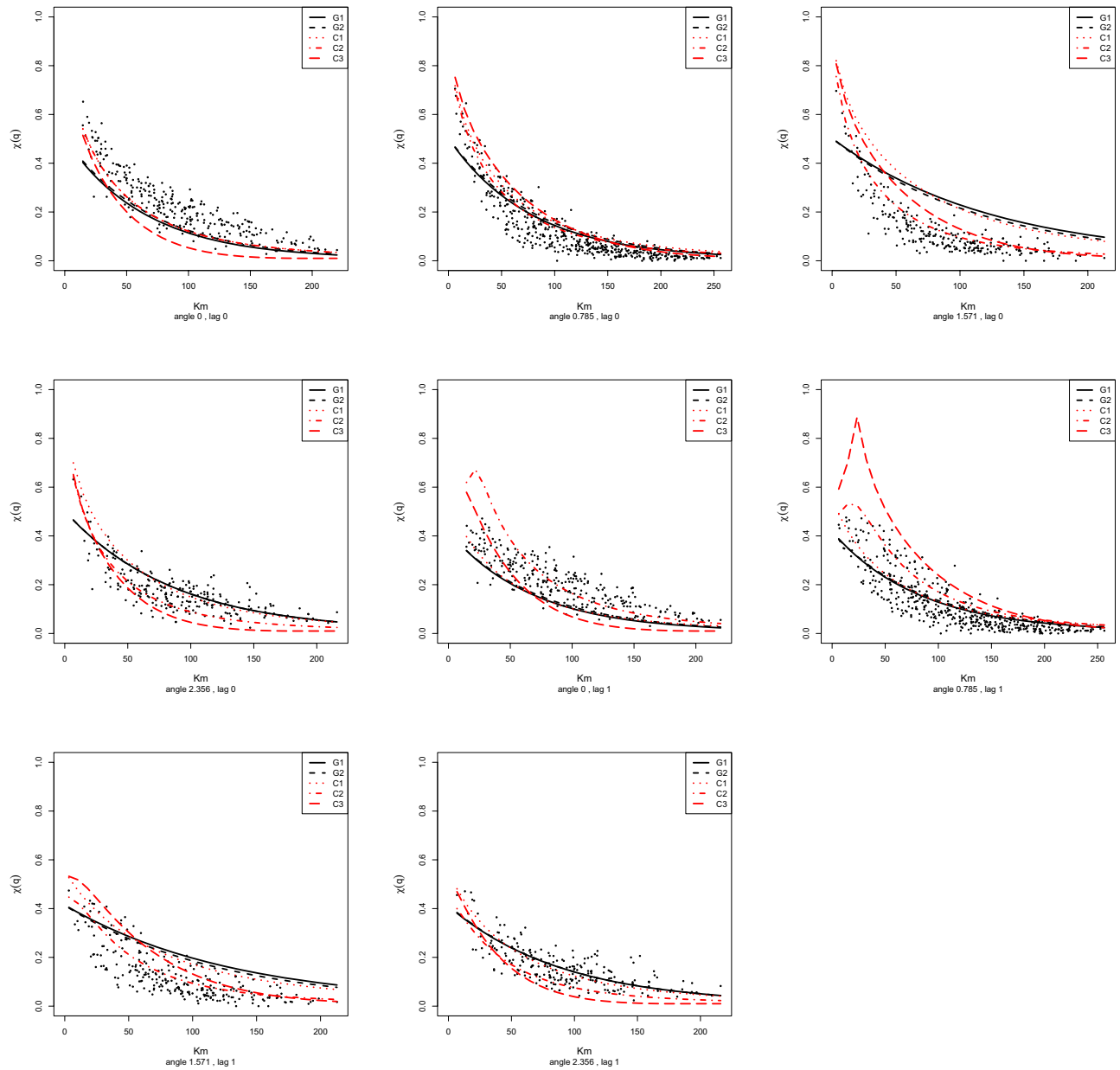
Figure 8. Estimated probabilities $\Pr(Z(s,t) > q | Z(s',t') > q)$ along different directions (expressed in radians) and at different temporal lags for the precipitation data. Dotted points correspond to empirical estimates. The value q is fixed to the empirical 99% quantile.

Table 3. Total root mean squared errors for the estimates of $\chi^*_{s_i;h}(q)$.

	RMSE(0)		RMSE(1)		RMSE(2)	
	$q_{0.99}$	$q_{0.995}$	$q_{0.99}$	$q_{0.995}$	$q_{0.99}$	$q_{0.995}$
G1	2.614	2.096	1.901	1.643	1.475	1.496
G2	2.605	2.072	1.907	1.626	1.477	1.480
C1	2.240	2.455	2.053	2.428	1.779	1.928

from clear evidence of asymptotic independence highlighted at an exploratory stage. It provides practical illustration of the high flexibility of our model and its capability to accurately predict extreme event probabilities for concomitant threshold exceedances in space and time. Based on meteorological knowledge about the precipitation processes in the study region, we had hoped to estimate a clear velocity effect. As a matter of fact, the fitted hierarchical model with velocity appeared to be only slightly superior to other models in some aspects. This interesting finding may also be interpreted as evidence for the highly fragmented structure arising in precipitation processes at small spatial and temporal scales.

Ongoing work aims to adapt the current latent process construction to the multivariate setting by considering constructions with Gamma factors common to several components, specifically structures with a hierarchical tree-based construction of the latent Gamma components, and extensions to asymptotic dependence using the above-mentioned Beta-scaling. Ultimately, such novelties could provide a flexible toolbox for multivariate space-time modeling with scenarios of partial or full asymptotic dependence.

Appendix A: Formulas for the Laplace Exponent of a Random Measure

The Laplace exponent of the random measure $\Gamma(\cdot)$ is defined as

$$\mathcal{L}(\phi) := -\log E\left(\exp\left\{-\int \phi(x)\Gamma(dx)\right\}\right)$$
$$= \int_{\mathcal{X}} \log\left\{1 + \frac{\phi(x)}{\beta}\right\}\alpha(dx),$$

where ϕ is any positive measurable function.
Consider $\phi = v\mathbf{1}_A(x)$. Then,

$$\mathcal{L}(\phi) = -\log E\left(\exp\{-v\Gamma(A)\}\right) = \int_A \log\left\{1 + \frac{v}{\beta}\right\}\alpha(dx)$$
$$= \alpha(A)\log\left\{1 + \frac{v}{\beta}\right\},$$

that is,

$$E\left(\exp\{-v\Gamma(A)\}\right) = \left(\frac{\beta}{v+\beta}\right)^{\alpha(A)}.$$

For bivariate analyses, choosing $\phi(x) = v_1\mathbf{1}_{A_1}(x) + v_2\mathbf{1}_{A_2}(x)$, yields

$$\mathcal{L}(\phi) = -\log E\left(\exp\{-v_1\Gamma(A_1) - v_2\Gamma(A_2)\}\right)$$
$$= -\log E\left(\exp\{-v_1\Gamma(A_1\backslash A_2) - (v_1+v_2)\Gamma(A_1\cap A_2)\right.$$
$$\left. -v_2\Gamma(A_2\backslash A_1)\}\right)$$
$$= \int_{A_1\backslash A_2} \log\left\{1 + \frac{v_1}{\beta}\right\}\alpha(dx)$$
$$+ \int_{A_1\cap A_2} \log\left\{1 + \frac{v_1+v_2}{\beta}\right\}\alpha(dx)$$
$$+ \int_{A_2\backslash A} \log\left\{1 + \frac{v_2}{\beta}\right\}\alpha(dx)$$
$$= \alpha(A_1\backslash A_2)\log\left\{1 + \frac{v_1}{\beta}\right\} + \alpha(A_1\cap A_2)\log\left\{1 + \frac{v_1+v_2}{\beta}\right\}$$
$$+ \alpha(A_2\backslash A_1)\log\left\{1 + \frac{v_2}{\beta}\right\}$$

and therefore

$$E(\exp\{-v_1\Gamma(A_1) - v_2\Gamma(A_2)\})$$
$$= \left(1 + \frac{v_1}{\beta}\right)^{-\alpha(A_1\backslash A_2)}\left(1 + \frac{v_1+v_2}{\beta}\right)^{-\alpha(A_1\cap A_2)}$$
$$\times \left(1 + \frac{v_2}{\beta}\right)^{-\alpha(A_2\backslash A_1)}.$$

Appendix B: Formulas for the Pairwise Censored Likelihood

Let $LP^{(1)}(v)$ and $LP^{(2)}_{x,x'}(v_1, v_2)$, $x \neq x'$ denote the univariate and bivariate Laplace transform of $\Lambda(A_x)$ that is,

$$LP^{(1)}(v) := E\left(e^{-v\Lambda(A_x)}\right) = \left(\frac{\beta}{v+\beta}\right)^{c_0},$$

and

$$LP^{(2)}_{x,x'}(v_1, v_2) := E\left(e^{-v_1\Lambda(A_x) - v_2\Lambda(A_{x'})}\right) = \left(\frac{\beta}{v_1+\beta}\right)^{c_1}$$
$$\times \left(\frac{\beta}{v_1+v_2+\beta}\right)^{c_2}\left(\frac{\beta}{v_2+\beta}\right)^{c_3}$$

with $c_0 = \alpha(A_x)$ $c_1 = \alpha(A_x\backslash A_{x'})$, $c_2 = \alpha(A_x\cap A_{x'})$, $c_3 = \alpha(A_{x'}\backslash A_x)$.

We obtain

$$\frac{\partial}{\partial v}LP^{(1)}(v)$$
$$= -c_0\beta^{c_0}(v+\beta)^{-c_0-1},$$

$$\frac{\partial}{\partial v_1}LP^{(2)}_{x,x'}(v_1, v_2)$$
$$= -\beta^{c_1+c_2+c_3}\left\{c_1(v_1+\beta)^{-c_1-1}(v_1+v_2+\beta)^{-c_2}(v_2+\beta)^{-c_3}\right.$$
$$\left. + c_2(v_1+\beta)^{-c_1}(v_1+v_2+\beta)^{-c_2-1}(v_2+\beta)^{-c_3}\right\},$$

$$\frac{\partial}{\partial v_2}LP^{(2)}_{x,x'}(v_1, v_2)$$
$$= -\beta^{c_1+c_2+c_3}\left\{c_3(v_1+\beta)^{-c_1}(v_1+v_2+\beta)^{-c_2}(v_2+\beta)^{-c_3-1}\right.$$
$$\left. + c_2(v_1+\beta)^{-c_1}(v_1+v_2+\beta)^{-c_2-1}(v_2+\beta)^{-c_3}\right\},$$

$$\frac{\partial}{\partial v_1\partial v_2}LP^{(2)}_{x,x'}(v_1, v_2)$$
$$= \beta^{c_1+c_2+c_3}\left\{c_1c_2(v_1+\beta)^{-c_1-1}(v_1+v_2+\beta)^{-c_2-1}\right.$$
$$\times (v_2+\beta)^{-c_3} + c_1c_3(v_1+\beta)^{-c_1-1}(v_1+v_2+\beta)^{-c_2}$$
$$\times (v_2+\beta)^{-c_3-1} + c_2(c_2+1)(v_1+\beta)^{-c_1}(v_1+v_2+\beta)^{-c_2-2}$$
$$\times (v_2+\beta)^{-c_3} + c_2c_3(v_1+\beta)^{-c_1}(v_1+v_2+\beta)^{-c_2-1}$$
$$\left.\times (v_2+\beta)^{-c_3-1}\right\}.$$

Supplementary Materials

The supplementary material contains the code to simulate and estimate the model of Section 2. This is a snapshot of the repository https://github.com/cgaetan/Gamma-GPD. Real data cannot be freely distributed, but the Readme.md files contains instructions for requesting them.

Acknowledgments

Funding

References

Bacro, J. N., Gaetan, C., and Toulemonde, G. (2016), "A Flexible Dependence Model for Spatial Extremes," *Journal of Statistical Planning and Inference*, 172, 36–52. [556]

Barndorff-Nielsen, O. E., Lunde, A., Shepard, N., and Veraat, A. E. D. (2014), "Integer-Valued Trawl Processes: A Class of Stationary Infinitively Divisible Processes," *Scandinavian Journal of Statistics*, 41, 693–724. [556]

Bevilacqua, M., Gaetan, C., Mateu, J., and Porcu, E. (2012), "Estimating Space and Space-Time Covariance Functions: A Weighted Composite Likelihood Approach," *Journal of the American Statistical Association*, 107, 268–280. [560]

Bortot, P., Coles, S., and Tawn, J. (2000), "The Multivariate Gaussian Tail Model: An Application to Oceanographic Data," *Journal of the Royal Statistical Society*, Series C, 49, 31–49. [556,563]

Bortot, P., and Gaetan, C. (2014), "A Latent Process Model for Temporal Extremes," *Scandinavian Journal of Statistics*, 41, 606–621. [556,557]

——— (2016), "Latent Process Modelling of Threshold Exceedances in Hourly Rainfall Series," *Journal of Agricultural, Biological, and Environmental Statistics*, 21, 531–547. [557,562]

Carlstein, A. (1986), "The Use of Subseries Values for Estimating the Variance of a General Statistic From a Stationary Sequence," *The Annals of Statistics*, 14, 1171–1179. [560]

Carreau, J., and Bouvier, C. (2016), "Multivariate Density Model Comparison for Multi-site Flood-Risk Rainfall in the French Mediterranean Area," *Stochastic Environmental Research Risk Assessement*, 30, 1591–1612. [555]

Casson, E., and Coles, S. G. (1999), "Spatial Regression Models for Extremes," *Extremes*, 1, 449–468. [555]

Christakos, G. (2017), *Spatiotemporal Random Fields Theory and Applications*, Amsterdam: Elsevier. [564]

Coles, S., Heffernan, J., and Tawn, J. (1999), "Dependence Measures for Extreme Value Analyses," *Extremes*, 2, 339–365. [559]

Cooley, D., Nychka, D., and Naveau, P. (2007), "Bayesian Spatial Modeling of Extreme Precipitation Return Levels," *Journal of the American Statistical Association*, 102, 824–840. [555]

Cox, D. R., and Isham, V. (1988), "A Simple Spatial-Temporal Model of Rainfall," *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 415, 317–328. [556]

Davis, R. A., Klüppelberg, C., and Steinkohl, C. (2013a), "Max-Stable Processes for Modeling Extremes Observed in Space and Time," *Journal of the Korean Statistical Society*, 42, 399–414. [556]

——— (2013b), "Statistical Inference for Max-Stable Processes in Space and Time," *Journal of the Royal Statistical Society*, Series B, 75, 791–819. [556,560]

Davis, R. A., and Mikosch, T. (2008), "Extreme Value Theory for Space-Time Processes With Heavy-Tailed Distributions," *Stochastic Processes and Their Applications*, 118, 560–584. [556]

Davison, A. C., and Gholamrezaee, M. M. (2012), "Geostatistics of Extremes," *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 468, 581–608. [555,556,560]

Davison, A. C., Huser, R., and Thibaud, E. (2013), "Geostatistics of Dependent and Asymptotically Independent Extremes," *Journal of Mathematical Geosciences*, 45, 511–529. [555,556,563]

Davison, A. C., Padoan, S. A., and Ribatet, M. (2012), "Statistical Modelling of Spatial Extremes," *Statistical Science*, 27, 161–186. [555]

de Haan, L., and Ferreira, A. (2006), *Extreme Value Theory: An Introduction*, New York: Springer. [557]

Ferguson, T. S. (1973), "A Bayesian Analysis of Some Nonparametric Problems," *The Annals of Statistics*, 1, 209–230. [557]

Ferreira, A., and de Haan, L. (2014), "The Generalized Pareto Process; With a View Towards Application and Simulation," *Bernoulli*, 20, 1717–1737. [555]

Gaetan, C., and Grigoletto, M. (2007), "A Hierarchical Model for the Analysis of Spatial Rainfall Extremes," *Journal of Agricultural Biological and Environmental Sciences*, 12, 434–449. [555]

Genz, A., and Bretz, F. (2009), *Computation of Multivariate Normal and t Probabilities*, New York: Springer. [565]

Hughes, G. B., and Chraibi, M. (2012), "Calculating Ellipse Overlap Areas," *Computing and Visualization in Science*, 15, 291–301. [558]

Huser, R., and Davison, A. C. (2014), "Space-Time Modelling of Extreme Events," *Journal of the Royal Statistical Society*, Series B, 76, 439–461. [556,559,560]

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. [556]

——— (2018), "Penultimate Modeling of Spatial Extremes: Statistical Inference for Max-Infinitely Divisible Processes," arXiv no. 1801.02946. [556]

Huser, R., and Wadsworth, J. L. (2019), "Modeling Spatial Processes With Unknown Extremal Dependence Class," *Journal of the American Statistical Association*, 114, 414–434. [556]

Joe, H. (1997), *Multivariate Models and Dependence Concepts*, London: Chapman & Hall. [557]

Kabluchko, Z., Schlather, M., and de Haan, L. (2009), "Stationary Max-Stable Fields Associated to Negative Definite Functions," *The Annals of Probability*, 37, 2042–2065. [555]

Morris, S. A., Reich, B. J., Thibaud, E., and Cooley, D. (2017), "A Space-Time Skew-t Model for Threshold Exceedances," *Biometrics*, 73, 749–758. [556]

Nieto-Barajas, L. E., and Huerta, G. (2017), "Spatio-Temporal Pareto Modelling of Heavy-Tail Data," *Spatial Statistics*, 20, 92–109. [556]

Noven, R. C., Veraart, A. E., and Gandy, A. (2015), "A Latent Trawl Process Model for Extreme Values," arXiv no. 1511.08190. [556]

Opitz, T. (2013), "Extremal t Processes: Elliptical Domain of Attraction and a Spectral Representation," *Journal of Multivariate Analysis*, 122, 409–413. [555]

——— (2016), "Modeling Asymptotically Independent Spatial Extremes Based on Laplace Random Fields," *Spatial Statistics*, 16, 1–18. [556]

——— (2017), "Spatial Random Field Models Based on Lévy Indicator Convolutions," arXiv no. 1710.06826. [556]

Opitz, T., Bacro, J.-N., and Ribereau, P. (2015), "The Spectrogram: A Threshold-Based Inferential Tool for Extremes of Stochastic Processes," *Electronic Journal of Statistics*, 9, 842–868. [555]

Politis, D. N., and Romano, J. P. (1994), "The Stationary Bootstrap," *Journal of the American Statistical Association*, 89, 1303–1313. [562]

Reich, B. J., and Shaby, B. A. (2012), "A Hierarchical Max-Stable Spatial Model for Extreme Precipitation," *The Annals of Applied Statistics*, 6, 1430–1451. [555,565]

Reiss, R., and Thomas, M. (2007), *Statistical Analysis of Extreme Values* (3rd ed.), Basel: Birkhäuser. [557]

Renard, B., and Lang, M. (2007), "Use of a Gaussian Copula for Multivariate Extreme Value Analysis: Some Case Studies in Hydrology," *Advances in Water Resources*, 30, 897 – 912. [563]

Rodriguez-Iturbe, I., Cox, D. R., and Isham, V. (1987), "Some Models for Rainfall Based on Stochastic Point Processes," *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 410, 269–288. [556]

Sang, H., and Gelfand, A. (2009), "Hierarchical Modeling for Extreme Values Observed Over Space and Time," *Environmental and Ecological Statistics*, 16, 407–426. [555,556]

Schlather, M. (2002), "Models for Stationary Max-Stable Random Fields," *Extremes*, 5, 33–44. [555,556]

Sibuya, M. (1960), "Bivariate Extreme Statistics," *Annals of the Institute of Statistical Mathematics*, 11, 195–210. [559]

Smith, R. L. (1990), "Max-Stable Processes and Spatial Extremes," Preprint, University of Surrey. [555]

Tawn, J., Shooter, R., Towe, R., and Lamb, R. (2018), "Modelling Spatial Extreme Events With Environmental Applications," *Spatial Statistics*, 28, 39–58. [555]

Thibaud, E., Mutzner, R., and Davison, A. C. (2013), "Threshold Modeling of Extreme Spatial Rainfall," *Water Resources Research*, 49, 4633–4644. [555]

Thibaud, E., and Opitz, T. (2015), "Efficient Inference and Simulation for Elliptical Pareto Processes," *Biometrika*, 102, 855–870. [555]

Varin, C., and Vidoni, P. (2005), "A Note on Composite Likelihood Inference and Model Selection," *Biometrika*, 52, 519–528. [560]

Wadsworth, J., and Tawn, J. (2012), "Dependence Modelling for Spatial Extremes," *Biometrika*, 99, 253–272. [556]

Wolpert, R. L., and Ickstadt, K. (1998a), "Poisson/Gamma Random Fields for Spatial Statistics," *Biometrika*, 85, 251–267. [556,557,558]

——— (1998b), "Simulation of Lévy Random Fields," in *Practical Nonparametric and Semiparametric Bayesian Statistics*, eds. D. Dey, P. Müller, and D. Sinha, New York: Springer, pp. 227–242. [560]