Received: 20 May 2024 | Revised: 19 June 2024 | Accepted: 4 July 2024

DOI: 10.1002/sta4.715

ORIGINAL ARTICLE

WILEY

A non-stationary factor copula model for non-Gaussian spatial data

Sagnik Mondal[1] | Pavel Krupskii[2] | Marc G. Genton[1]

[1]Statistics Program, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

[2]School of Mathematics and Statistics, University of Melbourne, Melbourne, Australia

Correspondence
Marc G. Genton, Statistics Program, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia.
Email: marc.genton@kaust.edu.sa

Funding information
King Abdullah University of Science and Technology

Abstract

We introduce a new copula model for non-stationary replicated spatial data. It is based on the assumption that a common factor exists that controls the joint dependence of all the observations from the spatial process. As a result, our proposal can model tail dependence and tail asymmetry, unlike the Gaussian copula model. Moreover, we show that the new model can cover a full range of dependence between tail quadrant independence and tail dependence. Although the log-likelihood of the model can be obtained in a simple form, we discuss its numerical computational issues and ways to approximate it for drawing inference. Using the estimated copula model, the spatial process can be interpolated at locations where it is not observed. We apply the proposed model to temperature data over the western part of Switzerland, and we compare its performance with that of its stationary version and with the Gaussian copula model.

KEYWORDS
factor copula model, non-Gaussian model, non-stationarity, replicated spatial data, tail asymmetry, tail dependence

1 | INTRODUCTION

Traditionally, many geostatistical applications have used Gaussian random fields for modelling spatial data. This bias toward Gaussian models is mainly due to the ease of inference based on Gaussian assumptions. Indeed, a mean vector and a covariance matrix can completely characterize a Gaussian model. Moreover, the marginal and conditional distributions of Gaussian random vectors are readily available in explicit forms. These are the main reasons behind the popularity of Gaussian random fields in geostatistical applications. However, Gaussian random fields also have some shortcomings. For example, the Gaussian model assumes symmetry for all the marginals and fails to capture potential outliers in the data. Furthermore, the Gaussian model is not appropriate when data show some signs of extremal dependence. Therefore, there is a need for more flexible spatial models that address these shortcomings.

In the statistical literature, different types of non-Gaussian random fields have been introduced to tackle the shortcomings of the Gaussian model. We discuss a few broad classes of non-Gaussian random fields next. A popular non-Gaussian modelling approach for analysing spatial data is the trans-Gaussian random field, obtained by taking a non-linear monotone transformation of a Gaussian random field. Examples of such models are log-normal random fields (Cressie, 1993; De Oliveira, 2006), Box–Cox transformation (De Oliveira et al., 1997), square-root transformation (Johns et al., 2003), and power transformations (Allcroft & Glasbey, 2003). For these models, one has to posit the possible transformation responsible for making the data non-Gaussian. If the correct transformation is used, the data can be made Gaussian by taking the inverse transformation, and inference can be made based on the Gaussian model. Instead of looking for this inverse transformation, one can model the original data directly and draw inference on the original scale. The Tukey *g*-and-*h* random field introduced by Xu and Genton (2017) is an example of such a non-Gaussian model where inference can be drawn directly on the original scale and the transformation is estimated from the data. Another approach for constructing non-Gaussian models is by mixing the Gaussian process with a random variable or with one or more independent

random processes. For instance, Palacios and Steel (2006) and Fonseca and Steel (2011) introduced a non-Gaussian model by mixing the scale parameter of a Gaussian model at each location. Ma (2009b) proposed a method to construct non-Gaussian random fields by multiplying a random scale factor with a Gaussian random field yielding elliptically contoured random fields. Ma (2009a) later introduced χ^2 random fields by summing the squares of m independent Gaussian processes. Yin and Craigmile (2018) presented the heteroscedastic asymmetric spatial process by mixing two dependent Gaussian random fields and transforming the resulting mixed process. Various non-Gaussian distributions can be used to construct non-Gaussian random fields. For instance, the skew-normal distribution used by Kim and Mallick (2004), Zhang and El-Shaarawi (2010), Genton and Zhang (2012), and Rimstad and Omre (2014). The t-distribution and the skew-t distribution have also been extended to random fields by Røislien and Omre (2006) and Bevilacqua et al. (2021). Marchenko and Genton (2010) created a non-Gaussian random field using log-skew-elliptical distribution. Zareifard and Khaledi (2013) proposed a non-Gaussian random field by scale mixing a unified skew-Gaussian process.

A crucial problem with the Gaussian model is that it fails to adequately model data that show signs of extremal dependence. In some geostatistical applications, such as studies related to extreme natural events, sometimes the extremal dependence is of particular interest. To capture the extremal dependence, if there is any, we need suitable models. The families of non-Gaussian random fields that are mentioned so far are not designed to capture potential extremal dependence. While trans-Gaussian random fields have no tail dependence, mixing distributions and t-distributions can handle tail dependence. Models based on the theory of copulas can be useful for constructing non-Gaussian spatial models capable of capturing extremal dependence. A copula is a p-variate distribution function of a p-variate random vector distributed over the $[0,1]^p$ hypercube with marginals distributed uniformly over the interval $[0,1]$. Sklar (1959) showed that for any p-dimensional distribution F with marginal distribution functions $F_1,\ldots,F_p$ there exists a p-variate copula C such that

$$F(x_1,\ldots,x_p) = C\{F_1(x_1),\ldots,F_p(x_p)\},\ (x_1,\ldots,x_p)^\top = \boldsymbol{x} \in \mathbb{R}^p.$$

Moreover, Sklar (1959) also showed that any combination of any p-dimensional copula C with any p univariate marginal distributions $F_1,\ldots,F_p$ results in a p-variate joint distribution function.

The theory of copulas has been used in constructing non-Gaussian spatial models. For example, Bárdossy (2006) introduced a non-Gaussian spatial model based on the chi-squared copula. Furthermore, Bárdossy and Li (2008) proposed a v-transformed copula, by taking a nonmonotonic transformation of multivariate normal variables and Bárdossy (2011) used another copula-based approach for modelling non-Gaussian groundwater data. The copula models based on the v-transformed copula can handle marginal and joint skewness but fail to model tail dependence. Moreover, the likelihood functions of these models are difficult to compute in high dimensions. The construction of flexible models for handling non-Gaussian spatial data with extremal dependence can be achieved by using vine copulas. Gräler and Pebesma (2011) and Gräler (2014) introduced non-Gaussian spatial models using vine copula and Erhardt et al. (2015) used a C-vine copula to model non-Gaussian spatial data. For these models, the composite likelihood method has been used for parameter estimation. These models are very capable of modelling tail dependence but lack interpretability. Krupskii et al. (2018) introduced the factor copula model for modelling stationary non-Gaussian processes and later Castro-Camilo and Huser (2020) generalized it for non-stationary non-Gaussian processes. The factor copula model is easy to interpret and can model replicated spatial data that show signs of tail dependence and tail asymmetry. Krupskii and Genton (2017, 2019) extended the idea of the factor copula model to spatio-temporal and multivariate non-Gaussian processes, respectively.

The factor copula model introduced by Krupskii and Genton (2019) is in the multivariate setting. In the univariate case, that model observed at location $\boldsymbol{s} \in \mathbb{R}^d$ is of the following form:

$$W(\boldsymbol{s}) = Z(\boldsymbol{s}) + \alpha_0^U \mathcal{E}_0^U + \alpha^U \mathcal{E}^U(\boldsymbol{s}) - \alpha_0^L \mathcal{E}_0^L - \alpha^L \mathcal{E}^L(\boldsymbol{s}), \tag{1}$$

with $\alpha_0^U > 0, \alpha^U > 0, \alpha_0^L > 0, \alpha^L > 0, Z(\boldsymbol{s})$ is a Gaussian process with mean 0 and with some correlation function, and $\mathcal{E}_0^U, \mathcal{E}^U(\boldsymbol{s}), \mathcal{E}_0^L$ and $\mathcal{E}^L(\boldsymbol{s})$ are mutually independent unit exponential random variables, $\mathcal{E}(1)$. Here, we term $\mathcal{E}_0^U$ and $\mathcal{E}_0^L$ as the upper and the lower common factor, respectively, and $\mathcal{E}^U(\boldsymbol{s})$ and $\mathcal{E}^L(\boldsymbol{s})$ as the upper and the lower independent factor, respectively. Due to the use of both common and independent factors for both upper and lower tails, this model is appropriate for modelling data with upper and/or lower tail dependence. In this work, we use the one-sided version of $W(\boldsymbol{s})$ from Equation (1) and generalize it further for accommodating non-stationary datasets as well. Thus, the new model is of the following form:

$$W(\boldsymbol{s}) = Z(\boldsymbol{s}) + \alpha_0(\boldsymbol{s})\mathcal{E}_0 + \alpha\mathcal{E}(\boldsymbol{s}), \tag{2}$$

with $\alpha_0(\boldsymbol{s}) > 0, \alpha > 0, Z(\boldsymbol{s})$ is a Gaussian process with mean 0 and with some correlation function (e.g., exponential or Matérn), and $\mathcal{E}_0 \sim \mathcal{E}(1), \mathcal{E}(\boldsymbol{s})$ is a white noise process with $\mathcal{E}(1)$ marginals, and $\mathcal{E}_0$ and $\mathcal{E}(\boldsymbol{s})$ are independently distributed. The copula governed by the model in Equation (2) is tail-asymmetric with a stronger upper tail than the lower tail as we are only considering the upper common factor from Equation (1) in the model in Equation (2). As a result, the model in Equation (2) is appropriate for datasets with upper tail dependence. Lower tail dependence can be modelled using the reflected copula of (2). Although models that can handle both lower and upper tail dependence are also desirable, in practice,

we mostly see datasets with either lower or upper tail dependence. Hence, the proposed model can be applied to many real datasets. The model in Equation (2) is a generalized version of the model proposed by Castro-Camilo and Huser (2020) for which $\alpha = 0$ and the model proposed by Krupskii et al. (2018) for which $\alpha = 0$ and $\alpha_0(\boldsymbol{s})$ is constant. Moreover, Castro-Camilo and Huser (2020) introduced non-stationarity in the model by proposing locally stationary models and allowing the parameters to change smoothly with locations, whereas the model in Equation (2) assumes global non-stationarity.

The proposed model has some similarities with spatial convolution-type models with non-stationary kernels. The main difference is that a convolution process involves an integral over a Lévy process with independent increments whereas in the proposed model, we use a finite number of additive exponential factors (one common and n independent factors where n is the number of spatial locations). Both methods can handle non-stationary non-Gaussian data; however, even stationary non-Gaussian convolution models are not computationally tractable because no likelihood is available in closed form; see Krupskii and Huser (2022) and references therein.

In this paper, Section 2 focuses on the properties of the model in Equation (2) and covers new important tail properties of this model. We present how to draw inference based on the model in Section 3, where we discuss how to estimate the model parameters with the exact and the Vecchia approximated log-likelihood, and how to make predictions at new locations based on the model. In Section 4, we justify our methods for drawing inference by a simulation study. Finally, in Section 5, we apply our model to a spatial dataset of daily average temperatures in Switzerland.

2 | THE NON-STATIONARY FACTOR COPULA MODEL

2.1 | The model and its copula

Suppose the process $W(\boldsymbol{s})$ from Equation (2) has been observed at n locations $\boldsymbol{s}_1,...,\boldsymbol{s}_n$ and write $\boldsymbol{W} = (W_1,...,W_n)^\top$, where

$$W_i = W(\boldsymbol{s}_i) = Z(\boldsymbol{s}_i) + \alpha_0(\boldsymbol{s}_i)\mathcal{E}_0 + \alpha\mathcal{E}(\boldsymbol{s}_i) = Z_i + \alpha_{0i}\mathcal{E}_0 + \alpha\mathcal{E}_i,\ i \in \{1,...,n\}, \tag{3}$$

and $\boldsymbol{Z} = (Z_1,...,Z_n)^\top$ follows a multivariate Gaussian distribution, with zero mean and a covariance matrix $\boldsymbol{\Sigma_Z}$ obtained from some correlation function. Moreover, $\mathcal{E}_0, \mathcal{E}_1,...,\mathcal{E}_n$ are i.i.d. $\mathcal{E}(1)$ random variables that are independent of $\boldsymbol{Z}$. From the definition of $\boldsymbol{W}$, we can derive the correlation between the variables at two locations W_{i_1} and W_{i_2}:

$$\mathbb{C}\text{orr}(W_{i_1}, W_{i_2}) = \{\mathbb{C}\text{orr}(Z_{i_1}, Z_{i_2}) + \alpha_{0i_1}\alpha_{0i_2}\} / \prod_{j \in \{i_1, i_2\}} \sqrt{1 + \alpha_{0j}^2 + \alpha^2},\ i_1, i_2 = 1,...,n.$$

Consequently, we can see that when $\mathbb{C}\text{orr}(Z_{i_1}, Z_{i_2}) = 1, \mathbb{C}\text{orr}(W_{i_1}, W_{i_2}) = \frac{1+\alpha_{0i_1}\alpha_{0i_2}}{\prod_{j \in \{i_1 i_2\}} \sqrt{1+\alpha_{0j}^2+\alpha^2}} \leq 1$. This suggests that even if the latent process at two locations is perfectly correlated, it does not mean the process $W(\boldsymbol{s})$ at those two locations will also be perfectly correlated. Moreover, when $\mathbb{C}\text{orr}(Z_{i_1}, Z_{i_2}) = 0, \mathbb{C}\text{orr}(W_{i_1}, W_{i_2}) \neq 0$. This is due to the presence of the common factor $\mathcal{E}_0$ at every location.

In the next proposition, we provide the closed-form expressions of the marginal cumulative distribution function (cdf) and the marginal probability density function (pdf) of the process $W(\boldsymbol{s})$.

Proposition 1. Consider the process $W(\boldsymbol{s})$ observed at n locations as given in Equation (3). Then the cdf of W_j is

$$F_{W_j}(w) = \Phi(w) - \frac{1}{\alpha - \alpha_{0j}}\left\{\alpha \exp\left(\frac{0.5}{\alpha^2} - \frac{w}{\alpha}\right)\Phi\left(w - \frac{1}{\alpha}\right) - \alpha_{0j}\exp\left(\frac{0.5}{\alpha_{0j}^2} - \frac{w}{\alpha_{0j}}\right)\Phi\left(w - \frac{1}{\alpha_{0j}}\right)\right\}, \tag{4}$$

where $\Phi(\cdot)$ is the cdf of the univariate standard normal distribution, $j = 1,...,n$.

Proof. Because $\mathcal{E}_0$ and $\mathcal{E}_j$ are independent $\mathcal{E}(1)$ random variables, the pdf of $X_j = \alpha_{0j}\mathcal{E}_0 + \alpha\mathcal{E}_j$ is obtained as

$$f_{X_j}(x) = (\alpha - \alpha_{0j})^{-1}\left\{\exp\left(-\frac{x}{\alpha}\right) - \exp\left(-\frac{x}{\alpha_{0j}}\right)\right\}.$$

The cdf of W_j then can be written as

$$F_{W_j}(w) = \mathbb{P}(Z_j + X_j \le w) = \int_0^\infty \Phi(w-x) f_{X_j}(x)\mathrm{d}x.$$

The proof can be completed using the following identity:

$$\int_0^\infty \Phi(w-x)\exp\left(-\frac{x}{\alpha}\right)\mathrm{d}x = \alpha\Phi(w) - \alpha\exp\left(\frac{0.5}{\alpha^2} - \frac{w}{\alpha}\right)\Phi\left(w - \frac{1}{\alpha}\right).$$

□ □ ■

From the cdf of W_j, we can find the expression of its pdf, $f_{W_j}(w)$, by differentiating Equation (4) with respect to w. Moreover, let $f_{\boldsymbol{W}}(\boldsymbol{w})$ be the pdf of $\boldsymbol{W}$ from Equation (3). Although the exact mathematical expression cannot be found explicitly, we can have a simplified form of $f_{\boldsymbol{W}}(\boldsymbol{w})$.

Proposition 2. The pdf of $\boldsymbol{W}$ is of the following form:

$$f_{\boldsymbol{W}}(\boldsymbol{w}) = \left(\frac{1}{\alpha}\right)^n \int_{\mathbb{R}_+} \exp\left\{-0.5\boldsymbol{w}(v_0)^\top \boldsymbol{\Sigma}_{\boldsymbol{Z}}^{-1}\boldsymbol{w}(v_0) + 0.5\boldsymbol{y}(v_0)^\top \boldsymbol{H}^{-1}\boldsymbol{y}(v_0)\right\}\Phi_n(\boldsymbol{0}; -\boldsymbol{y}(v_0), \boldsymbol{H})\exp(-v_0)\mathrm{d}v_0, \tag{5}$$

where $\Phi_n(\cdot;\boldsymbol{\mu},\boldsymbol{\Sigma})$ is the cdf of an n-variate Gaussian random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, $\boldsymbol{w}(v_0) = (w_1 - \alpha_{01}v_0, \ldots, w_n - \alpha_{0n}v_0)^\top$, $\boldsymbol{y}(v_0) = \boldsymbol{H}\boldsymbol{t}(v_0)$, $\boldsymbol{H} = (1/\alpha^2)\boldsymbol{\Sigma}_{\boldsymbol{Z}}$, $\boldsymbol{t}(v_0) = \{s(v_0)_1\alpha - 1, \ldots, s(v_0)_n\alpha - 1\}^\top$, and $\boldsymbol{s}(v_0) = \{s(v_0)_1, \ldots, s(v_0)_n\}^\top = \boldsymbol{\Sigma}_{\boldsymbol{Z}}^{-1}\boldsymbol{w}(v_0)$.

Proof. Let $\boldsymbol{W}^* = (W_1^*, \ldots, W_n^*)^\top$, $W_j^* = Z_j + \alpha\mathcal{E}_j, j \in \{1, \ldots, n\}$, and $\boldsymbol{\alpha}_0 = (\alpha_{01}, \ldots, \alpha_{0n})^\top$. The pdf of $\boldsymbol{W}^*$ is

$$\begin{aligned}
f_{\boldsymbol{W}^*}(\boldsymbol{w}) &= \int_{\mathbb{R}_+^n} \phi_n(\boldsymbol{w}; \boldsymbol{\alpha}_v, \boldsymbol{\Sigma}_{\boldsymbol{Z}})\exp\left(-\sum_{j=1}^n v_j\right)\mathrm{d}\boldsymbol{v},\ \boldsymbol{\alpha}_v = \alpha(v_1, \ldots, v_n)^\top, \\
&= \frac{1}{\left(\sqrt{2\pi}\right)^n}\frac{1}{\sqrt{\det(\boldsymbol{\Sigma}_{\boldsymbol{Z}})}}\int_{\mathbb{R}_+^n} \exp\left[-\frac{1}{2}\left\{(\boldsymbol{w} - \boldsymbol{\alpha}_v)^\top \boldsymbol{\Sigma}_{\boldsymbol{Z}}^{-1}(\boldsymbol{w} - \boldsymbol{\alpha}_v) + 2\sum_{j=1}^n v_j\right\}\right]\mathrm{d}\boldsymbol{v} \\
&= \exp(-0.5\boldsymbol{w}^\top \boldsymbol{\Sigma}_{\boldsymbol{Z}}^{-1}\boldsymbol{w} + 0.5\boldsymbol{y}^\top \boldsymbol{H}^{-1}\boldsymbol{y})\left(\prod_{j=1}^n \frac{1}{\alpha}\right) \\
&\times \int_{\mathbb{R}_+^n} \frac{1}{\left(\sqrt{2\pi}\right)^n \det(\boldsymbol{H})^{1/2}}\exp\left\{-\frac{1}{2}(\boldsymbol{v} - \boldsymbol{y})^\top \boldsymbol{H}^{-1}(\boldsymbol{v} - \boldsymbol{y})\right\}\mathrm{d}\boldsymbol{v} \\
&= \exp(-0.5\boldsymbol{w}^\top \boldsymbol{\Sigma}_{\boldsymbol{Z}}^{-1}\boldsymbol{w} + 0.5\boldsymbol{y}^\top \boldsymbol{H}^{-1}\boldsymbol{y})\left(\tfrac{1}{\alpha}\right)^n \Phi_n(\boldsymbol{0}; -\boldsymbol{y}, \boldsymbol{H}),
\end{aligned}$$

where $\phi_n(\cdot;\boldsymbol{\mu},\boldsymbol{\Sigma})$ is the pdf of an n-variate Gaussian random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{t}$, $\boldsymbol{H} = (1/\alpha^2)\boldsymbol{\Sigma}_{\boldsymbol{Z}}$, $\boldsymbol{t} = (s_1\alpha - 1, \ldots, s_n\alpha - 1)^\top$, and $\boldsymbol{s} = (s_1, \ldots, s_n)^\top = \boldsymbol{\Sigma}_{\boldsymbol{Z}}^{-1}\boldsymbol{w}$. The proof can be completed using the stochastic representation of $\boldsymbol{W} = \boldsymbol{W}^* + \mathcal{E}_0\boldsymbol{\alpha}_0$. □ □ ■

To compute $f_{\boldsymbol{W}}(\cdot)$, we use the expression of $f_{\boldsymbol{W}}(\cdot)$ from Equation (5) and perform the integration numerically using the Gauss–Laguerre quadrature method (Stroud & Secrest, 1966).

Next, suppose the cdf of $\boldsymbol{W}$ is $F_{\boldsymbol{W}}(\cdot)$. Then the copula corresponding to the distribution of $\boldsymbol{W}$ is

$$C_{\boldsymbol{W}}(\boldsymbol{u}) = F_{\boldsymbol{W}}\{F_{W_1}^{-1}(u_1), \ldots, F_{W_n}^{-1}(u_n)\}, \tag{6}$$

where $\boldsymbol{u} = (u_1, \ldots, u_n)^\top$ and $u_i \in [0,1], i = 1, \ldots, n$. Although $F_{W_i}^{-1}(\cdot)$ can be easily obtained numerically because of the simplified form of $F_{W_i}(\cdot)$, we cannot compute $C_{\boldsymbol{W}}(\cdot)$ as the explicit functional form of $F_{\boldsymbol{W}}(\cdot)$ is not known and numerically obtaining it is very challenging, especially when n is not small. Nevertheless, we can compute the copula density corresponding to the distribution of $\boldsymbol{W}$ that is

$$c_{\boldsymbol{W}}(\boldsymbol{u}) = \frac{f_{\boldsymbol{W}}\{F_{W_1}^{-1}(u_1), \ldots, F_{W_n}^{-1}(u_n)\}}{f_{W_1}\{F_{W_1}^{-1}(u_1)\} \times \ldots \times f_{W_n}\{F_{W_n}^{-1}(u_n)\}},\ \boldsymbol{u} = (u_1, \ldots, u_n)^\top \in [0,1]^n. \tag{7}$$

Instead of modelling directly with $W(\boldsymbol{s})$ from Equation (3), here, we perform the modelling using its copula. In this way, we can have arbitrary marginal distributions for $W(\boldsymbol{s})$, keeping the joint dependence the same. As a result, we can have a more flexible model.

2.2 | Tail order of the underlying copula

Tail dependence properties of the copula C_W linking realizations $W(s_1)$ and $W(s_2)$ of the process $W(s)$ as defined in (3) have been studied in Krupskii and Genton (2019) (see their proposition 1). Let $\delta_j = \alpha_{0j}/\alpha, j = 1,2$. The copula C_W has upper tail dependence if $\delta_1 > 1$ and $\delta_2 > 1$. In this section, we show that C_W allows for a full range of dependence between tail quadrant independence and tail dependence. The new results related to the tail order coefficient of $W(s)$ were not explored in Krupskii and Genton (2019). Let κ_U be the upper tail order of C_W such that

$$C_W^R(u,u) \sim u^{\kappa_U}\ell(u), u \to 0^+,$$

for some slowly varying function $\ell(u)$, where

$$C_W^R(u_1,u_2) = -1 + u_1 + u_2 + C_W(1-u_1, 1-u_2)$$

is the reflected copula. Here, $\kappa_U = 2$ corresponds to tail quadrant independence, and if $1 < \kappa_U < 2$, then the copula C_W is said to have intermediate tail dependence (such as that of the normal copula). The strongest dependence is achieved when $\kappa_U = 1$; in particular, $\kappa_U = 1$ if C_W has upper tail dependence. As the next proposition shows, κ_U cannot exceed 2 in our model. This is not the case for the Gaussian copula which can model situations with $\kappa_U > 2$. Note that this constraint is not very restrictive because for real data with positive dependence, $\kappa_U \le 2$.

Proposition 3. Let $\delta_L = \min(\delta_1,\delta_2)$ and $\delta_U = \max(\delta_1,\delta_2)$, and assume $\delta_L < 1$. The upper tail order of the copula C_W is given by the formula:

$$\kappa_U = \begin{cases} 2-\delta_L, & \delta_L < 1, \delta_U > 1, \\ 1+(1-\delta_L)/\delta_U, & \delta_L < 1, \delta_U < 1, \delta_L + \delta_U > 1, \\ 2, & \delta_L + \delta_U \le 1. \end{cases}$$

Proof. We consider the case $\delta_2 < \delta_1 < 1$. Other cases are considered analogously. It follows that

$$1 - C_W(1-u,1-u) \sim 2u - u^{\kappa_U}\ell(u), u \to 0^+ \text{ and } C_W(1-u,1-u) = F_W\left(F_1^{-1}(1-u), F_2^{-1}(1-u)\right),$$

where F_W is the joint cdf of $(W(s_1), W(s_2))^\top$, and F_j is the marginal cdf of $W(s_j), j = 1,2$. From (3), we find that

$$F_W(z_1,z_2) = \int_{\mathbb{R}_+^3} \Phi_2(z_1 - \alpha_{10}v_0 - \alpha v_1, z_2 - \alpha_{20}v_0 - \alpha v_2; \rho)\exp(-v_0 - v_1 - v_2)\mathrm{d}v_2\mathrm{d}v_1\mathrm{d}v_0.$$

Using the integration by parts formula with respect to v_1 and v_2, we get

$$\begin{aligned} F_W(z_1,z_2) &= \int_0^\infty \{I_0(v_0) - I_1(v_0) - I_2(v_0) + I_{12}(v_0)\}\mathrm{d}v_0, \text{ where} \\ I_0(v_0) &= \Phi_2(z_1 - \alpha_{10}v_0, z_2 - \alpha_{20}v_0; \rho)\exp(-v_0), \\ I_1(v_0) &= \Phi_2(z_1 - \alpha_{10}v_0 - \rho/\alpha, z_2 - \alpha_{20}v_0 - 1/\alpha; \rho)\exp\{(\delta_2 - 1)v_0 + 0.5/(\alpha)^2 - z_2/\alpha\}, \\ I_2(v_0) &= \Phi_2(z_1 - \alpha_{10}v_0 - 1/\alpha, z_2 - \alpha_{20}v_0 - \rho/\alpha; \rho)\exp\{(\delta_1 - 1)v_0 + 0.5/(\alpha)^2 - z_1/\alpha\}, \\ I_{12}(v_0) &= \Phi_2(z_1 - \alpha_{10}v_0 - 1/\alpha - \rho/\alpha, z_2 - \alpha_{20}v_0 - \rho/\alpha - 1/\alpha; \rho)\exp\{(\delta_{12} - 1)v_0 + 0.5(\rho_{12}^*)^2 - z_1/\alpha - z_2/\alpha\}, \end{aligned}$$

with $\delta_{12} = \delta_1 + \delta_2$ and $(\rho_{12}^*)^2 = 2(1+\rho)/\alpha^2$. The marginal distribution of W_j is

$$\begin{aligned} F_j(z) &= \Phi(z) - \Big[\alpha_{j0}\exp\{-z/\alpha_{j0} + 0.5/(\alpha_{j0})^2\}\Phi(z - 1/\alpha_{j0}) \\ &\quad - \alpha\exp\{-z/\alpha + 0.5/(\alpha)^2\}\Phi(z - 1/\alpha)\Big]/(\alpha_{j0} - \alpha), j = 1,2. \end{aligned}$$

Let $z_j = F_j^{-1}(1 - 1/n)$. It implies that $z_j = \alpha\log n + 0.5/\alpha - \alpha\log(1-\delta_j) + \epsilon_j, \epsilon_j = O(n^{1-1/\delta_j})$, where

$$\frac{1}{n}\left(1-\exp\left(-\frac{\epsilon_j}{\alpha_j}\right)\right)=-\frac{\delta_j}{1-\delta_j}n^{-1/\delta_j}\exp\left(\frac{0.5}{\alpha_{j0}^2}-\frac{0.5}{\alpha\alpha_{j0}}+\frac{1}{\delta_j}\log(1-\delta_j)-\frac{\epsilon_j}{\alpha_{j0}}\right).$$

Note that $\Phi_2(z_1-\alpha_{10}v_0,z_2-\alpha_{20}v_0;\rho)=\Phi(z_1-\alpha_{10}v_0)+o(1/n^2)$ if $v_0<(1/\delta_2-\epsilon)\log n$ for any $\epsilon>0$ because $z_2-\alpha_{20}v_0>\alpha\epsilon\log n+O(1)$. Furthermore, $\Phi_2(z_1-\alpha_{10}v_0,z_2-\alpha_{20}v_0;\rho)=o(1/n^2)$ and $\Phi(z_1-\alpha_{10}v_0)=o(1/n^2)$ if $v_0>(1/\delta_1+\epsilon)\log n$ because $z_1-\alpha_{10}v_0<-\alpha\epsilon\log n+O(1)$. It implies that the second argument in the bivariate normal cdf can be ignored when integrating $I_0(v_0)$ over $v_0>0$. The same reasoning applies to $I_1(v_0), I_2(v_0)$ and $I_{12}(v_0)$. We can therefore calculate $F_W(z_1,z_2)$ using the following formula for $b>0$ which can be obtained using integration by parts:

$$\int_0^\infty \Phi(a-bv_0)\exp(-cv_0)\mathrm{d}v_0=\frac{1}{c}\Phi(a)-\frac{1}{c}\Phi\left(a-\frac{c}{b}\right)\exp\left(\frac{c^2}{2b^2}-\frac{ac}{b}\right).$$

Let

$$A_0=\exp\left(\frac{0.5}{\alpha_{10}^2}-\frac{z_1}{\alpha_{10}}\right)=n^{-1/\delta_1}\cdot\exp\left(\frac{0.5}{\alpha_{10}^2}-\frac{0.5}{\alpha_1\alpha_{10}}+\frac{1}{\delta_1}\log(1-\delta_1)-\frac{\epsilon_1}{\alpha_{10}}\right)$$

and

$$A_j=\exp\left(0.5/\alpha^2-z_j/\alpha\right)=\frac{1}{n}(1-\delta_j)\cdot\exp\left(-\frac{\epsilon_j}{\alpha}\right),j=1,2.$$

Note that

$$\frac{1}{1-\delta_1}A_1-\frac{\delta_1}{1-\delta_1}A_0=\frac{1}{n}.$$

We find that

$$\begin{aligned}F_W(z_1,z_2)=\ & 1-A_0-\frac{1}{1-\delta_1}A_1-\frac{1}{1-\delta_2}A_2+C_{01}A_0^{1-\delta_1}A_1+C_{02}A_0^{1-\delta_2}A_2\\ &+C_{12}A_1A_2+C_{012}A_0^{1-\delta_{12}}A_1A_2+o(1/n^2),\end{aligned}$$

where

$$C_{01}=\frac{1}{1-\delta_1}\cdot\exp\left(\frac{1-\delta_1}{\alpha\alpha_{10}}-\frac{\delta_1(1-\delta_1)}{2\alpha_{10}^2}\right),C_{02}=\frac{1}{1-\delta_2}\cdot\exp\left(\frac{\rho(1-\delta_2)}{\alpha\alpha_{10}}-\frac{\delta_2(1-\delta_2)}{2\alpha_{10}^2}\right),$$

$$C_{12}=\frac{1}{1-\delta_{12}}\exp\left(\frac{2\rho}{\alpha^2}\right),C_{012}=-\frac{1}{1-\delta_{12}}\exp\left(\frac{2\rho}{\alpha^2}+\frac{(1-\delta_{12})^2}{2\alpha_{10}^2}-\left(\frac{1}{\alpha}-\frac{\rho}{\alpha}\right)\frac{1-\delta_{12}}{\alpha_{10}}\right).$$

Note that $A_0^{1-\delta_2}A_2=O\left(n^{-(1+(1-\delta_2)/\delta_1)}\right),A_1A_2=O(n^{-2}),A_0^{1-\delta_{12}}A_1A_2=O\left(n^{-(1+(1-\delta_2)/\delta_1)}\right)$, and $A_2=\frac{1}{n}(1-\delta_2)(1-\epsilon_2/\alpha+o(\epsilon_2))=\frac{1}{n}(1-\delta_2)+O(n^{-1/\delta_2})$ where $1/\delta_2>1+(1-\delta_2)/\delta_1$. Finally, one can show that

$$C_{01}A_0^{1-\delta_1}A_1=\frac{1}{1-\delta_1}A_0\Rightarrow A_0+\frac{1}{1-\delta_1}A_1-C_{01}A_0^{1-\delta_1}A_1=\frac{1}{1-\delta_1}A_1-\frac{\delta_1}{1-\delta_1}A_0=\frac{1}{n}.$$

It implies that $C_W(1-1/n,1-1/n)=1-2/n+O(n^{-1+(1-\delta_2)/\delta_1})+O(n^{-2})$, and the upper tail order of C_W is $\kappa_U=\min\{1+(1-\delta_2)/\delta_1,2\}$. □ □ ■

Similar to the model proposed by Huser and Wadsworth (2019), the transition between asymptotic dependence and asymptotic independence for our model takes place in the interior of the parameter space. This is a very desirable property for the model as the asymptotic tail behaviour can be estimated from the data. Furthermore, the model can handle the full range of dependence, from asymptotic tail dependence with $\kappa_U = 1$ to tail quadrant independence with $\kappa_U = 2$.

3 | MAXIMUM LIKELIHOOD ESTIMATION AND INTERPOLATION

3.1 | The likelihood function

To estimate the copula parameters corresponding to the model in Equation (3), we rely on maximizing the log-likelihood function. Consider a random sample of size N observed from the model in Equation (3) as

$$w_{ij} = w_i(\boldsymbol{s}_j) = z_i(\boldsymbol{s}_j) + \alpha_0(\boldsymbol{s}_j)v_{0i} + \alpha v_i(\boldsymbol{s}_j),\ i \in \{1,\ldots,N\}, j \in \{1,\ldots,n\}. \tag{8}$$

Moreover, we assume that the samples from each location, that is, $\boldsymbol{w}_j = (w_{1j},\ldots,w_{Nj})^\top$, are not necessarily from the cdf F_{W_j} from Equation (4), but from any arbitrary distribution. To compute the copula likelihood, we need to transform the data to a uniform scale. To do that, we use the nonparametric uniform scores, defined as $u_{ij} = \{\text{rank}(w_{ij}) - 0.5\}/N$. Moreover, let $\boldsymbol{z}_i = (z_{i1},\ldots,z_{in})^\top$ and $z_{ij} = F^{-1}_{W_j}(u_{ij},\boldsymbol{\theta}_F), i = 1,\ldots,N, j = 1,\ldots,n$.

As we use the data transformed to uniform scores, they are assumed to be from $\mathcal{U}(0,1)$ approximately, however, they are not mutually independent anymore. The extent of dependence vanishes as the number of replicates, N, goes to infinity and the maximum likelihood estimates (MLE) obtained by maximizing the pseudo-log-likelihood are consistent and asymptotically normal, provided the copula is correctly specified. From Equation (7), the pseudo-log-likelihood is

$$\ell(\boldsymbol{z}_1,\ldots,\boldsymbol{z}_N) = \sum_{i=1}^{N}\left[\log\{f_{\boldsymbol{W}}(\boldsymbol{z}_i;\boldsymbol{\theta}_F,\boldsymbol{\theta}_\Sigma)\} - \sum_{j=1}^{n}\log\{f_{W_j}(z_{ij};\boldsymbol{\theta}_F)\}\right], \tag{9}$$

where $\boldsymbol{\theta}_\Sigma$ is the parameter vector which corresponds to the correlation function for constructing $\boldsymbol{\Sigma}_{\boldsymbol{Z}}$ and $\boldsymbol{\theta}_F$ is the parameter vector compiling the rest of the parameters used for computing $f_{\boldsymbol{W}}$ in Equation (5).

3.2 | Vecchia approximation

The computation of the full log-likelihood given in Equation (9) involves the computation of $f_{\boldsymbol{W}}$ from Equation (5). This involves computing $\Phi_n(\cdot;\boldsymbol{\mu},\boldsymbol{\Sigma})$, the n-dimensional cdf of a Gaussian random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ and n is the same as the number of locations in our study. Whenever n goes beyond 20 (which is often the case in applications) the exact computation of $\Phi_n(\cdot;\boldsymbol{\mu},\boldsymbol{\Sigma})$ in our setting cannot be done accurately. This is due to the fact that for our model, we have seen in computing the log-likelihood at the true parameters when $n > 20$ that $-\boldsymbol{y}(v_0)$ $(\forall v_0 > 0)$ from Equation (5) is a large positive vector and as a result the true value of $\Phi_n(\boldsymbol{0}; -\boldsymbol{y}(v_0),\boldsymbol{H})$ is very close to zero and hence is difficult to compute accurately. As a result, for applications with the number of locations of more than 20, we compute the log-likelihood using the Vecchia approximation (Vecchia, 1988). Consider an n-dimensional joint copula density function $c_{1:n}(\cdot)$ written as

$$\begin{aligned} &c_{1:n}(u_1,\ldots,u_n) = c_{n|1:(n-1)}(u_n|u_1,\ldots,u_{n-1})c_{(n-1)|1:(n-2)}(u_{n-1}|u_1,\ldots,u_{n-2})\ldots c_{3|1:2}(u_3|u_1,u_2)c_{1:2}(u_1,u_2) \\ &\Rightarrow \log\{c_{1:n}(u_1,\ldots,u_n)\} = \log\{c_{1:2}(u_1,u_2)\} + \sum_{j=3}^{n}\log\{c_{j|1:(j-1)}(u_j|u_1,\ldots,u_{j-1})\}, \end{aligned} \tag{10}$$

where $u_j \in [0,1], j = 1,\ldots,n$. The idea of the Vecchia approximation is, instead of conditioning on all the variables in the last expression, to condition only on a few variables. Suppose the variables $u_1,\ldots,u_n$ are ordered in some particular way as $u_{[1]},\ldots,u_{[n]}$, and we only use the previous n_0 $(2 < n_0 \ll n)$ ranked variables for conditioning the jth variable, $j = 3,\ldots,n$. In this way the copula log-likelihood from Equation (10) can be rewritten as

$$
\begin{aligned}
\log\{c_{1:n}(u_1,\ldots,u_n)\} &= \log\{c_{1:2}(u_1,u_2)\} + \sum_{j=3}^{n} \log\{c_{j|1:(j-1)}(u_j|u_1,\ldots,u_{j-1})\} \\
&= \log\{c_{[1]:[2]}(u_{[1]},u_{[2]})\} + \sum_{j=3}^{n} \log\{c_{[j]|[1]:[j-1]}(u_{[j]}|u_{[1]},\ldots,u_{[j-1]})\} \\
&= \log\{c_{[1]:[2]}(u_{[1]},u_{[2]})\} + \sum_{j=3}^{n_0} \log\{c_{[j]|[1]:[j-1]}(u_{[j]}|u_{[1]},\ldots,u_{[j-1]})\} + \sum_{j=n_0+1}^{n} \log\{c_{[j]|[1]:[j-1]}(u_{[j]}|u_{[1]},\ldots,u_{[j-1]})\} \\
&\approx \log\{c_{[1]:[2]}(u_{[1]},u_{[2]})\} + \sum_{j=3}^{n_0} \log\{c_{[j]|[1]:[j-1]}(u_{[j]}|u_{[1]},\ldots,u_{[j-1]})\} + \sum_{j=n_0+1}^{n} \log\{c_{[j]|[j-n_0]:[j-1]}(u_{[j]}|u_{[j-n_0]},\ldots,u_{[j-1]})\} \\
&= \sum_{j=n_0+1}^{n} \left[\log\{c_{[j-n_0]:[j]}(u_{[j-n_0]},\ldots,u_{[j]})\} - \log\{c_{[j-n_0]:[j-1]}(u_{[j-n_0]},\ldots,u_{[j-1]})\}\right] + \log\{c_{[1]:[n_0]}(u_{[1]},\ldots,u_{[n_0]})\}.
\end{aligned}
$$

Moreover, using the setup used to write the log-likelihood in Equation (9), the Vecchia approximation of the pseudo-log-likelihood is

$$
\begin{aligned}
\ell_V(\boldsymbol{z}_1,\ldots,\boldsymbol{z}_N) &= \sum_{i=1}^{N}\sum_{j=n_0+1}^{n} \log\{f_{[j-n_0]:[j]}(z_{[j-n_0]},\ldots,z_{[j]};\boldsymbol{\theta}_F,\boldsymbol{\theta}_{\boldsymbol{\Sigma}})\} \\
&\quad - \sum_{i=1}^{N}\sum_{j=n_0+2}^{n} \log\{f_{[j-n_0]:[j-1]}(z_{[j-n_0]},\ldots,z_{[j-1]};\boldsymbol{\theta}_F,\boldsymbol{\theta}_{\boldsymbol{\Sigma}})\} - \sum_{i=1}^{N}\sum_{j=1}^{n} \log\{f_{W_j}(z_{ij};\boldsymbol{\theta}_F)\},
\end{aligned} \tag{11}
$$

where $f_{[j-n_0]:[j]}(z_{[j-n_0]},\ldots,z_{[j]})$ is the corresponding (n_0+1)-dimensional marginal pdf of $f_{\boldsymbol{W}}(\cdot)$ from Equation (5). For computing $\ell_V(\boldsymbol{z}_1,\ldots,\boldsymbol{z}_N)$ in Equation (11), we need only to compute $\Phi_{n_0+1}(\boldsymbol{\mu},\boldsymbol{\Sigma})$ which enables us to fit the copula for applications with the number of locations being more than 20. Hence, for applications with more than 20 locations, it is not possible to evaluate the exact pseudo-log-likelihood and the Vecchia approximation of the pseudo-log-likelihood gives us an avenue to estimate the model parameters.

3.3 | Conditional copula and interpolation

Using a sample as in the setup given in Equation (8), we can estimate the copula parameters either by maximizing the full log-likelihood in Equation (9) or the Vecchia approximation of the log-likelihood in Equation (11). From the estimated copula, we can predict the spatial process at new locations. Suppose the estimated copula parameters based on the samples at n locations are $\hat{\boldsymbol{\theta}}_F$ and $\hat{\boldsymbol{\theta}}_{\boldsymbol{\Sigma}}$. For a given data vector $(u_1,\ldots,u_n)^\top$, the conditional copula at a new location $\boldsymbol{s}_0$ is

$$
\hat{C}_{0|n}(u_0|u_1,\ldots,u_n) = \int_0^{u_0} c_{n+1}(u_1,\ldots,u_n,u;\hat{\boldsymbol{\theta}}_F,\hat{\boldsymbol{\theta}}_{\boldsymbol{\Sigma}})\mathrm{d}u / c_n(u_1,\ldots,u_n;\hat{\boldsymbol{\theta}}_F,\hat{\boldsymbol{\theta}}_{\boldsymbol{\Sigma}}),
$$

where $c_{n+1}(\cdot)$ and $c_n(\cdot)$ are the corresponding $n+1$ and n dimensional copula densities. The copula densities can be computed from Equation (7). Similar to the computational issues of the full log-likelihood, the computation of $c_n(\cdot)$ will involve computing $\Phi_n(\cdot;\boldsymbol{\mu},\boldsymbol{\Sigma})$, which becomes problematic when n is more than 20. We suggest using the estimated conditional copula $\hat{C}_{0|n_0}(u_0|u_{(1)},\ldots,u_{(n_0)})$, where $(u_{(1)},\ldots,u_{(n_0)})^\top$ are n_0 observations from n_0 nearest neighbors of $\boldsymbol{s}_0$, and $n_0 \ll n$. From the estimated conditional copula, we can compute different quantities such as the conditional mean $\hat{m}_0$ and conditional pth-quantile $\widehat{q_{0p}}$ as

$$
\hat{m}_0 = \int_0^1 u_0 \mathrm{d}\hat{C}_{0|n}(u_0|u_1,\ldots,u_n),\ \widehat{q_{0p}} = \hat{C}^{-1}_{0|n}(p|u_1,\ldots,u_n).
$$

These quantities are computed numerically and $\hat{C}_{0|n}(u_0|u_1,\ldots,u_n)$ can also be replaced by $\hat{C}_{0|n_0}(u_0|u_{(1)},\ldots,u_{(n_0)})$ in order to make computation feasible. Note that the estimates are all in copula scale. To get the estimates in the original scale, we need to transform the estimates using the inverse of the estimated marginal cdf at the prediction location $\boldsymbol{s}_0$.

4 | SIMULATION STUDY

We provide a simulation study for comparing the estimates obtained from maximizing the exact log-likelihood and the Vecchia approximated log-likelihood. For this study, we simulate observations from the process $W(\boldsymbol{s})$, where $\boldsymbol{s} \in [0,1] \times [0,1]$, with $\alpha_0(\boldsymbol{s}) = \exp(b + a_1 s_1 + a_2 s_2)$. Moreover, we assume that the correlation function corresponding to the latent Gaussian process is the Matérn correlation function of the following form:

$$\rho_Z(\boldsymbol{s}_1,\boldsymbol{s}_2)=\mathbb{C}\mathrm{orr}\{Z(\boldsymbol{s}_1),Z(\boldsymbol{s}_2)\}=\frac{1}{\Gamma(\nu)2^{\nu-1}}\left(\frac{\|\boldsymbol{s}_1-\boldsymbol{s}_2\|}{\phi}\right)^{\nu}\mathcal{K}_\nu\left(\frac{\|\boldsymbol{s}_1-\boldsymbol{s}_2\|}{\phi}\right), \tag{12}$$

where $\|\boldsymbol{s}_1-\boldsymbol{s}_2\|$ is the distance between locations $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$, $\nu>0$ is the smoothness parameter, $\phi>0$ is the range parameter, $\Gamma(\cdot)$ is the gamma function, and $\mathcal{K}_\nu(\cdot)$ is the modified Bessel function of the second kind of order ν. We select 20 locations randomly on $[0,1]\times[0,1]$, and at those locations, we randomly generate a Gaussian process with the Matérn correlation function using $\nu=0.5$ and $\phi=0.167$. With the generated latent Gaussian observations, we generate the target process $W(\boldsymbol{s})$ using $b=0, a_1=1, a_2=-1$, and $\alpha=1.5$. Based on samples of size $N=$ **100**, **200**, and **500**, we compute the exact log-likelihood and the Vecchia approximated log-likelihood using $n_0=5$ neighbours and maximize them to get the maximum likelihood estimates (MLE). For computing the Vecchia approximated log-likelihood, we have used the Morton ordering scheme (Morton, 1966) and no-ordering scheme where we do not order the observations. Boxplots in Figure 1 show MLEs based on 100 simulations. The boxplots suggest that the Vecchia approximation yields comparable estimates to the exact log-likelihood, usually for larger sample sizes. In addition, there is not much difference between the estimates obtained from two different ordering schemes for the Vecchia approximation indicating that the ordering scheme does not play a big role on the estimates. This finding is similar to what Stein et al. (2004) have observed in their paper. To justify the last claim, in Table 1, we present the mean squared error (MSE) of the estimates. The MSE of the parameter estimates obtained from the Vecchia approximation for two different ordering schemes do not show any general patterns like one is always higher or lower than the other, which makes both of them very similar. Moreover, as we expected, the MSE of the estimates from the exact log-likelihood is smaller for all sample sizes when compared with that of the Vecchia approximated log-likelihood and the difference becomes smaller as the sample size increases.

5 | TEMPERATURE DATA ANALYSIS

In this section, we analyse the spatial dependence structure of the daily mean temperature in Switzerland using the model from Equation (2). The data was initially used by Davison et al. (2013). We consider the data from 1 May 2011 to 30 September 2011, 153 days in total, that is, the daily average temperature during the summer months to ensure the weather pattern remains the same across the region. We analyse data from a total of 15 weather stations over the western part of Switzerland, with the maximum distance between two stations being 164.44 km. The area of the study is fairly large, with mountains, and that is why a non-stationary copula model as in Equation (2) may be more appropriate in this case. The goal of this study is to estimate the copula model parameters and then to interpolate data at new locations. For estimating the parameters, we use data from 12 randomly chosen weather stations and test the validity of the model on the data from the remaining three weather stations.

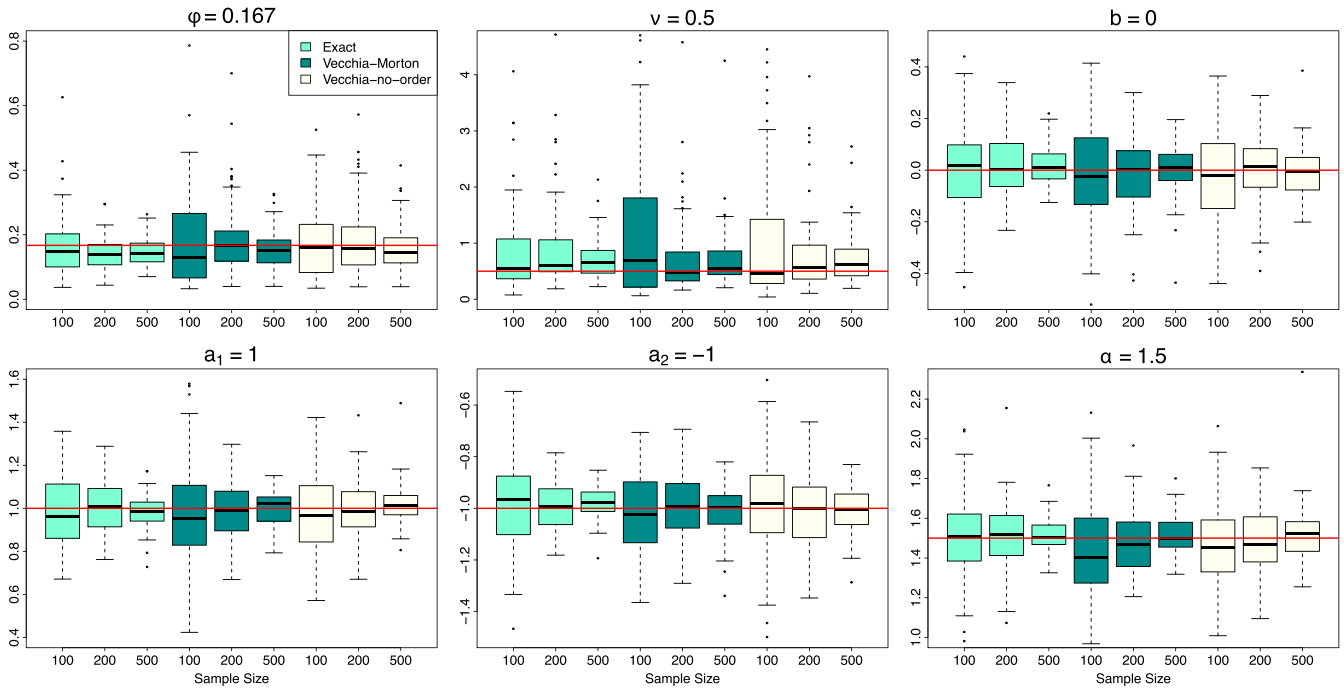

FIGURE 1 Boxplots of the parameter MLEs based on 100 simulations. The true values of each parameter are indicated by the red line in each plot.

TABLE 1 The MSE (based on 100 simulations) of the MLEs of all the parameters, obtained from the exact log-likelihood and the Vecchia approximated log-likelihood with Morton ordering and with no ordering, are given below.

	ϕ	ν	b	a_1	a_2	α
Exact (100)	0.0077	1.5725	0.0300	0.0274	0.0297	0.0383
Vecchia Morton (100)	0.3086	3.5437	0.0392	0.0501	0.0265	0.0676
Vecchia no-order (100)	0.3146	2.6219	0.0293	0.0387	0.0366	0.0378
Exact (200)	0.0030	0.8059	0.0135	0.0122	0.0094	0.0234
Vecchia Morton (200)	0.0109	1.3949	0.0170	0.0169	0.0181	0.0243
Vecchia no-order (200)	0.0145	1.8526	0.0145	0.0161	0.0188	0.0250
Exact (500)	0.0021	0.1561	0.0055	0.0056	0.0037	0.0066
Vecchia Morton (500)	0.0032	0.8590	0.0081	0.0158	0.0074	0.0264
Vecchia no-order (500)	0.0041	0.3706	0.0089	0.0051	0.0073	0.0166

Note: The sample sizes N are given in parentheses.

5.1 | Marginal modelling

A subset of the dataset we are using has been already analysed by Krupskii et al. (2018). Similar to Krupskii et al. (2018), we remove the serial dependence by fitting an autoregressive-moving-average (ARMA) model with lags up to one. Moreover, we include a quadratic trend in the model. Therefore, the model for the marginals is

$$temp_{\boldsymbol{s},t} = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 temp_{\boldsymbol{s},t-1} + \beta_4 \epsilon_{\boldsymbol{s},t-1} + \epsilon_{\boldsymbol{s},t}, \tag{13}$$

where $temp_{\boldsymbol{s},t}$ is the average temperature measured at location $\boldsymbol{s}$ on day $t, \boldsymbol{s} \in \{1,...,12\}$ and $t \in \{2,...,153\}$. We assume that the marginal distributions of $\epsilon_{\boldsymbol{s},t}$ are the same for all $\boldsymbol{s}, \boldsymbol{s} \in \{1,...,12\}$ and $t \in \{2,...,153\}$. The parameters $\beta_i, i = \{0,1,...,4\}$ are obtained by minimizing the sum of squared errors. We found that the skew-t distribution proposed by Azzalini and Capitanio (2003) fits well to the marginals of $\epsilon_{\boldsymbol{s},t}$. Moreover, the Ljung–Box test at the 5% level of significance indicates that $\epsilon_{\boldsymbol{s},t}$ are uncorrelated over time t, up to lag 20, for all locations $\boldsymbol{s} \in \{1,...,12\}$. Hence, $\epsilon_{\boldsymbol{s},t}$ can be treated as a random sample of size 152 from a spatial process collected over 12 locations.

5.2 | Dependence structure of the data

Next, we study the dependence structure by some preliminary analysis of the dataset. We transform the estimated residuals $\hat{\epsilon}_{\boldsymbol{s},t}$ from the model given in Equation (13) to uniform scores by $u_{\boldsymbol{s},t} = \{\text{rank}(\hat{\epsilon}_{\boldsymbol{s},t}) - 0.5\}/152$, for $\boldsymbol{s} \in \{1,...,12\}$ and $t \in \{2,...,153\}$. For any $\boldsymbol{s} \in \{1,...,12\}, u_{\boldsymbol{s},t}, t \in \{2,...,153\}$ has an approximate $\mathcal{U}(0,1)$ distribution. From here, we can transform the uniform scores to the normal scores using the inverse of $\Phi(\cdot)$, and we write $z_{\boldsymbol{s},t} = \Phi^{-1}(u_{\boldsymbol{s},t})$. We draw the scatterplots for a few pairs of locations in Figure 2 to check if the dependence structure is similar to that of a bivariate Gaussian distribution. The scatterplots of normal scores for these pairs of locations show sharp tails and also show stronger dependence in the lower tail compared with the upper tail. This suggests that the data have asymmetric and tail dependence. Hence, the data show a dependence structure which cannot be captured by a Gaussian copula.

We justify these claims using the tail-weighted dependence measures Q_L and Q_U proposed by Krupskii and Joe (2015) for observations at each pair of locations using the reflected uniform scores $(1 - u_{\boldsymbol{s},t})$ to have stronger tail dependence in the upper tail compared with the lower tail. Moreover, for each pair of locations, we compute the tail-weighted dependence measures from simulated data from a Gaussian copula with a correlation same as the $(1 - u_{\boldsymbol{s},t})$. Because for a Gaussian copula, the lower and the upper tail-weighted dependence measures are the same, because of its tail-symmetry, we denote it as Q_N. If the dependence of the data at a pair of locations resembles a Gaussian copula, we expect Q_L, Q_U, and Q_N to be close to each other for all pairs. If Q_L (or Q_U) is higher (lower) than Q_N for some pairs, we expect the lower (or upper) tail dependence to be stronger (weaker) than that for a Gaussian copula. We compute

$$Q_k = \sum_{\boldsymbol{s}_i, \boldsymbol{s}_j, i<j} Q_k(1 - u_{\boldsymbol{s}_i,t}, 1 - u_{\boldsymbol{s}_j,t})/n_p, \; k = L, U, N,$$

where $n_p =$ number of distinct location pairs. For the training dataset, $Q_L = 0.60, Q_U = 0.87$, and $Q_N = 0.67$. These measures show stronger upper tail dependence than that of a normal copula and weaker tail dependence than that of a normal copula. Although the lower tail dependence is

weaker than that of a normal copula, it is still rather close to the normal copula. Hence, using the copula of the model from Equation (2) is more appropriate for modelling the dependence structure for this dataset, compared with the Gaussian copula.

5.3 | Joint dependence estimation and prediction

We fit the model $W(\boldsymbol{s})$ from Equation (2) to the reflected uniform scores assuming that they are independent over time. Moreover, we assume that the correlation structure of the latent Gaussian random field is the Matérn correlation function of the form given in Equation (12) and $\alpha_0(\boldsymbol{s}) = \exp(b + a_1 s_1 + a_2 s_2)$. We refer to this model as Model 1. For comparison, we also fit the stationary version of Model 1, that is, we assume $\alpha_0(\boldsymbol{s}) = \alpha_0$, and refer to it as Model 2. In addition, we fit the Gaussian copula model (Model 3) to this data as well, that is, $\alpha_0(\boldsymbol{s}) = 0, \alpha = 0$. We maximize the log-likelihood to estimate the model parameters. To assess model adequacy, we provide measures such as the average differences and the average absolute differences between the model-based estimates by simulation and the corresponding empirical estimates of the correlation coefficients ρ, Q_L, and Q_U. We denote them as $\Delta_\rho, \Delta_{Q_L}$, and Δ_{Q_U} for the average differences and as $|\Delta_\rho|, |\Delta_{Q_L}|$, and $|\Delta_{Q_U}|$ for the average absolute differences. The results are given in Table 2 for Model 1, Model 2, and Model 3 along with their corresponding Bayesian information criterion (BIC). We also provide the parameter estimates based on the three models in Table 3, as well as the standard errors for each estimate, which are computed based on 100 bootstrap samples.

The results in Table 2 suggest that Models 1 and 2 both capture the tail dependence very well compared with Model 3. Although Model 3 captures the joint dependence structure better than the other two models, the improvement is very small. The BIC shows that the non-stationary model is more appropriate for this dataset. The results in Table 3 reveal that the estimates of a_1 and a_2 from Model 1 are significantly away from 0, suggesting that the non-stationary model is more suitable for this dataset.

We now test the three models using the data from three testing locations (see maps in Figure 4). We make predictions at these three locations using the estimates of Models 1, 2, and 3 and check how the predicted values compare to the actual values by the mean absolute error measure. For any given location $\boldsymbol{s}_0$, we use the average values observed at the six nearest stations as starting values for $temp_{\boldsymbol{s}_0,t}, t = 1$. We compute the estimated medians of the variable on the uniform $(0,1)$ scale at location $\boldsymbol{s}_0$ and time $t+1$ using the uniform scores as discussed in Section 3.3, using $n_0 = 6$ (i.e., the 6 nearest neighbours from the training locations for each prediction locations). From there, we convert the estimated median to residuals using the inverse skew-t distribution. Then we get the prediction in the data scale using the marginal model in Equation (13). We

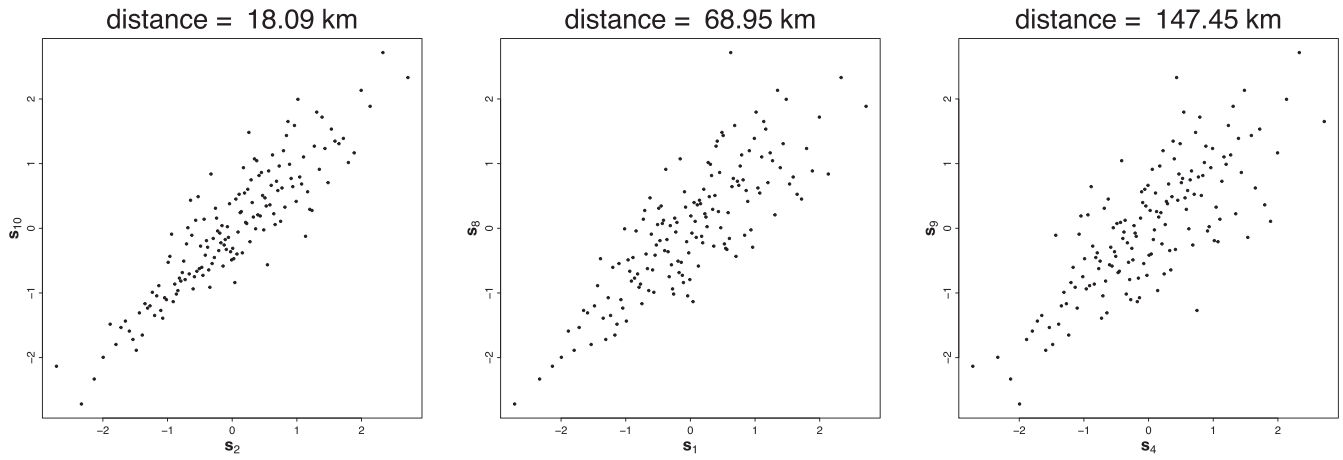

FIGURE 2 Scatter plots of normal scores for the average temperature data at three pairs of locations: $(\boldsymbol{s}_2, \boldsymbol{s}_{10}), (\boldsymbol{s}_1, \boldsymbol{s}_8)$, and $(\boldsymbol{s}_4, \boldsymbol{s}_9)$

TABLE 2 Estimates of $\Delta_\rho, |\Delta_\rho|, \Delta_{Q_L}, |\Delta_{Q_L}|, \Delta_{Q_U}$, and $|\Delta_{Q_U}|$ for Model 1, Model 2, and Model 3 are given here.

	Model 1	Model 2	Model 3
BIC	−2898	−2868	−2865
$\Delta_\rho / \lvert\Delta_\rho\rvert$	−0.03/0.06	−0.04/0.07	−0.01/0.06
$\Delta_{Q_L} / \lvert\Delta_{Q_L}\rvert$	−0.02/0.02	−0.05/0.05	0.06/0.06
$\Delta_{Q_U} / \lvert\Delta_{Q_U}\rvert$	−0.06/0.06	−0.05/0.05	−0.20/0.20

Note: We simulated data from the estimated Model 1, 2, and 3 to calculate these values using sample size $N = 50{,}000$.

repeat the process for all $t = 2,\ldots,153$ to get the predicted average temperatures for all the days at the location $\mathbf{s}_0$. The predicted time series, using Model 1 estimates, of the daily average temperature at the three testing locations are plotted along with the true time series in Figure 3. The mean absolute error of prediction, calculated using all three stations and all time points $t = 2,\ldots,153$, based on Model 1 is 1.54, that of Model 2 is 1.56, and that of Model 3 is 1.57, which suggests Model 1 is slightly outperforming Models 2 and 3 in terms of prediction for this dataset.

We also have estimated the 5% and the 95% quantiles of the average temperatures. To do that, we predict the observations at each location up to time $t-1$ using the estimated medians as described before, and for time t, we estimate the 5% and the 95% quantiles of the uniform scores using the methods described in Section 3.3. Then we add the transformed estimated uniform score quantiles using the inverse skew-t distribution to get the original scale counterparts. We have estimated the 5%, 50%, and 95% quantiles over the whole study region for 11 August 2011. The spatial predictions for 11 August 2011, based on Model 1 are given in Figure 4 along with the observed values in the training locations. Moreover, in Figure 4, we have provided the predicted quantiles based on Model 2 and Model 3. The estimates from Model 1 and Model 2 differ a bit for the 5% quantiles and are very similar for the median and the 95% quantiles. This suggests in terms of prediction that there is not much difference between Model 1 and Model 2. The difference between Model 1 and Model 3 estimates is higher compared with that between Model 1 and Model 2. Model 3 overestimates the 5% quantile and underestimates the 95% quantile compared with Model 1 in most of the locations.

6 | DISCUSSION

In this paper, we proposed a new non-stationary copula model for spatial data with replicates that can handle tail dependence and asymmetric dependence. We studied the tail order of the copula extensively. The parameters of the model can be estimated by maximizing the log-likelihood. Although the form of the joint copula density is simple, its computation becomes numerically challenging when the number of locations in the problem goes beyond 20. The Vecchia approximation of the log-likelihood in those cases gives a feasible alternative for estimating the parameters. Moreover, using a spatial dataset of average temperature in Switzerland, we showed how spatial interpolation can be carried out based on this model.

As the exact joint copula density computation becomes infeasible as the dimension increases, the exact parametric inference also becomes challenging. Perhaps this problem can be avoided if the Bayesian paradigm is considered for this model. This is one possible avenue of future research. Moreover, we have seen that the model cannot be applied when we have tail dependence in both upper and lower tails. For that, similar

TABLE 3 Estimates of the different model parameters along with their standard errors in parentheses.

Model	ϕ	ν	b	a_1	a_2	α
Model 1	254.96(98.38)	0.14(0.02)	−0.72(0.69)	40.35(4.62)	−3.84(1.27)	0.55(0.04)
Model 2	190.86 (26.59)	0.23(0.02)	0.40(0.35)	–	–	0.55(0.30)
Model 3	423.70 (14.03)	0.05(0.03)	–	–	–	–

Note: The standard errors are computed based on 250 bootstrap samples.

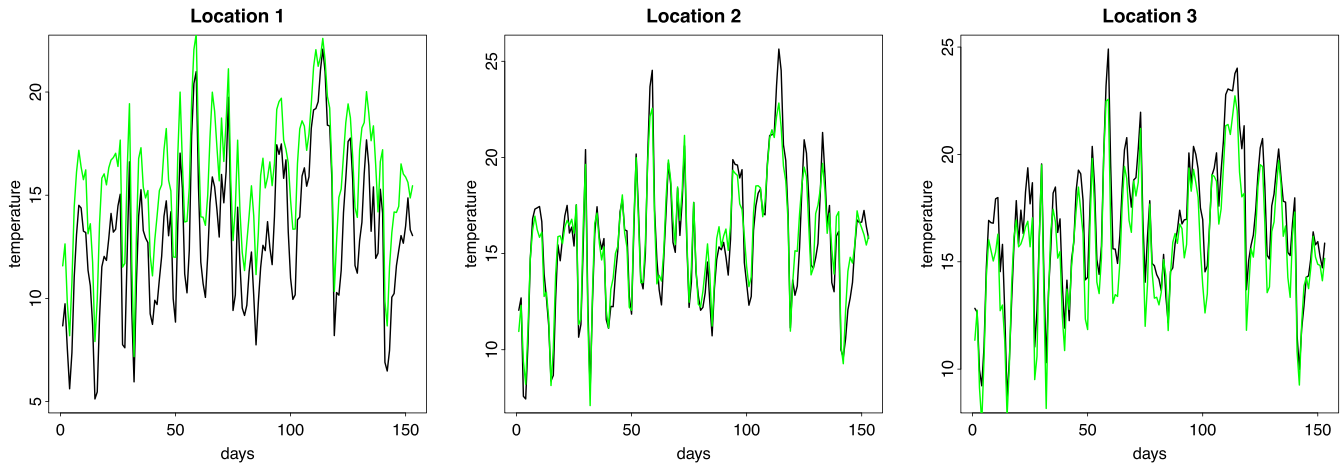

FIGURE 3 Based on Model 1, the predicted (in green) daily average temperature for three testing locations along with their corresponding observed values (in black) from 1 May 2011 to 30 September 2011 are plotted.

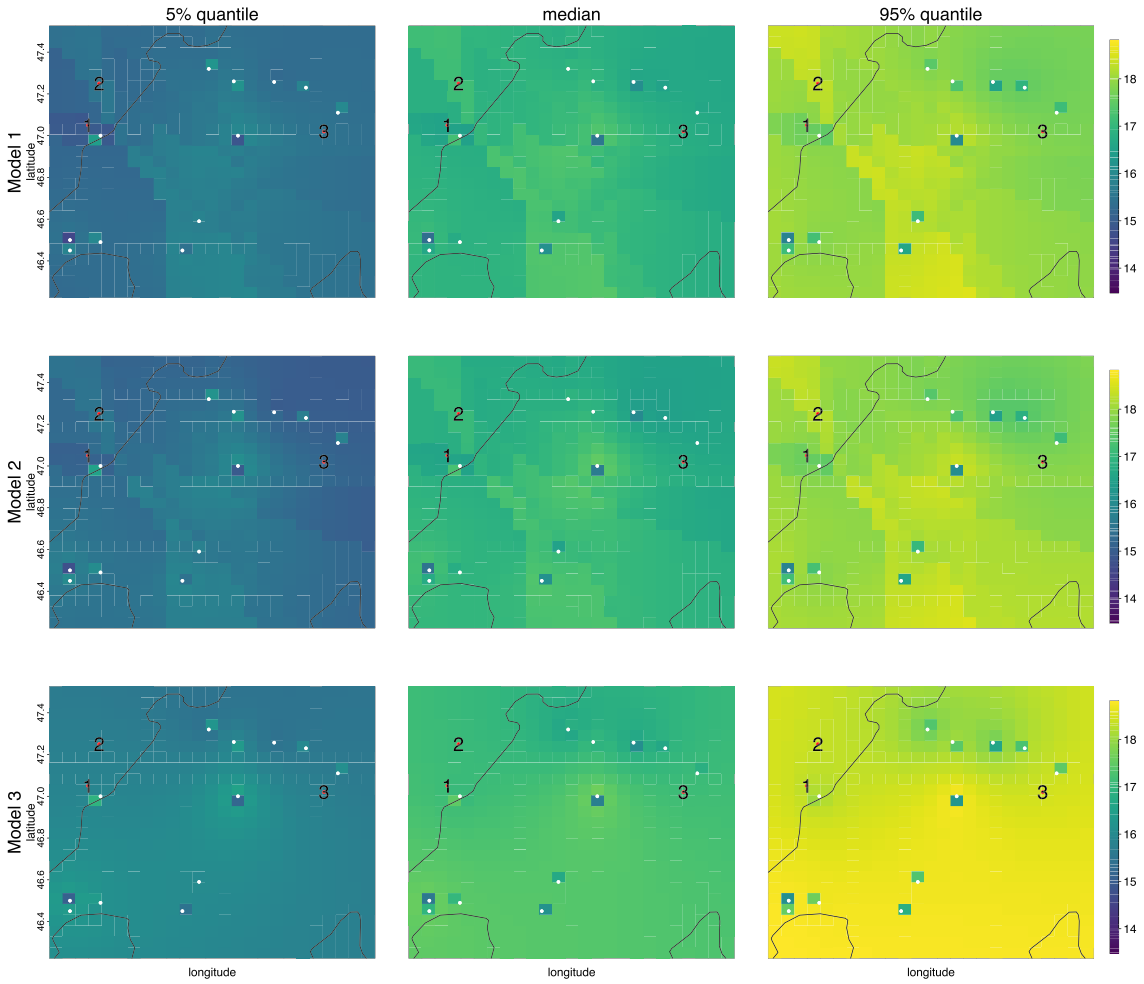

FIGURE 4 The estimated 5%, 50% (median) and 95% quantiles using Model 1 for average daily temperatures in the area of study for 11 August 2011 in the first row. The 12 training locations with observed temperature data are plotted in white dots and three testing locations are plotted in red dots. In the second and the third row, we plot the predicted 5%, 50% (median) and 95% quantiles based on Model 2 and Model 3, respectively.

to Krupskii and Genton (2019), we need to add an exponential common factor and independent exponential factor to the model to control the lower tail dependence. However, in that case, the joint copula density becomes very hard to work with because of its complex structure. A Bayesian framework may be considered here as well. Another possible future research idea could be to extend the proposed model to the space-time case like in Krupskii and Genton (2017) or with the Lagrangian framework. Furthermore, a direction to explore is to compute the copula and the copula density in Equations (6) and (7) using neural networks. This might enable us to compute the exact log-likelihoods for dimensions more than 20 as well as might speed up the modelling process greatly. One limitation of our proposed model is that it cannot capture full independence when the coefficients for common and independent factors are assumed to be positive. This can possibly be addressed by allowing them to be zeros as well. In that case, the related quantities such as the cdfs and pdfs have to be reformulated in order for them to be valid at the limiting points. This can be another new direction for further research. Finally, we have proposed the non-stationary model by assuming the coefficients of the common factor to be a function of space. This assumption can be made for the coefficients of the independent factors as well. That will yield a more general yet more complex non-stationary model. Codes for fitting the proposed model are available here.

AUTHOR CONTRIBUTIONS

All authors contributed to conceptualization, implementation, and manuscript writing.

ACKNOWLEDGEMENTS

The research in this article was funded by the King Abdullah University of Science and Technology (KAUST) in Saudi Arabia.

CONFLICT OF INTEREST STATEMENT

The authors declare no potential conflict of interests.

DATA AVAILABILITY STATEMENT

Data can be obtained from the authors of https://link.springer.com/article/10.1007/s11004-013-9469-y on request and are not publicly available.

ORCID

Pavel Krupskii https://orcid.org/0000-0001-9658-735X
Marc G. Genton https://orcid.org/0000-0001-6467-2998

REFERENCES

Allcroft, D. J., & Glasbey, C. A. (2003). A latent Gaussian Markov random-field model for spatiotemporal rainfall disaggregation. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *52*(4), 487–498.
Azzalini, A., & Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew-t distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *65*(2), 367–389.
Bárdossy, A. (2006). Copula-based geostatistical models for groundwater quality parameters. *Water Resources Research*, *42*(11), W11416.
Bárdossy, A. (2011). Interpolation of groundwater quality parameters with some values below the detection limit. *Hydrology and Earth System Sciences*, *15*(9), 2763–2775.
Bárdossy, A., & Li, J. (2008). Geostatistical interpolation using copulas. *Water Resources Research*, *44*(7), W07412.
Bevilacqua, M., Caamaño-Carrillo, C., Arellano-Valle, R. B., & Morales-Oñate, V. (2021). Non-Gaussian geostatistical modeling using (skew)-*t* processes. *Scandinavian Journal of Statistics*, *48*(1), 212–245.
Castro-Camilo, D., & Huser, R. (2020). Local likelihood estimation of complex tail dependence structures, applied to US precipitation extremes. *Journal of the American Statistical Association*, *115*(531), 1037–1054.
Cressie, N. (1993). *Statistics for spatial data*: John Wiley & Sons.
Davison, A. C., Huser, R., & Thibaud, E. (2013). Geostatistics of dependent and asymptotically independent extremes. *Mathematical Geosciences*, *45*, 511–529.
De Oliveira, V. (2006). On optimal point and block prediction in log-Gaussian random fields. *Scandinavian Journal of Statistics*, *33*(3), 523–540.
De Oliveira, V., Kedem, B., & Short, D. A. (1997). Bayesian prediction of transformed Gaussian random fields. *Journal of the American Statistical Association*, *92*(440), 1422–1433.
Erhardt, T. M., Czado, C., & Schepsmeier, U. (2015). R-vine models for spatial time series with an application to daily mean temperature. *Biometrics*, *71*(2), 323–332.
Fonseca, T. C. O., & Steel, M. F. J. (2011). Non-Gaussian spatiotemporal modelling through scale mixing. *Biometrika*, *98*(4), 761–774.
Genton, M. G., & Zhang, H. (2012). Identifiability problems in some non-Gaussian spatial random fields. *Chilean Journal of Statistics*, *3*(2), 171–179.
Gräler, B. (2014). Modelling skewed spatial random fields through the spatial vine copula. *Spatial Statistics*, *10*, 87–102.
Gräler, B., & Pebesma, E. (2011). The pair-copula construction for spatial data: A new approach to model spatial dependency. *Procedia Environmental Sciences*, *7*, 206–211.
Huser, R., & Wadsworth, J. L. (2019). Modeling spatial processes with unknown extremal dependence class. *Journal of the American Statistical Association*, *114*(525), 434–444.
Johns, C. J., Nychka, D., Kittel, T. G. F., & Daly, C. (2003). Infilling sparse records of spatial fields. *Journal of the American Statistical Association*, *98*(464), 796–806.
Kim, H.-M., & Mallick, B. K. (2004). A Bayesian prediction using the skew Gaussian distribution. *Journal of Statistical Planning and Inference*, *120*(1-2), 85–101.
Krupskii, P., & Genton, M. G. (2017). Factor copula models for data with spatio-temporal dependence. *Spatial Statistics*, *22*(1), 180–195.
Krupskii, P., & Genton, M. G. (2019). A copula model for non-Gaussian multivariate spatial data. *Journal of Multivariate Analysis*, *169*, 264–277.
Krupskii, P., & Huser, R. (2022). Modeling spatial tail dependence with Cauchy convolution processes. *Electronic Journal of Statistics*, *16*(2), 6135–6174.
Krupskii, P., Huser, R., & Genton, M. G. (2018). Factor copula models for replicated spatial data. *Journal of the American Statistical Association*, *113*(521), 467–479.
Krupskii, P., & Joe, H. (2015). Tail-weighted measures of dependence. *Journal of Applied Statistics*, *42*(3), 614–629.
Ma, C. (2009a). χ^2 random fields in space and time. *IEEE Transactions on Signal Processing*, *58*(1), 378–383.
Ma, C. (2009b). Construction of non-Gaussian random fields with any given correlation structure. *Journal of Statistical Planning and Inference*, *139*(3), 780–787.
Marchenko, Y. V., & Genton, M. G. (2010). Multivariate log-skew-elliptical distributions with applications to precipitation data. *Environmetrics*, *21*(3-4), 318–340.
Morton, G. M. (1966). A computer oriented geodetic data base and a new technique in file sequencing.
Palacios, M. B., & Steel, M. F. J. (2006). Non-Gaussian Bayesian geostatistical modeling. *Journal of the American Statistical Association*, *101*(474), 604–618.
Rimstad, K., & Omre, H. (2014). Skew-Gaussian random fields. *Spatial Statistics*, *10*, 43–62.
Røislien, J., & Omre, H. (2006). T-distributed random fields: A parametric model for heavy-tailed well-log data. *Mathematical Geology*, *38*(7), 821–849.
Sklar, M. (1959). Fonctions de répartition a *n* dimensions et leurs marges. *Institute of Statistics of the University of Paris*, *8*, 229–231.
Stein, M. L., Chi, Z., & Welty, L. J. (2004). Approximating likelihoods for large spatial data sets. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *66*(2), 275–296.
Stroud, A. H., & Secrest, D. (1966). *Gaussian quadrature formulas*: Prentice-Hall.

Vecchia, A. V. (1988). Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, *50*(2), 297–312.
Xu, G., & Genton, M. G. (2017). Tukey *g*-and-*h* random fields. *Journal of the American Statistical Association*, *112*(519), 1236–1249.
Yin, J., & Craigmile, P. F. (2018). Heteroscedastic asymmetric spatial processes. *Stat*, *7*(1), e206.
Zareifard, H., & Khaledi, M. J. (2013). Non-Gaussian modeling of spatial data using scale mixing of a unified skew-Gaussian process. *Journal of Multivariate Analysis*, *114*, 16–28.
Zhang, H., & El-Shaarawi, A. (2010). On spatial skew-Gaussian processes and applications. *Environmetrics*, *21*(1), 33–47.

AUTHOR BIOGRAPHIES

Sagnik Mondal is a data scientist at Victoria's Secret in Bangalore, India. His research interests are multivariate non-Gaussian distributions, non-Gaussian random fields, and time series forecasting.

Pavel Krupskii is a senior lecturer in the School of Mathematics and Statistics at Melbourne University, Australia. His research centers on models for multivariate data, with a focus on copula models, spatial data, and multivariate extremes.

Marc G. Genton is Al-Khawarizmi Distinguished Professor of Statistics at the King Abdullah University of Science and Technology, Saudi Arabia. His research interests include statistical analysis, modeling, prediction, and uncertainty quantification of spatio-temporal data, with applications in environmental and climate science, as well as renewable energies.

How to cite this article: Mondal, S., Krupskii, P., & Genton, M. G. (2024). A non-stationary factor copula model for non-Gaussian spatial data. *Stat*, *13*(3), e715. https://doi.org/10.1002/sta4.715