Flexible Modeling of Multivariate Spatial Extremes

Yan Gong\(^1\) and Raphaël Huser\(^1\)

June 24, 2022

Abstract

We develop a novel multi-factor copula model for multivariate spatial extremes, which is designed to capture the different combinations of marginal and cross-extremal dependence structures within and across different spatial random fields. Our proposed model, which can be seen as a multi-factor copula model, can capture all possible distinct combinations of extremal dependence structures within each individual spatial process while allowing flexible cross-process extremal dependence structures for both upper and lower tails. We show how to perform Bayesian inference for the proposed model using a Markov chain Monte Carlo algorithm based on carefully designed block proposals with an adaptive step size. In our real data application, we apply our model to study the upper and lower extremal dependence structures of the daily maximum air temperature (TMAX) and daily minimum air temperature (TMIN) from the state of Alabama in the southeastern United States. The fitted multivariate spatial model is found to provide a good fit in the lower and upper joint tails, both in terms of the spatial dependence structure within each individual process, as well as in terms of the cross-process dependence structure. Our results suggest that the TMAX and TMIN processes are quite strongly spatially dependent over the state of Alabama, and moderately cross-dependent. From a practical perspective, this implies that it may be worthwhile to model them jointly when interest lies in computing spatial risk measures that involve both quantities.

Keywords: Asymptotic dependence and independence; Copula model; Coregionalization; MALA; Multivariate spatial extremes; Tail asymmetry.

---

\(^1\)Statistics Program, Computer, Electrical and Mathematical Sciences and Engineering (CEMSE) Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia. E-mail: yan.gong@kaust.edu.sa, raphael.huser@kaust.edu.sa.
1 Introduction

The statistical modeling of multivariate spatial extremes is a fundamental but challenging task in geophysical, ecological, environmental, and climate sciences, in order to quantify the probability and compound risks of natural hazards such as floods, droughts, heat waves, and strong winds. In a related context, model-based statistical approaches can also help understand and disentangle the probabilistic behavior of anthropogenic processes, such as air pollutant concentrations and their impact on the environment. In practice, multiple variables are often observed simultaneously from monitors or generated by climate or geophysical models at different locations, and it is often the co-occurrence of extremes from these multiple variables that lead to the biggest impacts (Zscheischler et al., 2020; Raymond et al., 2020). For example, high temperatures combined with strong wind might increase wild fire risk (Littell et al., 2016; Ruffault et al., 2018); the heat index, which may be used as a measure of climate “livability”, is large when both temperature and relative humidity are simultaneously high; similarly, a heat wave is defined by the Australian Bureau of Meteorology as a multivariate event where “the maximum and the minimum temperatures are unusually hot over a three-day period at a location” (link), see also Baldwin et al. (2019) and Perkins-Kirkpatrick and Lewis (2020) for alternative heat wave definitions; and the formation of a hurricane is known to involve both warm ocean water, as well as moist, humid air in the same spatial region. Thus, it has become increasingly important to model such complex multivariate spatial extreme data.

In the literature of multivariate spatial extremes modeling, recent work has been done to extend spatial max-stable processes to the multivariate setting. In particular, Genton et al. (2015) proposed several multivariate spatial max-stable processes, generalizing the theory and application of max-stable processes to the multivariate context. Specifically, they introduced multivariate versions of (i) the Smith Gaussian extreme-value model (Smith, 1990), (ii) the Schlather extremal-Gaussian model (Schlather, 2002) and its extremal-$t$ model extension (Opitz, 2013), and (iii) the Brown–Resnick model (Brown and Resnick, 1977; Kabluchko et al., 2009). Later, Oesting et al. (2017) introduced a bivariate Brown–Resnick process, constructed from a customized pseudo cross-variogram, as a joint spatial model for real extreme observations and model-based forecasts, and designed to post-process and enhance probabilistic weather extremes forecasts. More recently, Vettori et al. (2019) developed nested tree-based multivariate max-stable processes, to study extreme concen-
trations of multiple pollutants over a spatial domain. The construction of these nested max-stable processes extends the spatial Reich–Shaby process (Reich and Shaby, 2012), which has a conditional independence representation given some latent positive stable random effects and is jointly max-stable (both across space and variables). Max-stable processes have been used in the above studies on the basis that they have a strong theoretical motivation and well-understood tail properties; however, max-stable processes are only justified for block maxima, and their asymptotic characterization entails a fairly rigid tail structure; see, e.g., Huser and Wadsworth (2022) for a detailed review. In particular, these processes are always asymptotically dependent (AD), which can be a strong limitation in some applications. This property means that if \( Z(s) \) is a process with common margins and infinite upper endpoint, then 
\[
\chi = \lim_{z \to \infty} \Pr\{Z(s_1) > z \mid Z(s_2) > z\} > 0
\]
for any pair of sites \( \{s_1, s_2\} \). However, empirical evidence suggests that asymptotic independence (AI) with \( \chi = 0 \) is equally important in practice, particularly over large domains or with data exhibiting short-range extremal dependence through localized extreme events (e.g., convective precipitation events, or wind gusts); see, e.g., Wadsworth and Tawn (2012), Opitz (2016), Hazra et al. (2021), Huser and Wadsworth (2022), and Zhang et al. (2022). Therefore, from a methodological and practical perspective, it is crucial to develop flexible multivariate spatial models that can capture AD or AI in their upper and/or lower tails, with a smooth transition in between, in order to model the full within- and cross-process dependence structures, while accurately capturing the joint behavior of extremes.

In the classical geostatistical framework, several types of multivariate spatial models have been proposed, including multivariate Gaussian processes (Kleiber and Nychka, 2012; Genton and Kleiber, 2015; Cressie and Zammit-Mangion, 2016; Gelfand, 2021) or multivariate \( t \)-processes (Kotz and Nadarajah, 2004; Opitz, 2013; Hazra et al., 2020). However, while the former are always AI, the latter are always AD for all pairs of sites; thus, both lack tail flexibility. In addition, both Gaussian and \( t \)-processes are tail-symmetric, and there has been little work on constructing multivariate spatial models with flexible asymmetric tail dependencies.

In a recent related work, Krupskii and Genton (2019) proposed an exponential factor copula model for non-Gaussian multivariate spatial data, which allows fast likelihood-based inference. In their model, the obtained extremal dependence structure within or across fields can capture tail asymmetry, but it always exhibits AD, while AI lies on the boundary of the parameter space, which is a crucial practical limitation.
To capture complex extremal dependence structure in multivariate spatial extremes, we propose in this paper a novel multi-factor copula model for multivariate spatial data that can capture all possible distinct combinations of extremal dependence types within each individual spatial process while allowing for flexible cross-process extremal dependence structures, for both the upper and lower tails. Our new model builds upon Huser and Wadsworth (2019), who proposed a spatial extremes model for the upper tail of a single variable, and it extends the bivariate copula model of Gong and Huser (2022), used to estimate time-varying tail dependencies in bivariate financial time series, to the multivariate spatial extremes setting. Rewriting our proposed model as a Bayesian hierarchical model with multiple latent variables, we here perform Bayesian inference using a customized Markov chain Monte Carlo (MCMC) algorithm based on carefully designed block proposals with an adaptive step size. Specially, we implement adaptive Metropolis–Hastings updates for hyperparameters and use the Metropolis-adjusted Langevin algorithm (MALA) to update the high-dimensional vector of latent variables.

We apply our model to study the lower and upper extremal dependence structures of the daily maximum air temperature (TMAX) and daily minimum air temperature (TMIN) in the state of Alabama in the southeastern United States (US). Since simultaneous temperature extremes (both high and low extremes) can severely affect human health (Vicedo-Cabrera et al., 2021; Mitchell, 2021), the environment, ecosystems, and energy consumption (Seneviratne et al., 2012), it is crucial to study the co-occurrence probability of such extreme events over space, and we here exploit our model to assess the risk of jointly low/high daily temperatures in Alabama, a state known for its subtropical climate (due to its proximity to the Gulf of Mexico), which is prone to thunderstorms, tropical storms, and hurricanes. In other words, we use our proposed methodology to jointly model the spatial extremal dependence within and across TMAX and TMIN random fields over the whole state of Alabama, US.

This paper is organized as follows. In Section §2, we present the model construction and derive the model’s extremal dependence properties. In Section §3, we describe the proposed Bayesian inference procedure, which is carefully designed for our model’s hierarchical specification. A simulation study is discussed in Section §4 and the results of the real-data application are presented in Section §5. In Section §6, we discuss our results and findings, and conclude with some perspectives on future research.
2 Modeling

In this section, we describe the construction of our proposed multivariate spatial model with flexible extremal dependence structure in both tails. The dependence (i.e., copula) model is designed to capture the different combinations of marginal and cross-extremal dependence structures within and across different spatial random fields. We present the construction steps in detail and derive the extremal dependence properties. Furthermore, we demonstrate the flexibility of the developed model in a bivariate spatial setting.

2.1 Model construction

We define the $p$-dimensional multivariate random process $X(s) = (X_1(s), \ldots, X_p(s))^\top$ on the spatial domain $S \subset \mathbb{R}^D$, with finite-dimensional realizations at the $d$ spatial locations $\{s_1, \ldots, s_d\} \subset S$, as

$$X_i(s) = \alpha_i \{ \gamma_i [\delta^U R^U_0 + (1 - \delta^U) R^U_i] - (1 - \gamma_i) [\delta^L R^L_0 + (1 - \delta^L) R^L_i] \} + W_i(s), \quad i = 1, \ldots, p,$$

where $\alpha_i > 0$, and $\gamma_i, \delta^U, \delta^L \in [0, 1]$ are dependence parameters. For simplicity, we write $X_{ij} := X_i(s_j)$ and $X_i := (X_{i1}, \ldots, X_{id})^\top$, and similarly for $W_{ij}$ and $W_i$, $i = 1, \ldots, p$, $j = 1, \ldots, d$. The (spatially-constant) random variables $R^U_0, R^L_0, \{R^U_i\}, \{R^L_i\} \overset{i.i.d.}{\sim} \text{Exp}(1)$ are mutually independent and exponentially distributed, while $W(s) = (W_1(s), \ldots, W_p(s))^\top$ is a multivariate random field displaying AI in both tails, with exponentially-decaying tails, and defined more precisely below. Note that although the latent random factors $\{R^U_i\}$ and $\{R^L_i\}$ are field-specific, $R^U_0$ and $R^L_0$ are shared among the $p$ different random fields, which induces cross-dependence (and potentially cross-extremal dependence). Our proposed model (1) is carefully parametrized, so that the dependence parameters play different roles and control the model’s flexibility in different ways, thus ensuring that it remains theoretically identifiable. In particular, $\alpha_i$ controls the overall “weight” of the spatially-constant common factors with respect to the non-trivial spatially correlated random field $W_i(s)$, thus mainly controlling the “overall” spatial dependence strength of the $i$-th field while also impacting its extremal dependence type; $\gamma_i$ captures the extent of tail asymmetry for the $i$-th field, with values $\gamma_i > 0.5$ producing “right-skewed” dependencies with larger mass in the joint upper tail than in the joint lower tail, and vice versa when $\gamma_i < 0.5$; and $\delta^U/\delta^L$ (which are here common to all fields) control the strength of the upper/lower tail cross-dependence, as $\delta^U$ (respectively $\delta^L$) determines the balance between the field-specific random effect $R^U_i$ (respectively $R^L_i$) and the shared random effect $R^U_0$ (respectively, $R^L_0$).
We demonstrate the model’s flexibility using simulation examples in §2.3.

Here, for simplicity, we assume that the processes \( W_i(s), i = 1, \ldots, p \), are cross-correlated trans-Gaussian processes with standard Laplace \((0, 1)\) margins, i.e., \( W_{ij} = W_i(s_j) \sim F_W \), where \( F_W(w) = \frac{1 + \text{sign}(w)\{1 - \exp(-|w|)\}}{2}, w \in \mathbb{R} \) and the variables \( W_{ij} \) are driven by a Gaussian copula. The cross-dependence between the process \( W_i(s) \) may be constructed using the Gaussian Linear Model of Coregionalization (LMC) (Bourgault and Marcotte, 1991), although there are also other possibilities. Specifically, we first define the Gaussian random vector

\[
(W_1^T, \ldots, W_p^T)^T = L(W_1^*T, \ldots, W_p^*T)^T,
\]

where \( W_i^* \sim \mathcal{N}_d(0, \Sigma_i^*) \) are independent \( d \)-dimensional Gaussian vectors with mean zero, variance one, and spatial correlation structure \( \Sigma_i^* \); the matrix \( L \) is a lower triangular \( p \times p \) matrix such that \( \Sigma = LL^\top \) is the cross-covariance matrix of \( (W_1^T, \ldots, W_p^T)^T \). Finally, we obtain \( W_i = (W_{i1}, \ldots, W_{id})^T \) by back-transforming the correlated Gaussian vector \( W_i^* \) to the Laplace scale, i.e.,

\[
W_i = F_W^{-1}\{\Phi(W_i^*)\}.
\]

More details about the properties of \( (W_1^T, \ldots, W_p^T)^T \) are given in Appendix §A.2.

Our proposed dependence model is defined by extracting the copula from the multivariate spatial process (1). We denote the joint cumulative distribution function of the multivariate spatial \((p \times d)\)-dimensional random vector \( X = (X_1^T, \ldots, X_p^T)^T \) as \( X \sim F_X \). In addition, we define the marginally uniform random vector \( U = (U_1, \ldots, U_p)^T \), with \( U_i = (U_{i1}, \ldots, U_{id})^T \) and \( U_{ij} = F_i^X(X_{ij}) \sim \text{Unif}(0, 1) \) for \( i = 1, \ldots, p \), \( j = 1, \ldots, d \), where \( F_i^X \) is the marginal cumulative distribution function of \( X_i \), which can be obtained in the closed form (see Appendix §A.1 for the precise mathematical expressions). Thus, the associated copula of \( X \) (i.e., the joint distribution function of the vector \( U \)) and its density can be obtained respectively as

\[
C^X(u_1, \ldots, u_p) = F^X\{(F_1^X)^{-1}(u_1), \ldots, (F_p^X)^{-1}(u_p)\},
\]

\[
c^X(u_1, \ldots, u_p) = \frac{f^X\{(F_1^X)^{-1}(u_1), \ldots, (F_p^X)^{-1}(u_p)\}}{f_1^X\{(F_1^X)^{-1}(u_1)\} \times \cdots \times f_p^X\{(F_p^X)^{-1}(u_p)\}},
\]

where \( f_i^X \) and \( f^X \) are the marginal and joint densities of \( X_i \) and \( X \), respectively, and the functions are applied componentwise. Note that the quantile function \((F_i^X)^{-1}\) is not available in closed form, but it can be approximated relatively quickly and accurately using standard numerical root-finding algorithms.
2.2 Extremal dependence properties

The proposed model is designed to capture different combinations of marginal and cross-extremal dependence structures within and across the different random fields. In particular, the model can capture all the possible distinct combinations of extremal dependence types marginally, while allowing high flexibility in capturing cross-process extremal dependence for both upper and lower tails.

The model can be viewed as a particular type of random scale construction (Engelke et al., 2019; Huser and Wadsworth, 2019), and the marginal and cross-extremal dependencies of model (1) may be classified as reported in Table 1. See Appendix §B for the proof. In the bivariate spatial setting (i.e., with \( p = 2 \)), there are a number of \( 2 \times 2 \times 2 = 8 \) different combinations of all possible marginal and cross-extremal dependence scenarios for each tail (AI or AD for the first process \( \times \) AI or AD for the second process \( \times \) AI or AD for the cross-extremal dependence). The proposed model can capture five out of these eight cases and has a separate control on the joint upper tail and joint lower tail. For example, the process can capture AD in the upper tail and AI in the lower tail, or vice versa. The three cases that the proposed model cannot capture include the cases when some of the marginal extremal dependencies are AI but the cross-extremal dependence is AD. In other words, when the extremal cross-dependence is AD, both processes must also be AD individually. In practice, this is usually not a major limitation, as the cross-process dependence is often found to be weaker in applications than the within-process dependence.

The conditions for asymptotic dependence (AD) in the marginal and cross-extremal dependence structures of our model (1) are given in Table 1. Conditions for asymptotic independence (AI) are found by symmetry. In the table, the indices are \( i, i_1, i_2 = 1, \ldots, p \), \( i_1 \neq i_2 \), and \( j_1, j_2 = 1, \ldots, d \).

| \( (X_i(s_{j_1}), X_i(s_{j_2}))^\top \) (within-field dependence) | Upper Tail | Lower Tail |
|---|---|---|
| \( \max(\delta^U, 1 - \delta^U) > 1/(\alpha_i \gamma_i) \) | \( \max(\delta^L, 1 - \delta^L) > 1/(\alpha_i(1 - \gamma_i)) \) |

| \( (X_{i_1}(s_{j_1}), X_{i_2}(s_{j_2}))^\top \) (cross-dependence) | | |
|---|---|
| \( \delta^U > \max\{(\alpha_{i_1} \gamma_{i_1})^{-1}, (\alpha_{i_2} \gamma_{i_2})^{-1}, 1/2\} \) | \( \delta^L > \max\{(\alpha_{i_1} (1 - \gamma_{i_1}))^{-1}, (\alpha_{i_2} (1 - \gamma_{i_2}))^{-1}, 1/2\} \) |

2.3 Illustration

We illustrate the flexibility of our proposed model in the bivariate spatial case with \( p = 2 \) by simulation. We focus specifically on the spatial and cross-extremal dependence structures of \( X(s) = (X_1(s), X_2(s))^\top \), with \( X_1(s) \) and \( X_2(s) \) defined as in (1) where we choose an isotropic exponential correlation matrix for \( W_i^+(s) \), i.e., \( \Sigma_i := c_i(H) = \exp(-H/\lambda_i) \), for \( i = 1, 2 \), \( \lambda_i > 0 \) represent the field-specific spatial range...
parameters, and $H$ is the distance matrix with $(j_1, j_2)$ element $h_{j_1,j_2} = ||s_{j_1} - s_{j_2}||$, for $j_1, j_2 = 1, \ldots, d$. We specify the elements of the correlation matrix $\Sigma$ by taking $L_{11} = 1, L_{21} = \rho_{12}$, and $L_{22} = \sqrt{1 - \rho_{12}^2}$. Recall §2.1 for details. Here we collect all parameters in the vector $\theta = (\alpha_1, \alpha_2, \gamma_1, \gamma_2, \delta^U, \delta^L, \lambda_1, \lambda_2, \rho_{12})^\top \in (0, \infty)^2 \times [0, 1]^4 \times (0, \infty)^2 \times [-1, 1]$. Figure 1 shows realizations from our model in two scenarios. For each field, we generate $n = 2000$ replicates at two random locations within a unit square. The simulated data are plotted both on the original scale of $X$ (top row) and the uniform scale (bottom row). In Case 1 (left column), the upper tail spatial extremal dependence types (i.e., for each of the two individual fields) are AD and AI, respectively, while the upper tail cross-extremal dependence is AI, which we write concisely in the title of the corresponding panel as U: AD, AI–AI. The lower tail structure in Case 1 is chosen to be AD for each individual marginal field, and AI for the cross-dependence structure, i.e., we write it as L: AD, AD–AI. Similarly, using the same notation, the extremal dependence types of Case 2 (right column) are U: AI, AI–AI and L: AD, AD–AD. In both cases, we fix the following parameters: $\alpha_1 = 4, \alpha_2 = 5, \lambda_1 = 0.7$, and $\lambda_2 = 0.3$. However, in Case 1, we further specify, $\gamma_1 = 0.6, \gamma_2 = 0.2, \delta^U = 0.4, \delta^L = 0.7$ and $\rho_{12} = 0.7$, while in Case 2, $\gamma_1 = 0.4, \gamma_2 = 0.3, \delta^U = 0.4, \delta^L = 0.8$, and $\rho_{12} = -1$. The simulation shows the flexibility of our model to capture tail asymmetry and to separately control the spatial and cross-extremal dependences in both upper and lower tails.

3 Inference

3.1 Marginal modeling

Let $(Y_{1,1}^\top, \ldots, Y_{p,1}^\top)^\top, \ldots, (Y_{1,n}^\top, \ldots, Y_{p,n}^\top)^\top$ denote $n$ independent copies from a random vector $Y = (Y_1^\top, \ldots, Y_p^\top)^\top$ that shares the same copula (2) as the vector $X$ in (1) but possesses potentially different marginal distributions $F_1^Y, \ldots, F_p^Y$, here assumed to be continuous and the same across space, though it can be easily generalized to the case where each spatial process has non-stationary margins. We write $Y_{i,k} = (Y_{i1,k}, \ldots, Y_{id,k})^\top$ for $i = 1, \ldots, p$ and $k = 1, \ldots, n$. The joint distribution of $Y$ can be expressed as

$$F^Y(y_1, \ldots, y_p) = \Pr(Y_1 \leq y_1, \ldots, Y_p \leq y_p)$$

$$= F^X((F_1^X)^{-1}\{F_1^Y(y_1)\}, \ldots, (F_p^X)^{-1}\{F_p^Y(y_p)\}) = C^X(F_1^Y(y_1), \ldots, F_p^Y(y_p)),$$
Figure 1: Simulation results for Cases 1 and 2 (left and right columns, respectively). For each field, we generated \( n = 2000 \) replicates at two random locations within the unit square. The simulated data are plotted both on the original scale of \( X \) (top row) and the uniform scale (bottom row). In Case 1, the upper tail spatial extremal dependence types (i.e., for each of the two individual fields) are AD and AI, respectively, and the upper tail cross-extremal dependence is AI. We write it as U: AD, AI–AI. We adopt a similar notation for the lower tail case. Precisely, the parameter settings for both cases are \( \theta = (\alpha_1 = 4, \alpha_2 = 5, \gamma_1 = 0.6, \gamma_2 = 0.2, \delta_U = 0.4, \delta_L = 0.7, \lambda_1 = 0.7, \lambda_2 = 0.3, \rho_{12} = 0.7)^\top \) (Case 1) and \( \theta = (\alpha_1 = 4, \alpha_2 = 5, \gamma_1 = 0.4, \gamma_2 = 0.3, \delta_U = 0.4, \delta_L = 0.8, \lambda_1 = 0.7, \lambda_2 = 0.3, \rho_{12} = -1)^\top \) (Case 2).

where the functions \((F_X^X)_{-i}^i\) and \((F_Y^Y)_{-i}^i\), \(i = 1, \ldots, p\), are applied componentwise. To estimate the copula \( C^X \) from an observed random sample \((y_{1,1}, \ldots, y_{p,1})^\top, \ldots, (y_{1,n}, \ldots, y_{p,n})^\top\), where \((y_{1,k}, \ldots, y_{p,k})^\top = (y_{11,k}, \ldots, y_{1d,k}, \ldots, y_{p1,k}, \ldots, y_{pd,k})^\top, k = 1, \ldots, n\), we first need to transform the data to the standard
uniform scale. To achieve this goal, we may either estimate $F^Y_i$ using a parametric model, or more simply using the empirical distribution function given as

$$\hat{F}^Y_i(y) = \frac{1}{n+1} \sum_{j=1}^d \sum_{k=1}^n I(y_{ij,k} \leq y), \quad i = 1, \ldots, p.$$  

Note that this formula needs to be adjusted when the data have different marginal distributions at different sites. Then we define pseudo-uniform scores using their corresponding ranks (computed among all $d$ spatial observations and $n$ independent replicates, but separately for each field $i = 1, \ldots, p$), i.e.,

$$u_{ij,k} = \hat{F}^Y_i(y_{ij,k}) = \frac{\text{rank}(y_{ij,k})}{n+1}, \quad i = 1, \ldots, p, \quad j = 1, \ldots, d, \quad k = 1, \ldots, n.$$  

The joint distribution of the proposed model cannot be obtained in the closed form. However, by conditioning on the latent factors and model parameters, the copula can be expressed analytically in terms of the Gaussian copula, which is computationally convenient. This suggests adopting a Bayesian inference approach based on Markov chain Monte Carlo (MCMC) methods, by treating the unobserved common factors as latent variables to be updated at each iteration of the MCMC algorithm.

### 3.2 Bayesian inference

Conditioning on the model parameter vector $\theta$ and latent variables $R = (R^U_0, R^L_0, R^U_1, R^L_1, \ldots, R^U_p, R^L_p)^T$, we have that the copula $C^X$ of uniformly transformed observations from model (1) is

$$C^X(u_1, \ldots, u_p \mid \theta, R) = \Pr(U_1 \leq u_1, \ldots, U_p \leq u_p \mid \theta, R)$$

$$= \Pr(X_1 \leq (F^X_1)^{-1}(u_1), \ldots, X_p \leq (F^X_p)^{-1}(u_p) \mid \theta, R)$$

$$= \Phi_{pd}(\Phi^{-1}[F_W(w_1)], \ldots, \Phi^{-1}[F_W(w_p)]; 0_{pd}, \Sigma_{pd \times pd}), \quad (3)$$

where $w_i = (F^X_i)^{-1}(u_i) - \alpha_i \{\gamma_i[\delta^U R^U_0 + (1 - \delta^U)R^U_i] - (1 - \gamma_i)[\delta^L R^L_0 + (1 - \delta^L)R^L_i]\}$, for $i = 1, \ldots, p$, $\Phi^{-1}$ is the univariate Gaussian quantile function and $\Phi_{pd}(\cdot; 0_{pd}, \Sigma_{pd \times pd})$ denotes the multivariate Gaussian distribution in dimension $pd$ with mean zero and correlation matrix $\Sigma$. Therefore, with $n$ independent realizations of $u = (u_1, \ldots, u_p)^T$, we hereby perform Bayesian inference of our model by jointly updating $\theta$ and the $n$ independent copies of $R$. To simplify the sampling scheme, we first transform $\theta$ and $R$ into new variables that are supported on the whole real line. In particular, we reparametrize $\theta$ as $\theta^* = g(\theta)$, such that $\text{supp}(\theta^*) = \mathbb{R}^M$, where $M = 9$ (when no parameter is fixed) is the dimension of the parameter
space. Similarly, we reparametrize $R$ as $R^* = h(R)$, such that $\text{supp}(R^*) = \mathbb{R}^Q$, where $Q = 2(p + 1)n$ is the dimension of the latent variable space. We use here log-transforms for positive variables or parameters, and (possibly rescaled) logit transforms for bounded parameters. The joint posterior density of $\theta^*$ and the transformed latent variables $R^*$, for a single replicate, may then be expressed as

$$
\pi(\theta^*, R^* \mid u) \propto \pi(u \mid \theta^*, R^*) \times \pi(\theta^*, R^*)
$$

$$
\propto \pi(u \mid \theta, R) \times \pi(R^*) \times \pi(\theta^*),
$$

where $u = (u_1^T, \ldots, u_p^T)^T$ are the observed data, and the density $\pi(u \mid \theta^*, R^*)$ can be derived by differentiating (3) with respect to $u$. The density $\pi(R^*) = \pi(R)/\partial h(R)/\partial R^*$, derived from the change-of-variable formula, can be computed from $\pi(R)$, which is a product of standard exponential densities (as defined in the proposed model). The term $\pi(\theta^*)$ is the prior density of $\theta^*$ and can be specified similarly as $\pi(\theta^*) = \pi(\theta)/\partial g(\theta)/\partial \theta^*$, where $\pi(\theta)$ is the prior density of $\theta$ on its original scale. With $n$ independent replicates, the densities $\pi(u \mid \theta^*, R^*)$ (i.e., the likelihood function) and $\pi(R^*)$ are replaced by products of $n$ similarly defined terms, and the joint posterior density has to be modified accordingly. Precisely, the likelihood function with $n$ replicates is

$$
\pi(\{u_k\}_{k=1}^n \mid \theta, \{R_k\}_{k=1}^n) = \prod_{k=1}^n \pi(u_{1,k}, \ldots, u_{p,k} \mid \theta, R_k),
$$

where $u_k = (u_{1,k}^T, \ldots, u_{p,k}^T)^T$, $u_{i,k} = (u_{i1,k}^T, \ldots, u_{id,k}^T)^T$, denotes the $k$-th uniform data replicate associated with the $k$-th independent latent random factor $R_k$, and the log-likelihood thus becomes

$$
\ell(\{u_k\}_{k=1}^n \mid \theta, \{R_k\}_{k=1}^n) = \log \pi(\{u_k\}_{k=1}^n \mid \theta, \{R_k\}_{k=1}^n)
$$

$$
= \sum_{k=1}^n \log \pi(u_{1,k}, \ldots, u_{p,k} \mid \theta, R_k)
$$

$$
= \sum_{k=1}^n \left[ \log \phi_{pd}(\Phi^{-1}[F_W(w_{1,k})], \ldots, \Phi^{-1}[F_W(w_{p,k})] \mid \theta, R_k) \right.
$$

$$
\left. + \sum_{i=1}^p \sum_{j=1}^d \log f_{W}(w_{ij,k}) - \log f_{X}^X([F_i^{-1}(u_{ij,k})] - \log \phi(\Phi^{-1}[F_W(w_{ij,k})])) \right],
$$

where $w_{i,k} = (w_{i1,k}, \ldots, w_{id,k})^T$, $i = 1, \ldots p$, denotes the $k$-th replicate of $w_i$. Using the expressions above, the overall joint log posterior density can thus easily be computed, up to an additive constant, as

$$
\log \pi(\theta^*, \{R_k^*\}_{k=1}^n \mid \{u_k\}_{k=1}^n) \equiv \ell(\{u_k\}_{k=1}^n \mid \theta, \{R_k\}_{k=1}^n) + \sum_{k=1}^n \log \pi(\theta^* \mid \theta, \{R_k\}_{k=1}^n) + \log \pi(\theta^*).
$$
3.3 MCMC algorithm

We hereby introduce our proposed MCMC algorithm customized to fit model (1). First, we present the general Metropolis–Hastings algorithm procedure and then describe proposal distributions that are designed to sample efficiently from the joint posterior distribution of hyperparameters and latent variables. Let \( \pi(z) \) be the target density (e.g., the posterior density) of an MCMC sampler where \( z \) is an \( m \)-dimensional vector of variables we need to infer. The basic Metropolis–Hastings algorithm is constructed by sampling candidate values \( z' \) according to a proposal distribution that has density \( z' \sim q(\cdot | z) \) (also known as the Markov kernel), and to accept the proposed \( z' \) with probability \( \alpha = \min \left\{ \frac{\pi(z')}{\pi(z)}, \frac{q(z | z')}{q(z' | z)} \right\} \). Based on the detailed balance property, if the resulting Markov chain \( \{ z^{(t)} \}_{t=1}^N \) (where \( N \) is the total number of iterations) is irreducible (Robert and Casella, 2013), convergence to the target distribution \( \pi(z) \) is guaranteed from this acceptance-rejection procedure. The choice of suitable proposal distributions is crucial for obtaining good performances (i.e., good mixing of Markov chains and fast convergence). We now describe our approach to efficiently sample candidate values of hyperparameters and high-dimensional latent variables from their posterior density in our case. Our proposed methodology relies on three features:

(i) Random walk proposal for hyperparameters. To update the vector \( \theta^* \in \mathbb{R}^M \), we choose a random walk proposal \( \theta'^* \sim q_{RW}(\cdot | \theta^*), \) where \( q_{RW} \) is the density of the Gaussian distribution \( \mathcal{N}_m(\theta^*, \sigma^2_{RW} I_M) \), and \( I_M \) is the \( M \times M \) identity matrix and \( \sigma^2_{RW} > 0 \) is a tuning parameter that controls the step size.

(ii) Metropolis-Adjusted Langevin Algorithm proposal (MALA) for latent variables. To update the high-dimensional vector of latent variables \( R^* = \{ R^*_k \}_{k=1}^n \in \mathbb{R}^Q \), we choose a MALA block proposal \( R'^* \sim q_{MALA}(\cdot | R^*, \theta^*), \) which has the density of \( \mathcal{N}_m(R^* + \frac{\sigma^2_{MALA}}{2} \nabla_{R^*} \log \pi(\theta^*, R^* | u), \sigma^2_{MALA} I_Q) \), where \( \sigma^2_{MALA} > 0 \) is a tuning parameter and \( I_Q \) is the \( Q \times Q \) identity matrix (recall that \( Q = 2(p+1)n \)). The MALA proposal was introduced by Roberts and Tweedie (1996) and further studied by Roberts and Rosenthal (1998). One key advantage of MALA proposals is the additional information on the gradient \( \nabla_{R^*} \log \pi(\theta^* R^* | u) \), which allows exploring high-dimensional parameter spaces more efficiently.

(iii) Adaptive strategy. As suggested in the above proposals, the tuning parameters \( \sigma^2_{RW} \) and \( \sigma^2_{MALA} \) (denoted generically below by \( \sigma^2 \)) directly affect the step size and Metropolis–Hastings acceptance
The developed algorithm is summarized in Algorithm 1.

**Algorithm 1:** MCMC Algorithm.

| Input          | Data $u = \{u_k\}_{k=1}^n$; Initial parameters $\{\theta^{(0)}_k, R^{(0)}_k\}$, where $R^{(0)}_k = \{R^{(0)}_k\}_{k=1}^n$. |
|----------------|---------------------------------------------------------------------------------|
| Output         | MCMC chain $\{\theta^{(t)}_k, R^{(t)}_k\}_{t=1,...,N}$, where $R^{(t)}_k = \{R^{(t)}_k\}_{k=1}^n$. |
| for $t \leftarrow 1$ to $N$ do |                                                                              |
| // Block 1: updating $\theta^*$ with RW proposal |                                                                              |
| $\theta^{(p)} \sim q_{\text{RW}}(\cdot | \theta^{(t-1)});$ |                                                                              |
| $\alpha_{\text{RW}}(\theta^{(t-1)}, \theta^{(p)}) \leftarrow \min \left\{ 1, \frac{\pi(\theta^{(p)}, R^{(t-1)}_k | u)}{\pi(\theta^{(t-1)}, R^{(t-1)}_k | u)} \right\};$ Draw $U \sim U(0,1);$ |                                                                              |
| if $\alpha_{\text{RW}}(\theta^{(t-1)}, \theta^{(p)}) \leq U$ then |                                                                              |
| $\theta^{(t)} \leftarrow \theta^{(p)}$ |                                                                              |
| else |                                                                              |
| $\theta^{(t)} \leftarrow \theta^{(t-1)}$ |                                                                              |
| // Block 2: updating latent variables $R^*$ with MALA proposal |                                                                              |
| $R^{(p)} \sim q_{\text{MALA}}(\cdot | \theta^{(t)}, R^{(t-1)});$ |                                                                              |
| $\alpha_{\text{MALA}}(R^{(t-1)}, R^{(p)}) \leftarrow \min \left\{ 1, \frac{\pi(\theta^{(t)}, R^{(t)}_k | u)q_{\text{MALA}}(R^{(t-1)}_k | R^{(t)}_k, \theta^{(t)}_k)}{\pi(\theta^{(t-1)}, R^{(t-1)}_k | u)q_{\text{MALA}}(R^{(t)}_k | R^{(t-1)}_k, \theta^{(t-1)}_k)} \right\};$ Draw $V \sim U(0,1);$ |                                                                              |
| if $\alpha_{\text{MALA}}(R^{(t-1)}, R^{(p)}) \leq V$ then |                                                                              |
| $R^{(t)} \leftarrow R^{(p)}$ |                                                                              |
| else |                                                                              |
| $R^{(t)} \leftarrow R^{(t-1)}$ |                                                                              |
| // Adaptive strategy |                                                                              |
| if $t < N_b$ and $t \pmod{N_0} = 0$ then |                                                                              |
| update both $\sigma_{\text{RM}}$ and $\sigma_{\text{MALA}}$ according to (5); |                                                                              |
| end |                                                                              |
| end |                                                                              |
| Return $\{\theta^{(t)}_k, R^{(t)}_k\}_{t=1,...,N}$; |                                                                              |
4 Simulation study

4.1 Simulation setting and two-step estimation approach

In the simulation study, we mimic our real data application and generate \( n = 300 \) independent replicates at \( d = 25 \) randomly sampled locations on the unit square \([0,1]^2\) using the bivariate \( (p = 2) \) spatial model specified in §2.3. Here the true parameters are set to \( \alpha_1 = 4, \alpha_2 = 4, \gamma_1 = 0.4, \gamma_2 = 0.6, \delta^U = 0.8, \delta^L = 0.6, \lambda_1 = 0.6, \lambda_2 = 0.3, \) and \( \rho_{12} = -0.7 \), such that the resulting marginal and cross-extremal dependences of the upper and lower tails are as follows: U: AD, AD–AD and L: AD, AI–AI using the notation of §2.3.

Although all parameters of model (1) are theoretically identifiable, some may be weakly identifiable in practice as they all control various bulk and tail dependence characteristics. This means that all parameters may be difficult to estimate simultaneously. This is particularly challenging in relatively low sample sizes, and especially for estimating the parameters \( \delta^U \) and \( \delta^L \), which are mostly related to the cross-process tail structure. Therefore, to simplify inference, we suggest adopting a two-step inference approach, whereby \( \alpha_1, \alpha_2, \delta^U, \) and \( \delta^L \) are initially fixed to some reasonable candidate values, and all other five parameters are then estimated using the Bayesian algorithm outlined in §3. The different candidate models can finally be compared using various visual or more objective model diagnostics. Here, in this simulation, we only illustrate the second step for simplicity, i.e., we show that our Bayesian approach works well when \( \alpha_1, \alpha_2, \delta^U, \delta^L \) are fixed to their true values. Concerning the priors for \( \theta = (\gamma_1, \gamma_2, \lambda_1, \lambda_2, \rho_{12})^T \), we assign \( \text{Exp}(1) \) priors for all hyperparameters that have a domain at the positive real line, while we use uniform priors for parameters with bounded domains. Furthermore, we set the number of burn-in iterations to \( N_b = 10^5 \), the total number of iterations to \( N = 5 \times 10^5 \), and we monitor convergence of Markov chains through traceplots.

4.2 Extremal dependence diagnostics

To illustrate the performance of our Bayesian inference approach, we compare the bivariate extremal dependence measure, \( \chi \), estimated either empirically from the data or based on the fitted model, to the true \( \chi \) value obtained by simulating from (1) with a very large number of replicates. Let \( \chi_{11}^U(h,u), \chi_{22}^U(h,u), \) and \( \chi_{12}^U(h,u) \) denote the true sub-asymptotic upper tail dependence measure \( \chi \) between two variables from the first field, second field, and both fields, respectively, at distance \( h \) and threshold level \( u \in (0,1) \), i.e.,

\[
\chi_{i_1,i_2}^U(h,u) = \Pr(U_{i_1}(s_{j_1}) > u \mid U_{i_2}(s_{j_2}) > u) \text{ for } i_1,i_2 = 1,2, \text{ and } h = \|s_{j_1} - s_{j_2}\|, \ s_{j_1},s_{j_2} \in \mathcal{S}, \]

where
\( U_i(s) = F_i^X \{ X_i(s) \} \) denotes the \( i \)-th random field transformed to the uniform scale through its marginal distribution \( F_i^X \). By symmetry, a similar quantity, denoted \( \chi^L_{i,i}\_2(h,u) \), can be defined for the lower tail. We display the results by plotting them as a function of distance \( h \) and threshold \( u \).

### 4.2.1 Performance for fixed threshold

We first display the spatial extremal dependence estimates with respect to distance for the fixed thresholds \( u = 0.1 \) and \( u = 0.9 \) for the lower and upper tails, respectively. Figure 2 reports the results. Specifically, we plot the empirical estimates (black dots), model-based estimates (red lines) and true values (blue lines) of \( \chi^L_{11}(h,u) \), \( \chi^L_{22}(h,u) \), \( \chi^L_{12}(h,u) \) for fixed threshold \( u = 0.1 \), and \( \chi^U_{11}(h,u) \), \( \chi^U_{22}(h,u) \), \( \chi^U_{12}(h,u) \) for fixed threshold \( u = 0.9 \), as a function of distance \( h \) between sites. The envelopes displayed as gray shadows are pointwise 95% confidence intervals for the empirical estimates, and the envelopes displayed as pink shadows are 95% credible intervals for model-based estimates, obtained from the post-burn-in MCMC samples. We can see that while empirical estimates are very variable, our inference procedure performs very well at recovering the true tail dependence characteristics, with low uncertainty.

### 4.2.2 Performance for fixed distance

We then display estimates of the upper tail extremal dependence measures \( \chi^U_{11}(h,u) \), \( \chi^U_{22}(h,u) \), \( \chi^U_{12}(h,u) \) estimates with respect to the threshold level \( u \) for the upper tail and fixed distance \( h \). Figure 3 reports the results for short, median, and long distances (recall that the possible distances in the unit square ranges approximately from zero to 1.41). Results for the lower tail at different distances are given in the Supplementary Material. The conclusion is the same as for the fixed threshold case. From our simulation experiments, we can see that the proposed model can describe and capture the within-process and cross-extremal spatial dependences flexibly among different distances considered for both the upper and lower tails, and that our inference procedure works well.

### 5 Real data application to bivariate temperature extremes

#### 5.1 Context, data description and pre-processing

The state of Alabama, in the southeastern US, has humid subtropical climate and with its proximity to the Gulf of Mexico, it is also subject to extreme weather events like tropical storms, which include the
Figure 2: Empirical estimates (black dots), model estimates (red lines) and true values (blue lines) of $\chi^L_{11}(h, u = 0.1)$, $\chi^L_{22}(h, u = 0.1)$, $\chi^L_{12}(h, u = 0.1)$ (top row, left to right) for fixed threshold $u = 0.1$, and $\chi^U_{11}(h, u), \chi^U_{22}(h, u), \chi^U_{12}(h, u)$ (bottom row, left to right) for fixed threshold $u = 0.9$, as a function of distance $h$. The gray shadow envelopes are pointwise 95% confidence intervals for the empirical estimates. The pink shadow envelopes are 95% credible intervals sampled from the post-burn-in MCMC chains.

infamous Hurricanes Camille (1969) and Katrina (2005). Moreover, because of its warm and humid climate, heat-related health issues in Alabama have been an important concern for decades (Taylor and McGwin Jr, 2000; Wang et al., 2019; Carter et al., 2020; Wang et al., 2021). In our real data application, we apply the proposed model to study the lower and upper extremal dependence structures of the daily maximum (TMAX) and daily minimum (TMIN) air temperature data in Alabama, which are closely related to heat waves and tropical nights, which are extreme events severely affecting human health. We are interested in modeling jointly the spatial extremal dependence within and across the TMAX and TMIN processes, using the proposed model which can capture and help identify both tail dependence types for the upper and lower
Figure 3: Empirical estimates (black dots), model estimates (red lines) and true values (blue lines) of $\chi_{11}^{U}(h, u)$, $\chi_{22}^{U}(h, u)$, $\chi_{12}^{U}(h, u)$ (left to right) as a function of the threshold $u$, for fixed distance $h = 0.21, 0.71, 1.14$ (top to bottom), i.e., short, medium and long distance, respectively. The gray shadow envelopes are pointwise 95% confidence intervals for the empirical estimates. The pink shadow envelopes are 95% credible intervals sampled from the post-burn-in MCMC chains.
Figure 4: Map of Alabama in the Southeastern United States, with the monitoring stations. The size of each circle is proportional to the mean value of TMAX (left, yellow circles) and TMIN (right, blue circles).

tails. The complete daily dataset is available from January 1st, 2010, to May 29th, 2021, giving a total of 3554 observations at \( d = 18 \) monitoring sites in Alabama.

To focus on a rather extreme year, we here consider data from 2020, which contains the most active Atlantic hurricane season on record with 31 tropical or subtropical cyclones including 14 hurricanes, partly caused by the “La Niña” phenomenon that developed in the summer months of 2020. The year 2020 contains \( n = 271 \) daily observations with complete data at the 18 monitoring stations. The maximum distance between two stations is 493.71 km. Figure 4 displays the map of the study region and monitoring stations, as well as the annual mean value of TMAX and TMIN for 2020. This figure shows that there is a clear correspondence between TMAX and TMIN values, and that high values tend to be spatially clustered and co-located (with larger values close to the South, near the coast), but it is unclear whether, and to what extent, this is due to similar marginal distributions or to the dependence structure.

To disentangle marginal and dependence characteristics of the data, and to fit our proposed multivariate spatial copula model, we first need to retrieve stationary residuals that are uniformly distributed marginally. To this end, we preprocess our data site by site separately, first removing their seasonal pattern and then using the rank-based empirical distribution function to standardize the data to the uniform scale (details are
Figure 5: Residuals for each station obtained after removing the seasonalities of the standardized TMAX and TMIN (in uniform scale) for the year 2020.

given in the Supplementary Material. For 2020, bivariate scatterplots of residuals obtained at each station are displayed in Figure 5. Even though the sample size is rather small, we can see that most pairs are moderately dependent and mildly asymmetric, with usually slightly stronger dependence in the upper tail (i.e., high daily temperature maxima with high daily minima). This suggests that our proposed model might be appropriate for capturing these dependence features.

5.2 Model selection and diagnostics

We fit the bivariate spatial model specified in §2.3 to the data for the year 2020. For the inference, we follow the Bayesian procedure presented in §3, but because the sample size is rather small (here, \( n = 271 \)), we adopt the two-step approach outlined in the simulation study in §4. Specifically, we fix \( \alpha_1, \alpha_2 \in \{0.5, 2.5\} \) and \( \delta^U, \delta^L \in \{0.3, 0.6, 0.9\} \), and then apply our customized MALA-based MCMC algorithm of §3 to estimate the remaining parameters, for each \( \{\alpha_1, \alpha_2, \delta^U, \delta^L\} \) parameter configuration. This means that in total, we have 81 different candidate model specifications and we then compare the fitted models based on the estimated \( \chi \) metric. Among the 81 candidate models, our “best” model, which provides the best fit in terms of bivariate
extremal dependence structures has \( \alpha_1 = 2, \alpha_2 = 2, \gamma_1 = 0.39, \gamma_2 = 0.60, \delta^U = 0.9, \delta^L = 0.9, \lambda_1 = 888.8, \lambda_2 = 868.6, \) and \( \rho_{12} = 0.06 \), which suggests that the most plausible extremal dependence structures for the upper and the lower tails of the bivariate TMAX/TMIN spatial process might be respectively of the following types: U: AI, AD–AI and L: AD, AI–AI, using the notation introduced previously. In other words, while TMAX appears to be (spatially) asymptotically independent in the upper tail but asymptotically dependent in the lower tail, the opposite is true for TMIN, thus inducing asymptotic independence in both tails for the cross-dependence between TMAX and TMIN.

Then, we compare the model performance at finite threshold levels and different distances by plotting the extremal dependence measures \( \chi_{11}(h,u), \chi_{22}(h,u), \) and \( \chi_{12}(h,u) \), estimated both empirically and based on our best fitted model, as a function of distance \( h \) and threshold \( u \). Figure 6 shows the model performance for the upper tail, as a function of distance \( h \) for a fixed threshold \( u = 0.9 \) (top), and as a function of the threshold \( u \) for a fixed distance \( h = 28.39 \) (bottom), i.e., for two close-by sites. The results for medium and large distances, i.e., \( h = 232.73 \) and \( h = 432.79 \), are reported in the Supplementary Material. Overall, the plots show that the within- and cross-extremal dependence structures are captured accurately at all threshold levels and spatial distances, and that the model-based extremal dependence estimates have considerably lower uncertainty than their empirical counterparts. Furthermore, our results indicate that the spatial dependence in each individual field is quite strong, but the cross-dependence structure is rather weak (but not inexistent).

Using the estimated model parameters, we can then simulate synthetic data at a finer resolution over the study region. Figure 7 shows one simulation result from the fitted model, displayed on the uniform scale. This simulation illustrates again that both TMAX and TMIN processes are large-scale phenomena, with TMIN being more strongly dependent (spatially) than TMAX in the upper tail, and that they are also moderately cross-dependent, as peaks and troughs appear to be co-located.

6 Discussion

In this paper, we have proposed a novel multi-factor copula model for multivariate spatial data, with appealing tail dependence properties. The model is, by design, able to distinctly capture the different combinations of marginal and cross-extremal dependence types within and across different spatial random fields, with a
Figure 6: Empirical estimates (blue dots), and model estimates (red lines) of \( \chi_{11}^U(h,u = 0.9) \), \( \chi_{22}^U(h,u = 0.9) \), \( \chi_{12}^U(h,u = 0.9) \) (left to right) as a function of distance \( h \) for fixed \( u = 0.9 \) (top), and as a function of threshold \( u \) for fixed \( h = 28.39 \) (bottom). The envelopes (blue shadows) are pointwise 95% confidence intervals for the empirical estimates. The envelopes (pink shadows) are 95% credible intervals sampled from the post-burn-in MCMC chains.

smooth transition between tail dependence classes in the interior of the parameter space. To the best of our knowledge, this is the first multivariate spatial model of this kind, contrasting with all other multivariate spatial models from the published literature, which either lack tail flexibility or have restrictive symmetries. Thanks to its flexibility and ability to handle tail asymmetry, our proposed model may be used in a wide range of real-world applications. Although we mainly focused in our study on accurately capturing the joint lower and upper tails, we also stress that the proposed model was fitted to the entire dataset (i.e., not only to one tail), and that its flexibility allows one to accurately capture the full distribution (bulk and tails) of multivariate spatial data. Furthermore, inference can be performed using a Bayesian approach, which
exploits the natural hierarchical representation of our model described by a simple Gaussian copula at the data level, and multiple exponential random factors at the latent level. We adopted a MALA-type algorithm with adaptive block proposals to boost its efficiency and enhance the convergence of Markov chains to their stationary distribution. In case of high-dimensional data, computations of the required gradients can be run in parallel to accelerate the inference procedure. To further speed-up computations, Aicher et al. (2019) proposed the Stochastic gradient Langevin dynamics (SGLD) MCMC algorithm for models with latent factors, which we plan to explore in future research.

Finally, in our application, we investigated the spatial joint tail dependence structure of TMAX and TMIN in Alabama, US, and interestingly, our model revealed different asymptotic dependence types for each process and each tail. Although this finding is subject to uncertainty, it has important implications on environmental and heat-related health concerns, and it would be interesting to confirm our study in future research using a larger dataset.
Acknowledgments

This publication is based upon work supported by the King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) under Awards No. OSR-CRG2017-3434 and No. OSR-CRG2020-4394.

References

Aicher, C., Ma, Y.-A., Foti, N. J. and Fox, E. B. (2019) Stochastic gradient MCMC for state space models. *SIAM Journal on Mathematics of Data Science* 1(3), 555–587.

Baldwin, J. W., Dessy, J. B., Vecchi, G. A. and Oppenheimer, M. (2019) Temporally compound heat wave events and global warming: an emerging hazard. *Earth’s Future* 7(4), 411–427.

Bourgault, G. and Marcotte, D. (1991) Multivariable variogram and its application to the linear model of coregionalization. *Mathematical Geology* 23(7), 899–928.

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

Carter, A. W., Zaïtchik, B. F., Gohlke, J. M., Wang, S. and Richardson, M. B. (2020) Methods for estimating wet bulb globe temperature from remote and low-cost data: A comparative study in Central Alabama. *GeoHealth* 4(5), e2019GH000231.

Cressie, N. and Zammit-Mangion, A. (2016) Multivariate spatial covariance models: a conditional approach. *Biometrika* 103(4), 915–935.

Engelke, S., Opitz, T. and Wadsworth, J. (2019) Extremal dependence of random scale constructions. *Extremes* 22(4), 623–666.

Gelfand, A. E. (2021) Multivariate spatial process models. *Handbook of Regional Science* pp. 1985–2016.

Genton, M. G. and Kleiber, W. (2015) Cross-covariance functions for multivariate geostatistics. *Statistical Science* 30(2), 147–163.

Genton, M. G., Padoan, S. A. and Sang, H. (2015) Multivariate max-stable spatial processes. *Biometrika* 102(1), 215–230.

Gong, Y. and Huser, R. (2022) Asymmetric tail dependence modeling, with application to cryptocurrency market data. *The Annals of Applied Statistics* 16, 1822–1847.

Hazra, A., Huser, R. and Bolin, D. (2021) Realistic and fast modeling of spatial extremes over large geographical domains. arXiv preprint 2112.10248.

Hazra, A., Reich, B. J. and Staicu, A.-M. (2020) A multivariate spatial skew-t process for joint modeling of extreme precipitation indexes. *Environmetrics* 31(3), e2602.

Huser, R. and Wadsworth, J. L. (2019) Modeling spatial processes with unknown extremal dependence class. *Journal of the American Statistical Association* 114(525), 434–444.

Huser, R. and Wadsworth, J. L. (2022) Advances in statistical modeling of spatial extremes. *Wiley Interdisciplinary Reviews: Computational Statistics* 14(1), e1537.

Kabluchko, Z., Schlather, M. and De Haan, L. (2009) Stationary max-stable fields associated to negative definite functions. *The Annals of Probability* 37(5), 2042–2065.
Kleiber, W. and Nychka, D. (2012) Nonstationary modeling for multivariate spatial processes. *Journal of Multivariate Analysis* **112**, 76–91.

Kotz, S. and Nadarajah, S. (2004) *Multivariate t-Distributions and Their Applications*. Cambridge University Press.

Krupskii, P. and Genton, M. G. (2019) A copula model for non-Gaussian multivariate spatial data. *Journal of Multivariate Analysis* **169**, 264–277.

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

Littell, J. S., Peterson, D. L., Riley, K. L., Liu, Y. and Luce, C. H. (2016) A review of the relationships between drought and forest fire in the United States. *Global change biology* **22**(7), 2353–2369.

Mitchell, D. (2021) Climate attribution of heat mortality. *Nature Climate Change* **11**(6), 467–468.

Oesting, M., Schlather, M. and Friederichs, P. (2017) Statistical post-processing of forecasts for extremes using bivariate Brown-Resnick processes with an application to wind gusts. *Extremes* **20**(2), 309–332.

Opitz, T. (2013) Extremal t processes: Elliptical domain of attraction and a spectral representation. *Journal of Multivariate Analysis* **122**, 409–413.

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

Perkins-Kirkpatrick, S. and Lewis, S. (2020) Increasing trends in regional heatwaves. *Nature communications* **11**(1), 1–8.

Raymond, C., Horton, R. M., Zscheischler, J., Martius, O., AghaKouchak, A., Balch, J., Bowen, S. G., Camargo, S. J., Hess, J., Kornhuber, K., Oppenheimer, M., Ruane, A. C., Wahl, T. and White, K. (2020) Understanding and managing connected extreme events. *Nature Climate Change* **10**(7), 611–621.

Reich, B. J. and Shaby, B. A. (2012) A hierarchical max-stable spatial model for extreme precipitation. *The Annals of Applied Statistics* **6**(4), 1430–1451.

Robert, C. and Casella, G. (2013) *Monte Carlo Statistical Methods*. Springer Science & Business Media.

Roberts, G. O., Gelman, A., Gilks, W. R. *et al.* (1997) Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability* **7**(1), 110–120.

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.

Ruffault, J., Curt, T., Martin-StPaul, N. K., Moron, V. and Trigo, R. M. (2018) Extreme wildfire events are linked to global-change-type droughts in the northern Mediterranean. *Natural Hazards and Earth System Sciences* **18**(3), 847–856.

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

Seneviratne, S. I., Nicholls, N., Easterling, D., Goodess, C. M., Kanae, S., Kossin, J., Luo, Y., Marengo, J., McInnes, K., Rahimi, M., Reichstein, M., Sorteberg, A., Vera, C. and Zhang, X. (2012) Changes in climate extremes and their impacts on the natural physical environment. In *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change (IPCC)*, eds C. B. Field, V. Barros, T. F. Stocker, D. Qin, D. J. Dokken, K. L. Ebi, M. D. Mastrandrea, K. J. Mach, G.-K. Plattner, S. K. Allen, M. Tignor and P. M. Midgley, pp. 109–230. Cambridge, UK and New York, NY, USA: Cambridge University Press.
Sibuya, M. (1960) Bivariate extreme statistics, I. *Annals of the Institute of Statistical Mathematics* **11**(3), 195–210.

Smith, R. L. (1990) Max-stable processes and spatial extremes. Unpublished manuscript.

Taylor, A. J. and McGwin Jr, G. (2000) Temperature-related deaths in Alabama. *Southern Medical Journal* **93**(8), 787–792.

Vettori, S., Huser, R. and Genton, M. G. (2019) Bayesian modeling of air pollution extremes using nested multivariate max-stable processes. *Biometrics* **75**(3), 831–841.

Vicedo-Cabrera, A. M., Scovronick, N., Sera, F., Royé, D., Schneider, R., Tobias, A., Astrom, C., Guo, Y., Honda, Y., Hondula, D. M., Abrutzky, R., Tong, S., de Sousa Zanotti Stagliorio Coelho, M., Nascimento, P. H. *et al.* (2021) The burden of heat-related mortality attributable to recent human-induced climate change. *Nature Climate Change* **11**(6), 492–500.

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

Wang, S., Johnson, E., Tyson, S. and Gohlke, J. M. (2021) Heat-health behavior change during summer 2020 in African American Alabama residents. *American Journal of Public Health* **111**(8), 1443–1447.

Wang, S., Richardson, M. B., Wu, C. Y., Cholewa, C. D., Lungu, C. T., Zaitchik, B. F. and Gohlke, J. M. (2019) Estimating occupational heat exposure from personal sampling of public works employees in Birmingham, Alabama. *Journal of Occupational and Environmental Medicine* **61**(6), 518.

Yadav, R., Huser, R. and Opitz, T. (2021) Spatial hierarchical modeling of threshold exceedances using rate mixtures. *Environmetrics* **32**(3), e2662.

Yadav, R., Huser, R. and Opitz, T. (2022) A flexible Bayesian hierarchical modeling framework for spatially dependent peaks-over-threshold data. *Spatial Statistics* **51**, 100672.

Zhang, Z., Huser, R., Opitz, T. and Wadsworth, J. (2022) Modeling spatial extremes using normal mean-variance mixtures. *Extremes* **25**, 175–197.

Zscheischler, J., Martius, O., Westra, S., Bevacqua, E., Raymond, C., Horton, R. M., van den Hurk, B., AghaKouchak, A., Jézéquel, A., Mahecha, M. D., Maraun, D., Ramos, A. M., Ridder, N. N., Thiery, W. and Vignotto, E. (2020) A typology of compound weather and climate events. *Nature Reviews Earth & Environment* **1**(7), 333–347.
A Basic model properties

A.1 Marginal distribution functions
Recall the model construction

\[ X_i(s) = \alpha_i \{ \gamma_i \delta^U R_{0i} + (1 - \delta^U) R_{1i}^U \} - (1 - \gamma_i) \delta^L R_{0i} + (1 - \delta^L) R_{1i}^L \} + W_i(s). \]

Let

\[ T_i := \alpha_i \{ \gamma_i \delta^U R_{0i} + (1 - \delta^U) R_{1i}^U \} - (1 - \gamma_i) \delta^L R_{0i} + (1 - \delta^L) R_{1i}^L \}
\[ = \beta_{i0}^U R_{0i}^U + \beta_{i1}^U R_{1i}^U - \beta_{i0}^L R_{0i}^L - \beta_{i1}^L R_{1i}^L, \]

where \( \beta_{i0}^U = \alpha_i \gamma_i \delta^U, \beta_{i1}^U = \alpha_i \gamma_i (1 - \delta^U), \beta_{i0}^L = \alpha_i (1 - \gamma_i) \delta^L, \) and \( \beta_{i1}^L = \alpha_i (1 - \gamma_i) (1 - \delta^L). \) Then the model can be written as

\[ X_i(s) = T_i + W_i(s), \]

and it is straightforward to derive that the marginal distribution function of \( X_i \) can be expressed as

\[
F_i^X(x) = \begin{cases} 
1/2 \exp(x) & (\beta_{i0}^L)^5 \exp(x/\beta_{i0}^L) \\
(\beta_{i0}^U + 1)(\beta_{i1}^U - 1)(\beta_{i1}^L - 1) + (\beta_{i0}^L + 1)(\beta_{i0}^L + \beta_{i1}^L)(\beta_{i0}^L + \beta_{i1}^L)(\beta_{i0}^L - \beta_{i1}^L) & \text{for } \beta_{i0}^L \neq 1, \beta_{i1}^L \neq 1, \beta_{i0}^U \neq \beta_{i1}^U, \quad w < 0 \end{cases}
\]

\[
1 - 1/2 \exp(-x) - (\beta_{i0}^L)^5 \exp(-x/\beta_{i0}^L) \\
(\beta_{i0}^U - 1)(\beta_{i0}^L + 1)(\beta_{i1}^L + 1) - (\beta_{i0}^L + 1)(\beta_{i0}^L + \beta_{i1}^L)(\beta_{i0}^L + \beta_{i1}^L)(\beta_{i0}^L - \beta_{i1}^L) & \text{for } \beta_{i0}^U \neq 1, \beta_{i1}^L \neq 1, \beta_{i0}^L \neq \beta_{i1}^L, \quad w \geq 0.
\]

All other cases can be derived as limits.

A.2 The covariance structure of the bivariate \((W_1(s), W_2(s))^\top\) process
Recall the model construction and notation in Section 2.1. The components of the mean vector are

\[ \text{E}(W_1^*) = \text{E}(W_2^*) = 0_d, \]
\[ \text{E}(W_1^2) = \text{E}(W_2^2) = \rho_{12} \text{E}(W_1^*) + \sqrt{1 - \rho_{12}^2} \text{E}(W_2^*) = 0_d, \]

and the covariance matrix \( \Sigma_{2d \times 2d} \) can be constructed as

\[
\begin{bmatrix}
\Sigma'_{11} & \Sigma'_{12} \\
\Sigma'_{12} & \Sigma'_{22}
\end{bmatrix},
\]

where the elements of the block covariance matrix are

\[
\Sigma_{1,jj} = \text{var}(W_1^*) = \text{var}(W_2^*) = 1,
\]
\[
\Sigma_{1,jk} = \text{cov}(W_1^*, W_{1k}^*) = \Sigma_{1,kj} = \text{cov}(h_{jk}),
\]
\[
\Sigma'_{2,jj} = \text{var}(W_2^*), \quad \Sigma'_{2,jk} = \text{cov}(W_{2j}^*, W_{2k}^*) = \rho_{21} \text{var}(W_{2j}^*) + (1 - \rho_{21}^2) \text{var}(W_{2k}^*) = 1,
\]
\[
\Sigma'_{2,jk} = \text{cov}(W_{2j}^*, W_{2k}^*) = \text{cov}(\rho_{12} W_{1j}^* + \sqrt{1 - \rho_{12}^2} W_{2j}^*) = \rho_{12} \text{var}(W_{1j}^*) + (1 - \rho_{12}^2) \text{var}(W_{2j}^*) = 1.
\]

26
\[ \sum_{i,j}^{X} = \text{cov}(W_{ij}, W_{jk}) + (1 - \rho_{12}^2) \text{cov}(W_{2j}, W_{2k}) \]
\[ = \rho_{12}^2 c_1(h_{jk}) + (1 - \rho_{12}^2)c_2(h_{jk}), \]
\[ \sum_{i,j}^{X} = \text{cov}(W_{ij}, W_{jk}) + (1 - \rho_{12}^2) \text{cov}(W_{2j}, W_{2k}) \]
\[ = \rho_{12}^2 c_1(h_{jk}) + (1 - \rho_{12}^2)c_2(h_{jk}), \]

and \( j, k = 1, \ldots, d. \)

**B Proof of tail dependence properties**

We start with the formal proof of the spatial tail dependence properties of each individual process \( X_i(s) \).

Following Appendix §A.1, the model can be written as

\[ X_i(s) = T_i + W_i(s), \]

and the cumulative distribution function of \( T_i \) can be expressed in closed form, namely

\[
F_{T_i}(t) = \begin{cases} 
A_{\alpha} \exp(t/\beta_{\alpha}^L) - A_{\gamma} \exp(t/\beta_{\gamma}^U), & t < 0, \\
1 - B_{\alpha} \exp(-t/\beta_{\alpha}^L) - B_{\gamma} \exp(-t/\beta_{\gamma}^U), & t \geq 0.
\end{cases}
\]

where

\[
A_{\alpha} = \frac{(\beta_{\alpha}^L)^3}{(\beta_{\alpha}^L + \beta_{\alpha}^U)(\beta_{\alpha}^L + \beta_{\alpha}^0)(\beta_{\alpha}^L - \beta_{\alpha}^U)}, \quad A_{\gamma} = \frac{(\beta_{\gamma}^L)^3}{(\beta_{\gamma}^L + \beta_{\gamma}^U)(\beta_{\gamma}^L + \beta_{\gamma}^0)(\beta_{\gamma}^L - \beta_{\gamma}^U)},
\]

\[
B_{\alpha} = \frac{(\beta_{\alpha}^L)^3}{(\beta_{\alpha}^L + \beta_{\alpha}^U)(\beta_{\alpha}^0 + \beta_{\alpha}^L)(\beta_{\gamma}^0 + \beta_{\gamma}^L)(\beta_{\gamma}^L - \beta_{\gamma}^U)}, \quad B_{\gamma} = \frac{(\beta_{\gamma}^L)^3}{(\beta_{\gamma}^L + \beta_{\gamma}^U)(\beta_{\gamma}^L + \beta_{\gamma}^0)(\beta_{\gamma}^L - \beta_{\gamma}^U)}.
\]

By taking the exponential on both components of \( X_i(s) \), we obtain \( \tilde{X}_i(s) = \tilde{T}_i \tilde{W}_i(s) \), where \( \tilde{T}_i = \exp(T_i) \) and \( \tilde{W}_i(s) = \exp(W_i(s)) \), for \( i = 1, \ldots, p \). Then, for \( t > 1 \), we have

\[
\Pr(\tilde{T}_i > t) = \Pr(T_i > \log t) = B_{\alpha} t^{-1/\beta_{\alpha}^0} + B_{\gamma} t^{-1/\beta_{\gamma}^U} \xrightarrow{t \to \infty} \begin{cases}
B_{\alpha} t^{-1/\beta_{\alpha}^0}, & \text{if } \beta_{\alpha}^U > \beta_{\gamma}^U, \\
B_{\gamma} t^{-1/\beta_{\gamma}^U}, & \text{if } \beta_{\alpha}^U < \beta_{\gamma}^U.
\end{cases}
\]

Therefore, the value \( \max(\beta_{\alpha}^U, \beta_{\gamma}^U) \) determines the decay rate of \( \Pr(\tilde{T}_i > t) \), and \( \tilde{T}_i \) is regularly-varying at infinity with index \( -\alpha_{\tilde{T}_i} \), with \( \alpha_{\tilde{T}_i} = 1/\max(\beta_{\alpha}^U, \beta_{\gamma}^U) \).

Now, for \( w > 1 \), \( \Pr(\tilde{W}_i > w) = \Pr(W_i > \log w) = 1/(2w) \), which implies that \( \tilde{W}_i \) is regularly-varying at infinity with index \( -1 \). Moreover, \( \Pr(\tilde{W}_i > 0) = 1 \), for \( i = 1, \ldots, p \). Let \( \varepsilon > 0 \) and define \( \bar{\varepsilon} = \varepsilon/\alpha_{\tilde{T}_i} > 0 \),

\[
E(\tilde{W}_i^{\alpha_{\tilde{T}_i}^{\bar{\varepsilon}}}) = \int_0^\infty \Pr(\tilde{W}_i^{\alpha_{\tilde{T}_i}^{\bar{\varepsilon}}} > w)dw = \int_0^\infty \Pr(\tilde{W}_i > w^{1/(\alpha_{\tilde{T}_i}(1+\bar{\varepsilon}))})dw
\]

The first integral is bounded above by 1 and the second integral is finite if and only if \( 1/\{\alpha_{\tilde{T}_i}(1+\bar{\varepsilon})\} > 1 \), i.e., \( \alpha_{\tilde{T}_i} < 1/(1+\bar{\varepsilon}) \).
Then, by letting $\varepsilon \to 0$, we conclude from Table 2 of Engelke et al. (2019) that when $\alpha_{\tilde{T}_i} < 1$, or equivalently when $\max(\beta_{U_0}, \beta^U) = \alpha_{\tilde{T}_i} \max(\delta^U, 1-\delta^U) > 1$, the vector $(X_i(s_{j_1}), X_i(s_{j_2}))^T$ (which has the same copula structure as $(\tilde{X}_i(s_{j_1}), \tilde{X}_i(s_{j_2}))^T$) is AD. On the other hand, when $\alpha_{\tilde{T}_i} > 1$, because $(W_i(s_{j_1}), W_i(s_{j_2}))^T$ is Gaussian with correlation (and thus has $\chi^U = 0$ for correlations less than one, according to Sibuya (1960) and Ledford and Tawn (1996)), we deduce from Proposition 5 of Engelke et al. (2019) that the vector $(X_i(s_{j_1}), X_i(s_{j_2}))^T$ has $\chi^U = 0$, i.e., it is AI. The case $\alpha_{\tilde{T}_i} = 1$ can be deduced by applying Proposition 6(3c) of Engelke et al. (2019). Similar arguments hold for the lower tail.

For the cross-dependence structure, we argue by analogy. The vector $(X_{i_1}(s_{j_1}), X_{i_2}(s_{j_2}))^T$ has the same dependence structure as the vector $(\tilde{X}_{i_1}(s_{j_1}), \tilde{X}_{i_2}(s_{j_2}))^T$, for $i, i_1, i_2 = 1, \ldots, p$ and $j, j_1, j_2 = 1, \ldots, d$. Both components share the random variables $R_{U_0}$ and $R_{L_0}$, which are the only terms that can induce AD in the upper and lower tails, respectively, provided their respective coefficients are large enough. The term $R_{U_0}$ is the dominating term in the upper tail decay of both processes when we have $\beta_{i_10}^U > \max\{\beta_{i_1}^U, 1\}$ and $\beta_{i_20}^U > \max\{\beta_{i_2}^U, 1\}$ simultaneously, i.e., $\delta^U > 1 - \delta^U$ (i.e., $\delta^U > 1/2$), $\alpha_{i_1} \gamma_{i_1} \delta^U > 1$ and $\alpha_{i_2} \gamma_{i_2} \delta^U > 1$. The result follows. A similar argument holds for the lower tail. To summarize, we display the marginal and cross tail dependence properties in Table 1.