Spatial hierarchical modeling of threshold exceedances using rate mixtures

Rishikesh Yadav[1], Raphaël Huser[1] and Thomas Opitz[2]

December 9, 2019

Abstract

We develop new flexible univariate models for light-tailed and heavy-tailed data, which extend a hierarchical representation of the generalized Pareto (GP) limit for threshold exceedances. These models can accommodate departure from asymptotic threshold stability in finite samples while keeping the asymptotic GP distribution as a special (or boundary) case and can capture the tails and the bulk jointly without losing much flexibility. Spatial dependence is modeled through a latent process, while the data are assumed to be conditionally independent. We design penalized complexity priors for crucial model parameters, shrinking our proposed spatial Bayesian hierarchical model toward a simpler reference whose marginal distributions are GP with moderately heavy tails. Our model can be fitted in fairly high dimensions using Markov chain Monte Carlo by exploiting the Metropolis-adjusted Langevin algorithm (MALA), which guarantees fast convergence of Markov chains with efficient block proposals for the latent variables. We also develop an adaptive scheme to calibrate the MALA tuning parameters. Moreover, our models avoid the expensive numerical evaluations of multifold integrals in censored likelihood expressions. We demonstrate our new methodology by simulation and application to a dataset of extreme rainfall episodes that occurred in Germany. Our fitted model provides a satisfactory performance and can be successfully used to predict rainfall extremes at unobserved locations.

Keywords: Bayesian hierarchical modeling; extended generalized Pareto distribution; extreme event; Markov chain Monte Carlo; penalized complexity prior; precipitation extremes.

[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: rishikesh.yadav@kaust.edu.sa; raphael.huser@kaust.edu.sa

[2]INRAE, UR546 Biostatistics and Spatial Processes, 228, Route de l'Aérodrome, CS 40509, 84914 Avignon, France. E-mail: thomas.opitz@inra.fr

1 Introduction

Atmospheric, meteorologic, hydrologic, and land surface processes among others are prone to extreme episodes, which are believed to increase in frequency and severity in the context of a changing climate and global warming (see, e.g., Witze, 2018; Power and Delage, 2019). To quantify environmental risk, e.g., associated with weather variables or air pollution, or to attribute certain extreme events to human influence (Risser and Wehner, 2017), it is crucial to model and predict the univariate behavior of extreme events, while accounting for spatial dependence. Extreme-Value Theory (EVT) provides a natural methodological framework to tackle this problem; see, e.g., the review papers Davison *et al.* (2012), Davison and Huser (2015) and Davison *et al.* (2019).

In this paper, we first study a hierarchical construction that leads to new univariate tail models, which extend the classical EVT approach by gaining flexibility at finite levels. We then exploit this hierarchical construction in a Bayesian framework for the modeling of spatial extremes by embedding a latent process with spatial dependence, while keeping the desired unconditional marginal distributions.

Under mild conditions, classical univariate EVT suggests using the generalized Pareto (GP) distribution for modeling extreme events defined as high threshold exceedances (Davison and Smith, 1990). More precisely, let Y be a random variable following a distribution F with finite or infinite upper endpoint $y_F = \sup\{y \in \mathbb{R} : F(y) < 1\}$. Then, for a wide class of distributions F, high threshold exceedances $(Y - u) \mid Y > u$ may be *asymptotically* approximated as

$$\Pr(Y - u \leq y \mid Y > u) = \frac{F(u+y) - F(u)}{1 - F(u)} \approx H_{\tau,\xi}(y) = 1 - (1 + \xi y/\tau)^{-1/\xi}, \qquad (1)$$

as the threshold u converges to y_F, where $H_{\tau,\xi}(y)$ denotes the GP distribution function with scale parameter $\tau > 0$ and shape parameter $\xi \in \mathbb{R}$ (also called tail index), defined over

$\{y > 0 : 1 + \xi y/\tau > 0\}$. In other words, $1 - F(y) \approx \zeta_u\{1 - H_{\tau,\xi}(y - u)\}$, for $y > u$ large, where $\zeta_u = \Pr(Y > u)$. When $\xi = 0$, the distribution (1) is interpreted as the limit $\xi \to 0$, and we obtain the exponential distribution function $H_{\tau,0}(y) = 1 - \exp(-y/\tau)$, $y > 0$. When $\xi < 0$, the support is bounded, whereas when $\xi \geq 0$, it is unbounded. Short, light, and heavy tails correspond to $\xi < 0$, $\xi = 0$ and $\xi > 0$, respectively, with ξ controlling the tail weight.

In practice, the choice of a good threshold u should reflect the transition around which the asymptotic regime takes place for the tail approximation (1) to be valid. This implies a bias-variance trade-off, as a high threshold u leads to a good approximation (low bias) but yields a small number of exceedances (high variance), and vice versa for a low threshold. Experience shows that automatic threshold selection procedures are not always reliable. It is often difficult to find a good, natural and interpretable threshold, and parameter estimates are often sensitive to this choice (Scarrott and MacDonald, 2012). This has motivated the development of *sub-asymptotic* models for extremes, which are more flexible than the asymptotic GP distribution at finite levels, while keeping a GP-like behavior in the tail; see, e.g., Frigessi *et al.* (2003), Carreau and Bengio (2009), Papastathopoulos and Tawn (2013) and Naveau *et al.* (2016), among others. With sub-asymptotic tail models, parameter estimates are usually less sensitive to the threshold, the choice of which then becomes less crucial for inference, and we can thus set lower thresholds. Ideally, this allows us to describe also the distribution of moderate and low values through an appropriate parametrization, which partly or fully separates control over bulk and tail properties. Scarrott and MacDonald (2012) provide a review of models describing jointly the bulk and the tail of the distribution.

In this paper, we propose a novel Bayesian hierarchical modeling framework for sub-asymptotic threshold exceedances. It has an intuitive interpretation, permits fully Bayesian inference, and can naturally incorporate covariate information and be extended to the spatial setting. Several Bayesian hierarchical models were already proposed in the literature to

model threshold exceedances; see, e.g., Cooley *et al.* (2007), Opitz *et al.* (2018). However, unlike the latent Gaussian models proposed therein, our construction makes sure that the unconditional distribution (obtained after integrating out latent random effects) remains of the desired form, with the GP distribution as a particular case. Precisely, we extend the characterization of the GP distribution as an exponential mixture with rate parameter following a gamma distribution. Let $\text{Exp}(\lambda)$ denote the exponential distribution with rate $\lambda > 0$, $\text{Gamma}(\alpha, \beta)$ denote the gamma distribution with rate $\alpha > 0$ and shape $\beta > 0$, i.e., with density $g(y) = \{\Gamma(\beta)\}^{-1}\alpha^{\beta}y^{\beta-1}\exp(-\alpha y)$, $y > 0$, and $\text{GP}(\tau, \xi)$ denote the GP distribution with scale $\tau > 0$ and shape ξ as defined in (1). Then we have

$$\left.\begin{array}{rl} Y \mid \Lambda & \sim \text{Exp}(\Lambda) \\ \Lambda & \sim \text{Gamma}(\alpha, \beta) \end{array}\right\} \Rightarrow Y \sim \text{GP}(\alpha/\beta, 1/\beta); \tag{2}$$

see Reiss and Thomas (2007), Bortot and Gaetan (2014), Bortot and Gaetan (2016) and Bacro *et al.* (2019). In other words, exponentially-decaying tails become heavier by making their rate parameter Λ random. By integrating out the latent variable Λ in the hierarchical construction in (2), we obtain the GP distribution for the data Y. Our new tail models (detailed further in §2 below) are constructed as in (2), but we modify the top and/or lower levels of the hierarchy in order to gain in flexibility, while keeping the GP distribution with $\xi \geq 0$ as a special or boundary case. Moreover, we penalize departure from the GP distribution in the Bayesian framework by specifying penalized complexity (PC) priors (Simpson *et al.*, 2017) designed to shrink complex models toward simpler counterparts, thus preventing overfitting. This avoids estimating unreasonable tail models, and guarantees that the fitted distribution will not be too far away from the GP distribution which is supported by asymptotic theory, unless the data provide strong evidence that a different sub-asymptotic behavior prevails. In other words, our proposed models are constrained to remain in the "neighborhood" of moderately heavy-tailed GP distributions. Our modeling approach based on extensions of (2) is general and can potentially generate a wide variety of new models

with light and heavy tails and various behaviors in the bulk. Below, we mainly focus on a parsimonious extension of (2), which assumes a gamma distribution in both levels of the hierarchy, although we also discuss other possible models with interesting tail properties.

For spatial modeling, we incorporate spatial dependence at the latent level, while assuming that the data are conditionally independent given the latent process. Specifically, we assume that the observed spatial process $Y(\boldsymbol{s})$, $\boldsymbol{s} \in \mathcal{S} \subset \mathbb{R}^2$, may be described analogously to (2) with a hierarchical representation in terms of a latent spatially structured process $\Lambda(\boldsymbol{s})$, such that the data $Y(\boldsymbol{s}_1)$ and $Y(\boldsymbol{s}_2)$ at any two distinct locations $\boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathcal{S}$, $\boldsymbol{s}_1 \neq \boldsymbol{s}_2$, are independent given $\Lambda(\boldsymbol{s}_1)$ and $\Lambda(\boldsymbol{s}_2)$. This conditional independence assumption is common in Bayesian hierarchical models (Banerjee *et al.*, 2014; Cooley *et al.*, 2007; Opitz *et al.*, 2018) and is, to some extent, akin to using a "nugget effect" in classical geostatistics (Cressie, 1993), which captures measurement errors or unstructured local variations in the data. Here, we make this assumption mainly to keep the model simple and identifiable, and for computational convenience in order to efficiently handle the censoring of non-extreme values in our inference procedure. More precisely, to fit our models to threshold excesses $Y(\boldsymbol{s}) > u(\boldsymbol{s})$ for some moderately high threshold $u(\boldsymbol{s})$, we design a generic Markov chain Monte Carlo (MCMC) sampler that efficiently exploits the hierarchical representation. Low values such that $Y(\boldsymbol{s}) \leq u(\boldsymbol{s})$ are treated as censored and imputed by simulation in our MCMC algorithm. Thanks to the conditional independence assumption, multivariate censoring can be conveniently reduced to univariate site-by-site imputations. This computational benefit is significant, and it contrasts with the high computational burden due to multivariate censoring in peaks-over-threshold inference for most spatial extremes models (Wadsworth and Tawn, 2014; Huser *et al.*, 2017; Huser and Wadsworth, 2019; Castro-Camilo and Huser, 2019). This approach allows us to tackle higher spatial dimensions more easily. Furthermore, to efficiently sample from the posterior distribution and accelerate the mixing of MCMC

chains, we update the large number of latent parameters jointly (in one block) based on the Metropolis-adjusted Langevin algorithm (MALA, see Roberts and Tweedie (1996) and Roberts and Rosenthal (1998)) and we adaptively calibrate the MALA tuning parameters to obtain suitable acceptance rates. Our fully Bayesian algorithm can easily handle missing values and be used for prediction at unobserved locations. Finally, in spite of the conditional independence assumption, non-trivial extremal dependence structures may be obtained by carefully specifying the dependence structure of the latent process $\Lambda(\boldsymbol{s})$. This contrasts with classical latent Gaussian models for extremes (Cooley *et al.*, 2007; Opitz *et al.*, 2018), which are limited for capturing strong extremal dependence.

The paper is organized as follows. In §2, we define our hierarchical construction of univariate distributions and characterize their tail behavior. Spatial hierarchical modeling is developed in §3, and we describe Bayesian inference and our MCMC implementation using latent variables in §4. An extensive simulation study is reported in §5 showing that our approach works well in various scenarios. We use our approach to model extreme episodes in daily precipitation measurements recorded at 150 sites in Germany in §6. Concluding remarks are given in §7.

2 Hierarchical models for threshold exceedances

2.1 Univariate tail properties in rate mixture constructions

For flexible sub-asymptotic tail modeling, we seek to replace the distributions in the hierarchical representation (2) of the GP distribution by more general parametric families that contain the exponential-gamma mixture as a special (or boundary) case.

Specifically, we construct new rate mixture models for the data $Y \sim F$ as follows. We consider a family of distributions $F_Y(\cdot;\lambda)$ with rate parameter λ and supported in $[0,+\infty)$, and an independent latent random variable $\Lambda \geq 0$, such that $Y \mid \Lambda \sim F_Y(\cdot;\Lambda)$. Equivalently,

we have the following ratio representation, which is useful for simulation and inference:

$$Y \mid \Lambda \overset{D}{=} \frac{\tilde{Y}}{\Lambda}, \quad \text{with} \quad \Lambda \geq 0 \perp\!\!\!\perp \tilde{Y} \geq 0, \quad \tilde{Y} \sim F_Y(\cdot; 1), \tag{3}$$

where " $\overset{D}{=}$ " means equality in distribution. The (unconditional) upper tail behavior of Y is determined by the interplay between the upper tail of $\tilde{Y}$ and the lower tail of Λ (i.e., the upper tail of $1/\Lambda$). We now shortly discuss two general and particularly interesting scenarios. Recall that a positive function $r(\cdot)$ is regularly varying at infinity with index $a \in \mathbb{R}$ if $r(tx)/r(t) \to x^a$ as $t \to \infty$ and $x > 0$; when $a = 0$, $r(\cdot)$ is slowly varying at infinity.

In the first scenario, we assume that $1/\Lambda$ in (3) has power-law tail decay, i.e., its distribution is regularly varying with index $-a < 0$, such that $\Pr(1/\Lambda > y) = r_0(y)y^{-a}$, $y > 0$, and $r_0(\cdot) > 0$ is a slowly varying function. If the distribution $F_Y(\cdot; 1)$ in (3) has a lighter upper tail than that of $1/\Lambda$, such that $\mathbb{E}(\tilde{Y}^{a+\varepsilon}) < \infty$ for some $\varepsilon > 0$ with $\tilde{Y} \sim F_Y(\cdot; 1)$, then Breiman's Lemma (Breiman, 1965) implies that

$$1 - F(y) = \Pr(Y > y) \sim \mathbb{E}(\tilde{Y}^a) \Pr(1/\Lambda > y), \quad y \to \infty. \tag{4}$$

Therefore, the heavier-tailed random factor $1/\Lambda$ in (3) dominates the tail behavior of Y in this case, while the lighter-tailed random factor $\tilde{Y}$ contributes to extreme survival probabilities only through a scaling factor.

In the second scenario, we assume that both $\tilde{Y}$ and $1/\Lambda$ in (3) have tails of Weibull type, which are lighter than power-law tails. Formally, we assume that there exist regularly varying functions $\tilde{r}$, r_Λ (with any index of regular variation), rate parameters $\tilde{\alpha}, \alpha_\Lambda > 0$ and shape parameters (also referred to as Weibull indexes) $\tilde{\eta}, \eta_\Lambda > 0$ such that

$$\Pr(\tilde{Y} > y) = \tilde{r}(y) \exp(-\tilde{\alpha} y^{\tilde{\eta}}), \quad \Pr(1/\Lambda > y) = r_\Lambda(y) \exp(-\alpha_\Lambda y^{\eta_\Lambda}). \tag{5}$$

Then, the variable Y constructed as in (3) also has a tail of Weibull type, with representation $\Pr(Y > y) = r_Y(y) \exp(-\alpha_Y y^{\eta_Y})$ similarly to (5). Its Weibull index is $\eta_Y = (\tilde{\eta}\eta_\Lambda)/(\tilde{\eta} + \eta_\Lambda) <$

$\min(\tilde{\eta}, \eta_\Lambda)$, such that the tail of Y always has a slower decay rate than that of each random factor $\tilde{Y}$ and $1/\Lambda$, while its rate parameter α_Y is given by

$$\alpha_Y = \tilde{\alpha}^{1-b}\alpha_\Lambda^b \left\{ \left(\frac{\alpha_\Lambda}{\tilde{\alpha}}\right)^b + \left(\frac{\tilde{\alpha}}{\alpha_\Lambda}\right)^{1-b} \right\}, \quad b = \frac{\tilde{\eta}}{\tilde{\eta} + \eta_\Lambda};$$

see Arendarczyk and Debicki (2011). In the following Sections 2.2 and 2.3, we exploit the rate mixture construction (3) and we propose new sub-asymptotic univariate tail models. In §2.2, our proposed model is heavy-tailed with the GP distribution as a special case and we focus on it for spatial modeling in Sections 3–6, while in §2.3, our proposed model is of Weibull type and has the GP distribution as a limiting boundary case.

2.2 Gamma-gamma model

Replacing the exponential distribution of $F_Y(\cdot; 1)$ in (2) by a gamma distribution yields the hierarchical gamma-gamma model, which may be written as

$$Y \mid \Lambda \sim \text{Gamma}(\Lambda, \beta_1), \quad \Lambda \sim \text{Gamma}(\alpha, \beta_2), \quad \alpha, \beta_1, \beta_2 > 0; \tag{6}$$

i.e., $F_Y(\cdot; \lambda)$ is the $\text{Gamma}(\lambda, \beta_1)$ distribution. The model (6) simplifies to the GP distribution obtained in (2) when $\beta_1 = 1$. The distribution of Y corresponds to a rescaled F_{ν_1,ν_2} distribution with degrees of freedom $\nu_1 = 2\beta_1$ and $\nu_2 = 2\beta_2$, and scaling factor $\alpha\beta_1/\beta_2$, such that $Y \sim (\alpha\beta_1/\beta_2)Z$, with $Z \sim F_{2\beta_1,2\beta_2}$. Its density is

$$f(y) = \alpha^{-\beta_1} \frac{\Gamma(\beta_1 + \beta_2)}{\Gamma(\beta_1)\Gamma(\beta_2)} \left(1 + \frac{y}{\alpha}\right)^{-(\beta_1+\beta_2)} y^{\beta_1 - 1}, \qquad y > 0. \tag{7}$$

The r-th moment of Y is finite whenever $\beta_2 > r$, and is given as

$$\mathbb{E}(Y^r) = \frac{\alpha^r\, \Gamma(\beta_1 + \beta_2)\, \Gamma(\beta_1 + r)\, \Gamma(\beta_2 - r)}{\Gamma(\beta_2)\, \Gamma(\beta_1)\, \Gamma(\beta_1 + \beta_2 + 2)}, \qquad \beta_2 > r.$$

From (4), or directly from (7), we deduce that the gamma-gamma model has a heavy power-law tail. The tail index of the limiting GP distribution (1) is equal to $\xi = 1/\beta_2$, and hence β_2 determines the tail decay rate of the distribution function F of Y. Further details are provided in the Supplementary Material.

2.3 Model extension with Weibull-type tail behavior

The gamma-gamma model (6) yields heavy tails (i.e., with a positive tail index, $\xi > 0$) and thus has a relatively slow power-law tail decay. For data with a light upper tail (i.e., with a tail index equal to zero, $\xi = 0$), we now discuss a flexible model extension based on the hierarchical construction (3), which provides a faster tail decay than the gamma-gamma model, while keeping the heavy-tailed GP distribution on the boundary of the parameter space. Specifically, we propose the following hierarchical model:

$$\begin{aligned} Y^{1/k} \mid \Lambda &\sim \text{Gamma}(\Lambda, \beta_1), \quad k, \beta_1 > 0, \\ \Lambda &\sim \text{GIG}(\alpha/2, b, \beta_2), \quad (\alpha, b, \beta_2) \in D_{\text{GIG}}, \end{aligned} \tag{8}$$

where the latent rate parameter Λ is assumed to follow the generalized inverse Gaussian (GIG) distribution with parameters $\alpha/2$, b and β_2, and where D_{GIG} denotes its parameter space. More precisely, the $\text{GIG}(a, b, \beta)$ density is $g(y) = (a/b)^{\beta/2}\{2K_\beta(\sqrt{ab})\}^{-1} y^{\beta-1} \exp\{-(ay + b/y)/2\}$, $a \geq 0$, $b \geq 0$, $y > 0$, with parameter constraints on β given by $-\infty < \beta < \infty$ if $a, b > 0$, by $\beta > 0$ if $b = 0$ and $a > 0$, and by $\beta < 0$ if $a = 0$ and $b \geq 0$, and where K_β denotes the modified Bessel function of second kind with parameter β. The GIG distribution has an exponentially decaying tail (i.e., Weibull-type tail with Weibull index one).

Model (8) can also be represented as the ratio $Y = (\tilde{Y}/\Lambda)^k$ with $\tilde{Y} \sim \text{Gamma}(1, \beta_1)$ independent of $\Lambda \sim \text{GIG}(\alpha/2, b, \beta_2)$. This model generalizes the gamma-gamma construction in (6), which is on the boundary of the parameter space with $b = 0$, $k = 1$ and $\beta_2 > 0$. Hence, the model captures a wide range of tail behaviors, from very light tails to relatively heavy tails. Specifically, when $b > 0$, the random variables $\tilde{Y}^k$ and $1/\Lambda^k$ have Weibull-type tails, with Weibull indexes both equal to $1/k$. From (5), we deduce that Y has Weibull index $\eta_Y = (1/k^2)/(2/k) = 1/(2k) > 0$. Thus, when $b > 0$, this model can capture any Weibull tail with any Weibull index, while when $b = 0$, it can capture any power-law tail with any positive tail index $k/\beta_2 > 0$, thanks to Breiman's Lemma (4).

3 A Bayesian spatial gamma-gamma model

3.1 Bayesian hierarchical modeling framework

Accounting for spatial dependence is important for a variety of reasons, even if the precise estimation of the extremal dependence structure is of secondary importance. First, this allows to borrow strength across locations to reduce the uncertainty and improve the estimation of marginal distributions and of high quantiles. Second, a proper spatial model is needed whenever prediction at unobserved locations is required.

Let $Y(\boldsymbol{s})$, $\boldsymbol{s} \in \mathcal{S} \subset \mathbb{R}^2$, be the spatial process of interest, and assume that we observe it at finite set of locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_d \in \mathcal{S}$. There are different approaches to model the dependence structure of $\boldsymbol{Y} = (Y_1, \ldots, Y_d)^T$, where $Y_j = Y(\boldsymbol{s}_j) \sim F_{Y_j}$. One possibility is to directly bind together the components $Y_1, \ldots, Y_d$ through a *copula* model (i.e., a multivariate distribution with standard uniform margins) without assuming any hierarchical structure; we call this the *copula approach*. Let $C_{\boldsymbol{Y}}$ denote the underlying copula of the data, which is unique if $\boldsymbol{Y}$ has a continuous distribution. Then $\boldsymbol{Y}$ has distribution function $F(y_1, \ldots, y_d) = C_{\boldsymbol{Y}}\{F_{Y_1}(y_1), \ldots, F_{Y_d}(y_d)\}$ and density $f(y_1, \ldots, y_d) = c_{\boldsymbol{Y}}\{F_{Y_1}(y_1), \ldots, F_{Y_d}(y_d)\} \prod_{j=1}^{d} f_{Y_j}(y_j)$, where $c_{\boldsymbol{Y}}$ is the copula density and $f_{Y_1}, \ldots, f_{Y_d}$ are the marginal densities. To model threshold exceedances with respect to a fixed threshold vector $\boldsymbol{u} = (u_1, \ldots, u_d)^T$, it is common to censor observations Y_j falling below a corresponding threshold u_j. In the copula approach, the likelihood contribution of an observation $\boldsymbol{y} = (y_1, \ldots, y_d)^T$ such that $y_j \geq u_j$, $j = 1, \ldots, j_0$, and $y_j < u_j$, $j = j_0 + 1, \ldots, d$, is

$$\frac{\partial^{j_0}}{\partial y_1, \ldots, \partial y_{j_0}} F(\boldsymbol{y}) \bigg|_{y_{j_0+1}=u_{j_0+1}, \ldots, y_d=u_d} = \int_{-\infty}^{u_{j_0+1}} \cdots \int_{-\infty}^{u_d} f(y_1, \ldots, y_d) \mathrm{d}y_{j_0+1} \ldots \mathrm{d}y_d. \quad (9)$$

When $j_0 < d$, multivariate distribution functions must be calculated to evaluate (9). For many copula models (e.g., Gaussian, Student's t), this requires expensive multivariate numerical integrations. Then, the computational cost can become prohibitively high if the

dimension d is large, and accuracy issues may arise.

Here, we instead privilege the hierarchical approach, where the process $Y(\boldsymbol{s})$ is assumed to be conditionally independent given a latent process $\Lambda(\boldsymbol{s})$ with spatial dependence. Let $\Lambda_j = \Lambda(\boldsymbol{s}_j)$, $j = 1, \ldots, d$, and $\boldsymbol{\Lambda} = (\Lambda_1, \ldots, \Lambda_d)^T$. We assume that, conditional on $\boldsymbol{\Lambda}$, an observation Y_j is independent of the other observations $Y_{j'}$, $j' \neq j$. The vector of latent variables $\boldsymbol{\Lambda}$, on the other hand, is specified through a copula $C_{\boldsymbol{\Lambda}}$. This *hierarchical approach* with a *latent copula* separates the observed process $\boldsymbol{Y}$ from the latent variables $\boldsymbol{\Lambda}$. Although the conditional independence assumption may be seen as a restriction, it still permits to capture a wide range of dependence structures (see §3.4), and it has significant computational benefits. By augmenting the data $\boldsymbol{Y}$ with the latent variables $\boldsymbol{\Lambda}$, the censored likelihood (9) can be formulated in terms of univariate censored terms for the conditionally independent components $Y_j \mid \Lambda_j$, while no censoring is required for $\boldsymbol{\Lambda}$. In §4.2, we exploit this latent variable approach for fully Bayesian inference using Markov chain Monte Carlo (MCMC).

We now describe our proposed Bayesian spatial modeling framework in more detail. Let $\boldsymbol{\Theta} = (\boldsymbol{\Theta}_{\boldsymbol{Y}}^T, \boldsymbol{\Theta}_{\boldsymbol{\Lambda}}^{\text{mar}^T}, \boldsymbol{\Theta}_{\boldsymbol{\Lambda}}^{\text{cop}^T})^T$ be the vector of unknown hyperparameters, where $\boldsymbol{\Theta}_{\boldsymbol{Y}}$ controls the conditional distribution of observations, and $\boldsymbol{\Theta}_{\boldsymbol{\Lambda}} = (\boldsymbol{\Theta}_{\boldsymbol{\Lambda}}^{\text{mar}^T}, \boldsymbol{\Theta}_{\boldsymbol{\Lambda}}^{\text{dep}^T})^T$ contains parameters for the latent process $\Lambda(\boldsymbol{s})$, with $\boldsymbol{\Theta}_{\boldsymbol{\Lambda}}^{\text{mar}}$ controlling marginal distributions and $\boldsymbol{\Theta}_{\boldsymbol{\Lambda}}^{\text{dep}}$ controlling the dependence structure. Our general spatial hierarchical construction is specified as

$$
\begin{aligned}
Y_j \mid \boldsymbol{\Lambda}, \boldsymbol{\Theta}_{\boldsymbol{Y}} &\overset{\text{ind}}{\sim} F_Y(\cdot\,; \Lambda_j, \boldsymbol{\Theta}_{\boldsymbol{Y}}), \qquad j = 1, \ldots, d, \\
\boldsymbol{\Lambda} \mid \boldsymbol{\Theta}_{\boldsymbol{\Lambda}} &\sim C_{\boldsymbol{\Lambda}} \left\{ F_{\Lambda_1}(\cdot; \boldsymbol{\Theta}_{\boldsymbol{\Lambda}}^{\text{mar}}), \ldots, F_{\Lambda_d}(\cdot; \boldsymbol{\Theta}_{\boldsymbol{\Lambda}}^{\text{mar}}); \boldsymbol{\Theta}_{\boldsymbol{\Lambda}}^{\text{dep}} \right\}, \\
\boldsymbol{\Theta} &\sim \pi(\boldsymbol{\Theta}),
\end{aligned}
\tag{10}
$$

where $C_{\boldsymbol{\Lambda}}$ refers to the spatial copula of $\boldsymbol{\Lambda}$, $F_{\Lambda_j}(\cdot; \boldsymbol{\Theta}_{\boldsymbol{\Lambda}}^{\text{mar}})$ denotes the marginal distribution of Λ_j, $j = 1, \ldots, d$, and $\pi(\boldsymbol{\Theta})$ is the prior distribution of the parameter vector $\boldsymbol{\Theta}$. The joint distribution of $\boldsymbol{Y}$, $\boldsymbol{\Lambda}$, and $\boldsymbol{\Theta}$ can be decomposed into conditional distributions as

$\pi(\boldsymbol{Y}, \boldsymbol{\Lambda}, \boldsymbol{\Theta}_{\boldsymbol{Y}}, \boldsymbol{\Theta}_{\boldsymbol{\Lambda}}) = \pi(\boldsymbol{Y} \mid \boldsymbol{\Lambda}, \boldsymbol{\Theta}_{\boldsymbol{Y}})\ \pi(\boldsymbol{\Lambda} \mid \boldsymbol{\Theta}_{\boldsymbol{\Lambda}})\ \pi(\boldsymbol{\Theta})$, where, $\pi(\cdot)$ denotes a generic (conditional) distribution. The joint posterior distribution $\pi(\boldsymbol{\Lambda}, \boldsymbol{\Theta} \mid \boldsymbol{Y})$ of latent variables $\boldsymbol{\Lambda}$ and hyperparameters $\boldsymbol{\Theta}$ is then proportional to $\pi(\boldsymbol{Y}, \boldsymbol{\Lambda}, \boldsymbol{\Theta})$, and the posterior distribution of hyperparameters $\boldsymbol{\Theta}$ is obtained by integrating out the latent parameters $\boldsymbol{\Lambda}$, i.e.,

$$\pi(\boldsymbol{\Theta} \mid \boldsymbol{Y}) = \int_{\boldsymbol{\Lambda}} \pi(\boldsymbol{\Lambda}, \boldsymbol{\Theta} \mid \boldsymbol{Y})\ \mathrm{d}\boldsymbol{\Lambda}. \tag{11}$$

The dimension of the integration domain in (11) can be very high. We solve this issue in §4 by implementing an MCMC algorithm, where the latent variables $\boldsymbol{\Lambda}$ are imputed and updated at each iteration.

3.2 A spatial gamma-gamma model

We now present a specific Bayesian hierarchical model of the form (10) that is based on the gamma-gamma construction (6) with a latent Gaussian copula process for the spatial dependence in $\boldsymbol{\Lambda}$. Here, $\boldsymbol{\Theta}_Y = \beta_1 > 0$, $\boldsymbol{\Theta}_{\boldsymbol{\Lambda}}^{\text{dep}}$ contains the correlation parameters of the latent process $\Lambda(\boldsymbol{s})$, and $\boldsymbol{\Theta}_{\boldsymbol{\Lambda}}^{\text{mar}} = (\alpha, \beta_2)^T \in (0, \infty)^2$. We focus here on the isotropic exponential correlation function $\sigma(h) = \exp(-h/\rho),\ h \geq 0$, with range $\rho > 0$, so that $\boldsymbol{\Theta}_{\boldsymbol{\Lambda}}^{\text{dep}} = \rho$, although other correlation functions are also possible. We write Φ for the univariate standard normal distribution and Φ_ρ for the zero mean and unit variance multivariate normal distribution associated to any collection of d sites $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_d$, and parametrized by the correlation matrix $\boldsymbol{\Sigma}(\rho)$ with entries $\boldsymbol{\Sigma}_{i_1 i_2} = \sigma(\|\boldsymbol{s}_{i_1} - \boldsymbol{s}_{i_2}\|) = \exp(-\|\boldsymbol{s}_{i_2} - \boldsymbol{s}_{i_1}\|/\rho)$, $1 \leq i_1, i_2 \leq d$. The Gamma(α, β) distribution function is denoted by $\Gamma(\,\cdot\,; \alpha, \beta)$. Following (10), we define the gamma-gamma hierarchical model with latent Gaussian copula as

$$\begin{aligned} Y_j \mid \boldsymbol{\Lambda}, \boldsymbol{\Theta}_{\boldsymbol{Y}} &\overset{\text{ind}}{\sim} \Gamma(\cdot; \Lambda_j, \beta_1), \quad j = 1, \ldots, d, \\ \boldsymbol{\Lambda} \mid \boldsymbol{\Theta}_{\boldsymbol{\Lambda}} &\sim \Phi_\rho[\Phi^{-1}\{\Gamma(\,\cdot\,; \alpha, \beta_2)\}, \ldots, \Phi^{-1}\{\Gamma(\,\cdot\,; \alpha, \beta_2)\}], \\ \boldsymbol{\Theta} &\sim \pi(\boldsymbol{\Theta}) = \pi(\alpha) \times \pi(\beta_1) \times \pi(\beta_2) \times \pi(\rho). \end{aligned} \tag{12}$$

While a Gaussian copula is specified in (12), other copula models with stronger tail dependence (e.g., the elliptically symmetric Student's t copula with $\nu > 0$ degrees of freedom and dispersion matrix $\boldsymbol{\Sigma}(\rho)$) are also possible. The implied extremal dependence structure is discussed in §3.4. Covariate information may be included in various ways into the model parameters. Here, because α describes the scale of the marginal distribution of the process $Y(\boldsymbol{s})$, a natural approach is to use a log-linear specification of the form $\log \alpha = \log \alpha_0 + \alpha_1 x_1 + \ldots + \alpha_p x_p$ for some known covariates $x_1, \ldots, x_p$. We next describe our choice of prior distributions for the hyperparameters of the gamma-gamma model (12).

3.3 Prior distributions for hyperparameters

Appropriate prior distributions in the model (12) need to be designed for the components of the hyperparameter vector $\boldsymbol{\Theta}$, and special care is required for the shape parameters β_1 and β_2, which represent the "distance" to the GP model and the upper tail decay rate, respectively. A possible choice is to select an informative prior distribution $\pi(\beta_1)$ for $\beta_1 > 0$ that shrinks our mixture model (6) towards the GP with $\beta_1 = 1$, believed to be valid in the limit. This can be achieved through the concept of penalized-complexity (PC) priors (Simpson *et al.*, 2017). PC priors assume a constant-rate exponential distribution for the square root of the Kullback-Leibler divergence with respect to a simpler reference model. Let $\gamma(\cdot; \lambda, \beta_1)$ be the Gamma(λ, β_1) density and $\gamma(\cdot; \lambda, 1)$ be the Exp(λ) density. The (asymmetric) Kullback-Leibler divergence of $\gamma(\cdot; \lambda, \beta_1)$ with respect to $\gamma(\cdot; \lambda, 1)$ is

$$\begin{aligned} \mathrm{KLD}\left\{\gamma(\cdot;\lambda,\beta_1) \| \gamma(\cdot;\lambda,1)\right\} &= \int_0^\infty \log\left\{\frac{\gamma(y;\lambda,\beta_1)}{\gamma(y;\lambda,1)}\right\}\gamma(y;\lambda,\beta_1)\mathrm{d}y, \\ &= (\beta_1 - 1)\psi(\beta_1) - \log\{\Gamma(\beta_1)\}, \end{aligned} \tag{13}$$

where $\psi(\beta_1) = \mathrm{d}\log\{\Gamma(\beta_1)\}/\mathrm{d}\beta_1$ denotes the polygamma function of order 0, also known as the digamma function. From (13), the derivative of the Kullback-Leibler divergence is $\frac{\mathrm{d}}{\mathrm{d}\beta_1}\mathrm{KLD}\{\gamma(\cdot;\lambda,\beta_1)\|\gamma(\cdot;\lambda,1)\} = (\beta_1 - 1)\psi'(\beta_1)$ where $\psi'(\beta_1) = \mathrm{d}\psi(\beta_1)/\mathrm{d}\beta_1$ is the polygamma

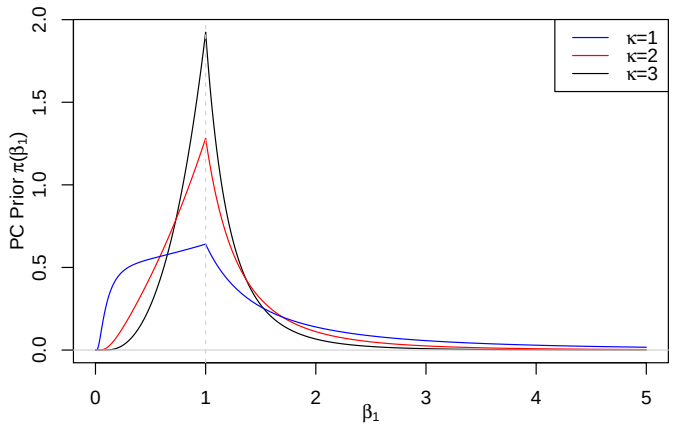

Figure 1: PC prior $\pi(\beta_1)$, $\beta_1 > 0$, as derived in (14), for penalty rates $\kappa = 1, 2, 3$ (blue, red and black curves, respectively).

function of order 1. Writing $d(\beta_1) = \sqrt{2\mathrm{KLD}(\beta_1)}$, we deduce that the corresponding PC prior is a mixture of two densities defined over $0 < \beta_1 < 1$ and $\beta_1 > 1$, i.e.,

$$\begin{aligned}\pi(\beta_1) &= \frac{\kappa}{2}\exp\{-\kappa\, d(\beta_1)\}\left|\frac{\mathrm{d}}{\mathrm{d}\beta_1}d(\beta_1)\right| \\ &= \frac{\kappa}{2}\exp\{-\kappa\sqrt{2(\beta_1-1)\psi(\beta_1) - 2\log\{\Gamma(\beta_1)\}}\}\left|\frac{(\beta_1-1)\psi'(\beta_1)}{\sqrt{2(\beta_1-1)\psi(\beta_1) - 2\log\{\Gamma(\beta_1)\}}}\right|,\end{aligned} \tag{14}$$

for $\beta_1 > 0$, and where $\kappa > 0$ is a predetermined penalty rate. The PC prior (14) is displayed in Figure 1 for $\kappa = 1, 2, 3$. As expected, the mass is concentrated near $\beta_1 = 1$.

The prior distribution for β_2 is more conveniently constructed through the reparametrization given by the tail index $\xi = 1/\beta_2$. It makes sense to prevent very heavy tails by shrinking ξ towards zero (i.e., β_2 towards infinity), which corresponds to an exponential GP distribution in (1). Opitz *et al.* (2018) derived the PC-prior for ξ, which may be written as

$$\pi(\xi) = \sqrt{2}\kappa\exp\left\{-\sqrt{2}\kappa\,\xi(1-\xi)^{-1/2}\right\}(1-\xi/2)(1-\xi)^{-3/2}, \quad 0 < \xi < 1, \tag{15}$$

where κ is the penalty rate. As (15) is compactly supported over the interval $(0, 1)$, it prevents infinite-mean models. A change of variables establishes that the PC prior for β_2 is

$$\pi(\beta_2) = \sqrt{2}\kappa\exp\left[-\sqrt{2}\kappa\{\beta_2(\beta_2-1)\}^{-1/2}\right](\beta_2 - 1/2)\{\beta_2(\beta_2-1)\}^{-3/2}, \quad \beta_2 > 1. \tag{16}$$

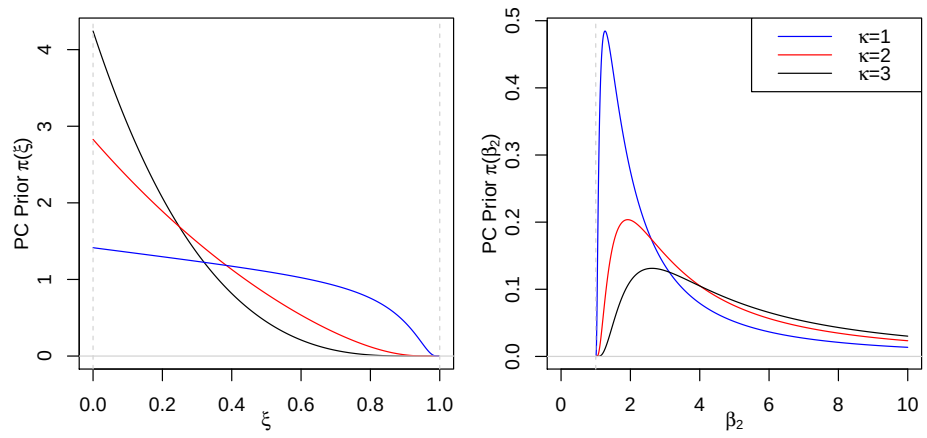

Figure 2: PC priors $\pi(\xi)$, $0 < \xi < 1$ (left) and $\pi(\beta_2)$, $\beta_2 > 1$ (right) as derived in (15) and (16), respectively, for penalty rates $\kappa = 1, 2, 3$ (blue, red and black curves, respectively).

Both PC priors (15) and (16) are illustrated in Figure 2 for $\kappa = 1, 2, 3$.

We specify vague priors for the other hyperparameters. More explicitly, we choose a gamma distribution with mean 1 and variance 100 for the correlation range ρ and a Gaussian distribution with mean 0 and variance 100 for the covariate parameters $\log \alpha_0, \alpha_1, \ldots, \alpha_p$.

3.4 Joint tail behavior

The upper-tail dependence in the hierarchical model with latent copula (10) is determined by the interplay between the lower joint tail of $\boldsymbol{\Lambda} = (\Lambda_1, \ldots, \Lambda_d)^T$ and the univariate upper tail of the distribution $F_Y(\cdot; 1)$ of the independent random variables $\tilde{Y}_j$, $j = 1, \ldots, d$, in (3).

We here provide more details for the case where $1/\Lambda(\boldsymbol{s})$ has regularly varying distribution with positive tail index ξ, and $\tilde{Y}$ is lighter-tailed such that $\mathbb{E}(\tilde{Y}^{1/\xi+\varepsilon}) < \infty$ for some $\varepsilon > 0$, which includes the gamma-gamma model (12). If, in addition, the multivariate distribution $F_{1/\boldsymbol{\Lambda}}$ of $1/\boldsymbol{\Lambda}$ is regularly varying at infinity (Resnick, 1987), we have

$$\frac{1 - F_{1/\boldsymbol{\Lambda}}(t\boldsymbol{y})}{1 - F_{1/\boldsymbol{\Lambda}}(t\mathbf{1})} \to V_{1/\boldsymbol{\Lambda}}(\boldsymbol{y}), \quad \boldsymbol{y} > \mathbf{0}, \quad t \to \infty,$$

where $\mathbf{1} = (1, \ldots, 1)^T \in \mathbb{R}^d$ and $V_{1/\boldsymbol{\Lambda}}(\boldsymbol{y})$ is some positive limit function. Theorem 3 of

Fougeres and Mercadier (2012) then implies multivariate regular variation of $F_{\boldsymbol{Y}}$, i.e.,

$$\frac{1-F_{\boldsymbol{Y}}(t\boldsymbol{y})}{1-F_{\boldsymbol{Y}}(t\boldsymbol{1})} \to V_{\boldsymbol{Y}}(\boldsymbol{y}) = \int_0^\infty \dots \int_0^\infty V_{1/\boldsymbol{\Lambda}}(\boldsymbol{y}/\boldsymbol{x}) \prod_{j=1}^{d} F_Y(\mathrm{d}x_j; 1), \qquad \boldsymbol{y} > \boldsymbol{0}, \quad t \to \infty. \quad (17)$$

The functions $V_{1/\boldsymbol{\Lambda}}$ and $V_{\boldsymbol{Y}}$ are homogeneous of order $-1/\xi$, i.e., $V_{1/\boldsymbol{\Lambda}}(t\boldsymbol{y}) = t^{-1/\xi} V_{1/\boldsymbol{\Lambda}}(\boldsymbol{y})$ and $V_{\boldsymbol{Y}}(t\boldsymbol{y}) = t^{-1/\xi} V_{\boldsymbol{Y}}(\boldsymbol{y})$ for positive values of t and $\boldsymbol{y}$. Equation (17) fully characterizes the extremal dependence structure of the process $Y(\boldsymbol{s})$ resulting from the construction (10) in the heavy-tailed case. Let $Y_1 \sim F_{Y_1}$ and $Y_2 \sim F_{Y_2}$, then a summary of the extremal dependence strength is the coefficient $\chi = \lim_{u \to 1} \chi(u)$, with $\chi(u) = \Pr\{Y_1 > F_{Y_1}^{-1}(u) \mid Y_2 > F_{Y_2}^{-1}(u)\}$. It can be shown that $\chi = 2 - V_{\boldsymbol{Y}}[\{V_{\boldsymbol{Y}}(\infty, 1)\}^\xi, \{V_{\boldsymbol{Y}}(1, \infty)\}^\xi]$, where $\boldsymbol{Y} = (Y_1, Y_2)^T$. The pair of variables $\boldsymbol{Y}$ is called asymptotically independent if $\chi = 0$ and asymptotically dependent if $\chi > 0$. The case of asymptotic independence corresponds to a $V_{\boldsymbol{Y}}$ function that is a sum of separate terms for the components, i.e., $V_{\boldsymbol{Y}}(y_1, y_2) = c(y_1^{-1/\xi} + y_1^{-1/\xi})$ with a constant $c > 0$. From (17), it follows that $\boldsymbol{Y}$ is asymptotically independent if and only if $1/\boldsymbol{\Lambda} = (1/\Lambda_1, 1/\Lambda_2)^T$ is asymptotically independent. Therefore the gamma-gamma model (12) with latent Gaussian copula is asymptotically independent. When the Student t copula with $\nu > 0$ degrees of freedom is used instead, the process $Y(\boldsymbol{s})$ becomes asymptotically dependent (despite the conditional independence assumption at the data level).

To illustrate the dependence strength of the gamma-gamma model (12), we compute $\chi(u)$, with $u \in (0, 1)$, by simulation for different parameter values. The left panel of Figure 3 shows $\chi(u)$ obtained at spatial distance 0.5 using a Gaussian copula with correlation range $\rho = 1$, and the other hyperparameters set to $\alpha = 1$, $\beta_1 = 0.5, 1, 5, 50, 100$, $\beta_2 = 2.5$ (i.e., $\xi = 0.4$). The right panel of Figure 3 shows $\chi(u)$ obtained using a Student t latent copula with range $\rho = 1$ and degrees of freedom $\nu = 0.5, 1, 5, 10, \infty$ (Gaussian), and the other hyperparameters set to $\alpha = 1$, $\beta_1 = 50$, $\beta_2 = 2.5$. These plots demonstrate that our hierarchical modeling approach can capture various joint tail decay rates and extremal dependence structures.

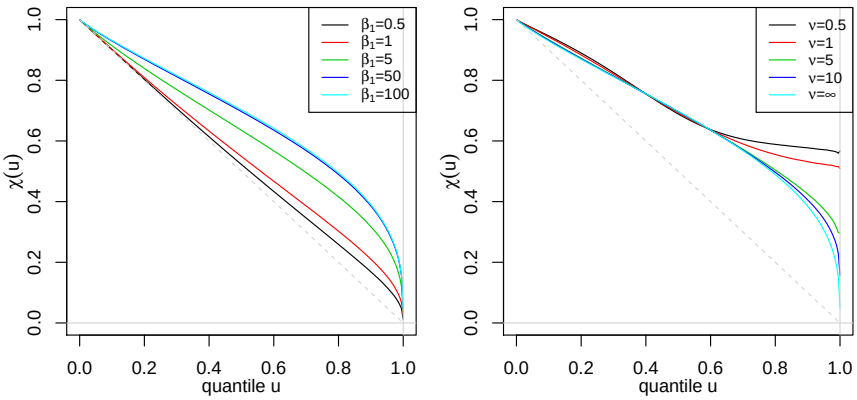

Figure 3: Plot of $\chi(u)$ against the quantile $u \in (0,1)$ for the gamma-gamma model (12) at spatial distance 0.5 with latent Gaussian copula (left) and parameters $\alpha = 1, \beta_1 = 0.5, 1, 5, 50, 100, \beta_2 = 2.5$ and $\rho = 1$, and Student t copula (right) with degrees of freedom $\nu = 0.5, 1, 5, 10, \infty$ (Gaussian) and $\alpha = 1$, $\beta_1 = 50$, $\beta_2 = 2.5$ and $\rho = 1$. The dashed diagonal gray lines correspond to exact independence.

4 Simulation-based Bayesian inference

4.1 General strategy

We use Markov chain Monte Carlo (MCMC) sampling to generate a representative posterior sample of the hyperparameter vector $\boldsymbol{\Theta}$ and the latent variables $\boldsymbol{\Lambda}$ involved in the hierarchical model (10), conditional on observed data. As we fit the model to threshold exceedances, we first describe the censored likelihood mechanism with latent variables in §4.2, focusing on the gamma-gamma model (12). Then in §4.3 we develop an efficient MCMC sampler by using the Metropolis-adjusted Langevin algorithm (MALA) for generating MCMC block proposals that ensure a relatively fast exploration of the high-dimensional parameter space of $\boldsymbol{\Lambda}$. Also, we propose an adaptive algorithm to tune the calibration parameters of the MALA and random walk proposals for an appropriate convergence rate.

4.2 Censored likelihood with latent variables

We suppose that observed data $Y_i(\boldsymbol{s}_j)$, $i = 1, \ldots, n$, $j = 1, \ldots, d$, are composed of n independent time replicates of the d components of a random vector $\boldsymbol{Y} = \{Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_d)\}^T$

indexed by locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_d$. We discuss here inference for the spatial hierarchical gamma-gamma model with latent Gaussian copula in (12), although little would change for other hierarchical models of the form (10). We write $y_{ij} = Y_i(\boldsymbol{s}_j), \lambda_{ij} = \Lambda_i(\boldsymbol{s}_j)$, $i = 1, \ldots, n$, $j = 1, \ldots, d$, and use the symbols ϕ and ϕ_ρ for the univariate and multivariate Gaussian densities corresponding to Φ and Φ_ρ, respectively. Given a data vector $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{id})^T$ and a fixed threshold vector $\boldsymbol{u}_i = (u_{i1}, \ldots, u_{id})^T \in [0, \infty)^d$, we introduce the exceedance indicator vector $\boldsymbol{e}_i = (e_{i1}, \ldots, e_{id})^T$ with $e_{ij} = 1$ if $y_{ij} > u_{ij}$ and $e_{ij} = 0$ otherwise. If $u_{ij} = 0$, no censoring is applied to the value y_{ij}, on the other hand, if $u_{ij} = \infty$ then the observation y_{ij} is treated as fully censored. This may be used to handle missing data and prediction at unobserved locations. We now give the augmented censored likelihood contribution of $\boldsymbol{y}_i$ where we consider both $\boldsymbol{\Theta}$ and $\boldsymbol{\lambda}_i = (\lambda_{i1}, \ldots, \lambda_{id})^T$ as parameters. The density of observations (y_{ij}, e_{ij}) conditional on λ_{ij} is $f_c(y_{ij}, e_{ij}; \lambda_{ij}, \beta_1) = \Gamma(u_{ij}; \lambda_{ij}, \beta_1)$ if $e_{ij} = 0$ and $f_c(y_{ij}, e_{ij}; \lambda_{ij}, \beta_1) = \gamma(y_{ij}; \lambda_{ij}, \beta_1)$ if $e_{ij} = 1$, where $\gamma(\,\cdot\,; \lambda_{ij}, \beta_1)$ is the gamma density with parameters λ_{ij} and β_1. The augmented censored likelihood contribution for the data vector $(\boldsymbol{y}_i^T, \boldsymbol{e}_i^T)^T$ is thus

$$
\begin{aligned}
L(\boldsymbol{\Theta}, \boldsymbol{\lambda}_i; \boldsymbol{y}_i, \boldsymbol{e}_i) = & \prod_{j=1}^{d} f_c(y_{ij}, e_{ij}; \lambda_{ij}, \beta_1) \\
& \times \phi_\rho[\Phi^{-1}\{\Gamma(\lambda_{i1}; \alpha, \beta_2)\}, \ldots, \Phi^{-1}\{\Gamma(\lambda_{id}; \alpha, \beta_2)\}] \times \prod_{j=1}^{d} \frac{\gamma(\lambda_{ij}; \alpha, \beta_2)}{\phi[\Phi^{-1}\{\Gamma(\lambda_{ij}; \alpha, \beta_2)\}]},
\end{aligned} \tag{18}
$$

where the first line refers to the observation model and the second line to the latent model. The overall augmented censored likelihood is

$$
L_n(\boldsymbol{\Theta}, \boldsymbol{\lambda}; \boldsymbol{y}, \boldsymbol{e}) = \prod_{i=1}^{n} L(\boldsymbol{\Theta}, \boldsymbol{\lambda}_i; \boldsymbol{y}_i, \boldsymbol{e}_i), \tag{19}
$$

where $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1^T, \ldots, \boldsymbol{\lambda}_n^T)^T$, $\boldsymbol{y} = (\boldsymbol{y}_1^T, \ldots, \boldsymbol{y}_n^T)^T$, and $\boldsymbol{e} = (\boldsymbol{e}_1^T, \ldots, \boldsymbol{e}_n^T)^T$. Notice that thanks to data augmentation and to the conditional independence assumption, only univariate censoring is required, hence facilitating computations.

4.3 Metropolis–Hastings MCMC algorithm with adaptive MALA and random walk proposals

We implement a Metropolis–Hastings MCMC algorithm to sample from the posterior distribution of hyperparameters $\boldsymbol{\Theta}$ and latent variables $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1^T, \ldots, \boldsymbol{\lambda}_n^T)^T$. More precisely, we update the parameters $\boldsymbol{\Theta}$ and $\boldsymbol{\lambda}$ in two separate blocks for a predetermined number of iterations, in order to construct a Markov chain whose stationary distribution is the posterior distribution of interest. To avoid invalid proposals or strong dependence between posterior samples of latent parameters or hyperparameters, we first apply the following reparametrization of the model: the latent parameters are log-transformed, i.e., $\tilde{\boldsymbol{\lambda}} = \log(\boldsymbol{\lambda})$, while the hyperparameters of the gamma-gamma model (12) are reparametrized (internally) as

$$\tilde{\alpha} = \log(\alpha\beta_1/\beta_2), \quad \tilde{\beta}_1 = \log(\alpha\beta_1^2/\beta_2), \quad \tilde{\beta}_2 = \log(1/\beta_2), \quad \tilde{\rho} = \log(\rho). \tag{20}$$

The reverse transformation is $\alpha = \exp(2\tilde{\alpha})\exp(-\tilde{\beta}_1)\exp(-\tilde{\beta}_2)$, $\beta_1 = \exp(-\tilde{\alpha})\exp(\tilde{\beta}_1)$, $\beta_2 = \exp(-\tilde{\beta}_2)$, $\rho = \exp(\tilde{\rho})$. To use this modified internal reparametrization, we correct the target posterior density $\pi_{\text{post}}(\boldsymbol{\Theta}, \boldsymbol{\lambda})$ through the determinant of the Jacobian matrix of the transformation; its value is $\exp(\tilde{\rho} + \tilde{\alpha} - 2\tilde{\beta}_2)$ for the hyperparameter transformation in (20).

Our proposed MCMC algorithm consists of the following steps: we iteratively propose candidate values for the transformed hyperparameters $\tilde{\boldsymbol{\Theta}}$ and latent parameters $\tilde{\boldsymbol{\lambda}}$ from some proposal densities $q_1(\tilde{\boldsymbol{\Theta}}' \mid \tilde{\boldsymbol{\Theta}})$ and $q_2(\tilde{\boldsymbol{\lambda}}' \mid \tilde{\boldsymbol{\lambda}})$, respectively, and we accept these candidates with probability

$$\min\left(1, \frac{L_n(\tilde{\boldsymbol{\Theta}}', \tilde{\boldsymbol{\lambda}}; \boldsymbol{y}, \boldsymbol{e})\, \pi(\tilde{\boldsymbol{\Theta}}')\, q_1(\tilde{\boldsymbol{\Theta}} \mid \tilde{\boldsymbol{\Theta}}')}{L_n(\tilde{\boldsymbol{\Theta}}, \tilde{\boldsymbol{\lambda}}; \boldsymbol{y}, \boldsymbol{e})\, \pi(\tilde{\boldsymbol{\Theta}})\, q_1(\tilde{\boldsymbol{\Theta}}' \mid \tilde{\boldsymbol{\Theta}})}\right), \qquad \min\left(1, \frac{L_n(\tilde{\boldsymbol{\Theta}}, \tilde{\boldsymbol{\lambda}}'; \boldsymbol{y}, \boldsymbol{e})\, q_2(\tilde{\boldsymbol{\lambda}} \mid \tilde{\boldsymbol{\lambda}}')}{L_n(\tilde{\boldsymbol{\Theta}}, \tilde{\boldsymbol{\lambda}}; \boldsymbol{y}, \boldsymbol{e})\, q_2(\tilde{\boldsymbol{\lambda}}' \mid \tilde{\boldsymbol{\lambda}})}\right),$$

for hyperparameters and latent parameters, respectively. The number of parameters (latent variables and hyperparameters) to be explored by the Markov chain is equal to $N = nd + l$, with $l = |\tilde{\boldsymbol{\Theta}}|$. In particular, it grows linearly with the sample size n and dimension d. To handle the high dimensionality of the vector of latent variables, we propose using the

Metropolis-adjusted Langevin algorithm (MALA), which exploits the gradient of the log-posterior density evaluated at the current parameter configuration to design an efficient multivariate Gaussian proposal density $q_2(\tilde{\boldsymbol{\lambda}}' \mid \tilde{\boldsymbol{\lambda}})$. Because the number of hyperparameters is moderate, we specify simple random walk proposals for $q_1(\tilde{\boldsymbol{\Theta}}' \mid \tilde{\boldsymbol{\Theta}})$. Specifically, we propose candidate values $\tilde{\boldsymbol{\Theta}}'$ and $\tilde{\boldsymbol{\lambda}}'$ consecutively as follows:

$$\tilde{\boldsymbol{\Theta}}' \mid \tilde{\boldsymbol{\Theta}} \sim \mathcal{N}(\tilde{\boldsymbol{\Theta}}, \tau_{\boldsymbol{\Theta}} I_l),$$
$$\tilde{\boldsymbol{\lambda}}' \mid \tilde{\boldsymbol{\Theta}}, \tilde{\boldsymbol{\lambda}} \sim \mathcal{N}(\tilde{\boldsymbol{\lambda}} + \tau_{\boldsymbol{\lambda}} \nabla_{\tilde{\boldsymbol{\lambda}}} \log \pi_{\text{post}}(\tilde{\boldsymbol{\Theta}}, \tilde{\boldsymbol{\lambda}}), 2\tau_{\boldsymbol{\lambda}} I_{nd}),$$

where I_l and I_{nd} are the identity matrices of dimensions $l \times l$ and $nd \times nd$, respectively, and $\tau_{\boldsymbol{\Theta}} > 0$ and $\tau_{\boldsymbol{\lambda}} > 0$ are step sizes controlling the variance of q_1 and q_2, respectively. In our proposed model, the gradient of the log-posterior density can be obtained in closed form, facilitating inference; see the details in the Supplementary Material.

We use two burn-in phases in our MCMC algorithm. During the initial burn-in phase, we adapt the tuning parameters $\tau_{\boldsymbol{\Theta}}$ and $\tau_{\boldsymbol{\lambda}}$ as follows. Let τ_{cur} denote the current value of either $\tau_{\boldsymbol{\Theta}}$ or $\tau_{\boldsymbol{\lambda}}$, P_{acc} be the current acceptance probability calculated from the last 500 iterations, and P_{tar} be a target acceptance probability. Every 500 iterations, we update τ_{cur} as $\tau_{\text{cur}} \mapsto \tau_{\text{new}} := \exp\{(P_{\text{acc}} - P_{\text{tar}})/\theta\}\, \tau_{\text{cur}}$, where θ controls the rate of change. Here, we set $P_{\text{tar}} = 0.57$ for $\tau_{\boldsymbol{\lambda}}$, which was found to be optimal for the MALA algorithm under independence assumptions (Roberts and Rosenthal, 1998), and $P_{\text{tar}} = 0.23$ for $\tau_{\boldsymbol{\Theta}}$, which usually works well for random walks. Moreover, we here fix θ to 0.4. In the second burn-in phase, we use the same adaptive scheme only if the acceptance probability drops out of the intervals $[0.5, 0.65]$ and $[0.15, 0.30]$ for MALA and random walk proposals, respectively, i.e., if it has not stabilized yet during the initial burn-in phase. In all simulation experiments and in the application in §6, we use a total of $500,000$ iterations, with $50,000$ iterations for the initial burn-in phase, and $250,000$ iterations for the second burn-in phase.

5 Simulation study

5.1 Simulation scenarios

In this section, we study the performance of our MCMC sampler under diverse simulation scenarios. Data are simulated from the gamma-gamma model (12) with latent copula model for $d \in \{20, 50, 100, 200\}$ spatial locations sampled at random in the unit square $[0, 1]^2$, with $n \in \{50, 100, 250\}$ temporal replicates, depending on the scenario. The latent process $\Lambda(\boldsymbol{s})$ has a Gaussian or Student's t copula with isotropic exponential correlation function $\sigma(\boldsymbol{s}_i, \boldsymbol{s}_j) = \exp\{-(\|\boldsymbol{s}_i - \boldsymbol{s}_j\|/\rho)\}$, $\rho > 0$. In some cases, we apply the censoring scheme presented in §4.2, using site-specific thresholds chosen as the empirical 75%-quantile. When performing spatial prediction, we set the censoring threshold to $+\infty$.

To be concise, we here only discuss the results for one scenario, and we report the results for all scenarios in the Supplementary Material. Specifically, we here consider $d = 100$ spatial locations (80 used for fitting, and 20 kept for spatial prediction), $n = 100$ temporal replicates, and we use a latent Gaussian copula with range $\rho = 1$. We assume that the marginal scale parameter depends on spatially-varying covariates in log-linear specification, such that $\alpha(\boldsymbol{s}) = \alpha_0 \exp\{\sum_{i=1}^{3} \alpha_i z_i(\boldsymbol{s})\}$, with covariates $\{z_i(\boldsymbol{s}),\ i = 1, 2, 3\}$ generated independently and uniformly in $[0, 1]$ and regression coefficients set to $\alpha_0 = \alpha_1 = \alpha_2 = \alpha_3 = 1$. The other hyperparameters are fixed to $\beta_1 = 5$ and $\beta_2 = 5$ (i.e., with tail index $\xi = 0.2$).

We use the adaptive MCMC algorithm developed in §4.3. We run two MCMC chains with different initial values in parallel to check the dependence on initial conditions, and we then calculate MCMC outputs by combining these two chains.

5.2 Results

Figure 4 displays the trace plots for all the hyperparameters and one selected latent variable, for two MCMC chains with different initial values. The results show that there is good mixing

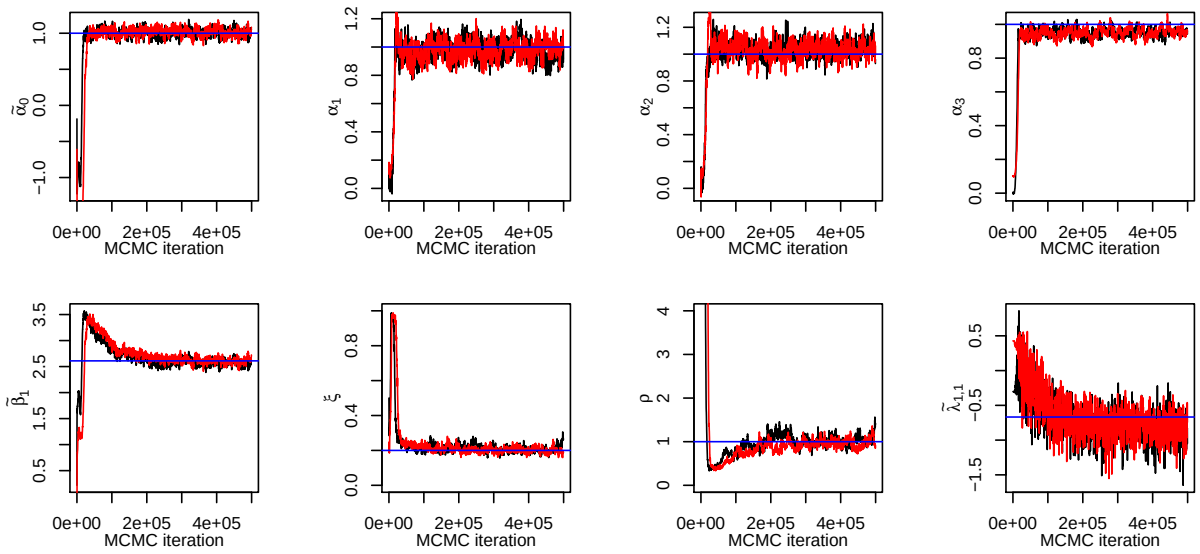

Figure 4: Trace plots for the simulation scenario detailed in §5.1, where the hyperparameters are set to $\alpha_0 = \alpha_1 = \alpha_2 = \alpha_3 = 1$, $\beta_1 = 5$, $\beta_2 = 5$ (i.e., $\xi = 0.5$), and $\rho = 1$. The rightmost plot in the second row corresponds to one chosen latent parameter, and the remaining 7 plots are for all the hyperparameters. The red and black chains are the two MCMC samples with different initial values, and the horizontal blue line represents the true values of the parameters. The total number of MCMC samples is $500,000$.

and that convergence occurs very quickly for the regression parameters α_0, α_1, α_2, and α_3, and relatively quickly for the other hyperparameters or latent variables, despite the large number of latent variables. Essentially in this scenario, all Markov chains seem to have converged after about $100,000$ to $150,000$ iterations. For all parameters, the true value lies well within the corresponding posterior distribution, confirming the good performance of our algorithm. The effective sample size per minute is between 3 and 8, see Table 1 in the Supplementary Material. The proposal variances, automatically tuned in our algorithm, converged to the values $\tau_{\boldsymbol{\Theta}} = 7.8 \times 10^{-5}$ for the hyperparameters (based on random walk proposals) and $\tau_{\boldsymbol{\lambda}} = 6.2 \times 10^{-4}$ for the latent variables (based on MALA proposals).

We then study the ability of our model to predict values at unobserved locations. To this end, we treat the data at 20 randomly selected sites as missing, and we compute the posterior predictive distributions for each individual missing observation, as well as for the

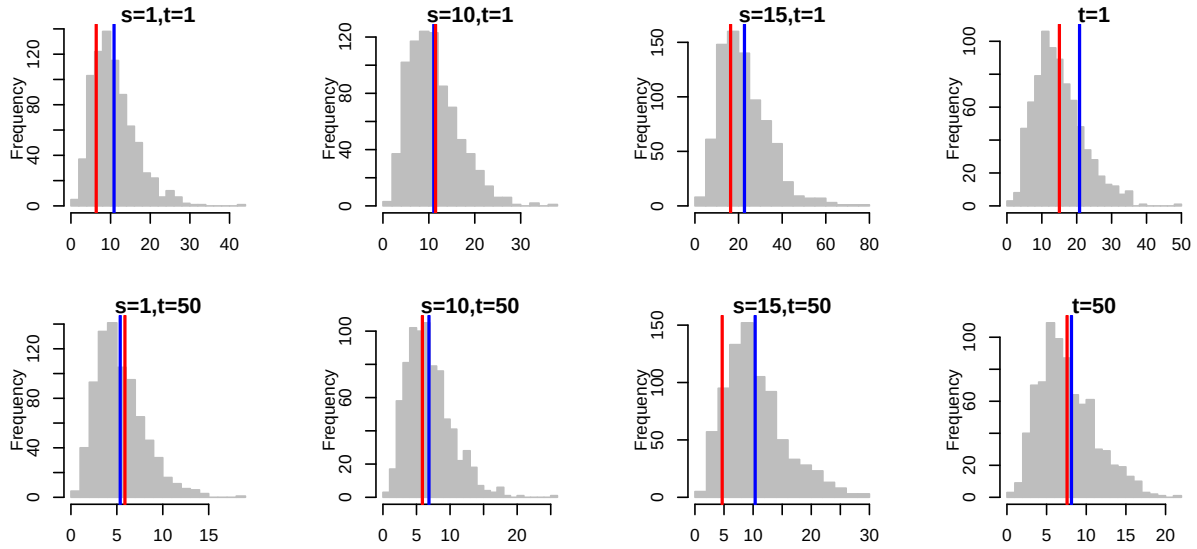

Figure 5: Posterior predictive histograms at time point 1 (top) and 50 (bottom), for sites 1, 10, and 15 (columns 1-3, respectively), and for the spatial average at the 20 missing sites (column 4), where "s" and "t" denote the site and time points, respectively. The vertical lines are the true (red) and the posterior mean (blue) values of the chain, respectively. The total number of MCMC samples is $500,000$, and all the histograms are based on the last $200,000$ MCMC samples, obtained after removing the first $300,000$ burn-in samples.

spatial average over the missing locations. Figure 5 displays the results at two different time points for three selected locations and for the spatial average. The true values lie well within the posterior predictive distributions, and are usually close to the corresponding predictive means. This suggests that our algorithm succeeds in performing spatial prediction. The great benefit of our Bayesian approach is that estimation based on censored data, and spatio-temporal prediction are performed simultaneously.

6 Application to precipitation extremes from Germany

6.1 Data description

We now study precipitation intensities observed in Germany, publicly available from the European Climate Assessment & Dataset project. The dataset reports daily precipitation

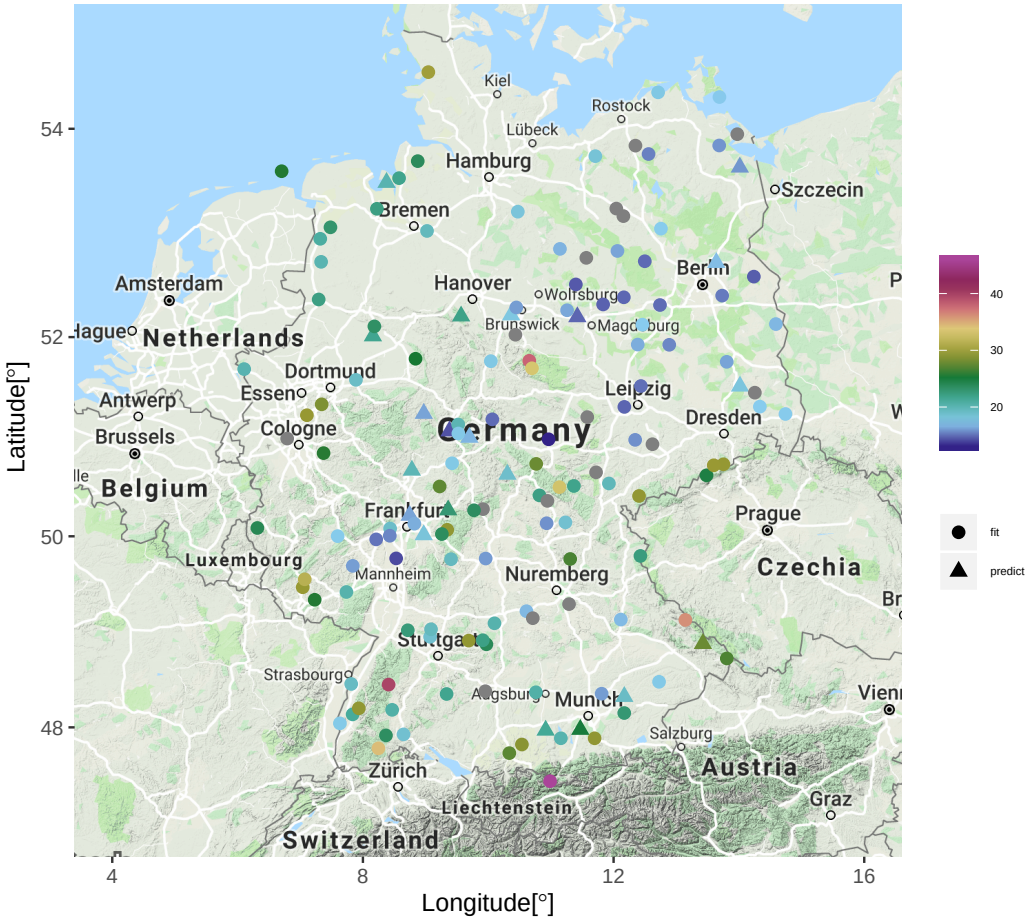

Figure 6: Mean precipitation [mm] at the study sites calculated over the days corresponding to the selected extreme episodes; see §6.2.

amounts observed at more than 5000 spatial locations during the period 1941 to 2018. We apply our hierarchical models to a subset of $d = 150$ locations with some missing data for the study period from 2009 to 2018. To avoid modeling complex seasonal nonstationarities, we consider the observations for the months of September to December (i.e., for the autumn season), resulting in $n = 1220$ temporal replicates. The site-specific mean precipitation intensities reported in Figure 6 show a tendency towards higher values in regions with higher altitudes. In a preliminary analysis of the tail behavior of the precipitation intensities, we fit the GP distribution to exceedances at each site with site-specific thresholds fixed at the 85% empirical quantile of positive precipitation intensities. The maximum likelihood

estimator, and the moment-based estimator of Dekkers *et al.* (1989) of the tail index, both provide systematically positive tail index estimates. This finding suggests that precipitation intensities are heavy-tailed as expected, and we proceed by fitting the spatial gamma-gamma model (12) to selected extreme precipitation episodes (see §6.3). The selection of extreme episodes and the modeling of their occurrences are described in §6.2.

6.2 Identifying and modeling extreme precipitation occurrences

The precipitation intensities are zero or very small for most of the days in the observation period, and we first extract extreme episodes (i.e., days) used to fit our spatial hierarchical model. Let $Y_t(\boldsymbol{s}_j)$ denote the precipitation intensities at time t and site $\boldsymbol{s}_j$, $j = 1, \ldots, d$. To select extreme episodes, we consider the spatially cumulated precipitation $S_t = \sum_{j=1}^{d} Y_t(\boldsymbol{s}_j)$, indexed by time t. We then defined extreme episodes as days t such that $S_t > \widehat{G}^{-1}(0.85)$, where $\widehat{G}$ is the empirical cumulative distribution function of the sample of S_t values. This scheme extracts 181 extreme episodes.

Let $E_{a,t}$ denote the binary sequence of 0 and 1 values representing the occurrence indicators of extreme episodes for the a-th year, with $a \in \{1, \ldots, 10\}$. That is, $E_{a,t} = 1$ if the t-th day of the year, with $t \in \{244, \ldots, 366\}$, was extreme in year a, and $E_{a,t} = 0$ otherwise. To capture temporal dependence, we model this time series through a logistic regression with a random effect defined as a first-order autoregressive Gaussian process, i.e.,

$$\begin{aligned} \log\left\{\frac{\Pr(E_{a,t}=1)}{1-\Pr(E_{a,t}=1)}\right\} &= \beta_{E,0} + W_{a,t}, \qquad (21)\\ W_{a,1} &\overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2/(1-\rho_E^2)), \quad a = 1, \ldots, 10,\\ W_{a,t} \mid W_{a,t-1} &= \rho_E W_{a,t-1} + \varepsilon_{a,t}, \quad t = 245, \ldots, 366,\\ \varepsilon_{a,t} &\overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2), \end{aligned}$$

with global intercept $\beta_{E,0}$, autoregression coefficient $\rho_E \in (-1, 1)$, and marginal variance of $W_{a,t}$, $\sigma_E^2 = \sigma_\varepsilon^2/(1-\rho_E^2) > 0$. We have explored a number of alternative models including

an additional seasonal trend component in the regression equation (21), but we could not detect any significant improvement with respect to model (21).

We now present the estimation results for model (21). While there would be no notable obstacles for MCMC-based estimation of this model, we here propose using the integrated nested Laplace approximation (INLA, Rue *et al.*, 2017) implemented in the `INLA` package of the statistical software `R`. It provides fast and "off-the-shelf" Bayesian inference for logistic regression models such as (21). We obtain the following parameter estimates, with 95% credible intervals in parentheses:

$$\widehat{\beta}_{E,0} = -2.3 \ (-2.7, -1.9), \quad \widehat{\rho}_E = 0.72 \ (0.55, 0.84), \quad \widehat{\sigma}^2_E = 1.7 \ (0.9, 2.9).$$

The estimated negative value of $\beta_{E,0}$ indicates that there is a higher tendency for non-extreme episodes, as expected, while the estimated value of ρ_E suggests that there is some non-negligible temporal persistence of zeros and ones. Figure 7 shows the fitted probability values for a selection of 4 years $(1, 2, 9, 10)$. The fitted models suggest relatively strong temporal autocorrelation: an extreme event at time t entails an extreme event at time $t+1$ with probability substantially above average.

6.3 Modeling extreme precipitation intensities

We now return to the modeling of (non-zero) precipitation intensities, and we fit the gamma-gamma model (12) with Gaussian copula to the time series of selected extreme episodes. The model structure and data dimension are similar to those considered in the simulation study in §5. We use standardized latitude, longitude, and altitude as covariates in the model. The dataset of extreme episodes is composed of $d = 150$ sites and $n = 181$ days, leading to $nd = 27,150$ latent variables in the model. The minimum and maximum distances between the selected sites are 6km and 819km, receptively. We use 20 sites for prediction and model validation (where the data are treated as completely missing). The remaining

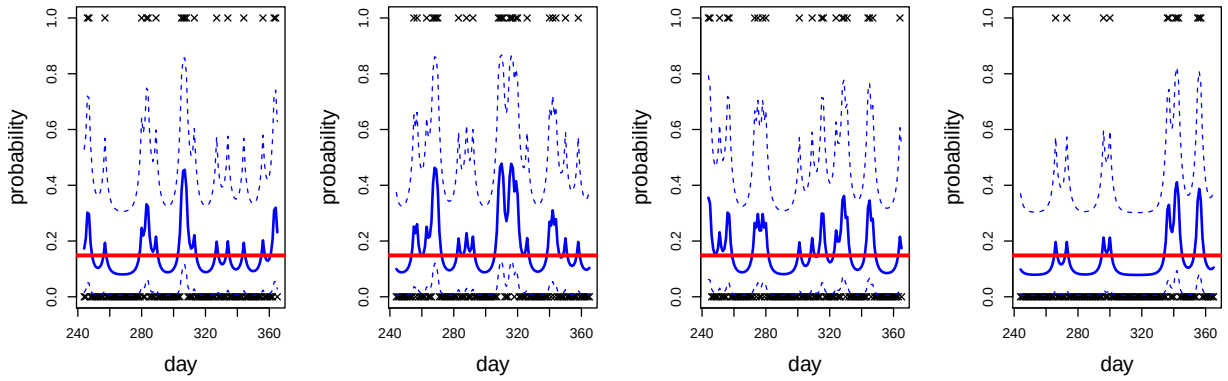

Figure 7: Posterior estimations (fitted probability values) of observing an extreme event at a given day according to model (21) for the years $1, 2, 9, 10$ (from left to right). Black $\times$ symbols indicate the observations $E_{a,t}$. Posterior means and 95% credible intervals of $\widehat{\Pr}(E_{a,t} = 1)$ are reported by continuous and dashed blue curves, respectively. The red line indicates the global empirical exceedance probability equal to 0.15.

130 sites are used for model fitting. The proportion of missing observations at each of the 130 training sites vary from 1% to 10%, for an average of 3%. Our goal is also to compare how our model performs at different marginal thresholds, in order to assess the flexibility of our "sub-asymptotic" modeling approach, and we consider empirical quantiles at three probability levels, namely 85%, 90% and 95%.

We fit the gamma-gamma model using the MCMC algorithm detailed in §4. We chose random initial values for all the hyperparameters, while we used the copula structure of latent variables defined in (12) to generate initial values for the latent parameters. The trace plots displayed in Figure 8 show two MCMC chains with different initial values for all the hyperparameters and the latent parameter $\tilde{\lambda}_{1,1} = \log(\lambda_{1,1})$ (site 1, day 1). The behavior of chains for the other latent variables is similar to $\tilde{\lambda}_{1,1}$. The chains mix satisfactorily and converge to their stationary distribution after about $150,000$–$200,000$ iterations. Figure 5 in the Supplementary Material shows trace plots of the tuning parameters $\tau_{\boldsymbol{\lambda}}$ and $\tau_{\boldsymbol{\Theta}}$ for the MALA and random walk proposals, respectively. The tuning parameters stabilize well before

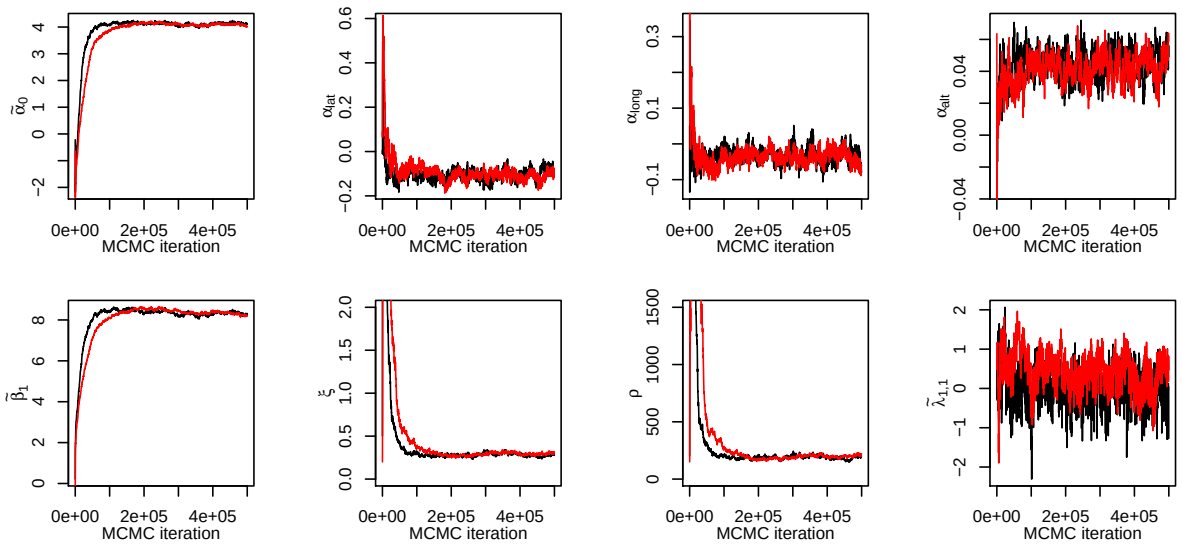

Figure 8: Trace plots of the 7 hyperparameters, and of a selected latent variable (lower right plot) in the data application. The marginal censoring threshold is here set to 95%. The red and black curves show two MCMC samples with different initial values. The total number of MCMC samples is $500,000$.

the end of the burn-in period of $300,000$ iterations, which illustrates the good performance of our adaptive MCMC algorithm.

Table 1 reports posterior mean estimates and two-sided 95% credible intervals for all hyperparameters and some latent parameters. Interestingly, the results are fairly consistent across marginal censoring thresholds, indicating that our sub-asymptotic model can flexibly accommodate departure from the asymptotic GP distribution at finite levels. As the conclusions are quite robust to the choice of the threshold, we now discuss the results for the 85% threshold. The effect of the three covariates (latitude, longitude, altitude) is always significant as the 95% credible intervals do not include 0. This result demonstrates the importance of including suitable geographical information in the scale parameter α of the distribution. The estimates for latitude ($\widehat{\alpha}_{\text{lat}} = -0.11$) and longitude ($\widehat{\alpha}_{\text{long}} = -0.05$) indicate that the South-Western part of Germany receives higher precipitation amounts than the North-Eastern part—a pattern that is clearly perceptible in the mean precipitation plot

in Figure 6. Moreover, a clear positive effect of higher altitude on precipitation amounts arises with an estimate of $\widehat{\alpha}_{\text{alt}} = 0.05$. The estimate of the shape parameter β_1 is around 60 and shows a huge difference with respect to the generalized Pareto model with $\beta_1 = 1$. This finding substantiates our claim that extensions to the generalized Pareto distribution are useful for capturing complex data behavior at sub-asymptotic levels. The estimated tail index $\widehat{\xi}$ at the 95% censoring threshold is about 0.3, which corresponds to quite heavy tails. The estimated range parameter ρ of the exponential correlation function is around $196km$, implying a correlation of approximately 0.95 *at the latent* level between two sites separated by 10km. This correlation decreases to approximately 0 between the two furthest sites. The quantile-quantile (QQ)-plots shown in Figure 9 confirm that our models provide a reasonable fit overall.

We now illustrate the spatial predictive performance of the model by the QQ-plots shown in Figure 9 and the histograms reported in Figure 7 of the Supplementary Material. Rows (from top to bottom) correspond to censoring thresholds 85%, 90% and 95%, respectively. QQ-plots are shown for a random selection of 4 sites in different columns, the two leftmost of which correspond to sites used for prediction only (i.e., treated as fully missing). We use them to assess if we appropriately predict marginal distributions at unobserved sites. The model quantiles are based on the distribution $(\widehat{\alpha}\widehat{\beta}_1/\widehat{\beta}_2)F_{2\widehat{\beta}_1,2\widehat{\beta}_2}$ where $\widehat{\alpha}_1$, $\widehat{\beta}_1$ and $\widehat{\beta}_2$ are posterior mean estimates. We obtain 95% uncertainty bands in the QQ-plots using a parametric bootstrap procedure accounting for missing values. Predictive performance turns out to be consistent for different thresholds with satisfactory goodness-of-fit for low and high quantiles. The histograms in Figure 7 of the Supplementary Material show posterior predictive distributions for four observed data values at a random selection of days and prediction sites. They exemplify how the spatial dependence in the model is used for predicting precipitation over space at unobserved locations (i.e., locations not used for fitting the model). True data

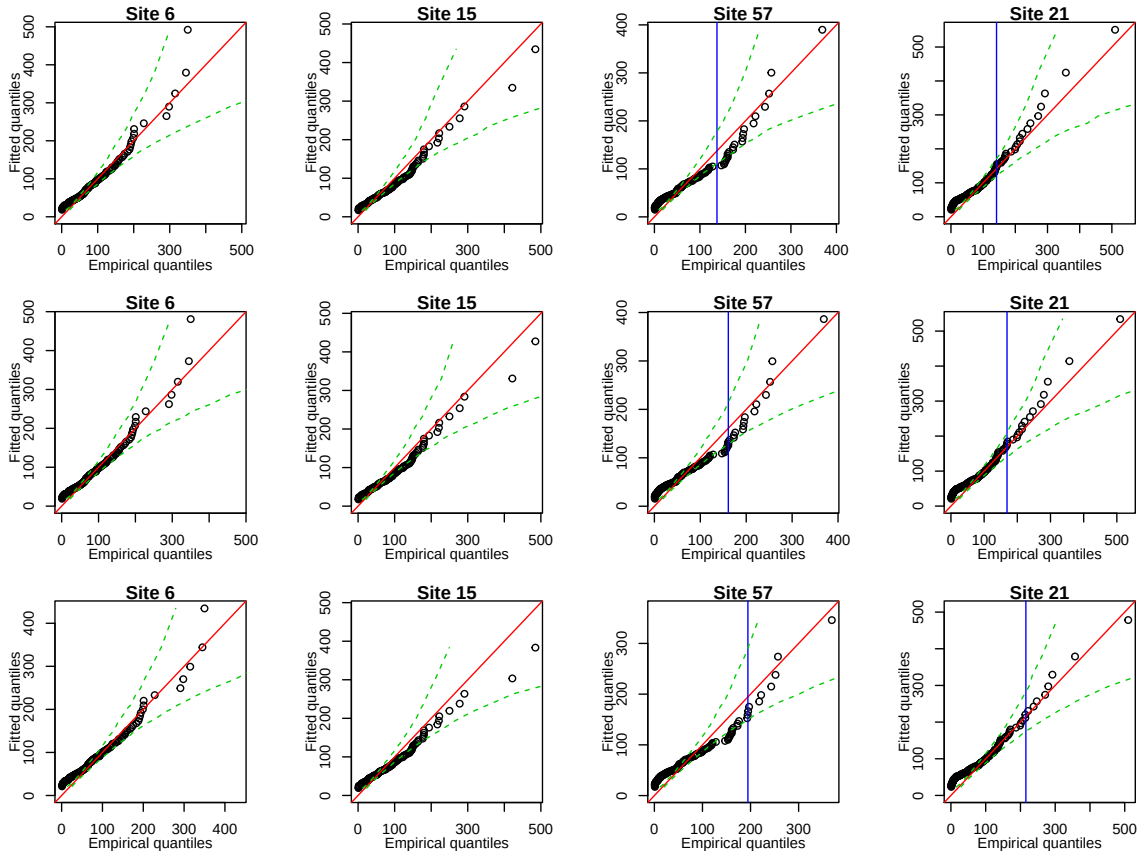

Figure 9: QQ-plots for four randomly selected sites. The first two columns correspond to prediction sites (with censoring threshold equal to $+\infty$), wile the last two columns correspond to sites used for fitting the model. Rows (from top to bottom) correspond to marginal censoring thresholds 85%, 90% and 95%, respectively. The vertical blue line indicates the threshold values. The QQ-plots are obtained by comparing the empirical data quantiles to the fitted quantiles obtained by plugging the posterior mean estimates into the F-distribution.

values always lie within histogram bars of the posterior predictive sample, with perfect coincidence of posterior predictive mean and observed value in the first column. We also checked several posterior predictive distributions for other data values and found similar results.

In summary, we conclude that our model appropriately captures spatial trend and dependence patterns in autumn precipitation data in Germany. Specifically, it provides useful probabilistic predictions for unobserved locations during extreme episodes.

7 Conclusion

We have proposed several univariate models extending the generalized Pareto limit model for threshold exceedances, and studied their tail properties. The high flexibility of these models suggests that they are good candidates for threshold-based modeling of moderate to large values. Our models are based on hierarchical constructions using latent processes for spatial dependence. In this way, they avoid the artificial and overly strong separation of marginal and dependence modeling often encountered in spatial extreme value analysis.

Our Bayesian estimation approach bears witness of the power of the MALA for simulation-based estimation in cases where the dimension of the latent model is comparable to the number of observations. Its use with censored data for exceedance-based spatial extreme value analysis allows bypassing high-dimensional numerical integration efficiently. Hence, our assumed model structure coupled with a data augmentation approach can be efficiently exploited to fit complex models in relatively high dimensions to censored threshold exceedances.

In our spatio-temporal precipitation data application, we have opted for a two-step modeling approach of extreme quantiles. Essentially, we first identify extreme episodes, defined here as days with large spatially aggregated values at all locations, and then separately model their occurrences and intensities. More precisely, in the first step, we model the binary time series of occurrences (1s) or non-occurrences (0s) of extreme episodes through logistic regression, and in the second step, we model the precipitation intensities of extreme episodes using a spatial hierarchical gamma-gamma model. This two-step approach has the benefit of reducing the number of latent variables included in our Bayesian model fitted in the second step, speeding up computations. Temporal dependence may be incorporated through covariate or random effect modeling in the logistic regression. In future research, it would be interesting to further extend the latent process of our spatial hierarchical model to take the spatio-temporal dependence of successive extreme precipitation intensities into account.

Acknowledgments

The data that support the findings of this study are openly available from the European Climate Assessment & Dataset project at https://www.ecad.eu/. This publication is based upon work supported by the King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) under Award No. OSR-CRG2017-3434.

References

Arendarczyk, M. and Debicki, K. (2011) Asymptotics of supremum distribution of a gaussian process over a weibullian time. *Bernoulli* **17**(1), 194–210.

Bacro, J.-N., Gaetan, C., Opitz, T. and Toulemonde, G. (2019) Hierarchical space-time modeling of asymptotically independent exceedances with an application to precipitation data. *Journal of the American Statistical Association*. In press.

Banerjee, S., Carlin, B. P. and Gelfand, A. E. (2014) *Hierarchical modeling and analysis for spatial data*. Second edition. CRC Press.

Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004) *Statistics of Extremes: Theory and Applications*. Wiley.

Bortot, P. and Gaetan, C. (2014) A Latent Process Model for Temporal Extremes. *Scandinavian Journal of Statistics* **41**(3), 606–621.

Bortot, P. and Gaetan, C. (2016) Latent process modelling of threshold exceedances in hourly rainfall series. *Journal of Agricultural, Biological and Environmental Statistics* **21**(3), 531–547.

Breiman, L. (1965) On some limit theorems similar to the arc-sin law. *Theory of Probability & Its Applications* **10**(2), 323–331.

Carreau, J. and Bengio, Y. (2009) A hybrid Pareto model for asymmetric fat-tailed data: the univariate case. *Extremes* **12**(1), 53–76.

Castro-Camilo, D. and Huser, R. (2019) Local likelihood estimation of complex tail dependence structures, applied to us precipitation extremes. *Journal of the American Statistical Association* pp. 1–29.

Cooley, D., Nychka, D. and Naveau, P. (2007) Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association* **102**(479), 824–840.

Cressie, N. A. C. (1993) *Statistics for spatial data*. Wiley Online Library.

Davison, A. C. and Huser, R. (2015) Statistics of Extremes. *Annual Review of Statistics and its Application* **2**, 203–235.

Davison, A. C., Huser, R. and Thibaud, E. (2019) Spatial extremes. In *Handbook of Environmental and Ecological Statistics*, eds A. E. Gelfand, M. Fuentes, J. A. Hoeting and R. L. Smith. Boca Raton: 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 Smith, R. L. (1990) Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **52**(3), 393–442.

Dekkers, A. L., Einmahl, J. H., De Haan, L. *et al.* (1989) A moment estimator for the index of an extreme-value distribution. *The Annals of Statistics* **17**(4), 1833–1855.

Fougeres, A.-L. and Mercadier, C. (2012) Risk measures and multivariate extensions of Breiman's theorem. *Journal of Applied Probability* **49**(2), 364–384.

Frigessi, A., Haug, O. and Rue, H. (2003) A dynamic mixture model for unsupervised tail estimation without threshold selection. *Extremes* **5**(3), 219–235.

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. and Wadsworth, J. L. (2019) Modeling spatial processes with unknown extremal dependence class. *Journal of the American Statistical Association* **114**, 434–444.

Naveau, P., Huser, R., Ribereau, P. and Hannart, A. (2016) Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection. *Water Resources Research* **52**(4), 2753–2769.

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.

Papastathopoulos, I. and Tawn, J. A. (2013) Extended generalised Pareto models for tail estimation. *Journal of Statistical Planning and Inference* **143**(1), 131–143.

Power, S. B. and Delage, F. c. P. D. (2019) Setting and smashing extreme temperature records over the coming century. *Nature Climate Change* **9**, 529–534.

Reiss, R.-D. and Thomas, M. (2007) *Statistical Analysis of Extreme Values*. Third edition. Basel: Birkhäuser.

Resnick, S. I. (1987) *Extreme values, regular variation and point processes*. Springer.

Risser, M. D. and Wehner, M. F. (2017) Attributable human-induced changes in the likelihood and magnitude of the observed extreme precipitation during Hurricane Harvey. *Geophysical Research Letters* **44**(24), 12457–12464.

Roberts, G. O. and Rosenthal, J. S. (1998) Optimal scaling of discrete approximations to langevin diffusions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **60**(1), 255–268.

Roberts, G. O. and Tweedie, R. L. (1996) Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli* **2**(4), 341–363.

Rue, H., Riebler, A., Sørbye, S. H., Illian, J. B., Simpson, D. P. and Lindgren, F. K. (2017) Bayesian computing with inla: a review. *Annual Review of Statistics and Its Application* **4**, 395–421.

Scarrott, C. and MacDonald, A. (2012) A Review of Extreme Value Threshold Estimation And Uncertainty Quantification. *REVSTAT* **10**(1), 33–60.

Simpson, D. P., Rue, H., Riebler, A., Martins, T. G. and Sørbye, S. H. (2017) Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science* **32**(1), 1–28.

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.

Witze, A. (2018) Why extreme rains are gaining strength as the climate warms. *Nature* **563**(7732), 458–460.

Table 1: Posterior mean (Post. mean) and 95% Credible Interval (CI) of hyperparameters and of several latent parameters, reported for censoring thresholds 85%, 90%, and 95%. The tail index parameter ξ is equal to $1/\beta_2$. The total number of MCMC iterations is $500,000$. Estimations are based on the last $200,000$ MCMC samples, obtained after removing the first $300,000$ burn-in samples.

Threshold	85%		90%		95%	
Parameter	Post. mean	95% CI	Post. mean	95% CI	Post. mean	95% CI
α_0	1.02	$(0.90, 1.13)$	1.08	$(0.93, 1.21)$	1.11	$(0.97, 1.24)$
α_{lat}	-0.11	$(-0.13, -0.08)$	-0.10	$(-0.13, -0.08)$	-0.11	$(-0.14, -0.07)$
α_{long}	-0.05	$(-0.07, -0.03)$	-0.04	$(-0.07, -0.02)$	-0.03	$(-0.07, 0)$
α_{alt}	0.05	$(0.05, 0.06)$	0.05	$(0.04, 0.06)$	0.04	$(0.03, 0.06)$
β_1	59.5	$(55.3, 62.6)$	59.0	$(54.1, 63.7)$	68.2	$(62.9, 71.8)$
ξ	0.34	$(0.32, 0.36)$	0.33	$(0.30, 0.35)$	0.29	$(0.27, 0.32)$
ρ	196	$(181, 215)$	206	$(188, 221)$	197	$(177, 218)$
$\lambda_{1,1}$	1.07	$(0.66, 1.71)$	1.53	$(0.83, 2.61)$	1.3	$(0.68, 2.21)$
$\lambda_{6,95}$	0.44	$(0.29, 0.63)$	0.41	$(0.27, 0.59)$	1.03	$(0.48, 1.79)$
$\lambda_{56,45}$	1.07	$(0.64, 1.77)$	0.93	$(0.55, 1.56)$	1.16	$(0.66, 1.91)$