Wang Huixia (Judy) ORCID iD: 0000-0002-5195-8564

Copula-Based Semiparametric Models for Spatio-Temporal Data

Yanlin Tang[1], Huixia Judy Wang[2,*], Ying Sun[3], and Amanda S. Hering[4]

[1]Key Laboratory of Advanced Theory and Application in Statistics and Data Science - MOE, School of Statistics, East China Normal University, Shanghai, 200062, China

[2]Department of Statistics, George Washington University, Washington, DC, 20052, USA

[3]Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal, 23955, Saudi Arabia

[4]Department of Statistical Science, Baylor University, Waco, TX, 76798, USA

email: judywang@gwu.edu

SUMMARY: The joint analysis of spatial and temporal processes poses computational challenges due to the data's high dimensionality. Furthermore, such data are commonly non-Gaussian. In this paper, we introduce a copula-based spatio-temporal model for analyzing spatio-temporal data and propose a semiparametric estimator. The proposed algorithm is computationally simple, since it models the marginal distribution and the spatio-temporal dependence separately. Instead of assuming a parametric distribution, the proposed method models the marginal distributions nonparametrically and thus offers more flexibility. The method also provides a convenient way to construct both point and interval predictions at new times and new locations, based on the estimated conditional quantiles. Through a simulation study and an analysis of wind speeds observed along the border between Oregon and Washington, we show that our method produces more accurate point and interval predictions for skewed data than those based on normality assumptions.

KEY WORDS: Copula; Markov process; Pseudo likelihood; Spatio-temporal.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi:10.1111/biom.13066

1. Introduction

The main challenge in processing spatio-temporal data is accurately modeling the finite-dimensional joint and conditional distributions of the spatio-temporal process. These distributions can be used to describe the spatio-temporal variability of a process, to obtain forecasts at new time points and predictions at new locations, and to provide the prediction uncertainties. The model must appropriately characterize the marginal distribution for a given location while accounting for both spatial and temporal dependencies, which can be computationally challenging even for time series observed at a modest number of locations.

Gaussian processes are commonly used to model spatio-temporal data (Storlie et al., 2017; Ludwig et al., 2017; Datta et al., 2017). However, the Gaussianity assumption is known to be inappropriate or insufficient for representing many real-world datasets. When spatio-temporal processes exhibit non-Gaussian features, a Gaussian process model may be fitted to transformed data. However, direct transformations of the data cannot guarantee Gaussianity and may alter other features of the process. For example, the wind speed data studied in Section 5 have non-Gaussian distributions and remain skewed even after a power transformation (Figure 3). Wallin and Bolin (2015) and Schmidt et al. (2017) illustrated that transforming data with either a square root or a logarithm induces a link between the mean and the covariance, resulting in a non-stationary process. Therefore, transforming the data is not a desirable approach. Real data applications call for non-Gaussian models.

One aim of our work is to provide temporal forecasts and spatial predictions. To make one-step-ahead forecasts, many authors have built separate models for individual sites (Gneiting et al., 2006; Patton, 2012; Kazor and Hering, 2015). Such models cannot be used for predictions at new sites. To make predictions at new locations, the most common approach is to apply a variation of kriging (Krige, 1951; Cressie and Wikle, 2011; Gräler et al., 2016). Kriging is the best linear predictor; however, if the normality assumption fails, then the best predictor may be nonlinear, and thus kriging may no longer be optimal for predictions with a skewed distribution, such as the wind speed data. Various approaches have been proposed to relax the Gaussian assumption; see for instance Zhang and El-Shaarawi (2010), Bai et al. (2014), Quessy et al. (2015), Schmidt et al. (2017), Kang and Song (2017). These methods are either not flexible enough or computationally intensive.

Spatio-temporal predictions of wind speed are particularly challenging because the best performing model may depend on the number of locations available, the spatial scale,

or the frequency of temporal averaging (for example, daily averages versus monthly averages); see some relevant studies in Malmberg et al. (2005), Luo et al. (2008), and Lenzi et al. (2017). Spatio-temporal models are generally complex, and their likelihood-based inference is computationally expensive even for Gaussian process models (Smith and Wilen, 2003; Derado et al., 2010; Berrocal et al., 2012; Lindström et al., 2013; Krupskii and Genton, 2017).

For the analysis of non-Gaussian spatio-temporal data, we propose a new copula-based spatio-temporal (COST) model and a semiparametric maximum pseudo-likelihood estimator. Copula-based estimation methods were also considered in Chen and Fan (2006), Chen et al. (2009), and Rémillard et al. (2012), but these were developed for time series data. Compared to existing methods, the proposed approach has the following advantages. First, the COST framework creates a simplified temporal and spatial dependence structure by separating the modeling of dependence from the marginal distributions, which reduces the computational challenges. Second, instead of assuming a parametric distribution, the proposed method models the marginal distributions nonparametrically, thereby increasing the model's flexibility. Third, it is straightforward to assess the uncertainty using the COST method. By generating multivariate random draws from the conditional prediction distribution and calculating the conditional quantiles, we can use the conditional median (which is more suitable for skewed distributions) for point forecasts/predictions and then use the conditional tail quantiles to construct prediction intervals. Finally, while in theory the COST method requires a large number of locations for spatial predictions, we find that the method performs very well even with a moderate number of irregularly spaced sites, as we demonstrate on a wind speed dataset in Section 5.

The rest of the paper is organized as follows. In Section 2, we introduce the COST model framework and the maximum pseudo-likelihood estimator of the copula parameters. In Section 3, we describe how COST can be used to construct one-step-ahead forecasts and new-location predictions. We assess the finite-sample performance of the COST method through a simulation study in Section 4 and apply the proposed method to an analysis of wind speed data in Section 5. Additional supporting information may be found online in the Supporting Information section at the end of the article.

2. Copula-based semiparametric model for spatio-temporal data

2.1 *COST model*

Let $Y(\mathbf{s},t)$ be a continuous spatio-temporal process of the location $\mathbf{s}\in\mathcal{S}\subset\mathbb{R}^2$ and time $t\in\mathcal{T}\subset\mathbb{N}$, and $\{Y(\mathbf{s}_i,t), i=1,\ldots,d, t=1,\ldots,n\}$ be the observed data generated from this process at d locations and n time points. For notational simplicity, we denote $Y_{t,i}=Y(\mathbf{s}_i,t)$ and $\mathbf{Y}_t=(Y_{t,1},\ldots,Y_{t,d})^T$. We assume that the data are generated from the following copula-based spatio-temporal (COST) model.

(i) First, the model assumes that the process is p-order Markovian in time, which implies that $P(\mathbf{Y}_{t+p}\leqslant\mathbf{y}_{t+p}|\mathbf{Y}_{t+p-1},\ldots,\mathbf{Y}_1)=P(\mathbf{Y}_{t+p}\leqslant\mathbf{y}_{t+p}|\mathbf{Y}_{t+p-1},\ldots,\mathbf{Y}_t)$, that is, the probabilistic properties of the process are completely determined by the joint distribution $H(\mathbf{y}_t,\ldots,\mathbf{y}_{t+p})=P\Big(Y_{t,1}\leqslant y_{t,1},\ldots,Y_{t,d}\leqslant y_{t,d},\ldots,Y_{t+p,1}\leqslant y_{t+p,1},\ldots,Y_{t+p,d}\leqslant y_{t+p,d}\Big)$, where $\mathbf{y}_t=(y_{t,1},\ldots,y_{t,d})^T$.

(ii) Second, by Sklar's theorem (Sklar, 1959), there exists a $\{(p+1)d\}$-dimensional copula C_{ST} such that

$$H(\mathbf{y}_t,\ldots,\mathbf{y}_{t+p})=C_{ST}\Big\{F(y_{t,1}|\mathbf{s}_1),\ldots,F(y_{t,d}|\mathbf{s}_d),\ldots,F(y_{t+p,1}|\mathbf{s}_1),\ldots,F(y_{t+p,d}|\mathbf{s}_d);\mathcal{B}_t,\boldsymbol{\theta}\Big\},\quad(1)$$

where $F(\cdot|\mathbf{s})$ is the marginal cumulative distribution function (CDF) for $\{Y(\mathbf{s},t),t=1,\ldots\}$ at location $\mathbf{s}$, $\mathcal{B}_t=\{\mathbf{s}_1,\ldots,\mathbf{s}_d\}\times\{t,t+1,\ldots,t+p\}$, and $\boldsymbol{\theta}$ is the copula parameter vector.

For ease of presentation, we focus on $p=1$ hereafter, but the proposed method can be adapted for general p with some modification.

Remark 1: The copula in the above COST model is a special case of the spatio-temporal field copula defined in Definition 1 of the Supporting Information S1.1 with $D=d$ and $N=p+1$. The copula links the marginals together to form the joint distribution, and it allows us to model the marginal distributions $F(\cdot|\mathbf{s}_i)$ separately from the spatio-temporal dependence structure, which is expressed in $C_{ST}(\cdot;\mathcal{B}_t,\boldsymbol{\theta})$. By the COST model, data can be generated by first generating $\mathbf{y}_1,\ldots,\mathbf{y}_p$ from $C_{ST}\Big\{F(y_{1,1}|\mathbf{s}_1),\ldots,F(y_{1,d}|\mathbf{s}_d),\ldots,F(y_{p,1}|\mathbf{s}_1),\ldots,F(y_{p,d}|\mathbf{s}_d),\mathbf{1}_d;\mathcal{B}_1,\boldsymbol{\theta}\Big\}$ and then generating $\mathbf{Y}_{t+p}$ sequentially from the conditional distribution $\mathbf{Y}_{t+p}|\mathbf{y}_t,\ldots,\mathbf{y}_{t+p-1}$.

2.2 *Choice of spatio-temporal copula*

Though any multivariate copula can be adopted in the COST model, for our purpose we need choose a copula that is convenient for specifying the spatio-temporal dependence. Two practical choices are the elliptic copula (Rémillard et al., 2012), including Gaussian

and Student t copulas as special cases, and the v-transformed multivariate normal (VTMN) copula (Bárdossy and Li, 2008), which can capture the asymmetry in the tail dependence. For both copulas, the dependence can be specified through a correlation matrix $\mathbf{R}$. Considering such copulas, we can rewrite $C_{ST}(\cdot;\mathcal{B}_t,\boldsymbol{\theta})$ as $C_{ST}(\cdot;\mathbf{R})$, and rewrite (1) for $p=1$ as

$$\begin{aligned} H(\mathbf{y}_1,\mathbf{y}_2) &= C_{ST}\{F(y_{1,1}|\mathbf{s}_1),\ldots,F(y_{1,d}|\mathbf{s}_d),F(y_{2,1}|\mathbf{s}_1),\ldots,F(y_{2,d}|\mathbf{s}_d);\mathbf{R}\} \\ &= \Psi_{2d}\{\Psi_1^{-1}(u_{1,1}),\ldots,\Psi_1^{-1}(u_{1,d}),\Psi_1^{-1}(u_{2,1}),\ldots,\Psi_1^{-1}(u_{2,d});\mathbf{R}\}, \end{aligned}$$

where $u_{t,i}=F(y_{t,i}|\mathbf{s}_i), t=1,2, i=1,\ldots,d$, and Ψ_{2d} and Ψ_1 are the joint and marginal distributions of the $(2d)$-dimensional distribution associated with the copula, respectively. In general, the marginals of Ψ_{2d} need not be the same, but here we assume that they are for simplicity. Using the Gaussian copula as an example, Ψ_{2d} is the joint CDF of a $(2d)$-dimensional normal distribution with unit variance and correlation matrix $\mathbf{R}$, and Ψ_1 is the CDF of $N(0,1)$; more details on the elliptic and VTMN copulas are given in the Supporting Information S1.2.

For the copula function to appropriately capture the spatio-temporal dependence, we must specify the structure of the matrix $\mathbf{R}$. In this paper, we assume second order stationarity so that the correlation between two locations only depends on the difference in their locations, not on the locations themselves. We adopt the non-separable spatio-temporal correlation matrix as in Gneiting (2002), which incorporates possible anisotropy in space (Zimmerman, 1993; Huser et al., 2017). To apply this to our wind speed data analysis, we let h_{ij} denote the scaled Mahalanobis distance between $\mathbf{s}_i$ and $\mathbf{s}_j$; that is, $h_{ij}^2=(\mathbf{s}_i-\mathbf{s}_j)^T\mathbf{V}^{-1}(\mathbf{s}_i-\mathbf{s}_j)$ with

$$\mathbf{V}=\begin{pmatrix}\cos(\zeta) & -\sin(\zeta)\\ \sin(\zeta) & \cos(\zeta)\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & \lambda^{-1}\end{pmatrix}\begin{pmatrix}\cos(\zeta) & -\sin(\zeta)\\ \sin(\zeta) & \cos(\zeta)\end{pmatrix}^T,$$

where $\lambda>0$ is the length ratio of the two principal axes, and $\zeta\in[0,\pi/2]$ is the angle with respect to the west-east direction. When $\lambda=1$, the Mahalanobis distance degenerates to the Euclidean distance. We let $X(\mathbf{s}_i,t)=\Psi_1^{-1}[F\{Y(\mathbf{s}_i,t)|\mathbf{s}_i\}]$, and assume that

$$\mathrm{corr}\{X(\mathbf{s}_i,t),X(\mathbf{s}_j,t+\delta)\}=\frac{1}{(1+b_1|\delta|^{2b_3})^{b_2}}\exp\Big\{-\frac{ch_{ij}^{2\gamma}}{(1+b_1|\delta|^{2b_3})^{b_2\gamma}}\Big\}, \tag{2}$$

where $b_1>0, c>0$ are the scale parameters of time and space, $b_3>0, 0<\gamma<1$ are the smoothness parameters, and $b_2\in(0,1]$ is the space-time interaction parameter. The purely spatial correlation function is of the powered exponential form, and the purely temporal correlation function belongs to the Cauchy class.

Since we consider Markov processes of order $p = 1$, we only need to consider the conditional distribution of $\mathbf{Y}_t$ given $\mathbf{Y}_{t-1}$, that is, $|\delta| = 0, 1$ in (2), where $\delta = 0$ describes the variance of the process, and $\delta = \pm 1$ describes the covariance between adjacent time points. Define $\eta = (1 + b_1)^{b_2} > 1$. The correlation matrix $\mathbf{R}$ can be written as

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{R}_{21} & \mathbf{R}_{22} \end{pmatrix}, \tag{3}$$

where $\mathbf{R}_{11} = \mathbf{R}_{22} = \left(\exp(-ch_{ij}^{2\gamma})\right)_{i,j=1}^{d}$, and $\mathbf{R}_{12} = \mathbf{R}_{21} = \left(\eta^{-1}\exp(-ch_{ij}^{2\gamma}/\eta^{\gamma})\right)_{i,j=1}^{d}$. By Theorem 2 and Example 1 from Gneiting (2002), (2) is a stationary correlation function on $\mathbb{R}^2 \times \mathbb{N}$; that is, $\text{corr}\{X(\mathbf{s}, t), X(\mathbf{s} + \mathbf{h}, t + \delta)\}$ only depends on $\mathbf{h}$ and δ (Gneiting, 2002). Consequently, the correlation matrix defined in (3) is nonnegative definite.

Let $\boldsymbol{\theta}$ be the vector of parameters involved in the copula; for example, $\boldsymbol{\theta} = (\zeta, \lambda, c, \gamma, \eta)$ for the Gaussian copula, and $\boldsymbol{\theta} = (\zeta, \lambda, c, \gamma, \eta, \upsilon)$ for the t copula with degrees of freedom (df) υ. Since $\mathbf{R}$ is determined by $\boldsymbol{\theta}$, we do not distinguish between $\mathbf{R}$ and $\boldsymbol{\theta}$ when the context is clear. The COST framework can easily accommodate nugget effects by introducing some additional parameters in the correlation matrix; see Supporting Information S2.

2.3 *Properties of the COST model*

We present the mixing and stationary properties implied by the COST model and the condition A.1 as below.

A.1 The spatio-temporal process $Y(\mathbf{s}, t)$ is induced by $C_{ST}(\cdot; \mathbf{R})$, where $\mathbf{R}$ is the correlation matrix defined in (3).

Proposition 1: Assuming that the COST model and A.1 hold, we have (i) for elliptic and VTMN copulas, with a given $\mathbf{s}$, $Y(\mathbf{s}, t)$ is stationary and β-mixing with the decay rate $\beta(t) = O\{\exp(-a_1 t)\}$ for some $a_1 > 0$; (ii) for Gaussian and VTMN copulas, with a given t, $Y(\mathbf{s}, t)$ is α-mixing with the mixing coefficient $\alpha(k) = O(k^{-a_2})$ for some $a_2 > 0$ large enough.

Supporting Information S3.1 gives the definitions of different types of mixings, and S3.2 provides the proof of Proposition 1.

Remark 2: In our numerical studies, we focus on elliptic copulas as the computation for VTMN is complicated. For the Gaussian copula, the dependence between two distant locations diminishes if the correlation decays with $\mathbf{h}$. However, the spatial dependence of t-copula does not converge to zero as the distance increases (Røislien and Omre, 2006).

Despite this limitation, we still consider the t copula in our numerical studies for several reasons. First, it is analytically tractable. Second, the rank correlation induced by the t copula decays with $\mathbf{h}$. More specifically, the prediction at new locations/times are based on the conditional t distribution with df increasing with the number of locations, which means that locations far away from those of interest play negligible roles in the prediction, similar to standard kriging. In addition, the t copula contains the Gaussian copula as a special case but provides more flexibility by leaving the df unspecified.

2.4 *Semiparametric maximum pseudo-likelihood estimator of* $\boldsymbol{\theta}$

Let $f(\cdot|\mathbf{s}_i)$ be the density function corresponding to $F(\cdot|\mathbf{s}_i)$. The conditional density of $\mathbf{Y}_t$, given $\mathbf{Y}_{t-1} = \mathbf{y}_{t-1} = (y_{t-1,1}, \ldots, y_{t-1,d})^T$ and evaluated at $\mathbf{y}_t = (y_{t,1}, \ldots, y_{t,d})^T$, is

$$f_{\mathbf{Y}_t|\mathbf{Y}_{t-1}}(\mathbf{y}_t|\mathbf{y}_{t-1}) = \frac{c_{ST}(\mathbf{u}_{t-1}, \mathbf{u}_t; \boldsymbol{\theta})}{c_S(\mathbf{u}_{t-1}; \boldsymbol{\theta})} \prod_{i=1}^{d} f(y_{t,i}|\mathbf{s}_i),$$

where $\mathbf{u}_t = \{F(y_{t,1}|\mathbf{s}_1), \ldots, F(y_{t,d}|\mathbf{s}_d)\}^T$, $c_{ST}(\mathbf{u}, \mathbf{v}; \boldsymbol{\theta}) = \partial^{2d} C_{ST}(\mathbf{u}, \mathbf{v}; \boldsymbol{\theta})/\partial\mathbf{u}\partial\mathbf{v}$, $c_S(\mathbf{u}; \boldsymbol{\theta}) = \partial^d C_S(\mathbf{u}; \boldsymbol{\theta})\partial\mathbf{u}$, and $C_S(\mathbf{u}; \boldsymbol{\theta}) = P(U_1 \leqslant u_1, \ldots, U_d \leqslant u_d, V_1 \leqslant 1, \ldots, V_d \leqslant 1; \boldsymbol{\theta}) = C_{ST}(\mathbf{u}, \mathbf{1}_d; \boldsymbol{\theta})$ is the d-dimensional purely spatial copula.

We propose a two-stage procedure for parameter estimation. In stage 1, at each location $\mathbf{s}_i, i = 1, \ldots, d$, we estimate $F(\cdot|\mathbf{s}_i)$ by the rescaled empirical CDF $\widetilde{F}(y|\mathbf{s}_i) = \frac{1}{n+1}\sum_{t=1}^{n} I(Y_{t,i} \leqslant y)$, and get the pseudo uniform variables $\tilde{\mathbf{u}}_t = \{\widetilde{F}(y_{t,1}|\mathbf{s}_1), \ldots, \widetilde{F}(y_{t,d}|\mathbf{s}_d)\}^T$, $t = 1, \ldots, n$. In stage 2, we construct the pseudo likelihood based on the pseudo variables and define

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \sum_{t=2}^{n} l(\tilde{\mathbf{u}}_{t-1}, \tilde{\mathbf{u}}_t; \boldsymbol{\theta}), \tag{4}$$

where $l(\mathbf{u}, \mathbf{v}; \boldsymbol{\theta}) = \log\{c_{ST}(\mathbf{u}, \mathbf{v}; \boldsymbol{\theta})/c_S(\mathbf{u}; \boldsymbol{\theta})\}$. Following the copula literature (Joe, 1997; Rémillard et al., 2012), we refer to the proposed two-stage estimator $\widehat{\boldsymbol{\theta}}$ as the maximum pseudo-likelihood estimator (MPLE). Under the COST model, condition A.1, and some additional regularity conditions, we can establish the consistency and asymptotic normality of $\widehat{\boldsymbol{\theta}}$ as $n \to \infty$. The additional conditions and proofs can be found in Supporting Information S3.3.

3. Forecasts and predictions

3.1 *Forecasts at future time points*

With the estimation $\widetilde{F}(\cdot|\mathbf{s}_i), i = 1, \ldots, d$, and $\widehat{\boldsymbol{\theta}}$ (consequently $\widehat{\mathbf{R}}$), we can obtain a one-step-ahead forecast at time $n + 1$. For a first-order Markov series, we forecast $\mathbf{Y}_{n+1}$ by

generating random draws from the conditional distribution of $\mathbf{Y}_{n+1}|\mathbf{Y}_n = \mathbf{y}_n$. To simplify notation, denote $F(\cdot|\mathbf{s}_i)$ by F_i, and $\widetilde{F}(\cdot|\mathbf{s}_i)$ by $\widetilde{F}_i$. Below, we describe how to generate from the conditional distribution based on the Gaussian copula, following Algorithm 1 in Supporting Information S4.1; the computing details for the t copula can be found in Supporting Information S4.2.

Define $\mathbf{X}_n = (X_{n,1}, \ldots, X_{n,d})^T$ and $\tilde{\mathbf{x}}_n = (\tilde{x}_{n,1}, \ldots, \tilde{x}_{n,d})^T$, where $X_{n,i} = \Phi^{-1}\{F_i(Y_{n,i})\}$, and $\tilde{x}_{n,i} = \Phi^{-1}\{\widetilde{F}_i(y_{n,i})\}$. We let $\mathbf{x}_{n+1}^{(m)} = \{x_{n+1,1}^{(m)}, \ldots, x_{n+1,d}^{(m)}\}^T, m = 1, \ldots, M$, be random draws from $\mathbf{X}_{n+1}|\mathbf{X}_n = \tilde{\mathbf{x}}_n \sim N(\widehat{\mathbf{B}}\tilde{\mathbf{x}}_n, \widehat{\mathbf{\Omega}})$, where $\widehat{\mathbf{B}} = \widehat{\mathbf{R}}_{21}\widehat{\mathbf{R}}_{11}^{-1}$, $\widehat{\mathbf{\Omega}} = \widehat{\mathbf{R}}_{22} - \widehat{\mathbf{R}}_{21}\widehat{\mathbf{R}}_{11}^{-1}\widehat{\mathbf{R}}_{12}$, then the random draws of $\mathbf{Y}_{n+1}|\mathbf{Y}_n = \mathbf{y}_n$ are generated by $\mathbf{y}_{n+1}^{(m)} = \left(\widetilde{F}_1^{-}[\Phi\{x_{n+1,1}^{(m)}\}], \ldots, \widetilde{F}_d^{-}[\Phi\{x_{n+1,d}^{(m)}\}]\right)^T$ where $\widetilde{F}_i^{-}$ is the generalized inverse of $\widetilde{F}_i$. The q-th estimated quantile of $Y_{n+1,i}|\mathbf{Y}_n = \mathbf{y}_n$ is

$$\widehat{Q}_q(Y_{n+1,i}|\mathbf{y}_n) = \widetilde{F}_i^{-}\left[\Phi\left\{\widehat{\mathbf{R}}_{12,i}^T\widehat{\mathbf{R}}_{11}^{-1}\tilde{\mathbf{x}}_n + \Phi^{-1}(q)\sqrt{1 - \widehat{\mathbf{R}}_{12,i}^T\widehat{\mathbf{R}}_{11}^{-1}\widehat{\mathbf{R}}_{12,i}}\right\}\right],$$

where $\widehat{\mathbf{R}}_{12,i}$ is the i-th column of $\widehat{\mathbf{R}}_{12}$. Thus, at each location i, we can construct a point forecast of $\mathbf{Y}_{n+1}$ by using either the sample mean of the random draws, $M^{-1}\sum_{m=1}^{M}\mathbf{y}_{n+1}^{(m)}$, or the conditional median estimator, $\widehat{Q}_{0.5}(Y_{n+1,i}|\mathbf{y}_n)$, which is more suitable for skewed distributions. We can also construct a 95% forecast interval using $(\widehat{Q}_{0.025}(Y_{n+1,i}|\mathbf{y}_n), \widehat{Q}_{0.975}(Y_{n+1,i}|\mathbf{y}_n))$. The random draws $\mathbf{y}_{n+1}^{(m)}$ are generated as a vector from the conditional distribution of $\mathbf{Y}_{n+1}|\mathbf{y}_n$. Therefore, using the random draws for probabilistic forecasts can appropriately capture the spatial dependence of the forecasts; see numerical evidences given in Section 4.2.

3.2 *Prediction at new locations*

Let $\mathbf{Y}_t(\underline{\mathbf{s}}^*) = (Y_{t,1}^*, \ldots, Y_{t,K}^*)^T$ be the unknown outcomes at K new locations, denoted by $\underline{\mathbf{s}}^* = (\mathbf{s}_1^*, \ldots, \mathbf{s}_K^*)$, at time t. To construct the point and interval predictions, we generate random draws from the conditional distribution of $\mathbf{Y}_t(\underline{\mathbf{s}}^*)|\{\mathbf{Y}_t = \mathbf{y}_t, \mathbf{Y}_{t-1} = \mathbf{y}_{t-1}\}$. We demonstrate with the Gaussian copula; computation details for the t copula are in Supporting Information S4.2.

Define the estimated $(2d + K) \times (2d + K)$ extended correlation matrix as

$$\widetilde{\mathbf{R}} = \begin{pmatrix} \widehat{\mathbf{R}} & \widehat{\mathbf{R}}_* \\ \widehat{\mathbf{R}}_*^T & \widehat{\mathbf{R}}_{**} \end{pmatrix},$$

where $\widehat{\mathbf{R}}_*$ and $\widehat{\mathbf{R}}_{**}$ are the estimated correlation matrices of dimension $2d \times K$ and $K \times K$, constructed similarly as $\widehat{\mathbf{R}}$. For COST with the Gaussian copula, we have

$$\mathbf{X}_t^*|\{\mathbf{X}_{t-1} = \mathbf{x}_{t-1}, \mathbf{X}_t = \mathbf{x}_t\} \sim N\left(\widehat{\mathbf{R}}_*^T\widehat{\mathbf{R}}^{-1}\mathbf{x}_{t-1:t},\ \widehat{\mathbf{R}}_{**} - \widehat{\mathbf{R}}_*^T\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{R}}_*\right),$$

where $\mathbf{x}_{t-1:t} = (\mathbf{x}_{t-1}^T, \mathbf{x}_t^T)^T$. We denote the M random draws from the above multivariate normal distribution as $\mathbf{x}_t^{*(m)}, m = 1, \ldots, M$, and the random draws from $\mathbf{Y}_t(\underline{\mathbf{s}}^*)|\mathbf{Y}_t = \mathbf{y}_t, \mathbf{Y}_{t-1} = \mathbf{y}_{t-1}$ as $\mathbf{y}_t^{*(m)} = \left(\widetilde{F}_1^{*-}[\Phi\{x_{t,1}^{*(m)}\}], \ldots, \widetilde{F}_K^{*-}[\Phi\{x_{t,K}^{*(m)}\}]\right)^T$, where $\widetilde{F}_k^*(\cdot)$ is the estimated marginal distribution of $\{Y_{\mathbf{s}_k^*,t}\}$, and $\widetilde{F}_k^{*-}$ is the corresponding generalized inverse. Then, the point prediction of $\mathbf{Y}_t(\underline{\mathbf{s}}^*)$ at a single time point $t \geqslant 2$ can be constructed by $\widehat{E}\{\mathbf{Y}_t(\underline{\mathbf{s}}^*)|\mathbf{Y}_t = \mathbf{y}_t, \mathbf{Y}_{t-1} = \mathbf{y}_{t-1}\} = M^{-1}\sum_{m=1}^M \mathbf{y}_t^{*(m)}$. For an interval prediction at the new location $\mathbf{s}_k^*$, we derive the q-th conditional quantile of $Y_{t,k}^*|\{\mathbf{Y}_t = \mathbf{y}_t, \mathbf{Y}_{t-1} = \mathbf{y}_{t-1}\}$ as

$$\widehat{Q}_q(Y_{t,k}^*|\mathbf{y}_t, \mathbf{y}_{t-1}) = \widetilde{F}_k^{*-}\Big[\Phi\Big\{\widehat{\mathbf{R}}_{*k}^T\widehat{\mathbf{R}}^{-1}\mathbf{x}_{t-1:t} + \Phi^{-1}(q)\sqrt{1 - \widehat{\mathbf{R}}_{**k}^T\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{R}}_{**k}}\Big\}\Big],$$

where $\widehat{\mathbf{R}}_{*k}, \widehat{\mathbf{R}}_{**k}$ are the k-th columns of $\widehat{\mathbf{R}}_*$ and $\widehat{\mathbf{R}}_{**}$, respectively. The 95% prediction interval can be constructed by $(\widehat{Q}_{0.025}(Y_{t,k}^*|\mathbf{y}_t, \mathbf{y}_{t-1}), \widehat{Q}_{0.975}(Y_{t,k}^*|\mathbf{y}_t, \mathbf{y}_{t-1}))$, and if the data are skewed, we can construct the point prediction using $\widehat{Q}_{0.5}(Y_{t,k}^*|\mathbf{y}_t, \mathbf{y}_{t-1})$. To estimate $F_k^*(\cdot) = F(\cdot|\mathbf{s}_k^*)$, we need to use information from neighboring sites, because no data have been observed at the new locations; see estimation details and the consistency in the Supporting Information S3.4.

4. Simulation study

4.1 *Simulation design and parameter estimation*

The data are generated from the COST model using Algorithm 1 (Supporting Information S4.1) with

$$\mathbf{R}_{11} = \mathbf{R}_{22} = \left(\exp\left(-c_0 h_{ij}^{2\gamma_0}\right)\right)_{i,j=1}^d, \ \mathbf{R}_{12} = \mathbf{R}_{21} = \left(\frac{1}{\eta_0}\exp\left(-\frac{c_0 h_{ij}^{2\gamma_0}}{\eta_0^{\gamma_0}}\right)\right)_{i,j=1}^d,$$

where h_{ij} is the Mahalanobis distance between $\mathbf{s}_i$ and $\mathbf{s}_j$. We consider four cases. In all cases, $(c_0, \gamma_0, \eta_0) = (1, 0.5, 1.5)$. In Cases 1–3, $(\zeta_0, \lambda_0) = (0, 1)$, that is, $h_{ij} = \|\mathbf{s}_i - \mathbf{s}_j\|$, so the true correlation structure is isotropic. In Case 4, $(\zeta_0, \lambda_0) = (0, 5)$, so the correlation structure is anisotropic. The copulas used for generating data in Cases 1–4 are Gaussian, t_3, t_{10}, and t_3, respectively. The time series at each location has a Gamma marginal distribution, with location-dependent shape and scale parameters, $\alpha(\mathbf{s}) = 2s_1 + s_2^2$ and $\beta(\mathbf{s}) = s_1 + s_2$, respectively, where s_1, s_2 are the location coordinates. For each case, we consider 1000 replicates.

To mimic the wind speed data in Section 5, we assume that the data are observed at $d = 9$ fixed locations, $\{\mathbf{s}_1, \ldots, \mathbf{s}_9\} = \{(1/6, 1/6), (1/6, 3/6), (1/6, 5/6), (3/6, 1/6), (3/6, 3/6), (3/6, 5/6), (5/6, 1/6), (5/6, 3/6), (5/6, 5/6)\}$, and we make predictions at

$K = 4$ new locations, $\{\mathbf{s}_1^*, \ldots, \mathbf{s}_4^*\} = \{(1/3, 1/3), (1/3, 2/3), (2/3, 1/3), (2/3, 2/3)\}$. We set $n = 2000$, using a burn-in sequence of length 3000 for stationarity when generating data at $d+K$ locations with a sequence of length $n+1$. We use $\{\mathbf{Y}_t = (Y_{t,1}, \ldots, Y_{t,d})^T, t = 1, \ldots, n\}$ to estimate $\boldsymbol{\theta}$, and use the rest as test data for verifying the forecast of $\mathbf{Y}_{n+1} = (Y_{n+1,1}, \ldots, Y_{n+1,d})^T$ at time $t = n + 1$ and prediction of $\mathbf{Y}_n^* = (Y_{n,1}^*, \ldots, Y_{n,K}^*)^T$ at new locations $\mathbf{s}_1^*, \ldots, \mathbf{s}_K^*$.

We consider two variations of COST: COST_G and COST_t, which use the Gaussian copula and t copula with unknown df, respectively. We also include GP and logGP methods, which assume a Gaussian first-order Markov process on the original and log-transformed data; see Supporting Information S4.3 for more details of GP. For all four methods, we assume an anisotropic correlation structure. For forecasts at future time points, we also include the copula-based semiparametric method from Chen and Fan (2006), denoted as CF, which analyzes the time series at each location separately using the t copula. Table 1 in Supporting Information S4.5 shows that COST_t is more flexible, and it in general gives similar or more efficient estimation of correlation parameters than COST_G.

4.2 *Forecasts at future time points*

We consider the following metrics to assess the forecasting performance of different methods at time $t = n+1$: (i) the mean squared error of the estimator of $E(Y_{n+1,j}|\mathbf{Y}_n) \doteq \mu_j$, that is, $E(\hat{\mu}_j - \mu_j)^2$ for $j = 1, \ldots, 9$; (ii) the empirical coverage percentages (ECPs) and mean lengths (MLs) of 95% forecast intervals; and (iii) the multivariate rank. Let $y_{n+1,j}^{(0)}$ be the observed response at time $n + 1$ and location $\mathbf{s}_j$. The multivariate rank is an assessment tool for probabilistic forecasts of vector-valued quantities (Gneiting et al., 2008). Here it measures the rank of $\mathbf{y}_{n+1} = (y_{n+1,1}^{(0)}, \ldots, y_{n+1,d}^{(0)})^T$, the observed verifying vector at time $n + 1$ among the M random draw from $\mathbf{Y}_{n+1}|\mathbf{Y}_n = \mathbf{y}_n$; see Supporting Information S4.4 for the definition of multivariate rank. If the probabilistic forecasts ensemble the verifying vectors well, the multivariate rank follows a uniform distribution. In Sections 4.2 and 4.3, we report the MSE of the estimated conditional mean instead of the mean forecast errors, $E\{\hat{\mu}_j - y_{n+1,j}^{(0)}\}^2$, because we find that the mean forecast errors are almost always dominated by the conditional variance of $\mathbf{Y}_{n+1}|\mathbf{Y}_n$.

Figure 1 shows that the GP and logGP methods give worse estimation than COST_t, likely due to their misspecified distributions. COST_G performs almost the same as COST_t in Case 1, but it is less efficient when the data are generated from the t copula in Cases 2–3. Taking spatial information into account clearly improves the efficiency of COST_t over its location-specific counterpart method CF. The prediction intervals from all methods have

coverage close to the nominal level. However, COST_t produces shorter intervals than the other methods except in Case 1, where COST_G, COST_t, and CF perform similarly.

[Figure 1 about here.]

Figure 2 presents histograms of the multivariate ranks across 1000 replicates for Case 1; results for Cases 2-4 are in Figure 1 of Supporting Information S4.5. The histograms of the multivariate ranks from the COST_G and COST_t forecasts have nearly uniform distributions, indicating that the random draws from the proposed predictive distribution ensemble the verifying vectors well, even when the copula function is misspecified. The histograms for the CF ranks are far from uniform, as the forecasts are obtained by analyzing each location separately and thus fail to capture the spatial dependence. The GP and logGP histograms also show clear deviation from uniform, reflecting the bias in ensembles obtained from Gaussian/log-Gaussian processes for non-normal data. In practice, it can be difficult to find a unique transformation to make data from all locations normal due to the non-stationarity, making it impossible to predict at new locations.

[Figure 2 about here.]

4.3 *Predictions at new locations*

To estimate the marginal distribution at a new location, we consider the kernel-type estimator with $m_0 = 4$ nearest neighbors, corresponding to setting $h_1 = \sqrt{2}/6$, for all four new locations; the kernel-type estimator can be found in Supporting Information S3.4. For the GP method, we estimate the mean and variance at a new location by averaging the sample means and variances of the four nearest neighbors. We do not include kriging predictions because kriging captures the prediction uncertainty based on the kriging variance, which is only a function of measurement density and spatial configuration of data points, while in COST, uncertainty estimation at the unsampled location is based on the full conditional distribution calculated as a conditional copula (Li et al., 2011). Furthermore, in our numerical analysis, the number of locations is small, so accurately characterizing the spatial dependence structure needed for kriging will be difficult, and with the large number of time points, space-time kriging is computationally infeasible.

Table 1 presents the ECP and ML of the 95% prediction intervals from the COST_t, GP, and logGP methods at four new locations. The prediction intervals from all methods have coverage close to the nominal level. However, the intervals from COST_t are consistently

shorter than those from GP and logGP. This, together with findings in Section 4.2, suggests that the proposed semiparametric method gives more accurate predictions and is more suitable for analyzing non-normal, spatio-temporal data than Gaussian-based methods.

[Table 1 about here.]

5. Analysis of the hourly wind speed data

The data contain the average hourly wind speeds recorded from January 1, 2012 to December 31, 2014 at twenty meteorological towers along the border between Oregon and Washington state, a region with large wind-energy resources. The data are maintained and archived by the Bonneville Power Administration and were presented previously in Kazor and Hering (2015). To minimize the imputation required to compensate for missing data, we focus on ten sites that have less than 3.5% missing data during each of the three years. The missing values were imputed using multiple linear regression of the wind speed on neighboring horizontal and vertical wind components. The dataset contains 26,304 hourly observations for each site.

First, we perform some preprocessing. We transfer the latitude and longitude of the ten sites into a Cartesian coordinate system, assigning the Hood River, OR, site as the origin with 100-kilometer axes. Our preliminary analysis at each site showed that the wind speed distribution is highly non-normal; see Figure 3 for an example. Applying the log transformation aligns the distribution most closely with normality, but it still differs. We apply COST_G and COST_t methods to the untransformed wind speed data, but apply the GP approach to the log-transformed data (denoted as logGP). Following the de-trending method in Zhu et al. (2014), we model and remove the diurnal trend with the average hourly wind speed in a rolling window so the detrended data can be treated as mean-zero processes.

[Figure 3 about here.]

We find that the first-order partial correlation of the time series at each location is around 0.9, and the largest second-order partial correlation is smaller than 0.2. Therefore, for computational convenience, we assume a first-order Markov process when fitting the data for the COST_G, COST_t and logGP methods. For the one-hour-ahead forecasts, we also include the CF method, which is the separate time series analysis method of Chen and Fan (2006). For all methods, we update the parameter estimation progressively in a rolling-window format: we incorporate the year 2014 data as it becomes available, with

a window length of 90 days, resulting in data vectors featuring n=2,160 time points, and update the parameters in each method every 24 hours. For COST_t, the estimated df ranges from 15.9 to 28.7, with mean 21.3 and standard deviation 3.0.

For short-term wind speed forecasting, Kazor and Hering (2015) proposed a regime-switching (RS) forecasting method, where the regimes are identified by a Gaussian mixture model (GMM). Within each regime, a multiple regression model is built with variables selected from among four lags for wind speed, and three lags for the cosine and sine of the wind direction across multiple locations. Compared to our proposed model that relies on only one regime and one lag for wind speed, the RS model is much more complex and data adaptive; thus, it provides excellent results for one-hour-ahead forecasts. Therefore, we use the forecast results from the RS method as a gold standard, where the median of the predictive distribution from the chosen regime is used for point forecasts.

Table 2 summarizes the mean absolute error (MAE) of different methods for one-hour-ahead point forecasts, based on hourly averages from year 2014, and the ECP and ML of the 95% prediction intervals. In terms of point forecasting, COST_G, COST_t and CF methods all perform comparably to the RS method and better than logGP, and COST_t is slightly better than COST_G. The forecast intervals from all four methods have coverage probabilities close to 95%, but those from logGP are generally wider. The multivariate rank histograms in Figure 4 indicate that the ensembles generated from both COST_G and COST_t capture the features of the observed spatio-temporal data well, but those from logGP and CF are either biased due to the violation of normality or failure to capture the spatial dependence.

[Table 2 about here.]

[Figure 4 about here.]

We now assess prediction performance at the Hood River testing site. The RS and CF methods are not included as they cannot be used for prediction at new sites. In terms of point prediction, the MAE of the logGP predictions is slightly smaller than COST_G and COST_t. However, the prediction intervals from logGP have much lower coverage probabilities and larger widths (ECP=73.8%, ML=15.39) than those of COST_G (ECP=94.8%, ML=14.83) and COST_t (ECP=94.9%, ML=14.48). The prediction intervals from COST_t are slightly narrower than those from COST_G.

6. Discussion

In our numerical studies, we considered Gaussian and t copulas, which are tail-symmetric copulas. The dependence in some studies may be asymmetric in the upper and lower tails. To increase the modeling flexibility, our future research may draw on ideas from spatial models developed to handle tail asymmetry, namely the factor copula models (Hua et al., 2017; Krupskii et al., 2018), vine models (Erhardt et al., 2015), and Tukey g-and-h random fields (Xu and Genton, 2017). Another alternative is to adopt asymmetric copulas in our proposed procedure, for instance, the skewed t-copula considered in Yoshiba (2018), which is a useful extension of the t copula. Extending these methods to analyze spatio-temporal data and conduct probabilistic predictions require further investigation both numerically and theoretically.

Acknowledgements

The authors would like to thank two reviewers, an associate editor, and the editor for constructive comments and helpful suggestions. The work is partially supported by King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research award OSR-2015-CRG4-2582, the National Science Foundation grant DMS-1712760, the IR/D program from the National Science Foundation, National Natural Science Foundation of China grants 11871376 and 11801355, Shanghai Pujiang Program 18PJ1409800, and Key Laboratory for Applied Statistics of MOE, Northeast Normal University 130028849. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Bai, Y., Kang, J., and Song, P. X. (2014). Efficient pairwise composite likelihood estimation for spatial-clustered data. *Biometrics* **70,** 661–670.

Bárdossy, A. and Li, J. (2008). Geostatistical interpolation using copulas. *Water Resources Research* **44,** W07412.

Berrocal, V., Gelfand, A., and Holland, D. (2012). Space-time data fusion under error in computer model output: An application to modeling air quality. *Biometrics* **68,** 837–848.

Chen, X. and Fan, Y. (2006). Estimation of copula-based semiparametric time series models. *Journal of Econometrics* **130,** 307–335.

Chen, X., Wu, W., and Yi, Y. (2009). Efficient estimation of copula-based semiparametric markov models. *The Annals of Statistics* **37,** 4214–4253.

Cressie, N. and Wikle, C. K. (2011). *Statistics for Spatio-Temporal Data*. John Wiley & Sons, Inc., Hoboken, NJ.

Datta, A., Banerjee, S., Finley, A., Hamm, N., and Schaap, M. (2017). Non-separable dynamic nearest neighbor gaussian process models for large spatio-temporal data with application to particulate matter analysis. *Annals of Applied Statistics* **10,** 1286–1316.

Derado, G., Bowman, D., and Kilts, C. (2010). Modeling the spatial and temporal dependence in fmri data. *Biometrics* **66,** 949–957.

Erhardt, T., Czado, C., and Schepsmeier, U. (2015). R-vine models for spatial time series with an application to daily mean temperature. *Biometrics* **71,** 323–332.

Gneiting, T. (2002). Nonseparable, stationary covariance functions for space-time data. *Journal of American Statistical Association* **97,** 590–600.

Gneiting, T., Larsen, K., Westrick, K., Genton, M., and Aldrich, E. (2006). Calibrated probabilistic forecasting at the stateline wind energy center: The regime-switching space-time method. *Journal of American Statistical Association* **101,** 968–979.

Gneiting, T., Stanberry, L., Grimit, E., Held, L., and Johnson, N. (2008). Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds (with discussion). *Test* **17,** 211–235.

Gräler, B., Pebesma, E., and Heuvelink, G. (2016). Spatio-temporal interpolation using gstat. *R Journal* **8,** 204–218.

Hua, L., Xia, M., and Basu, S. (2017). Factor copula approaches for assessing spatially dependent high-dimensional risks. *North American Actuarial Journal* **21,** 147–160.

Huser, R., Opitz, T., and Thibaud, E. (2017). Bridging asymptotic independence and dependence in spatial extremes using gaussian scale mixtures. *Spatial Statistics* **21,** 166–186.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman and Hall, London.

Kang, J. and Song, P. X. (2017). Composite likelihood estimation in copula models for longitudinal imaging data. Technical report, Department of Biostatistics, University of Michigan.

Kazor, K. and Hering, A. S. (2015). The role of regimes in short-term wind speed forecasting at multiple wind farms. *Stat* **4,** 271–290.

Krige, D. (1951). A statistical approach to some mine valuations and allied problems

on the witwatersrand. Master's thesis, University of Witwatersrand.

Krupskii, P. and Genton, M. G. (2017). Factor copula models for data with spatio-temporal dependence. *Spatial Statistics* **22,** 180–195.

Krupskii, P., Huser, R., and Genton, M. G. (2018). Factor copula models for replicated spatial data. *Journal of American Statistical Association* **113,** 467–479.

Lenzi, A., Pinson, P., Clemmensen, L. H., and Guillot, G. (2017). Spatial models for probabilistic prediction of wind power with application to annual-average and high temporal resolution data. *Stochastic Environmental Research and Risk Assessment* **31,** 1615–1631.

Li, J., Bardossy, A., Guenni, L., and Liu, M. (2011). A copula based observation network design approach. *Environmental Modelling & Software* **26,** 1349–1357.

Lindström, A., Green, M., Paulson, G., Smith, H., and Devictor, V. (2013). Rapid changes in bird community composition at multiple temporal and spatial scales in response to recent climate change. *Ecography* **36,** 313–322.

Ludwig, G., Chu, T., Zhu, J., Wang, H., and Koehler, K. (2017). Static and roving sensor data fusion for spatio-temporal hazard mapping with application to occupational exposure assessment. *The Annals of Applied Statistics* **11,** 139–160.

Luo, W., Taylor, M., and Parker, S. (2008). A comparison of spatial interpolation methods to estimate continuous wind speed surfaces using irregularly distributed data from england and wales. *International Journal of Climatology* **9,** 947–959.

Malmberg, A., Holst, U., and Holst, J. (2005). Forecasting near-surface ocean winds with Kalman filter techniques. *Ocean Engineering* **32,** 273–291.

Patton, A. (2012). A review of copula models for economic time series. *Journal of Multivariate Analysis* **110,** 4–18.

Quessy, J., Rivest, L., and Toupin, M. (2015). Semi-parametric pairwise inference methods in spatial models based on copulas. *Spatial Statistics* **14,** 472–490.

Rémillard, B., Papageorgiou, N., and Soustra, F. (2012). Copula-based semiparametric models for multivariate time series. *Journal of Multivariate Analysis* **110,** 30–42.

Røislien, J. and Omre, H. (2006). t-distributed random fields: A parametric model for heavy-tailed well-log data. *Mathematical Geology* **38,** 821–849.

Schmidt, A. M., Gonçalves, K. C. M., and Velozo, P. L. (2017). Spatiotemporal models for skewed processes. *Environmetrics* **28,** e2411.

Sklar, A. (1959). *Fonctions de Répartition À N Dimensions Et Leurs Marges*. Université Paris 8.

Smith, M. and Wilen, J. (2003). Economic impacts of marine reserves: the importance

of spatial behavior. *Journal of Environmental Economics and Management* **46,** 183–206.

Storlie, C. B., Reich, B. J., Rust, W. N., Ticknor, L. O., Bonnie, A. M., Montoya, A. J., and Michalak, S. E. (2017). Spatiotemporal modeling of node temperatures in supercomputers. *Journal of the American Statistical Association* **112,** 92–108.

Wallin, J. and Bolin, D. (2015). Geostatistical modelling using non-gaussian matérn fields. *Scandinavian Journal of Statistics* **42,** 872–890.

Xu, G. and Genton, M. G. (2017). Tukey g-and-h random fields. *Journal of American Statistical Association* **112,** 467–479.

Yoshiba, T. (2018). Maximum likelihood estimation of skew-t copulas with its applications to stock returns. *Journal of Statistical Computation and Simulation* **88,** 2489–2506.

Zhang, H. and El-Shaarawi, A. (2010). On spatial skew-Gaussian processes and applications. *Environmetrics* **21,** 33–47.

Zhu, X., Genton, M., Gu, Y., and Xie, L. (2014). Space-time wind speed forecasting for improved power system dispatch. *Test* **23,** 1–25.

Zimmerman, D. L. (1993). Another look at anisotropy in geostatistics. *Mathematical Geology* **25,** 453–470.

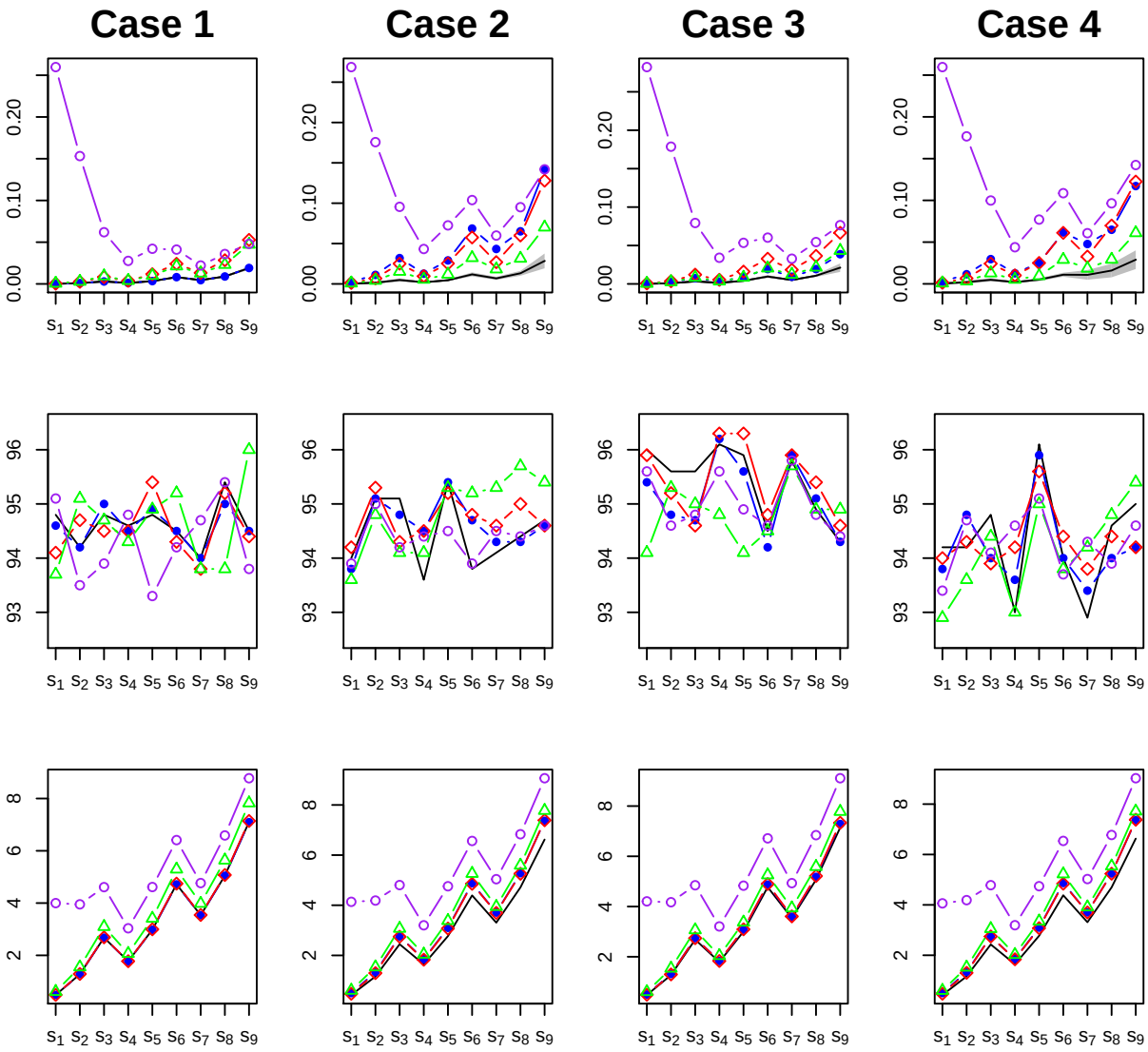

Figure 1. The mean squared error of $\widehat{E}(Y_{n+1,j}|\mathbf{Y}_n)$ at nine locations (top row), and the empirical coverage percentages (middle row) and mean lengths (bottom row) of 95% forecast intervals. COST_G: line with solid circle (blue), the proposed method assuming a Gaussian copula; COST_t: solid (black), the proposed method assuming a t copula; GP: line with triangle (green), the method assuming a Gaussian process; logGP: line with open circle (purple), the method assuming a Gaussian process after log transformation; CF: line with diamond (red), a separate time series analysis based on Chen and Fan (2006) with a t copula. The shaded areas in the first row correspond to the 95% pointwise confidence band for the mean squared error of the COST_t estimator. This figure appears in color in the electronic version of this article, and color refers to that version.

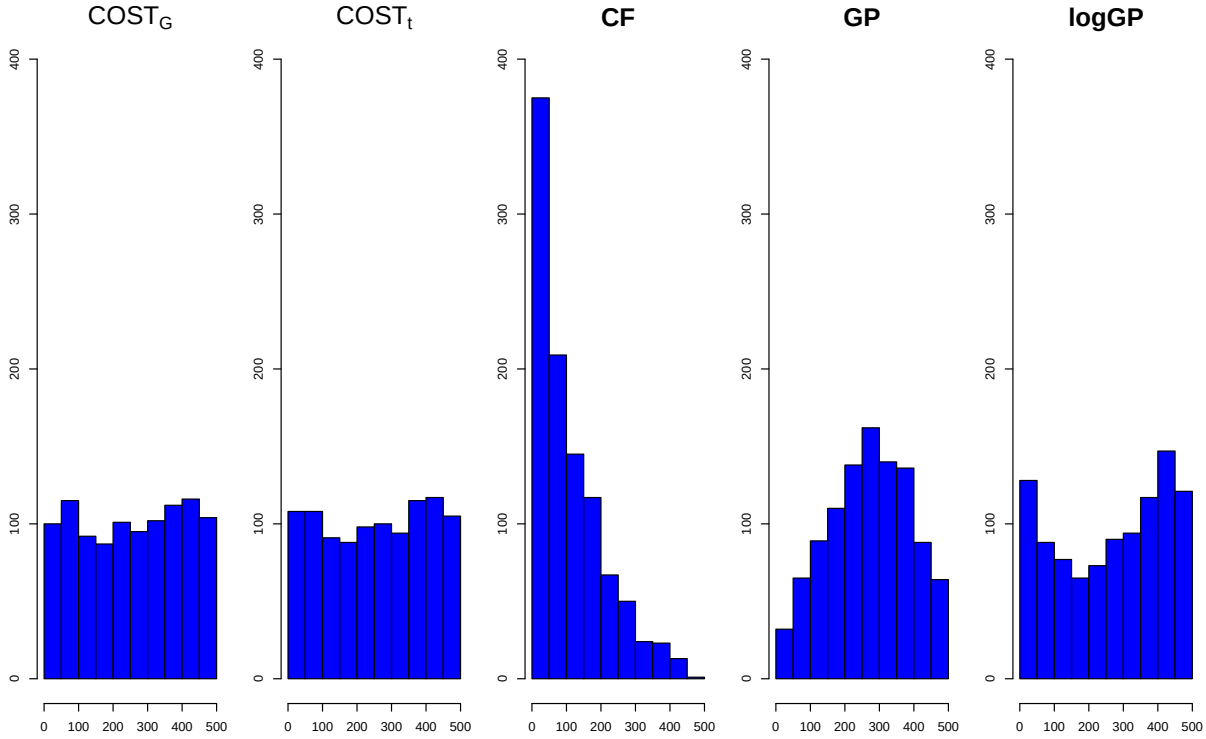

Figure 2. Histograms of multivariate ranks from Case 1. COST$_G$ and COST$_t$: the proposed method assuming Gaussian and t copulas, respectively; CF: a separate time series analysis based on Chen and Fan (2006) with a t copula; GP: the method assuming a Gaussian process; logGP: the method assuming a Gaussian process after log transformation. This figure appears in color in the electronic version of this article.

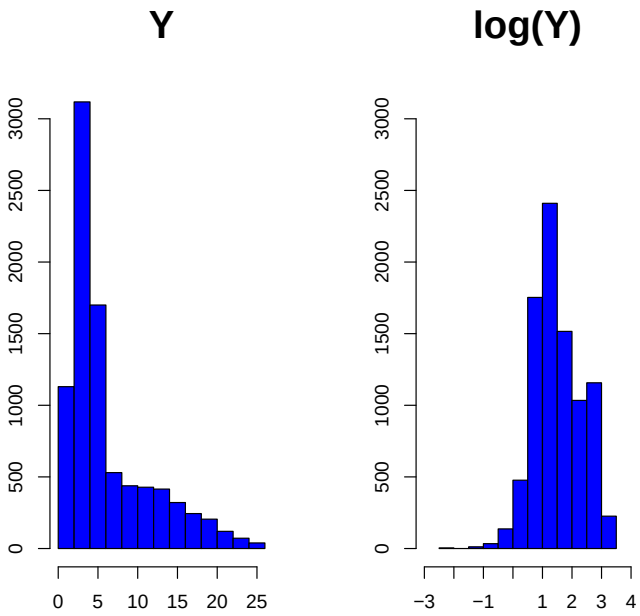

Figure 3. Histograms of the untransformed and log-transformed hourly wind speeds observed at the Biddle Butte site in 2014. This figure appears in color in the electronic version of this article.

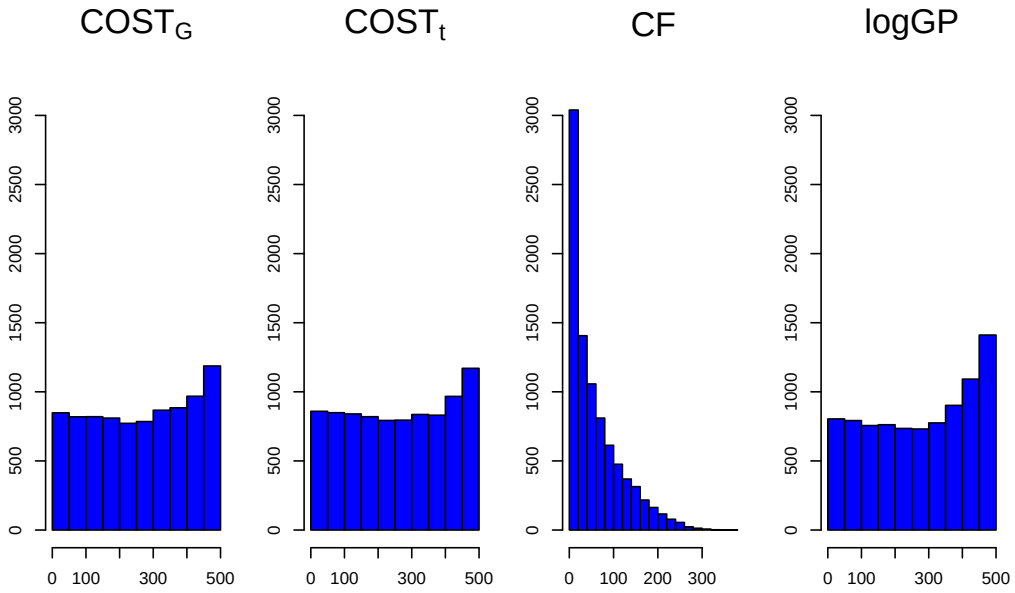

Figure 4. Histograms of multivariate ranks of nine sites over 8760 hours determined by the rolling-window method from the wind speed data analysis; for each hour, 500 ensembles were used to calculate the multivariate rank. COST_G: the proposed method with a Gaussian copula; COST_t: the proposed method with a t copula; CF: a separate time series forecast based on Chen & Fan (2006) with a t copula; logGP: the method based on a Gaussian process assumption, applied to the log-transformed data. This figure appears in color in the electronic version of this article.

Table 1

Empirical coverage percentage (ECP) and the mean length (ML) of the 95% prediction interval at four new locations.

Cases	Method	$\mathbf{s}_1^*$		$\mathbf{s}_2^*$		$\mathbf{s}_3^*$		$\mathbf{s}_4^*$	
		ECP	ML	ECP	ML	ECP	ML	ECP	ML
1	COST_t	95.5	1.23	96.7	2.09	96.8	2.29	96.3	3.33
	GP	94.1	1.29	94.4	2.15	95.9	2.38	95.1	3.45
	logGP	94.6	2.31	96.7	2.81	94.5	2.70	96.5	3.74
2	COST_t	93.8	1.12	96.3	1.92	96.7	2.14	96.1	3.10
	GP	93.2	1.27	94.1	2.12	95.1	2.34	93.8	3.40
	logGP	93.8	2.29	95.6	2.93	94.1	2.70	93.3	3.80
3	COST_t	94.6	1.15	97.0	2.01	95.9	2.23	98.0	3.27
	GP	94.7	1.27	94.4	2.12	95.3	2.35	94.6	3.40
	logGP	93.5	2.27	94.3	2.89	95.9	2.63	95.6	3.75
4	COST_t	94.7	1.50	97.6	2.55	97.4	2.86	96.7	4.12
	GP	95.3	1.65	94.7	2.75	95.1	3.05	93.9	4.41
	logGP	95.4	3.58	95.7	4.16	94.4	3.66	93.8	5.15

COST_t: the proposed method assuming a t copula; GP: the method assuming a Gaussian process; logGP: the method assuming a Gaussian process after log transformation.

Table 2

The mean absolute error (MAE) and corresponding standard error, empirical coverage percentages (ECPs), and mean lengths (MLs) of 95% forecast intervals, for the one-hour-ahead 2014 wind forecasts from different methods.

		Forecast site								
Method	Metric	BB	FG	HH	ME	NR	RO	SH	SU	TI
COST_G	MAE	0.908	0.502	0.979	0.888	0.915	1.206	1.041	1.499	0.655
	SE	(0.010)	(0.005)	(0.011)	(0.009)	(0.009)	(0.012)	(0.011)	(0.015)	(0.007)
COST_t	MAE	0.895	0.500	0.977	0.882	0.908	1.204	1.038	1.494	0.654
	SE	(0.010)	(0.005)	(0.011)	(0.009)	(0.009)	(0.012)	(0.011)	(0.015)	(0.007)
CF	MAE	0.841	0.512	1.000	0.881	0.877	1.194	1.049	1.485	0.679
	SE	(0.009)	(0.005)	(0.011)	(0.009)	(0.009)	(0.012)	(0.011)	(0.015)	(0.007)
logGP	MAE	1.109	0.507	1.027	0.936	0.983	1.337	1.061	1.585	0.688
	SE	(0.013)	(0.005)	(0.012)	(0.009)	(0.010)	(0.013)	(0.011)	(0.016)	(0.008)
RS	MAE	0.739	0.496	0.934	0.839	0.819	1.133	1.024	1.433	0.652
COST_G	ECP	95.0	91.1	92.9	95.7	97.8	96.0	91.7	95.6	90.4
	ML	7.33	2.52	6.02	5.40	6.75	8.81	5.09	9.26	3.46
COST_t	ECP	94.6	90.8	92.9	95.4	97.6	95.9	91.3	95.4	90.2
	ML	7.30	2.50	5.97	5.38	6.71	8.76	5.04	9.17	3.42
CF	ECP	93.2	93.7	93.4	93.9	93.6	93.0	93.3	93.1	93.8
	ML	5.92	2.82	6.09	4.77	4.77	7.25	5.49	7.95	4.05
logGP	ECP	96.2	91.1	93.4	95.2	97.3	94.9	92.4	94.5	89.4
	ML	9.40	2.88	6.87	7.10	8.69	13.57	6.41	12.91	3.99

COST_G: the proposed method with a Gaussian copula; COST_t: the proposed method with a t copula; CF: a separate time series forecast based on Chen and Fan (2006) with a t copula; logGP: the method based on a Gaussian process assumption, applied to the log-transformed data; RS: the GMM.r.s method from Kazor and Hering (2015). Values in parentheses are the corresponding standard errors of the mean absolute errors; the standard errors of RS were not reported in Kazor and Hering (2015). BB: Biddle Butte; FG: Forest Grove; HH: Horse Heaven; ME: Megler; NR: Naselle Ridge; RO: Roosevelt; SH: Shaniko; SU: Sunnyside; TI: Tillamook.