

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta

Factor copula models for data with spatio-temporal dependence

Pavel Krupskii^{*}, Marc G. Genton

CEMSE Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia

ARTICLE INFO

Article history:

Received 21 January 2017

Accepted 2 October 2017

Available online 10 October 2017

Keywords:

Copula

Heavy tails

Non-Gaussian random field

Spatial statistics

Tail asymmetry

Temporal dependence

ABSTRACT

We propose a new copula model for spatial data that are observed repeatedly in time. The model is based on the assumption that there exists a common factor that affects the measurements of a process in space and in time. Unlike models based on multivariate normality, our model can handle data with tail dependence and asymmetry. The likelihood for the proposed model can be obtained in a simple form and therefore parameter estimation is quite fast. Simulation from this model is straightforward and data can be predicted at any spatial location and time point. We use simulation studies to show different types of dependencies, both in space and in time, that can be generated by this model. We apply the proposed copula model to hourly wind data and compare its performance with some classical models for spatio-temporal data.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Flexible and tractable models for data are often required in real-world applications, but building such models can be a challenging task if the data have complex structures. One example of such data is measurements of a process taken in space and in time, such as daily temperature measurements obtained at different weather stations or concentrations of a certain air pollutant measured by balloons launched from different locations. The dependence between two measurements that are made at different locations and at different times is usually weaker with a larger distance and time lag. Classical models for data with spatio-temporal dependence often assume multivariate normality with a spatio-temporal covariance matrix; see, for example, [Gneiting \(2002\)](#), [Stein \(2005\)](#) and [Gneiting et al. \(2007\)](#) for a review of covariance functions.

^{*} Corresponding author.

E-mail addresses: pavel.krupskiy@kaust.edu.sa (P. Krupskii), marc.genton@kaust.edu.sa (M.G. Genton).

For non-Gaussian spatial data, [Bárdossy \(2006\)](#) introduced the chi-squared copula and [Bárdossy and Li \(2008\)](#) proposed a v-transformed copula. These copula models are obtained from a non-monotonic transformation of multivariate normal variables. They can handle dependence asymmetry but cannot be used for modeling data with tail dependence. Furthermore, the likelihood for these models is not tractable in high dimensions. To construct flexible distributions for spatial data and to do the interpolation, vine copulas can be used. [Gräler \(2014\)](#) used spatial vine copulas to model and interpolate data with very strong dependencies, and [Erhardt et al. \(2015\)](#) used C-vine copulas to model the spatial dependence structure locally. Parameters in their model can be estimated using the composite likelihood, and data can be interpolated at arbitrary spatial locations.

For data with spatio-temporal dependence, [de Luna and Genton \(2005\)](#) used vector autoregressive models with spatial structure for time-forward predictions in environmental applications, but these models are not computationally tractable if the innovation process is not Gaussian. [Stroud et al. \(2011\)](#) proposed a model for nonstationary spatio-temporal data in which the mean function at each time period is a locally-weighted mixture of linear regressions. The authors provided details for the Gaussian case but did not study the dependence properties of the proposed models in the general case. In practical applications, however, the multivariate normality assumption is not always suitable. For example, it would be unsuitable for data with strong joint dependence in the tails (i.e., when large/small values are simultaneously observed more often than predicted by the normal model), or for data with reflection asymmetry (i.e., when large values are simultaneously observed more often than small values, or vice versa). [Fonseca and Steel \(2011\)](#) introduced a model for spatio-temporal data that can handle heavy tails. However, the likelihood function in that model is not available in simple form, and it cannot handle dependence asymmetry. [Schmidt et al. \(2017\)](#) proposed a model for a skewed spatio-temporal process. Their model is based on the combination of Gaussian processes with purely spatial dependence structures and a purely temporal component. The joint density in that model is not possible to obtain in a simple form and it cannot handle data with tail dependence; see also the discussion by [Genton and Hering \(2017\)](#).

To overcome this problem, copulas can be used to construct flexible, multivariate distributions. A copula is a multivariate cumulative distribution function (cdf) with uniform $U(0, 1)$ marginals. [Sklar \(1959\)](#) showed that for any continuous d -dimensional cdf $F_{1,\dots,d}$ with univariate marginals F_1, \dots, F_d , there exists a unique copula $C_{1,\dots,d}$ such that $F_{1,\dots,d}(z_1, \dots, z_d) = C_{1,\dots,d}\{F_1(z_1), \dots, F_d(z_d)\}$ for any z_1, \dots, z_d . Copulas have been used in many different applications, such as modeling financial returns data ([Patton, 2006](#); [Krupskii and Joe, 2013](#)), hydrology data ([Genest and Favre, 2007](#)) and others.

Recently, [Krupskii et al. \(in press\)](#) introduced a copula model for spatial data with replicates and without temporal dependence. The model is based on the process

$$W(\mathbf{s}) = Z(\mathbf{s}) + V_0, \quad \mathbf{s} \in \mathbb{R}^d,$$

where Z is a Gaussian process and V_0 is a common factor that does not depend on Z or location \mathbf{s} . In this paper, we propose an extension of this model that is based on the process W measured in space and in time:

$$W(\mathbf{s}, t) = Z(\mathbf{s}, t) + \alpha(\mathbf{s}, t)\mathcal{E}_{\mathcal{P}(t)}, \quad \mathbf{s} \in \mathbb{R}^d, t \in \mathbb{R}_+. \quad (1)$$

Here $Z(\mathbf{s}, t)$ is a Gaussian process in space and in time with zero mean, unit variance and covariance matrix Σ_Z , $\alpha(\mathbf{s}, t)$ is a non-random function of space (\mathbf{s}) and time (t), $\mathcal{P}(t)$ is a Poisson process with intensity function $\Lambda(t)$ and $\mathcal{E}_t \sim_{i.i.d.} \text{Exp}(1)$ are exponential factors that do not depend on $Z(\mathbf{s}, t)$ or on location \mathbf{s} .

The factors $\mathcal{E}_{\mathcal{P}(t)}$ allow for tail dependence for the copula corresponding to the joint distribution of the process $W(\mathbf{s}, t)$ measured at different spatial locations and at different time points. The intensity function, $\Lambda(t)$, of the Poisson process $\mathcal{P}(t)$ controls the rate of decay of dependence over time. The exponential distribution of $\mathcal{E}_{\mathcal{P}(t)}$ allows one to obtain the joint copula density in this model (1) in closed form so that the model parameters can be efficiently estimated using the maximum likelihood approach.

The rest of this paper is organized as follows. In Section 2 we define the model (1) for data observed at different spatial locations and time points and study its dependence properties based on the covariance function of $Z(\mathbf{s}, t)$ and the choice of $\alpha(\mathbf{s}, t)$ and $\Lambda(t)$. In Section 3 we generate

various data sets to show the wide range of dependence structures that can be obtained from the proposed copula model. In Section 4 we give more details about maximum likelihood estimation and prediction. In Section 5 we apply the model to hourly wind data and compare the performance of the proposed model with some classical models for spatio-temporal data. Finally, Section 6 concludes with a discussion.

2. A factor copula model for spatio-temporal data

We use the following notation throughout this paper: Φ_{Σ} and ϕ_{Σ} are the joint cdf and its probability density function (pdf), respectively, for the multivariate normal random variable \mathbf{Z} with a covariance matrix Σ . Consider the process $W(\mathbf{s}, t)$, as defined in (1), measured at n different locations $\mathbf{s}_1, \dots, \mathbf{s}_n$ and at T different time points $t_1 < t_2 < \dots < t_T$. For simplicity, let $W_{i,j} := W(\mathbf{s}_i, t_j)$, $Z_{i,j} := Z(\mathbf{s}_i, t_j)$ and $\alpha_{i,j} := \alpha(\mathbf{s}_i, t_j)$. From (1) we have:

$$W_{i,j} = Z_{i,j} + \alpha_{i,j}\mathcal{E}_{\mathcal{P}(t_j)}, \quad i = 1, \dots, n, j = 1, \dots, T. \tag{2}$$

From the definition of the Poisson process $\mathcal{P}(t)$, we see that $\lambda_{j_1 j_2} := \text{pr}\{\mathcal{P}(t_{j_1}) = \mathcal{P}(t_{j_2})\} = \exp\left\{-\int_{t_{j_1}}^{t_{j_2}} \Lambda(t)dt\right\}$. Let $\Sigma_{\mathbf{Z}}$ and $\Sigma_{\mathbf{W}}$ be the covariance matrices of the vectors

$$\mathbf{Z} = (Z_{1,1}, \dots, Z_{n,1}, \dots, Z_{1,T}, \dots, Z_{n,T})^{\top} \text{ and } \mathbf{W} = (W_{1,1}, \dots, W_{n,1}, \dots, W_{1,T}, \dots, W_{n,T})^{\top},$$

respectively. It follows (with $\lambda_{j,j} = 1$) that

$$\Sigma_{\mathbf{W}}^{j_1 j_2} = \{\Sigma_{\mathbf{Z}}^{j_1 j_2} + \lambda_{j_1 j_2} \alpha_1(t_{j_1}) \alpha_1(t_{j_2})^{\top}\} / \{\alpha_2(t_{j_1}) \alpha_2(t_{j_2})^{\top}\}, \quad 1 \leq j_1, j_2 \leq T,$$

where $\alpha_1(t) = \{\alpha_{1,t}, \dots, \alpha_{n,t}\}^{\top}$, $\alpha_2(t) = \{\mathbf{1}_n + \alpha_1(t)^2\}^{1/2}$, $\mathbf{1}_n$ is a vector of ones of length n , and the superscripts j_1, j_2 denote the $n \times n$ block of the covariance matrix ($\Sigma_{\mathbf{Z}}$ or $\Sigma_{\mathbf{W}}$) corresponding to cross-covariances at time t_{j_1} and t_{j_2} , for different locations.

The correlation for the process $W(\mathbf{s}, t)$ is larger for smaller distances and smaller lags in time because $(\Sigma_{\mathbf{Z}}^{j_1 j_2})_{\mathbf{s}_1, \mathbf{s}_2}$ and λ_{j_1, j_2} become large when the quantities $\|\mathbf{s}_1 - \mathbf{s}_2\|$ and $|t_{j_1} - t_{j_2}|$ are small. Here we assume that the covariance matrix $\Sigma_{\mathbf{Z}}$ is parameterized in such a way that correlations are smaller for pairs of observations with larger distances and larger time lags. Most of the classical spatio-temporal covariance models satisfy this property.

Let $F^{\mathbf{W}}$ and $f^{\mathbf{W}}$ be the joint cdf and pdf of the vector \mathbf{W} , and $F_{i,j}^{\mathbf{W}}$ and $f_{i,j}^{\mathbf{W}}$ be the marginal cdf and pdf of $W_{i,j}$, respectively. The copula $C^{\mathbf{W}}$ and its density, $c^{\mathbf{W}}$, corresponding to the joint distribution, $F^{\mathbf{W}}$, can then be written as follows:

$$C^{\mathbf{W}}(\mathbf{u}) = F^{\mathbf{W}}(\mathbf{w}), \quad c^{\mathbf{W}}(\mathbf{u}) = f^{\mathbf{W}}(\mathbf{w}) / \prod_{i=1, j=1}^{n, T} f_{i,j}^{\mathbf{W}}(w_{ij}),$$

where $\mathbf{w} = (w_{11}, \dots, w_{n1}, \dots, w_{1T}, \dots, w_{nT})^{\top}$, $\mathbf{u} = (u_{11}, \dots, u_{n1}, \dots, u_{1T}, \dots, u_{nT})^{\top}$ and $w_{ij} = (F_{i,j}^{\mathbf{W}})^{-1}(u_{ij})$. Here, the copula $C^{\mathbf{W}}$ can be used for modeling data with arbitrary marginals (not necessarily those of the vector \mathbf{W}), thus allowing greater flexibility in the proposed model.

In the next section we show that the copula density $c^{\mathbf{W}}$ can be obtained in a simple form and therefore the maximum likelihood estimates can be obtained quite fast. Furthermore, the proposed copula has appealing tail properties as it allows one to control the strength of dependence in the upper tail. Classical measures of tail dependence are the lower and upper tail dependence coefficients, λ_L and λ_U , defined for a bivariate copula, $C_{1,2}$:

$$\lambda_L := \lim_{q \rightarrow 0} C_{1,2}(q, q)/q \in [0, 1] \quad \text{and} \quad \lambda_U := \lim_{q \rightarrow 0} \bar{C}_{1,2}(1 - q, 1 - q)/q \in [0, 1],$$

where $\bar{C}_{1,2}(u_1, u_2) := 1 - u_1 - u_2 + C_{1,2}(u_1, u_2)$ is the survival copula. If $\lambda_L > 0$ ($\lambda_U > 0$), then the copula $C_{1,2}$ is said to have lower (upper) tail dependence. For the Gaussian copula, $\lambda_L = \lambda_U = 0$, and therefore this copula may not be suitable for modeling data with strong dependence in the tails. At the same time, the copula $C^{\mathbf{W}}$ allows for upper tail dependence, and this dependence is weaker with a larger distance or with a larger time lag as the following proposition shows.

Proposition 1. Let $C_{1,j_1;2,j_2}^W$ be the bivariate copula corresponding to the joint distribution of the vector $(W_{1,j_1}, W_{2,j_2})^\top$ and let $C_{1,j_1;2,j_2}^{W^*}$ be the corresponding extreme-value copula (Segers, 2012). It follows that

$$C_{1,j_1;2,j_2}^W(u_1, u_2) = \{\text{HR}(u_1, u_2; \vartheta_{\text{HR}})\}^{\lambda_{j_1 j_2}} (u_1 u_2)^{1-\lambda_{j_1 j_2}},$$

where HR is a bivariate copula corresponding to the Hüsler–Reiss distribution (Hüsler and Reiss, 1989) with parameter ϑ_{HR} which depends on spatial locations $\mathbf{s}_1, \mathbf{s}_2$ and time points t_{j_1}, t_{j_2} :

$$\vartheta_{\text{HR}} = \vartheta_{\text{HR}}(t_{j_1}, t_{j_2}, \mathbf{s}_1, \mathbf{s}_2) = \frac{\left\{ \alpha_{1,j_1}^2 + \alpha_{2,j_2}^2 - 2\alpha_{1,j_1}\alpha_{2,j_2}(\Sigma_{\mathbf{Z}}^{j_1 j_2})_{\mathbf{s}_1, \mathbf{s}_2} \right\}^{1/2}}{\alpha_{1,j_1}\alpha_{2,j_2}}.$$

The proof is given in Appendix A.1. Let λ_U be the upper tail dependence coefficient for $C_{1,j_1;2,j_2}^W$. It implies that λ_U depends on spatial locations, $\mathbf{s}_1, \mathbf{s}_2$, and time points, t_{j_1}, t_{j_2} , and $\lambda_U = \lambda_U(t_{j_1}, t_{j_2}, \mathbf{s}_1, \mathbf{s}_2) = 2\lambda_{j_1 j_2} \Phi\{-\vartheta_{\text{HR}}(t_{j_1}, t_{j_2}, \mathbf{s}_1, \mathbf{s}_2)/2\}$. To simplify the notation, thereafter we use ϑ_{HR} and λ_U without showing the spatial locations and time points these two quantities depend on.

A special case of spatio-temporal isotropy can be obtained in the proposed model when $\alpha(\mathbf{s}, t) = \alpha > 0$ and $\Lambda(t) = \lambda > 0$ for any $\mathbf{s} \in \mathbb{R}^d$ and $t \in \mathbb{R}_+$. It implies that all marginal cdfs $F_{i,j}^W, i = 1, \dots, n, j = 1, \dots, T$, are the same and that $\Sigma_{\mathbf{W}} = (\Sigma_{\mathbf{Z}} + \alpha^2 \mathbf{1}_{nT} \mathbf{1}_{nT}^\top)/(1 + \alpha^2)$. The upper tail dependence coefficient in this model

$$\lambda_U = 2 \exp\{-\lambda|t_{j_1} - t_{j_2}|\} \Phi \left[-\frac{1}{\alpha} \left\{ \frac{1 - (\Sigma_{\mathbf{Z}}^{j_1 j_2})_{\mathbf{s}_1, \mathbf{s}_2}}{2} \right\}^{1/2} \right].$$

It follows that, if $Z(\mathbf{s}, t)$ is an isotropic Gaussian process both in space and in time, then $W(\mathbf{s}, t)$ is also an isotropic process in space and in time. For more general structures, one can select various functions $\alpha(\mathbf{s}, t)$ and $\Lambda(t)$; see Section 4.3 for more details.

3. Simulated examples

Simulating data from the proposed model is straightforward. To generate a vector \mathbf{W} as given by (2), a multivariate normal vector \mathbf{Z} with zero mean and the covariance matrix $\Sigma_{\mathbf{Z}}$ should be generated first. One then generates a Poisson process with the intensity function $\Lambda(t)$ at time points $t_1 < \dots < t_T$ and constructs the vector \mathbf{W} using (2). Finally, Krupskii et al. (in press) showed that $F_{i,j}^W(w) = \Phi(w) - \exp\{1/(2\alpha_{i,j}^2) - w/\alpha_{i,j}\} \Phi(w - 1/\alpha_{i,j})$ and therefore, to get the data with uniform $U(0, 1)$ marginals, one should use the probability integral transform: $u_{ij} = F_{i,j}^W(w_{ij}), i = 1, \dots, n, j = 1, \dots, T$.

Here we generate some data sets with standard normal marginals assuming $\alpha(\mathbf{s}, t) = 1$ and applying different functions $\Lambda(t)$ to show the flexibility of the proposed model. We discuss the choice of these functions in more detail in the next section. For illustration purposes, we use a simple, separable isotropic structure for the covariance matrix $\Sigma_{\mathbf{Z}}$ in all three cases. Namely, we assume

$$\text{cov}(Z_{1,1}, Z_{2,2}) = \exp(-\theta_{\text{SP}}\|\mathbf{s}_1 - \mathbf{s}_2\| - \theta_{\text{TM}}|t_1 - t_2|),$$

that is, the cross-covariance in space and in time only depends on the distance $d_s = \|\mathbf{s}_1 - \mathbf{s}_2\|$ and the time lag $d_t = |t_1 - t_2|$. In applications, more flexible nonseparable models for $\Sigma_{\mathbf{Z}}$ can be used if needed to increase the flexibility of the model in the middle of the joint distribution, whereas the functions $\alpha(\mathbf{s}, t)$ and $\Lambda(t)$ control the tail behavior of the copula C^W .

We use the following specifications (models):

- M1 $\alpha(\mathbf{s}, t) = 1, \Lambda(t) = 0.15, \theta_{\text{SP}} = 0.2, \theta_{\text{TM}} = 0.3$ (isotropic model; dependence quickly decreases in time);
- M2 $\alpha(\mathbf{s}, t) = 1, \Lambda(t) = 0.05, \theta_{\text{SP}} = 0.2, \theta_{\text{TM}} = 0.3$ (isotropic model; dependence slowly decreases in time).
- M3 $\alpha(\mathbf{s}, t) = 1, \Lambda(t) = 0.05 + 0.25t, \theta_{\text{SP}} = 0.2, \theta_{\text{TM}} = 0.1$ (tail dependence decreases very quickly in time).

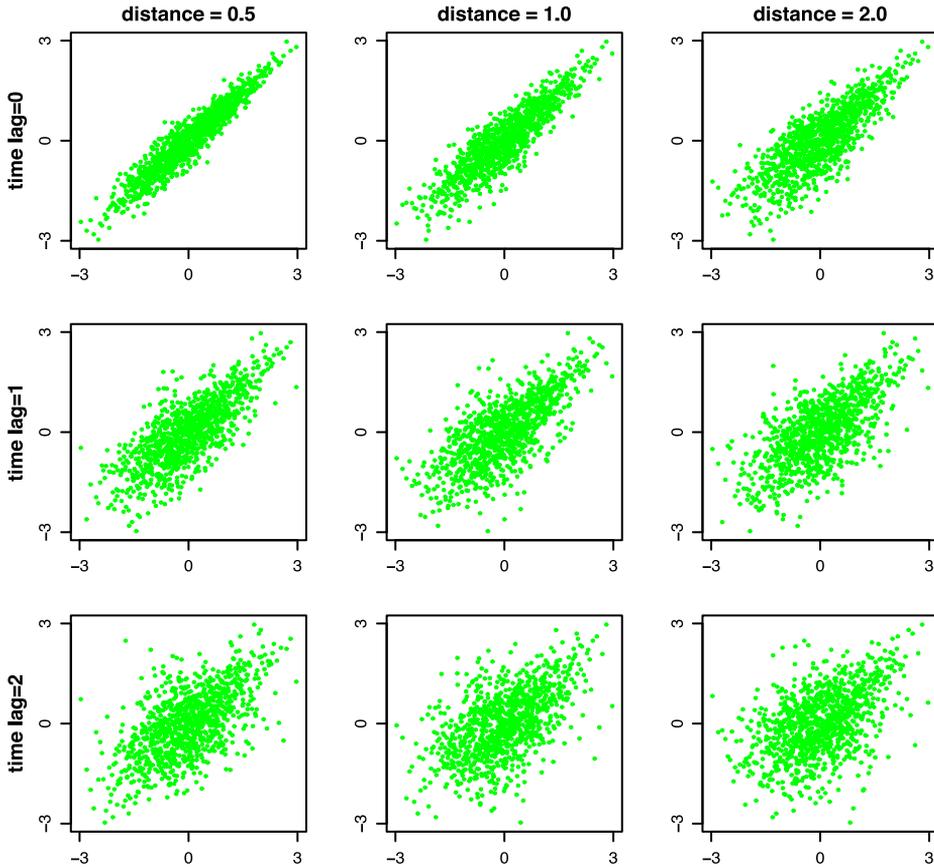


Fig. 1. Scatter plots of 1000 replicates of data generated from model M1: distance = 0.5 (left), distance = 1 (middle), distance = 2 (right); time lag = 0 (top), time lag = 1 (middle), time lag = 2 (bottom).

Fig. 1 shows bivariate scatter plots for different pairs of data generated from (isotropic) model M1, defined above, with time lags $d_t = 0, 1, 2$ and with distances $d_s = 0.5, 1, 2$. We can see that the dependence is stronger in the upper tail and weaker with a larger distance or a larger time lag. Because model M1 is isotropic in space and in time, the dependence between two realizations of the spatio-temporal process measured at two different locations only depends on the time lag and the distance between these locations.

Fig. 2 shows the Spearman's rho, S_ρ , and the upper tail dependence coefficient, λ_U , of the copula $C_{1,1;2,2}^W$ for $t_1 = 0$ and $t_2 = 1, \dots, 20$ and for distance $\|\mathbf{s}_1 - \mathbf{s}_2\| = 0.5, 2$. We see that the dependence (as measured by S_ρ and by λ_U) decreases more slowly in time for model M2 (normal line) than for model M1 (thick line) because of its smaller intensity λ for the Poisson process \mathcal{P}_t . This parameter controls the rate of decay of spatio-temporal dependence with time, especially for λ_U . In particular, as seen in Proposition 1, the corresponding limiting extreme-value copula, $C_{1,1;2,2}^W$, converges to independence at rate $\lambda_{1,2} = \exp\left\{-\int_0^{t_2} \Lambda(t)dt\right\}$. For model M3, the rate of decay of S_ρ is comparable to that for model M1; however, the rate of decay of λ_U is much larger. This is because the intensity function $\Lambda(t) = 0.05 + 0.25t$ is an increasing function of the time lag, t , and with larger t , $\Lambda(t)$ becomes very large. More flexible models can be obtained with different values of $\alpha(\mathbf{s}, t)$ for different locations, \mathbf{s} , and with nonseparable cross-covariance matrices Σ_Z . Of course, the choice of

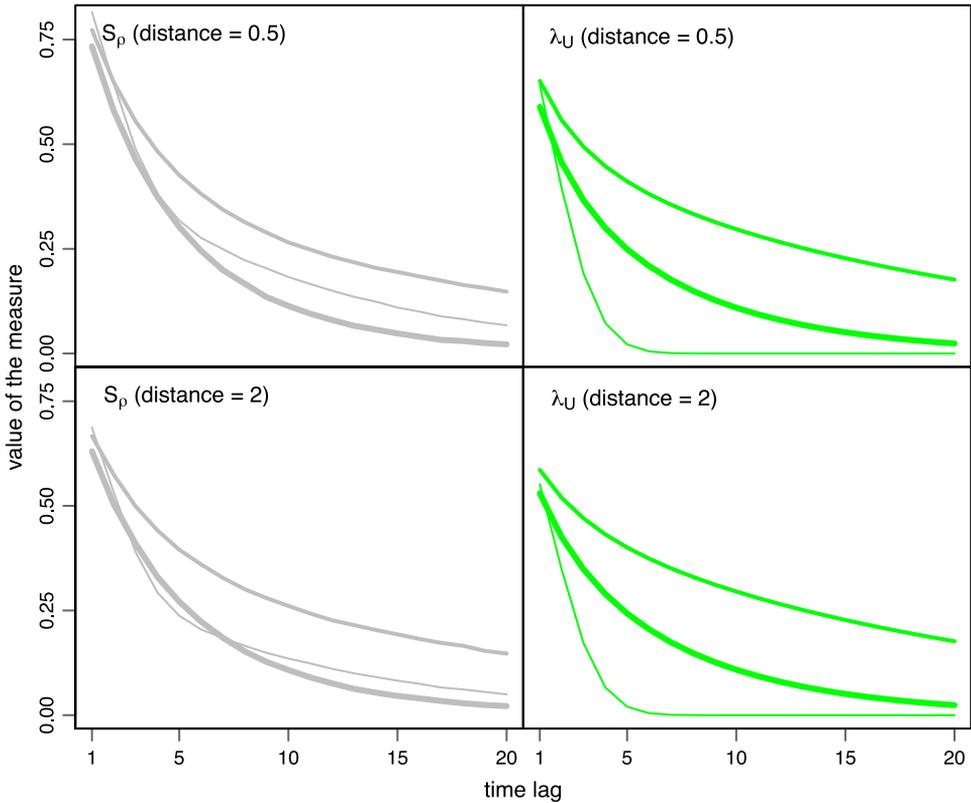


Fig. 2. Spearman's rho, S_ρ , and upper tail dependence coefficient, λ_U , for model M1 (thick line), model M2 (normal line) and model M3 (thin line) for different time lags and distance = 0.5 (top) and distance = 2 (bottom).

these functions, $\alpha(\mathbf{s}, t)$, $A(t)$, and the covariance matrix Σ_Z , depends on the particular application and structure of the data; see Section 4.3 for more guidelines.

4. Maximum likelihood estimates

In this section we obtain the formula for the copula density $c^{\mathbf{W}}$ and provide more details about maximum likelihood estimates.

4.1. The copula density

Following the notation in Section 2, $\mathbf{w}_t = (w_{1t}, \dots, w_{nt})^\top$ and $\boldsymbol{\alpha}_t = \{\alpha(\mathbf{s}_1, t), \dots, \alpha(\mathbf{s}_n, t)\}^\top$, $t = 1, \dots, T$. To compute the copula density, one needs to compute the joint density $f^{\mathbf{W}}(\mathbf{w})$.

Conditional on the factors $\mathcal{E}_{\mathcal{P}(t_j)}$, $j = 1, \dots, T$, the joint density $f^{\mathbf{W}}(\mathbf{w})$ is the multivariate normal density by construction. To obtain the unconditional density, one therefore needs to integrate the multivariate normal density with respect to the common factors. The number of these factors depends on how many jumps the process $\mathcal{P}(t)$ has, and on the times of these jumps. To describe the distribution of the jumps, we define the vector $\mathbf{j} = (j_1, \dots, j_T)^\top$ where $j_1 = 1$ and $j_{k+1} = j_k$ (the process $\mathcal{P}(t)$ has no jump for $j_k < t < j_{k+1}$) or $j_{k+1} = j_k + 1$ (the process $\mathcal{P}(t)$ has at least one jump for $j_k < t < j_{k+1}$) for any $k = 1, \dots, T - 1$. In particular, if no jumps occurred for $t_1 < t < t_T$, then $\mathbf{j} = (1, 1, \dots, 1)^\top$, and if at least one jump occurred for any $j_k < t < j_{k+1}$, $k = 1, \dots, T - 1$, then $\mathbf{j} = (1, 2, \dots, T)^\top$. There are 2^{T-1} possible vectors \mathbf{j} ; we denote this set of vectors \mathbf{j} as \mathcal{J} .

To compute $f^{\mathbf{W}}(\mathbf{w})$, one can use the formula of total probability: $f^{\mathbf{W}}(\mathbf{w}) = \sum_{\mathbf{j} \in \mathcal{J}} p(\mathbf{j}) I(\mathbf{j})$, where $I(\mathbf{j})$ is the conditional density $I(\mathbf{j}) = I(\mathbf{j}, \mathbf{w}) = f^{\mathbf{W}}(\mathbf{w}|\mathbf{j})$ and

$$p(\mathbf{j}) = \text{pr}\{\mathcal{P}(t_2) = j_2, \dots, \mathcal{P}(t_T) = j_T | \mathcal{P}(t_1) = 1\} = \prod_{k: j_{k-1}=j_k} \lambda_{k-1,k} \cdot \prod_{k: j_{k-1} \neq j_k} (1 - \lambda_{k-1,k}).$$

Because $|\mathcal{J}| = 2^{T-1}$, the joint density $f^{\mathbf{W}}$ is a weighted sum of 2^{T-1} terms of this type:

$$I(\mathbf{j}) = \int_{\mathbb{R}_+^{\max(\mathbf{j})}} \phi_{\Sigma_Z}(\mathbf{w}_1 - \alpha_1 v_{j_1}, \dots, \mathbf{w}_T - \alpha_T v_{j_T}) \exp\left(-\sum_{k=1}^{\max(\mathbf{j})} v_k\right) dv_1 \cdots dv_{\max(\mathbf{j})}.$$

In the integrand of $I(\mathbf{j})$, there are $\max(\mathbf{j})$ integration variables which we redefine as $v_1, \dots, v_{\max(\mathbf{j})}$ for simplicity. For example, if $\mathbf{j} = (1, 1, \dots, 1)^T$, then $\max(\mathbf{j}) = 1$ and $v_1 := v_{j_1} = v_{j_2} = \dots = v_{j_T}$. By combining terms with the same integration variables, $v_1, \dots, v_{\max(\mathbf{j})}$, we can see that

$$I(\mathbf{j}) = \int_{\mathbb{R}_+^{\max(\mathbf{j})}} \phi_{\Sigma_Z}(\tilde{\mathbf{w}}_1 - \tilde{\alpha}_1 v_1, \dots, \tilde{\mathbf{w}}_{\max(\mathbf{j})} - \tilde{\alpha}_{\max(\mathbf{j})} v_{\max(\mathbf{j})}) \exp\left(-\sum_{k=1}^{\max(\mathbf{j})} v_k\right) dv_1 \cdots dv_{\max(\mathbf{j})},$$

where $\tilde{\mathbf{w}}_k$ ($\tilde{\alpha}_k$) is a subvector of $\mathbf{w} = (\mathbf{w}_1^T, \dots, \mathbf{w}_T^T)^T$ ($\alpha = (\alpha_1^T, \dots, \alpha_T^T)^T$, respectively) that includes all \mathbf{w}_t (α_t) such that $j_t = k$, $k = 1, \dots, T$. In Appendix A.2, we show that integrals of this form can be obtained in closed form so that numerical integration is not required.

While there are 2^{T-1} terms that need to be calculated, they are all available in closed form, so the computationally more demanding part is in fact calculating the inverse quantities $w_{ij} = (F_{i,j}^{\mathbf{W}})^{-1}(u_{ij})$, $i = 1, \dots, n$, $j = 1, \dots, T$. These can be calculated just once and then used to compute all of the 2^{T-1} terms $I(\mathbf{j})$. The marginal distribution $F_{i,j}^{\mathbf{W}}(w) = \Phi(w) - \exp\{1/(2\alpha_{i,j}^2) - w/\alpha_{i,j}\} \Phi(w - 1/\alpha_{i,j})$ and, therefore, the inverse function can be calculated quite easily using numerical methods. As a result, the joint copula density $c^{\mathbf{W}}(\mathbf{u})$ can be calculated fairly quickly, at least for $T \leq 10$; see the next section for more details.

4.2. The log-likelihood function

Assume we have N replicates of i.i.d. data $(\xi_k)_{k=1}^N$, where $\xi_k = (\xi_{11,k}, \dots, \xi_{n1,k}, \dots, \xi_{1T,k}, \dots, \xi_{nT,k})^T$ has the joint distribution corresponding to the copula $C^{\mathbf{W}}$, $k = 1, \dots, N$. As we noticed earlier in Section 2, this vector can have arbitrary continuous univariate marginal cdfs; these need not be cdfs of W_{ij} , $F_{ij}^{\mathbf{W}}$, for $i = 1, \dots, n$ and $j = 1, \dots, T$. To estimate the parameters of $C^{\mathbf{W}}$, we need to transform the original data into uniform data. This can be done by estimating the marginal distributions of ξ_{ij} (values observed at the i th location and at time t_j), F_{ij}^{ξ} . The estimated marginal cdfs, \widehat{F}_{ij}^{ξ} can then be used to obtain uniform data:

$$u_{ij,k} = \widehat{F}_{ij}^{\xi}(\xi_{ij,k}), \quad i = 1, \dots, n, j = 1, \dots, T \text{ and } k = 1, \dots, N.$$

Alternatively, nonparametric ranks can be used to convert the original data to uniform scores for $i = 1, \dots, n$ and $j = 1, \dots, T$ as follows:

$$u_{ij,k} = \{\text{rank}(\xi_{ij,k}) - 0.5\} / N, \quad k = 1, \dots, N.$$

Let $\mathbf{u}_k = (u_{11}, \dots, u_{n1}, \dots, u_{1T}, \dots, u_{nT})^T$ and $\mathbf{u} = (\mathbf{u}_1^T, \dots, \mathbf{u}_N^T)^T$. The pseudo log-likelihood can then be written as follows:

$$\mathcal{L}(\mathbf{u}) = \sum_{k=1}^N \ln \left\{ \sum_{\mathbf{j} \in \mathcal{J}} p(\mathbf{j}; \theta_\lambda) I(\mathbf{j}, \mathbf{w}_k; \theta_\Sigma, \theta_\alpha) \right\} - \sum_{i,j,k=1}^{n,T,N} f_{i,j}^{\mathbf{W}}(u_{ij,k}; \theta_\alpha), \tag{3}$$

where the formula for $I(\mathbf{j}, \mathbf{w}_k) = I(\mathbf{j}, \mathbf{w}_k; \theta_\Sigma, \theta_\alpha)$ is given in the previous section, $\mathbf{w}_k = (w_{11,k}, \dots, w_{n1,k}, \dots, w_{1T,k}, \dots, w_{nT,k})^T$ and $w_{ij,k} = (F_{ij}^{\mathbf{W}})^{-1}(u_{ij,k})$. Here we assume that the functions $\alpha(\mathbf{s}, t)$, $\Lambda(t)$ and the covariance matrix Σ_Z are parameterized with vectors of parameters θ_α , θ_λ and θ_Σ , respectively. If the copula $C^{\mathbf{W}}$ is specified correctly and the number of replicates $N \rightarrow \infty$, then the parameter estimates $\hat{\theta}_\alpha$, $\hat{\theta}_\lambda$ and $\hat{\theta}_\Sigma$ obtained by maximizing the value of $\mathcal{L}(\mathbf{u})$ in (3) are asymptotically unbiased; see chapter 5.9 of Joe (2014) for details.

4.3. The choice of Σ_Z , $\alpha(\mathbf{s}, t)$ and $\Lambda(t)$

To estimate the parameters in model (2), one needs a parametric form for $\alpha(\mathbf{s}, t)$ and for $\Lambda(t)$. The former function models the change of spatial structure with time (and in space) and the latter function determines how quickly the dependence diminishes with time. Assume that Σ_Z is parameterized using an isotropic covariance function. We now consider some important cases:

- Anisotropic spatial structure that does not change in time. One can select $\alpha(\mathbf{s}, t) = \alpha(\mathbf{s})$ where $\alpha(\mathbf{s})$ can be parameterized depending on a particular application, for example, $\alpha(\mathbf{s}) = \alpha_0 + \alpha_1 s_x + \alpha_2 s_y$ where $\mathbf{s} = (s_x, s_y)^\top$ is a vector of geographical coordinates (longitude and latitude);
- Isotropic spatial structure (at a given time t) that changes in time. One can select $\alpha(\mathbf{s}, t) = \alpha(t)$. The function $\alpha(t)$ is parameterized depending on how quickly the spatial dependence changes in time, for example, $\alpha(t) = \alpha_0 \exp(-\alpha_1 t)$;
- Isotropic spatial structure (at a given time t) that does not change in space and in time. One can select $\alpha(\mathbf{s}, t) = \alpha_0 \geq 0$ (a nonnegative constant). If in addition $\Lambda(t) = \lambda > 0$, then we obtain a model with isotropic spatio-temporal dependence;
- Temporal dependence for a fixed time lag that changes in time. This can be modeled using the intensity function $\Lambda(t)$; for example, one can set $\Lambda(t) = \lambda t$ if the temporal dependence quickly decreases with time;
- Temporal dependence for a fixed time lag that is constant. This implies that the intensity function $\Lambda(t) = \lambda > 0$ is constant and the function $\alpha(\mathbf{s}, t) = \alpha(\mathbf{s})$ does not depend on time t .

Different types of asymmetric dependencies can be obtained as well in the proposed model. For example, for permutation asymmetry in space (in time), when the order of variables is important, one can use the function $\alpha(\mathbf{s}, t)$ such that $\alpha(\mathbf{s}_1, t) \neq \alpha(\mathbf{s}_2, t)$ for $\mathbf{s}_1 \neq \mathbf{s}_2$ ($\alpha(\mathbf{s}, t_1) \neq \alpha(\mathbf{s}, t_2)$ for $t_1 \neq t_2$, respectively). If the modeling process lacks full space–time symmetry, then one can select an asymmetric model for the covariance matrix Σ_Z ; see Gneiting (2002), Stein (2005), and Gneiting et al. (2007).

4.4. Interpolation in space and prediction in time

The estimated model (2) can be used for interpolating (predicting) data at new locations (time points). For a given set of locations $\mathbf{s}_1, \dots, \mathbf{s}_n$, when the time point $t > 0$ and a vector of uniform data $\mathbf{u}_t = (u_{1t}, \dots, u_{nt})^\top$, one can compute the joint density of $\mathbf{u}_{T,t} = (u_{n+1,T}, \mathbf{u}_t^\top)^\top$ where $u_{n+1,T}$ corresponds to the measurement of process in (1) at time $T > t$ and a spatial location \mathbf{s}_{n+1} (this can be one of the locations $\mathbf{s}_1, \dots, \mathbf{s}_n$ or a new location), on a uniform scale. This is possible if the cross-covariance matrix Σ_Z is parameterized using a spatio-temporal covariance function: with a new location and time point, one can recalculate Σ_Z corresponding to the joint distribution of the vector $\mathbf{u}_{T,t}$.

The conditional copula density is then given by

$$c_{n+1,T|t}(u_{n+1,T} | \mathbf{u}_t) = \frac{c_{n+1,T,t}(\mathbf{u}_{T,t}; \hat{\theta}_\alpha, \hat{\theta}_\lambda, \hat{\theta}_\Sigma)}{c_t(\mathbf{u}_t; \hat{\theta}_\alpha, \hat{\theta}_\Sigma)},$$

where c_t is the copula density of \mathbf{u}_t and $\hat{\theta}_\alpha, \hat{\theta}_\lambda, \hat{\theta}_\Sigma$ are parameter estimates. Note that c_t does not depend on θ_λ because c_t does not depend on $\Lambda(t)$:

$$c_t(\mathbf{u}_t; \theta_\alpha, \theta_\Sigma) = \frac{\int_{\mathbb{R}_+} \phi_{\Sigma_Z}(\mathbf{w}_t - \boldsymbol{\alpha}_t v_1) \exp(-v_1) dv_1}{\prod_{i=1}^n f_{i,t}^{\mathbf{W}}(w_{it})}.$$

By construction, the copula density $c_{n+1,T,t}$ is

$$c_{n+1,T,t}(\mathbf{u}_{T,t}) = \frac{\exp\left\{-\int_t^T \Lambda(t) dt\right\} I_1 + \left[1 - \exp\left\{-\int_t^T \Lambda(t) dt\right\}\right] I_2}{f_{n+1,T}^{\mathbf{W}}(w_{n+1,T}) \prod_{i=1}^n f_{i,t}^{\mathbf{W}}(w_{it})},$$

where

$$I_1 = \int_{\mathbb{R}_+^1} \phi_{\Sigma_Z} \left\{ (\mathbf{w}_t^\top, w_{n+1,T})^\top - (\boldsymbol{\alpha}_t^\top, \alpha_{n+1,T})^\top v_1 \right\} \exp(-v_1) dv_1$$

$$I_2 = \int_{\mathbb{R}_+^2} \phi_{\Sigma_Z} (\mathbf{w}_t - \boldsymbol{\alpha}_t v_1, w_{n+1,T} - \alpha_{n+1,T} v_2) \exp(-v_1 - v_2) dv_1 dv_2.$$

Both I_1 and I_2 can be obtained in closed form; see [Appendix A.2](#) for details.

The distribution of $u_{n+1,T}$ conditional on \mathbf{u}_t is

$$C_{n+1,T|t}(u_{n+1,T}|\mathbf{u}_t) = \int_0^{u_{n+1,T}} c_{n+1,T|t}(\tilde{u}|\mathbf{u}_t) d\tilde{u}.$$

This distribution, $C_{n+1,T|t}$, and its density, $c_{n+1,T|t}$, can be used to compute different quantities of interest, including the conditional expectation, \hat{m} , or the conditional median, $\hat{q}_{0.5}$:

$$\hat{m} := \int_0^1 \tilde{u} c_{n+1,T|t}(\tilde{u}|\mathbf{u}_t) d\tilde{u}, \quad \hat{q}_{0.5} := C_{n+1,T|t}^{-1}(0.5|\mathbf{u}_t).$$

One can similarly compute the distribution of $u_{n+1,T}$, conditional on data $(\mathbf{u}_{t_1}^\top, \dots, \mathbf{u}_{t_k}^\top)^\top$ observed at several time points in the past $t_1 < \dots < t_k < T$. It is not necessary to use all the spatial locations, $\mathbf{s}_1, \dots, \mathbf{s}_n$, to do interpolation or prediction. If the number of spatial locations is large, one can use 5–10 closest neighbors.

Note that \hat{m} and $\hat{q}_{0.5}$ are the interpolated values on the uniform scale. If $\hat{G}_{n+1,T}$ is the estimated univariate marginal distribution function for the process in (1) measured at time T and location \mathbf{s}_{n+1} , one can use $\hat{G}_{n+1,T}$ to convert these values to the original scale. For example, the predicted median on the original scale is $\hat{z}_{0.5} = \hat{G}_{n+1,T}^{-1}(\hat{q}_{0.5})$, and the predicted mean is $\hat{m}_z = \int_0^1 \hat{G}_{n+1,T}^{-1}(\tilde{u}) c_{n+1,T|t}(\tilde{u}|\mathbf{u}_t) d\tilde{u}$.

5. Empirical studies

In this section, we evaluate the performance of the algorithm in obtaining maximum likelihood estimates for simulated data sets and then apply the proposed model to hourly wind data. We also include classical models based on multivariate normal and Student- t distributions with spatio-temporal covariance matrices for comparison.

5.1. Maximum likelihood estimates for simulated data sets

In this section, we focus on spatio-temporal isotropic models with $\alpha(\mathbf{s}, t) = \alpha \geq 0$ and $\Lambda(t) = \lambda \geq 0$. In more general (non-isotropic) cases, the running time is usually 2–3 times slower since more parameters need to be estimated and more inverse functions $(F_{ij}^W)^{-1}(u_{ij})$ need to be calculated, $i = 1, \dots, n, j = 1, \dots, T$. We also assume that the matrix Σ_Z is a Kronecker product $\Sigma_Z = \Sigma_S \otimes \Sigma_T$. Here, Σ_S is an $n \times n$ matrix that models spatial covariance structure and Σ_T is a $T \times T$ matrix that models temporal covariance structure. Similar results can be obtained for other (nonseparable) models of Σ_Z and therefore we use the separable structure for simplicity. Note that Σ_W is not a separable covariance matrix even if Σ_Z is separable.

We further assume that Σ_S is a powered-exponential covariance matrix with the covariance function $C_S(\mathbf{h}) = \exp(-\alpha_S \|\mathbf{h}\|^\xi)$ and Σ_T is an exponential covariance matrix with the covariance function $C_T(u) = \exp(-\alpha_T |u|)$ where \mathbf{h} is a spatial lag, u is a temporal lag, $0 < \xi \leq 2, \alpha_S, \alpha_T \geq 0$.

Simulation 1. Let $T = 5$ and $t_i = i, i = 1, \dots, 5$. We generate 500 data sets from the model (2) with $(\alpha_S, \xi, \alpha_T, \alpha, \lambda)^\top = (0.3, 0.8, 0.1, 1.2, 0.16)^\top$. We use $n = 10$ locations that are generated randomly in $[-3, 3]^2 \subset \mathbb{R}^2$. For each data set, we compute maximum likelihood estimates and then calculate the bias and standard deviation for the obtained estimates. We repeat this simulation for data sets with $N = 50, 100, 200$ replicates; see [Table 1](#).

The bias and standard deviation are smaller with a larger sample size, as expected. The average running time on a Core i5-2410M CPU@2.3 GHz is 17 min for a data set with $N = 200$ replicates. The

Table 1

Bias and standard deviation for maximum likelihood estimates in the exponential common factor model with $N = 50, 100, 200$ replicates (based on 500 simulations). The true parameter is $(\alpha_S, \xi, \alpha_T, \alpha, \lambda)^\top = (0.3, 0.8, 0.1, 1.2, 0.16)^\top$.

N	Bias	Standard deviation
50	$(0.03, -0.25, 0.06, -0.31, -0.03)^\top$	$(0.07, 0.06, 0.02, 0.28, 0.07)^\top$
100	$(0.01, -0.19, 0.03, -0.26, -0.03)^\top$	$(0.05, 0.05, 0.01, 0.25, 0.03)^\top$
200	$(0.00, -0.13, 0.02, -0.20, -0.02)^\top$	$(0.03, 0.04, 0.01, 0.17, 0.03)^\top$

Table 2

$\Delta_\rho, |\Delta_\rho|, \Delta_L, |\Delta_L|, \Delta_U, |\Delta_U|$, AIC and BIC values for A1, A2 and A3. Original data are simulated from model (2) with the Pareto common factor.

Model	Δ_ρ	$ \Delta_\rho $	Δ_L	$ \Delta_L $	Δ_U	$ \Delta_U $	AIC	BIC
A1	-0.02	0.04	-0.02	0.10	0.06	0.07	-56 362	-56 346
A2	-0.01	0.04	0.40	0.40	-0.11	0.14	-52 996	-52 986
A3	-0.01	0.04	0.35	0.35	-0.15	0.17	-54 102	-54 089

estimates do not depend on the choice of starting points. We obtained similar results for data sets generated from the model with different sets of parameters $(\alpha_S, \xi, \alpha_T, \alpha, \lambda)^\top$.

Simulation 2. We generate a data set from model (2), however we assume that $\varepsilon_j \sim_{i.i.d.} \text{Pareto}(1, 2)$, that is, $\text{pr}(\varepsilon_j < r) = 1 - r^{-2}$ for $r \geq 1, j = 1, \dots, T$. With a Pareto factor and a time lag of zero, the variables are strongly dependent and, in fact, the upper tail dependence coefficient for the corresponding bivariate copula $\lambda_U = 1$ for any pair of variables, and λ_U does not depend on the distance between any two locations; see Krupskii and Genton (submitted for publication). We again assume $T = 5$ and $t_i = i, i = 1, \dots, 5$. We generate a data set for $(\alpha_S, \xi, \alpha_T, \alpha, \lambda)^\top = (0.3, 0.8, 0.1, 1.0, 0.16)^\top$ with 200 replicates and $n = 20$ spatial locations in $[-3, 3]^2 \subset \mathbb{R}^2$. We calculate maximum likelihood estimates assuming an exponential distribution for the common factor ε_j ; the estimated parameters are $\hat{\theta}_{MLE} = (0.38, 0.66, 0.14, 1.40, 0.13)^\top$.

To assess the goodness of fit of this misspecified model (2), we use the following measures of dependence applied to each pair (U_1, U_2) of multivariate data. We assume $U_i \sim U(0, 1)$, otherwise nonparametric ranks can be used to transform the data into uniform scores.

1. We use Spearman’s correlation $S_\rho = \text{cor}(U_1, U_2)$ to assess the fit in the middle of the distribution;
2. The tail-weighted measures of dependence are: $\varrho_L = \text{cor}\{(1 - 2U_1)^6, (1 - 2U_2)^6 | U_1 < 0.5, U_2 < 0.5\}$, $\varrho_U = \text{cor}\{(1 - 2U_1)^6, (1 - 2U_2)^6 | U_1 > 0.5, U_2 > 0.5\}$. These measures can be used to assess the fit in the lower and upper tails, respectively; see Krupskii and Joe (2015).

We calculate the theoretical values of S_ρ, ϱ_L and ϱ_U for the estimated model. To do this, we simulate 200,000 replicates from model (2) with parameters $\hat{\theta}_{MLE}$ (model A1) and estimate the abovementioned measures of dependence for each pair of variables of the simulated data set. We also compute empirical estimates of these measures for the original data set and then compute the difference between the empirical estimates and the theoretical estimates based on the misspecified model (2) with the exponential common factor. The average (absolute) differences of S_ρ, ϱ_L and ϱ_U for all different pairs of variables are denoted by $\Delta_\rho, \Delta_L, \Delta_U$ ($|\Delta_\rho|, |\Delta_L|, |\Delta_U|$), respectively. We also use a multivariate normal copula and a multivariate Student- t copula with separable covariance matrix Σ_Z (models A2 and A3, respectively) to fit the original data. We then calculate $\Delta_\rho, |\Delta_\rho|, \Delta_L, |\Delta_L|, \Delta_U, |\Delta_U|$ for A2 and A3. The results are presented in Table 2.

Models A1, A2 and A3 have different number of parameters (5, 3 and 4, respectively) and therefore we use the Akaike information criterion (AIC) and Bayesian information criterion (BIC) to compare these models. We can see that the spatio-temporal covariance structure is fitted well by all three models. However, models A2 and A3 fail to fit the data well in the tails. Both multivariate normal and Student- t copulas significantly underestimate the strength of dependence in the upper tail. The latter copula can handle tail dependence; however, this is a symmetric copula and therefore it is not suitable for modeling asymmetric dependence for the original data set. Model A1 fits the data

quite well even in the tails, although the distribution for the common factor is misspecified for this model. The good fit of A1 implies that the proposed model (2) with the exponential common factor can be appropriate for modeling asymmetric dependence and tail dependence for spatio-temporal data, and that the assumption about the distribution of \mathcal{E}_j is not very restrictive. In addition, with the exponential common factor, the likelihood estimation is fairly fast even in high dimensions, and the resulting model (2) has some appealing properties as discussed in Sections 2 and 3.

5.2. Application to wind data

In this section, we apply the proposed model (2) to hourly wind speed data measured at 10 weather stations located in the Netherlands. We use data measured from 9:00 to 23:00 because the average wind speed is higher during this time; we therefore expect dependencies among wind speeds measured at different stations to be stronger. We use wind measurements from February to August 2016, seven months in total, excluding the stormy autumn months because of possibly different wind patterns. For each station with spatial locations $\mathbf{s}_1, \dots, \mathbf{s}_{10}$, we compute the average wind speed measured at hours: 9, 10, 11; 12, 13, 14; 15, 16, 17; 18, 19, 20 and 21, 22, 23, to get five variables, $W_{i,1}, \dots, W_{i,5}, i = 1, \dots, 10$, with strong temporal dependence. We use the proposed copula model (2) to estimate the joint dependence of these variables, at different spatial locations. These variables are equally spaced in time but unequally spaced variables can be modeled as well if needed.

We treat different days as replicates; however, there is a weak temporal dependence between the variables W_{ij} measured on two consecutive days. We therefore use only every other day to remove any temporal dependence, 106 days in total. Let $W_{ij,n}$ be the value of W_{ij} measured at day $n = 1, \dots, 106$. For every location $i = 1, \dots, 10$ and time $j = 1, \dots, 5$, we compute the uniform scores:

$$U_{ij,n} = \frac{\text{rank}(W_{ij,n}) - 0.5}{106}, \quad i = 1, \dots, 10, j = 1, \dots, 5.$$

We transform the uniform scores to normal scores data; Fig. 3 shows bivariate scatter plots for some variables W_{ij} . These scatter plots can be used as a diagnostic tool to detect departures from normality (Nikoloulopoulos et al., 2012). Under the joint normality of the normal scores, the scatter plots should have an elliptical shape. However, we can see in Fig. 3 that the dependence is stronger in the upper tail, so the Gaussian copula may not be suitable for modeling the wind data. Also, the dependence is weaker where there is a larger distance between the stations, as expected for data with spatial dependence.

Some ties can be observed in the data due to rounding errors. However, adding small perturbations to the original data (in order to remove the ties) has no significant effect on the results. Kojadinovic and Yan (2010) showed that the randomization-based approach, when using pseudo-observations by randomly breaking the ties, can give satisfactory results when ties are present in data.

We use the following three models to fit the uniform scores data:

- B1 Gaussian copula (symmetric dependence, no tail dependence);
- B2 Student- t copula with ν degrees of freedom (symmetric dependence, tail dependence);
- B3 Factor copula model based on the process (1).

For all the three models we use the covariance matrix Σ_Z based on the nonseparable spatio-temporal cross-covariance function $\psi(\mathbf{h}, u) = \frac{1}{\theta_T |u|^{2+1}} \exp \left\{ - \left(\frac{\theta_S \|\mathbf{h}\|^2}{\theta_T |u|^{2+1}} \right)^\gamma \right\}$, with $\theta_S, \theta_T > 0$ and $0 < \gamma < 1$. Here, \mathbf{h} is a spatial lag and u is a temporal lag. For B3, we assume $\Lambda(t) = \lambda \geq 0$ (i.e., intensity is a constant) and $\alpha(\mathbf{s}, t) = \alpha_0 + \alpha_1 \text{LAT} + \alpha_2 \text{LON}$ where LAT and LON are spatial coordinates (latitude and longitude, respectively).

We estimate models B1, B2 and B3 using maximum likelihood. To assess the goodness of fit, we calculate the theoretical values of the Spearman’s rho, S_ρ , and the tail-weighted measures of dependence, ϱ_L and ϱ_U , and then compare these estimates with the empirical (non-parametric) estimates of these measures as we did in the previous section. Similarly, we compute $\Delta_\rho, \Delta_L, \Delta_U$ and $|\Delta_\rho|, |\Delta_L|, |\Delta_U|$ for the three models B1, B2 and B3; Table 3 shows the results.

Models B1, B2 and B3 have different number of parameters (3, 4 and 8, respectively) and we use AIC and BIC to compare these models. It is seen that both models B1 and B2 have a very bad fit in the

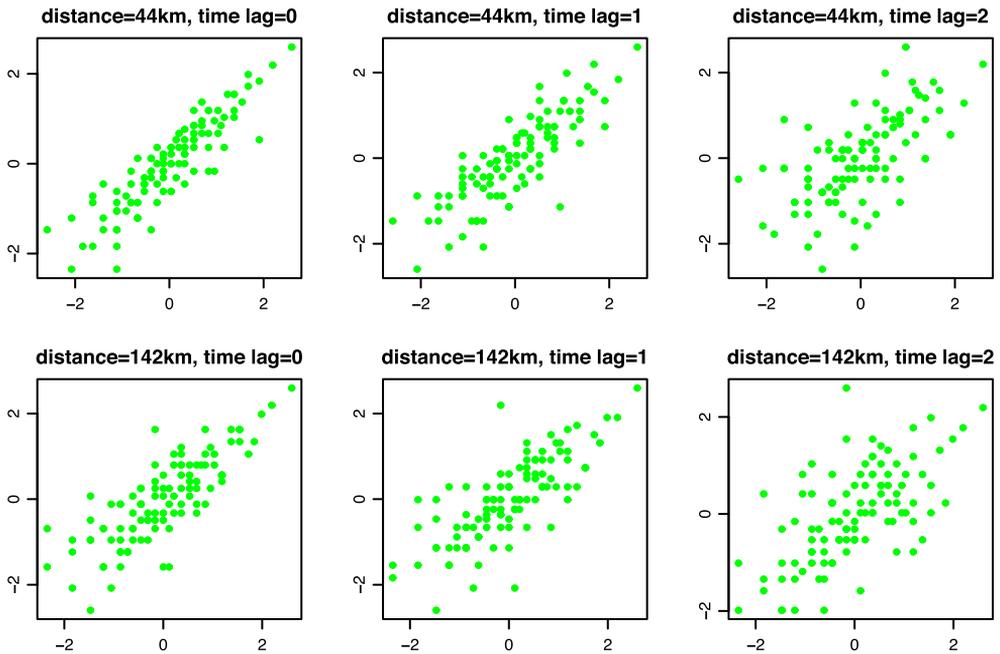


Fig. 3. Normal scores scatter plots for pair (W_{11}, W_{3j}) (top) and (W_{31}, W_{10j}) (bottom) with $j = 1$ (left), $j = 2$ (middle) and $j = 3$ (right).

Table 3

$\Delta_\rho, |\Delta_\rho|, \Delta_L, |\Delta_L|, \Delta_U, |\Delta_U|$, AIC and BIC values for models B1, B2 and B3 applied to the wind data.

Model	Δ_ρ	$ \Delta_\rho $	Δ_L	$ \Delta_L $	Δ_U	$ \Delta_U $	AIC	BIC
B1	0.24	0.24	0.04	0.16	0.33	0.34	−8819	−8811
B2	0.24	0.25	0.01	0.15	0.30	0.30	−8915	−8904
B3	0.04	0.08	0.03	0.15	−0.05	0.11	−9084	−9063

middle and in the upper tails of the joint distribution of the wind data, as indicated by the very large values of Δ_ρ and Δ_U , respectively. A positive sign implies that B1 and B2 significantly underestimate the strength of dependence in the upper tail and the overall dependence. This is because these are symmetric models and they cannot handle data with significant asymmetry. On the other hand, model B3 significantly improves the fit and has the lowest AIC value. The choice of spatio-temporal covariance function $\psi(\mathbf{h}, u)$ does not significantly change the fit of any of these models; in all cases model B3 is significantly better, both in terms of AIC and the goodness of fit.

Finally, we use the morning wind speed measurements, w_{i1} , to compute w_{i2}, \dots, w_{i5} , $i = 1, \dots, 10$. Here, we do not interpolate data but rather do prediction because the data are predicted for different time points. For simplicity, we predict data at the same locations and the predicted values (we use medians) can then be treated as the wind forecast at these particular locations. In general, data can be predicted at different locations and different time points as described in Section 4.4.

For prediction, the marginal distributions need to be estimated. We assume $W_{it} \sim \text{Gamma}(\text{scale} = k_1 + tk_2 + k_3\text{LAT}_i + k_4\text{LON}_i, \text{shape} = k_5 + tk_6)$, where $\text{LAT}_i, \text{LON}_i$ are spatial coordinates of the i th station (latitude and longitude, respectively). Parameter estimates $\hat{k}_1, \dots, \hat{k}_6$ are obtained using the same data as we used to fit the copula models B1, B2 and B3.

For a given day (to do prediction, we select days that were not used to fit the marginal distribution and the copula), the observed values w_{i1} are transformed into uniform data using the estimated

Table 4

δ_T (in 0.1 m per second) and $R\delta_T$ (shown in parentheses), $T = 2, \dots, 5$, for model B1 (based on the Gaussian copula, left) and model B3 (the proposed exponential factor copula model, right). Smaller values are shown in bold font.

Date	Time lags, B1				Time lags, B3			
	1	2	3	4	1	2	3	4
06.03	23(26%)	27(32%)	27(43%)	17(29%)	14(15%)	11(12%)	13(16%)	10(21%)
11.04	29(29%)	47(45%)	30(38%)	51(58%)	13(13%)	20(18%)	9(12%)	19(21%)
21.04	27(31%)	37(41%)	11(20%)	10(38%)	22(25%)	28(30%)	15(36%)	11(53%)
11.05	24(25%)	35(39%)	34(41%)	38(47%)	9(10%)	8(9%)	10(12%)	16(19%)
21.05	15(19%)	17(23%)	7(16%)	11(31%)	7(9%)	11(19%)	19(46%)	18(63%)
04.06	31(29%)	50(45%)	54(52%)	56(59%)	8(7%)	15(13%)	16(15%)	15(16%)
27.08	14(16%)	13(17%)	11(23%)	28(39%)	7(10%)	9(17%)	22(58%)	13(20%)

marginal model:

$$u_{i1} = \Gamma(w_{i1}; \text{scale} = \hat{k}_1 + \hat{k}_2 + \hat{k}_3 \text{LAT}_i + \hat{k}_4 \text{LON}_i, \text{shape} = \hat{k}_5 + \hat{k}_6), \quad (t = 1)$$

and the predicted values on the uniform scale are obtained using the predicted medians: $u_{iT|1} = C_{i,T|1}^{-1}(0.5|u_{1,1}, \dots, u_{10,1}; \hat{\theta}_\alpha, \hat{\theta}_\lambda, \hat{\theta}_\Sigma)$, $i = 1, \dots, 10$, $T = 2, \dots, 5$. The predicted medians are then transformed back to the original scale:

$$w_{iT|1} = \Gamma^{-1}(u_{iT|1}; \text{scale} = \hat{k}_1 + T\hat{k}_2 + \hat{k}_3 \text{LAT}_i + \hat{k}_4 \text{LON}_i, \text{shape} = \hat{k}_5 + T\hat{k}_6) \quad (T = 2, \dots, 5).$$

Here Γ and Γ^{-1} are the Gamma cdf and inverse cdf, respectively.

For $T = 2, \dots, 5$ (corresponding to time lags 1,...,4), we compute the mean absolute errors $\delta_T = 0.1 \sum_{i=1}^{10} |w_{iT|1} - w_{iT}|$ and the mean relative absolute errors $R\delta_T = 10 \sum_{i=1}^{10} |1 - w_{iT|1}/w_{iT}|$, measured in percents, where w_{iT} are the actual (observed) wind data at the i th station and time T . Table 4 shows the mean absolute errors for model B1 (assuming multivariate normal copula) and model B3 (the proposed model) for seven different days with strong winds in the morning (high measured values $w_{11}, \dots, w_{10,1}$).

We see that the proposed model B3 has smaller prediction errors than B1 based on the Gaussian copula, especially for the first two time lags. Model B3 improves the prediction by taking into account the strong spatio-temporal dependence between high wind speed measurements. At the same time, the prediction errors are comparable for B1 and B3 if we select days with weak to moderate winds in the morning. This implies that the classical multivariate normal model can adequately fit the wind data with moderate dependence but fails to account for strong upper tail dependence.

Finally, we compute the predicted medians for the mean wind speed as shown in Section 4.4 for a 25×25 uniform grid in the region located between 51.9° and 53.6° North and between 5.2° and 6.8° East. We select a day with strong winds in the morning, June 4th, and use the morning measurements for the ten stations, $W_{1,1}, \dots, W_{1,10}$, to predict wind speeds at different spatial locations in the afternoon (12–2 pm), corresponding to the time lag equals one. Fig. 4 shows the predicted medians for models B1 and B3.

It is seen that model B1 significantly underestimates the wind speeds and model B3 has smaller prediction errors. The proposed copula model (2) can be used to construct the maps of predicted medians at any time lags and any spatial locations. We tried other days and got similar results with strong winds in the morning. The difference between the two models, B1 and B3, is significantly smaller, when the predicted medians are calculated for days with no strong winds.

6. Discussion

In this paper, we proposed an extension of the factor copula model for replicated spatial data by Krupskii et al. (in press). This extended model can handle data with asymmetric dependence where the dependence is stronger in the joint upper tail. Simulation studies showed that different types of spatio-temporal dependencies can be obtained with this model, and the rate of decay of dependence in time can be controlled using the intensity function, $\lambda(t)$, as well as parameters of the covariance matrix Σ_Z . The likelihood function for the proposed model does not require a numerical multivariate

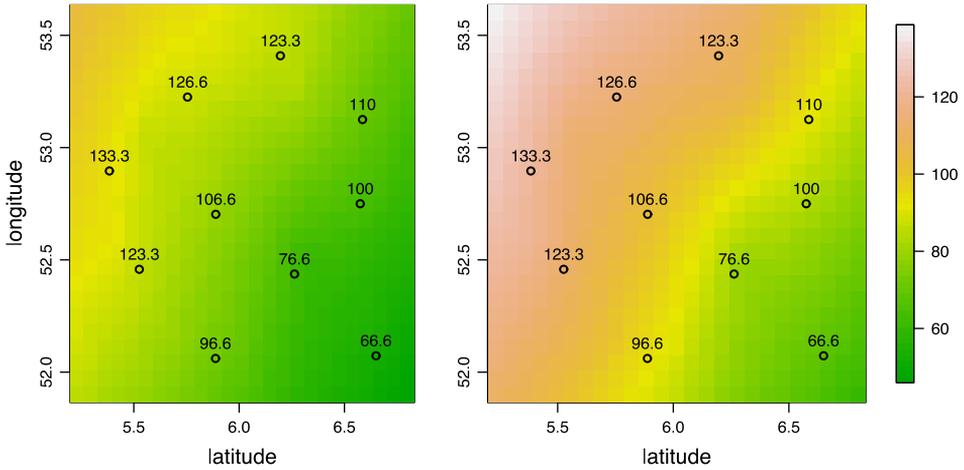


Fig. 4. Predicted medians for model B1 (left) and model B3 (right) for the wind speed data in the area of study (in 0.1 m/s), calculated for the period of 12–2pm on June 4th, 2016. The 10 stations with recorded wind data are shown as circles. The actual speed measurements (in 0.1 m/s), $W_{2,1}, \dots, W_{2,10}$, are shown next to each station.

integration and so estimation is fairly easy unless the number of time lags is very large. This copula model can be used with arbitrary univariate marginals, thus allowing greater flexibility in modeling spatio-temporal data.

Despite its flexibility, the proposed model requires replicates to estimate its parameters if the number of time lags is small. With a large number of time lags, estimation for this model becomes very difficult because of the exponentially growing number of terms to be calculated for the likelihood function. This model is ergodic in time but not in space because the purely spatial model of Krupskii et al. (in press) is not ergodic in space. In geostatistical applications, data often have only one replicate. One direction for future research is therefore to find a copula model for spatio-temporal data that does not require replicates for estimation. Another topic for future research is to define models for multivariate data with spatio-temporal dependence when different variables are measured repeatedly in time and at different spatial locations. Examples include weather data (temperature, pressure, wind speed) measured at different weather stations or concentrations of different pollutants measured by weather balloons launched from different sites.

Acknowledgment

This research was supported by the King Abdullah University of Science and Technology (KAUST).

Appendix

A.1. Proof of Proposition 1

Let $\mathcal{E}_1, \mathcal{E}_2 \sim_{i.i.d.} \text{Exp}(1)$ and let $\mathcal{E}_1, \mathcal{E}_2$ be independent of Z_{1,j_1}, Z_{2,j_2} . For $u_i \in (0, 1)$ define $w_i = F_{i,j_i}^{-1}(u_i)$, $i = 1, 2$. We have:

$$\begin{aligned} C_{1,j_1;2,j_2}^{\mathbf{W}}(u_1, u_2) &= \text{pr}(W_{1,j_1} < w_1, W_{2,j_2} < w_2) \\ &= \text{pr}(Z_{1,j_1} + \alpha_{1,j_1} \mathcal{E}_1 < w_1, Z_{2,j_2} + \alpha_{2,j_2} \mathcal{E}_1 < w_2) \text{pr}\{\mathcal{P}(t_{j_1}) = \mathcal{P}(t_{j_2})\} \\ &\quad + \text{pr}(Z_{1,j_1} + \alpha_{1,j_1} \mathcal{E}_1 < w_1, Z_{2,j_2} + \alpha_{2,j_2} \mathcal{E}_2 < w_2) \text{pr}\{\mathcal{P}(t_{j_1}) < \mathcal{P}(t_{j_2})\} \\ &= \lambda_{j_1,j_2} C_{1,1}^{\mathbf{W}}(u_1, u_2) + (1 - \lambda_{j_1,j_2}) C_{1,2}^{\mathbf{W}}(u_1, u_2), \end{aligned}$$

where $C_{1,k}^W$ is a copula corresponding to the joint distribution of the vector $(Z_{1,j_1} + \alpha_{1,j_1}\varepsilon_1, Z_{2,j_2} + \alpha_{2,j_2}\varepsilon_k)^\top, k = 1, 2$. Krupskii and Genton (submitted for publication) showed that the limiting extreme-value copula for $C_{1,1}^W(u_1, u_2)$ is

$$C_{1,1}^W(u_1, u_2) = \lim_{k \rightarrow 0} C_{1,1}^W(u_1^k, u_2^k)^{1/k} = \text{HR}(u_1, u_2; \vartheta_{\text{HR}}),$$

where

$$\vartheta_{\text{HR}} = \frac{\left\{ \alpha_{1,j_1}^2 + \alpha_{2,j_2}^2 - 2\alpha_{1,j_1}\alpha_{2,j_2}(\Sigma_Z^{j_1,j_2})_{s_1,s_2} \right\}^{1/2}}{\alpha_{1,j_1}\alpha_{2,j_2}}.$$

At the same time, $C_{1,2}^W(u_1, u_2) = \lim_{k \rightarrow 0} C_{1,2}^W(u_1^k, u_2^k)^{1/k} = u_1 u_2$, and

$$\begin{aligned} C_{1,j_1;2,j_2}^W(u_1, u_2) &= \lim_{k \rightarrow 0} \left\{ \lambda_{j_1,j_2} C_{1,1}^W(u_1^k, u_2^k) + (1 - \lambda_{j_1,j_2}) C_{1,2}^W(u_1^k, u_2^k) \right\}^{1/k} \\ &= \text{HR}(u_1, u_2; \vartheta_{\text{HR}}) \lim_{k \rightarrow 0} \left\{ 1 - (1 - \lambda_{j_1,j_2}) C^*(u_1, u_2; k) \right\}^{1/k}, \end{aligned}$$

where $C^*(u_1, u_2; k) = 1 - \frac{C_{1,2}^W(u_1^k, u_2^k)}{C_{1,1}^W(u_1^k, u_2^k)} = -k \ln \left\{ \frac{u_1 u_2}{\text{HR}(u_1, u_2; \vartheta_{\text{HR}})} \right\} + o(k)$ and therefore

$$\begin{aligned} C_{1,j_1;2,j_2}^W(u_1, u_2) &= \text{HR}(u_1, u_2; \vartheta_{\text{HR}}) \lim_{k \rightarrow 0} \left[1 + k(1 - \lambda_{j_1,j_2}) \ln \left\{ \frac{u_1 u_2}{\text{HR}(u_1, u_2; \vartheta_{\text{HR}})} \right\} \right]^{1/k} \\ &= \{ \text{HR}(u_1, u_2; \vartheta_{\text{HR}}) \}^{\lambda_{j_1,j_2}} (u_1 u_2)^{1 - \lambda_{j_1,j_2}}. \quad \square \end{aligned}$$

A.2. Closed-form formula for $I(\mathbf{j})$ from Section 4

Let $m = \max(\mathbf{j})$. It follows that

$$I(\mathbf{j}) = (2\pi)^{-nT/2} |\Sigma_Z|^{-1/2} \int_{\mathbb{R}_+^m} \exp\{h(v_1, \dots, v_m)\} dv_1 \cdots dv_m,$$

where

$$\begin{aligned} h(v_1, \dots, v_m) &= -\frac{1}{2} (\tilde{\mathbf{w}}_1 - \tilde{\alpha}_1 v_1, \dots, \tilde{\mathbf{w}}_m - \tilde{\alpha}_m v_m)^\top \Sigma_Z^{-1} (\tilde{\mathbf{w}}_1 - \tilde{\alpha}_1 v_1, \dots, \tilde{\mathbf{w}}_m - \tilde{\alpha}_m v_m) - \sum_{k=1}^m v_k \\ &= C_0 - \frac{1}{2} (v_1 - v_1^*, \dots, v_m - v_m^*)^\top \Sigma_*^{-1} (v_1 - v_1^*, \dots, v_m - v_m^*). \end{aligned} \quad (4)$$

By equating coefficients of the two quadratic functions of v_1, \dots, v_m in (4), we get:

$$\begin{aligned} \Sigma_*^{-1} &= \mathcal{A}^\top \Sigma_Z^{-1} \mathcal{A}, \quad v^* = \Sigma_* (-I_m + \mathcal{W}^\top \Sigma_Z^{-1} \mathcal{A}) \mathbf{1}_m, \\ C_0 &= -0.5 \mathbf{1}_m^\top (\mathcal{W} - \mathcal{A} v^*)^\top \Sigma_Z^{-1} (\mathcal{W} - \mathcal{A} v^*) \mathbf{1}_m - v_1^* - \dots - v_m^*, \end{aligned}$$

where \mathcal{A} and \mathcal{W} are $nT \times m$ matrices such that the k th column is equal to α_k^* (\mathbf{w}_k^* , respectively) where α_k^* (\mathbf{w}_k^*) is a vector α (\mathbf{w}) with all elements but α_k (\mathbf{w}_k) replaced by zeros. It implies that

$$I(\mathbf{j}) = (2\pi)^{(m-nT)/2} |\Sigma_Z|^{-1/2} |\Sigma_*|^{1/2} \exp(C_0) \Phi_{\Sigma_*}(v_1^*, \dots, v_m^*),$$

so that $I(\mathbf{j})$ has a closed form and no numerical integration is required. □

References

Bárdossy, A., 2006. Copula-based geostatistical models for groundwater quality parameters. *Water Resour. Res.* 42 (11).
 Bárdossy, A., Li, J., 2008. Geostatistical interpolation using copulas. *Water Resour. Res.* 44 (7).
 de Luna, X., Genton, M.G., 2005. Predictive spatio-temporal models for spatially sparse environmental data. *Statist. Sinica* 15, 547–568 Special issue in honor of George Tiao's retirement.

- Erhardt, T.M., Czado, C., Schepsmeier, U., 2015. Spatial composite likelihood inference using local C-vines. *J. Multivariate Anal.* 138, 74–88.
- Fonseca, T.S., Steel, M.F., 2011. Non-Gaussian spatiotemporal modelling through scale mixing. *Biometrika* 98 (4), 761–774.
- Genest, C., Favre, A.C., 2007. Everything you always wanted to know about copula modeling but were afraid to ask. *J. Hydrol. Eng.* 12, 347–368.
- Genton, M.G., Hering, A., 2017. discussion of “Spatiotemporal models for skewed processes”, by A. Schmidt, K. Goncalves, P. Velozo. *Environmetrics* 28, e2430.
- Gneiting, T., 2002. Nonseparable, stationary covariance functions for space-time data. *J. Amer. Statist. Assoc.* 97, 590–600.
- Gneiting, T., Genton, M.G., Guttorp, P., 2007. Geostatistical space-time models, stationarity, separability and full symmetry. In: Finkenstaedt, B., Held, L., Isham, V. (Eds.), *Statistics of Spatio-Temporal Systems*. In: *Monograph in Statistics and Applied Probability*, Chapman & Hall/CRC Press, Boca Raton.
- Gräler, B., 2014. Modelling skewed spatial random fields through the spatial vine copula. *Spat. Stat.* 10, 87–102.
- Hüsler, J., Reiss, R.D., 1989. Maxima of normal random vectors: between independence and complete dependence. *Statist. Probab. Lett.* 7 (2), 283–286.
- Joe, H., 2014. *Dependence Modeling with Copulas*. Chapman & Hall/CRC, Boca Raton, FL.
- Kojadinovic, I., Yan, J., 2010. Modeling multivariate distributions with continuous margins using the copula R package. *J. Statist. Softw.* 34.
- Krupskii, P., Genton, M.G., 2017. A copula-based linear model of coregionalization for non-Gaussian multivariate spatial data, (submitted for publication). <http://arxiv.org/abs/1603.03950>.
- Krupskii, P., Huser, R., Genton, M.G., 2017. Factor copula models for replicated spatial data. *J. Amer. Statist. Assoc.* (in press).
- Krupskii, P., Joe, H., 2013. Factor copula models for multivariate data. *J. Multivariate Anal.* 120, 85–101.
- Krupskii, P., Joe, H., 2015. Tail-weighted measures of dependence. *J. Appl. Stat.* 42 (3), 614–629.
- Nikoloulopoulos, A.K., Joe, H., Li, H., 2012. Vine copulas with asymmetric tail dependence and applications to financial return data. *Comput. Statist. Data Anal.* 56, 3659–3673.
- Patton, A., 2006. Modeling asymmetric exchange rate dependence. *Internat. Econom. Rev.* 47 (2), 527–556.
- Schmidt, A., Concalves, K., Velozo, P., 2017. Spatiotemporal models for skewed processes. *Environmetrics* 28, e2411.
- Segers, J., 2012. Max-stable models for multivariate extremes. *REVSTAT - Statist. J.* 10 (1), 61–82.
- Sklar, A., 1959. Fonctions de répartition à n dimensions et leurs marges. *Inst. Statist. Univ. Paris* 8, 229–231.
- Stein, M.L., 2005. Space-time covariance functions. *J. Amer. Statist. Assoc.* 100, 310–321.
- Stroud, J., Müller, P., Sansó, B., 2011. Dynamic models for spatio-temporal data. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 63, 673–689.