my journal manuscript No.
(will be inserted by the editor)

Student's-t process with spatial deformation for spatio-temporal data

Fidel Ernesto Castro Morales · **Dimitris N. Politis** · **Jacek Leskow** · **Marina Silva Paez**

Received: date / Accepted: date

Abstract Many models for environmental data that are observed in time and space have been proposed in the literature. The main objective of these models is usually to make predictions in time and to perform interpolations in space. Realistic predictions and interpolations are obtained when the process and its variability are well represented through a model that takes into consideration its peculiarities. In this paper, we propose a spatio-temporal model to handle observations that come from distributions with heavy tails and for which the assumption of isotropy is not realistic. As a natural choice for a heavy-tailed model, we take a Student's-t distribution. The Student's-t distribution, while being symmetric, provides greater flexibility in modeling data with kurtosis and shape different from the Gaussian distribution. We handle anisotropy through a spatial deformation method. Under this approach, the original geographic space of observations gets mapped into a new space where isotropy holds. Our main result is, therefore, an anisotropic model based on the heavy-tailed t distribution. Bayesian approach and the use of MCMC enable us to sample from the posterior distribution of the model parameters. In Section 2, we discuss the main properties of the proposed model. In section 3 we present a simulation study, showing its superiority over the traditional isotropic Gaussian model. In Section 4, we show the motivation that has led us to propose the t distribution-based anisotropic model – the real dataset of evaporation coming from the Rio Grande do Sul state of Brazil.

Keywords Student's-t process · spatio-temporal modeling · spatial deformation · Markov Chain Monte Carlo · heavy tails.

1 Introduction

In the previous half of a century, space-time models have been extensively explored, one of the main goals being the exploration of environmental phenomena. The leading applications include: spatial and spatio-temporal variation of the atmospheric pollution, the moisture content in the soil, relationships between the presence of certain

Address(es) of author(s) should be given

diseases and pollutants, and rainfall estimation. Model based approaches have been proposed to deal with this kind of data not only for the understanding of the phenomena themselves (including exploiting the influence the other variables might exert on them), but also for the purpose of performing predictions in time and interpolation in space.

A common approach to handle spatial correlation is by including in the model a spatial random effect (Kang and Cressie 2011). The usual choice is to select as a prior distribution a Gaussian process with a covariance function that is isotropic so that such covariance depends only on the distance between locations. Isotropic models are useful and very popular due to their ease in estimating model parameters. In the era of global warming, however, the climate models based on gaussianity have to be replaced by more realistic ones that reflect the heavy-tailed nature of climate. We now observe much more extreme behavior and atypical observations. The effects of El Niño and La Niña, for example, can cause significant changes in temperature and precipitation regimes depending on the location and time interval in which they operate (Ropelewski and Halpert 1987; Marengo et al. 2017).

Another example of this scenario is generated by the climatic phenomena known as cold fronts and heatwaves. When the objective is to study the temperature in places where, in a given period, these two climatic phenomena may occasionally occur, lower temperatures (produced by cold fronts) or higher temperatures (produced by heatwaves) will be observed influencing the average temperature of the region. An example of this case appears in South America, where the region's temperature can be modified by the occurrence of cold fronts from the south pole (Reboita et al. 2010). On the other hand, climatic and meteorological phenomena, topography, geographic location, proximity to the ocean, and the time interval to be considered in the study can also shape the covariance structure of some environmental variables. Several studies in the literature exemplify this situation, e.g., Morales and Vicini (2020) conclude that the spatial covariance structure of extreme rainfall frequency in the States of Maranhão and Píau, in Brazil, is anisotropic due mainly to the congruence of three types of climatic systems in this region. Other examples of anisotropic processes in this context can be seen in Sampson and Guttorp (1992); Damian et al. (2001); Schmidt and O'Hagan (2003); Bruno et al. (2008); Schmidt et al. (2011); Morales et al. (2013) and Fouedjio et al. (2015).

The main motivation for this work came from the spatiotemporal analysis of evaporation in the state of Rio Grande do Sul, Brazil. The dataset is measured through time in 11 monitoring stations, and our main goal is to propose a methodology that provides robust interpolation for points in the region of interest. In the preliminary descriptive analysis, we observed that the evaporation time series has extremely low and high measurements compared to the average value of the time series (see Figure 8 in Appendix A). Therefore, to assume that the data generating process has Gaussian distribution seems to be neither adequate nor realistic. Working under the normality assumption would be too restrictive, and the model would suffer from the lack of robustness in the presence of outliers. Another aspect that we consider in the analysis of these data is to use an anisotropic correlation function that allows to include the complex iteration of climatic phenomena (e.g., frontal systems that move from the Pacific Ocean, cyclones and cold fronts (Reboita et al. 2010)), topography and geographic location in this region. Therefore, our goal is to provide a working and relatively simple model that will be usable for the datasets with the presence of outliers from anisotropic spatial processes with heavy tails distribution. As a natural choice for a heavy-tailed model, we take a Student's-t distribution. The Student's-t distribution, while being symmetric, provides greater flexibility in modeling data with kurtosis and shape different from the Gaussian distribution. When the number of degrees of freedom in a Student's-t is large enough, it gets close to the normal one. Estimating the number of degrees of freedom one can tell if it is justifiable to drop the normality assumption in favor of a Student's-t process. We would recommend the choice of a Student's-t when its degrees of freedom are less than 30. In geostatistical research, the Student's-t distribution has already gained popularity, for example, in the case of modelling the soybean yield (see e.e. Assumpcao et al. (2011); do Prado et al. (2013) and Schemmer et al. (2017)).

In our paper, we propose a model that can be represented as a sum of two components. The first is a mean process which varies smoothly in time, while the second is a purely spatial component. The mean process incorporates time variation through a state space approach (West and Harrison 1997). The model assumes spatial dependency through the specification of a spatial correlation function for the purely spatial component, and it handles anisotropy through spatial deformation (Sampson and Guttorp 1992). Under this approach, the original geographic space of observations gets mapped into a new space where isotropy holds. Working under the Bayesian approach to inference, we propose efficient MCMC methods to sample from the posterior distribution of the unknown parameters in the model. Also, extra steps can be added to the algorithm to perform time forecast and interpolation in space. The idea of how to perform interpolation under the proposed anisotropic construction is simple: firstly a grid of points is built in the original space, then they are mapped into the deformed space, and the interpolation if performed there under the hypotheses of isotropy.

Our paper is organized as follows: Section 2 presents the proposed Student's-t variate distribution for non-homogeneous, heavy tailed, anisotropic space-time processes. This section also explains the inference method, the computational aspects, and the interpolation procedure. Section 3 presents a simulation study aiming to measure performance in terms of accuracy of parameter estimation, goodness-of-fit, and interpolation of the proposed model compared to simpler models. In Section 4, the spatio-temporal model is applied to the evaporation dataset. Section 5 presents conclusions and a discussion of possible extensions of our research.

2 Student's-t space-time models

Consider a geographic region of interest denoted by $G \subset \mathbb{R}^p$, and suppose that a random process is observed at T distinct moments in time and at n fixed geographic locations in G. Denote these locations by $\mathbf{s}_1, \ldots, \mathbf{s}_n$. Under the typical assumption that $p = 2$ we can define $\mathbf{s}_i = (x_i, y_i)'$, where x_i and y_i represent coordinates in the two-dimensional space.

We assume that the observations $\mathbf{Y}_t = (Y(\mathbf{s}_1,t),\ldots,Y(\mathbf{s}_n,t))', t = 1,\ldots,T$, can be explained by a sum of two independent components:

$$Y(\mathbf{s}_i,t) = \mu(\mathbf{s}_i,t) + W(\mathbf{s}_i,t),\ i = 1,\ldots,n,\ t = 1,\ldots,T,$$

with $\mu(\cdot,\cdot)$ representing the mean of the process and $W(\cdot,\cdot)$ representing deviations from the mean. In matrix notation, the random vector of observations $\mathbf{Y}$ can be written as

$$\mathbf{Y} = \mu + \mathbf{W},$$

where μ and $\mathbf{W}$ are $n \times T$ matrices defined as $\mu = (\mu_1,\ldots,\mu_T)$, with $\mu_t = (\mu(\mathbf{s}_1,t),\ldots,\mu(\mathbf{s}_n,t))'$; and $\mathbf{W} = (\mathbf{W}_1,\ldots,\mathbf{W}_T)$, with $\mathbf{W}_t = (W(\mathbf{s}_1,t),\ldots,W(\mathbf{s}_n,t))'$.

Different specifications could be used for the mean μ. One simple approach would be to consider a deterministic variation of the process in time and space through a linear combination of explanatory variables, in the form:

$$\mu(\mathbf{s}_i,t) = \mathbf{X}'_{i,t}\beta,$$

where $\mathbf{X}'_{i,t}$ is a vector of explanatory variables observed for the location $\mathbf{s}_i$ and time t, and β are the corresponding regression coefficients. A more flexible approach can incorporate space and time variation in the regression coefficients, such that

$$\mu(\mathbf{s}_i,t) = \mathbf{X}'_{i,t}\beta_{i,t},$$

possibly incorporating smooth variation in time and space in the coefficients β.

In this paper, we assume a stochastic variation in time through a state space model formulation (West and Harrison 1997) where we assume temporal variation in the regression coefficients. More specifically, we have that:

$$\text{Structural equation: } \mu(\mathbf{s}_i,t) = \mathbf{F}'_{it}\beta_t, \tag{1}$$
$$\text{System equation: } \beta_t = \mathbf{G}_t\beta_{t-1} + \mathbf{h}_t, \mathbf{h}_t \sim N(\mathbf{0},\mathbf{H}),$$

where $\mathbf{F}_{it}$ is a known vector which can include covariates, and $\mathbf{G}_t$ is a known transition matrix.

We assume that the elements $\mathbf{W_t}$ of the component $\mathbf{W}$ are independent, for $t = 1,\ldots,T$, and all the temporal dependence is incorporated into the mean process. We then assume a Student's-t process with smooth variation in space for each $\mathbf{W_t}$. We propose the use of the Student's-t distribution as it is a heavy tailed distribution which can be easily specified by its mean vector and covariance matrix, making it simple to perform model interpolation and forecast as will be presented later on this paper. Other alternative heavy tailed distributions could be considered, as the family of elliptic distributions proposed by De Bastiani et al. (2015). Note that the Student's-t distribution is a member of this family. We then have that:

$$\mathbf{W_t} \sim t_\nu(\mathbf{0}, \Sigma),$$

where $t_\nu(\mathbf{a}, \mathbf{A})$ denotes the multivariate Student's-t distribution with mean vector $\mathbf{a}$, covariance matrix $\mathbf{A}$ and ν degrees of freedom; $\mathbf{0}$ indicates a vector of 0's of appropriate size; Σ is defined as $\Sigma = \sigma_\xi^2 \mathbf{R}(\phi) + \sigma_\varepsilon^2 \mathbf{I}_n$, with $\mathbf{I}_n$ representing the identity matrix of order n. $\mathbf{R}(\phi)$ is a correlation matrix, such that $\mathbf{R}(\phi)[i,j] = \rho_\phi(\mathbf{s}_i, \mathbf{s}_j)$ for $i, j = 1, \ldots, n$, where ρ_ϕ is a valid correlation function parametrized by ϕ, which incorporates spatial dependence in the model. The variance parameter σ_ε^2 represents a nugget effect, while σ_ξ represents the variability of the component which incorporates the spatial dependence. More details on the multivariate t distribution can be seen in Roth (2013).

Another way of representing $\mathbf{W}_t$, which helps simplifying the algorithms that will be proposed for inference later on, is through the Normal-Gamma representation (see for example Carlin et al. (1992) and Chib and Ramamurthy (2014)) including a latent vector $\mathbf{U} = (U_1, \ldots, U_T)$, such that $\mathbf{W}_t = U_t^{-1/2} \mathbf{Z}_t$, in which $\mathbf{Z}_t \overset{iid}{\sim} N_n(\mathbf{0}, \Sigma)$ and $U_t \overset{iid}{\sim} G(\nu/2, \nu/2)$. Here, $N_n(\mathbf{a}, \mathbf{A})$ denotes the multivariate Normal distribution of dimension n with mean vector $\mathbf{a}$ and covariance matrix $\mathbf{A}$; and $G(\alpha, \beta)$ indicates the Gamma distribution with mean α/β and variance α/β^2. That way, we can write:

$$\mathbf{Y}_t = \mu_t + U_t^{-1/2} \mathbf{Z}_t, \qquad t = 1, \ldots, T.$$

To handle anisotropy, we follow the idea proposed by Sampson and Guttorp (1992), that considers a function $\mathbf{d}(\cdot)$ to map the original geographic coordinates from space G to the new space $D \subset \mathbb{R}^q$, $q \geq p$ (we assume $q = 2$), where the hypothesis of isotropy hold. That way, for each geographic location $\mathbf{s}_i$ in $G \subset \mathbb{R}^2$ there is a location $\mathbf{d}(\mathbf{s}_i) = \mathbf{d}_i = (d_{x_i}, d_{y_i})'$ in the new deformed space $D \subset \mathbb{R}^2$, $i = 1, \ldots, n$.

Denote by $\mathbf{S} = (\mathbf{s}_1, \ldots, \mathbf{s}_n)'$ the matrix of geographic coordinates in the original space with dimension $n \times 2$ and $\mathbf{d}^G = (\mathbf{d}_1, \ldots, \mathbf{d}_n)'$ the corresponding matrix on the deformed space D with the same dimension. Under isotropy at the new space D, $\rho_\phi(\mathbf{s}_i, \mathbf{s}_j)$ will only depend on $\mathbf{s}_i$ and $\mathbf{s}_j$ through the Euclidean distance between their corresponding locations in the deformed space D, i.e. $\mathbf{d}_i$ and $\mathbf{d}_j$. Particularly, we work with the exponential correlation function $\rho_\phi(\mathbf{s}_i, \mathbf{s}_j) = e^{-\phi|\mathbf{d}_i - \mathbf{d}_j|}$ and wave correlation function $\rho_\phi(\mathbf{s}_i, \mathbf{s}_j) = \frac{\phi \sin(|\mathbf{d}_i - \mathbf{d}_j|/\phi)}{|\mathbf{d}_i - \mathbf{d}_j|}$, but any other valid spatial correlation function could have been specified.

The models are completed with prior specifications for the unknown parameters, which will be described in detail in Section 2.1.

2.1 Prior specifications

Under the Bayesian point of view, to complete the model specification, we must specify prior distributions for all unknown model parameters. Here the unknown model parameters are the geographic coordinates $\mathbf{d}(\cdot)$ in D-space; the Student's-t degrees of freedom ν; the variance of the deformation σ_d^2, the variance of the purely spatial component σ_ξ^2; the variance of the error term σ_ε^2; the spatial correlation function parameter ϕ; the regression coefficients β; and the covariance matrix of the system equation $\mathbf{H}$. The specified prior distributions for these parameters will be presented below. Note that since the stochastic description of β is completely done in the model, they do not require further prior specification except for the prior specification of β_0.

- Coordinates in D-space

  We assume for $\mathbf{d}(\cdot)$ the prior specification proposed by Schmidt and O'Hagan (2003), given by

$$\mathbf{d}(\cdot) \sim GP(\mathbf{g}(\cdot), \sigma_\mathbf{d}^2 \rho_\psi(\cdot)),$$

  which is a Gaussian Process where $\mathbf{g}(\cdot)$ is the mean function, and $\sigma_\mathbf{d}^2 \rho_\psi(\cdot)$ is a covariance function given by the multiplication of a covariance matrix $\sigma_\mathbf{d}^2$, and a scalar function $\rho_\psi(\cdot)$. The mean function $\mathbf{g}(\cdot)$ typically depends on the locations $\mathbf{s}$. Particularly, we assume the identity function $\mathbf{g}(\mathbf{s}) = \mathbf{s}$. We assume $\sigma_\mathbf{d}^2$ to be a 2×2 diagonal matrix, which controls the variance of the deviation in each coordinate from the original space. Finally, $\rho_\psi(\cdot)$ is defined as Gaussian correlation function, such that $\rho_\psi(x) = \exp\{-\psi x^2\}$. It is not an easy task to estimate ψ, and so we will consider it to be known. Reasonable approximations for this parameter were proposed by Morales et al. (2013) and Schmidt and O'Hagan (2003), and they are given respectively by $\psi = -2\log(0.05)/max_{i,j=1,\ldots,n}(\mid s_i - s_j \mid)^2$, and $\psi = 1/(2 \times max_{i,j=1,\ldots,n}(\mid s_i - s_j \mid)^2)$.

  Through properties of the Gaussian process it is easy to see that $\mathbf{d}^G$ will follow a Normal matrix-variate distribution with mean $\mathbf{S}$, row covariance matrix $\sigma_\mathbf{d}^2$, and column covariance matrix $\mathbf{R}_d$, with $\mathbf{R}_d$ being the matrix whose elements are obtained through the Gaussian correlation function $\rho_\psi(\cdot)$. That way, the prior density probability function for the parameter $\mathbf{d}^G$ given $\sigma_\mathbf{d}^2$ is:

$$\pi(\mathbf{d}^G \mid \sigma_\mathbf{d}^2) \propto \mid \sigma_\mathbf{d}^2 \mid^{-n/2} \exp\left\{-\frac{1}{2} tr[(\mathbf{d}^G - \mathbf{S})'(\sigma_\mathbf{d}^2)^{-1}(\mathbf{d}^G - \mathbf{S})\mathbf{R}_d^{-1}]\right\}.$$

- Prior distribution for the degrees of freedom parameter ν

  The prior distribution for parameter ν is the same proposed by Cabral et al. (2012), that is:

$$\nu|\lambda \sim G(1, \lambda)$$
$$\lambda \sim U(a_\lambda, b_\lambda)$$

  with $U(a,b)$, where $0 < a < b < 1$, denoting the Uniform distribution in the interval (a,b).

– Prior distribution for the parameters σ^2_ξ, σ^2_ε and ϕ

The choice of prior distribution for the parameters σ^2_ξ, σ^2_ε and ϕ must be done carefully, as improper prior distributions for these parameters can lead to improper posterior distributions. That way, it is recommended to use informative priors for them. Thus, we adopt the prior distribution proposed by Schmidt and Gelfand (2003); Morales et al. (2013) for ϕ, that is: $\phi \sim G(a_\phi \eta, \eta)$, where $a_\phi = -2\log(0.05)/max(\mid \mathbf{s}_i - \mathbf{s}_j \mid)$.

The prior distributions for the variances σ^2_ξ, σ^2_ε are based on an alternative representation proposed by Yan et al. (2007) for matrix Σ. They propose writing

$$\Sigma = \sigma^2[(1-\kappa)\mathbf{R}(\phi) + \kappa \mathbf{I}_n],$$

such that $\sigma^2 = \sigma^2_\xi + \sigma^2_\varepsilon$ and $\kappa = \frac{\sigma^2_\xi}{\sigma^2}$. Thus, κ represents the fraction of the total variability of $\mathbf{Y}$ that corresponds to the measurement error. One advantage of this parametrization is that κ is limited. The prior distributions of σ^2 and κ proposed by Yan et al. (2007) are

$$\pi(\kappa) \propto \frac{\kappa^{a_\xi - 1}(1-\kappa)^{a_\varepsilon - 1}}{[b_\xi \kappa + b_\varepsilon(1-\kappa)]^{a_\xi + a_\varepsilon}}, \ \kappa \in (0,1),$$
$$\sigma^2 \mid \kappa \sim GI\left(a_\varepsilon + a_\xi, \frac{b_\xi}{1-\kappa} + \frac{b_\varepsilon}{\kappa}\right).$$

These prior distributions are induced by the prior distributions for σ^2_ξ and σ^2_ε, given by $\sigma^2_\xi \sim IG(a_\xi, b_\xi)$ and $\sigma^2_\varepsilon \sim IG(a_\varepsilon, b_\varepsilon)$, where $IG(a_{IG}, b_{IG})$ denotes the inverse Gamma distribution with mean $b_{IG}/(a_{IG}-1)$.

– Prior distribution for the parameter β_0

For the regression coefficient β_0 we assign a normal conjugate prior, given by

$$\beta_0 \sim N(\mathbf{m}_0, \mathbf{C}_0).$$

– Prior distribution for the covariance matrix $\mathbf{H}$

We assume a Inverted Wishart prior distribution for the covariance matrix $\mathbf{H}$, written as:

$$\mathbf{H} \sim WI^{-1}_{n_0}(\mathbf{S}_0).$$

We assume that the prior specifications given above are independent, and define $\theta = \{\beta, \nu, \lambda, \sigma^2, \kappa, \phi, \mathbf{d}^G, \sigma^2_{\mathbf{d}}, \mathbf{H}\}$ as the complete set of unknown model parameters. Thus, the prior distribution for θ can be written as:

$$\pi(\theta) = \pi(\beta)\pi(\nu \mid \lambda)\pi(\lambda)\pi(\sigma^2)\pi(\kappa)\pi(\phi)\pi(\mathbf{d}^G \mid \sigma^2_{\mathbf{d}})\pi(\sigma^2_{\mathbf{d}})\pi(\mathbf{H}). \quad (2)$$

2.2 Inference under the proposed spatial model

The likelihood function of θ given the observed processes $\{\mathbf{y} : \mathbf{y}_t,\, t = 1,\ldots,T\}$ can be written as:

$$L(\theta \mid \mathbf{y}) = \prod_{t=1}^{T} \frac{\Gamma(\frac{p+\nu}{2})}{\Gamma(\frac{\nu}{2})\pi^{p/2}} \nu^{-p/2} \mid \Sigma \mid^{-1/2} \left(1 + \frac{d(\mathbf{y}_t,\theta)}{\nu}\right)^{-\frac{p+\nu}{2}}, \tag{3}$$

where $d(\mathbf{y_t},\theta) = (\mathbf{y_t} - \mu_\mathbf{t})'\Sigma^{-1}(\mathbf{y_t} - \mu_\mathbf{t})$. Note that the augmented likelihood function including $\mathbf{U}$ presents a simpler form, as $(\mathbf{y_t}|\mathbf{U_t},\theta)$ follows a normal distribution. The joint likelihood of $(\mathbf{U},\theta)$ is given by

$$L(\mathbf{U},\theta \mid \mathbf{y}) \propto \prod_{t=1}^{T} \mid \mathbf{Q}_t \mid^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{y}_t - \mu_t)^T \mathbf{Q}_t^{-1}(\mathbf{y}_t - \mu_t)\right\}, \tag{4}$$

where $\mathbf{Q}_t = \mathbf{U}_t^{1/2}\Sigma_u \mathbf{U}_t^{1/2^T}$.

It is interesting as well to obtain the marginal distribution of the observed response in each time and location. We have that $Y(\mathbf{s}_i,t) = \mu(\mathbf{s}_i,t) + W(\mathbf{s}_i,t)$, and, under our specification in (2), considering that $\beta_0 \sim N(\mathbf{m}_0,\mathbf{C}_0)$, $\mu(\mathbf{s}_i,t)$ follows a normal distribution. After a little algebra we obtain that

$$E(\mu(\mathbf{s}_i,t)) = \mathbf{F}'_{it}\mathbf{G}^t\mathbf{m}_0,$$
$$V(\mu(\mathbf{s}_i,t)) = (\mathbf{F}'_{it}\mathbf{G}^t)\mathbf{C}_0(\mathbf{F}'_{it}\mathbf{G}^t)^T + \sum_{i=0}^{t-1}(\mathbf{F}'_{it}\mathbf{G}_t^i)\mathbf{H}(\mathbf{F}'_{it}\mathbf{G}_t^i)^T.$$

On the other hand, $\mathbf{W}_t = (W(\mathbf{s}_1,t),W(\mathbf{s}_2,t),\ldots,W(\mathbf{s}_n,t))^T \sim t_\nu(\mathbf{0},\Sigma)$. The marginal distribution of $W(\mathbf{s}_i,t)$ is a zero mean Student's-t distribution with ν degrees of freedom and variance given by $\nu/(\nu-2)\Sigma_i$, where Σ_i is the i^{th} element of the diagonal of matrix Σ. Therefore, the marginal distribution of $Y(\mathbf{s}_i,t)$ is a sum of a normal distribution and a Student's-t distribution. This is a symmetric distribution centered around its mean and with tails heavier than a normal distribution. The sum gets close to the normal distribution then the variance of the normal part is large when compared to the variance of the t part. Therefore, our method shows better performance comparing to a normal model when the variability of the observations in space is not much smaller than the variability in time.

Combining the likelihood function in (4) with the prior distribution of θ defined in (2), and the distribution of $\mathbf{U}|\theta$ we obtain the posterior distribution given by

$$\pi(\mathbf{U},\theta \mid \mathbf{y}) \propto L(\theta \mid \mathbf{y},\mathbf{U})\pi(\mathbf{U}|\theta)\pi(\theta). \tag{5}$$

The posterior distribution in (5) does not have a closed form so the MCMC methods are proposed to obtain samples of this distribution.

The proposed algorithms combine Gibbs sampling (when the full conditional is simple to sample from) and Metropolis-Hastings steps (Metropolis et al. 1953). In the next section, we present details of the implementation of the MCMC algorithm proposed for the approximation of (5).

2.3 Approximate Bayesian computation

We propose a MCMC algorithm based on Gibbs sampling and Metropolis-Hastings steps to generate from the posterior distribution of the model parameters. We consider the idea of data augmentation (Kuo and Yang 1996; Neal and Kypraios 2015; Chib and Ramamurthy 2014) under the Normal-Gamma representation of $\mathbf{W}_t$. Under the proposed model, the full conditional distributions for the parameters β, $\mathbf{H}$, σ^2 and $\mathbf{U}$ have a closed form and these parameters can be sampled through Gibbs Sampling. Their full conditionals are given below, where $\Theta_{-\theta}$ denotes the vector of parameters Θ excluding parameter θ:

- $\beta_0 \mid \mathbf{y}, \theta_{-\beta_0} \sim N(\mathbf{A}_0, \mathbf{B}_0)$,
  with $\mathbf{B}_0 = (\mathbf{G}_1' \Psi^{-1} \mathbf{G}_1 + \mathbf{C}_0^{-1})^{-1}$ and $\mathbf{A}_0 = \mathbf{B}_0(\mathbf{C}_0^{-1}\mathbf{m}_0 + \mathbf{G}_1'\Psi^{-1}\beta_1)$;

- $\beta_t \mid \mathbf{y}, \theta_{-\beta_t}, U_t \sim N(\mathbf{A}_t, \mathbf{B}_t)$,
  with $\mathbf{B}_t = (\Psi^{-1} + \mathbf{F}_t'\Sigma_t^{-1}\mathbf{F}_t + \mathbf{G}_{t+1}'\Psi^{-1}\mathbf{G}_{t+1})^{-1}$ and $\mathbf{A}_t = \mathbf{B}_t(\mathbf{F}_t'\Sigma_t^{-1}\mathbf{y}_t + \Psi^{-1}\mathbf{G}_t\beta_{t-1} + \mathbf{G}_{t+1}\Psi^{-1}\beta_{t+1})$,
  $t = 1, \ldots, T-1$, where $F_t = (F_{1t}, F_{2t}, \cdots, F_{nt})'$;

- $\beta_T \mid \mathbf{y}, \theta_{-\beta_T}, U_t \sim N(\mathbf{A}_T, \mathbf{B}_T)$,
  with $\mathbf{B}_T = (\Psi^{-1} + \mathbf{F}_T\Sigma_t^{-1}\mathbf{F}_T)^{-1}$ and $\mathbf{A}_T = \mathbf{B}_T(\Psi^{-1}\mathbf{G}_T\beta_{T-1} + \mathbf{F}_T'\Sigma_t^{-1}\mathbf{y}_T)$;

- $\mathbf{H} \mid \theta_{-\Psi} \sim WI_{n_0^*}^{-1}(\mathbf{S}_0^*)$,
  with $n_0^* = n_0 + T$ and $\mathbf{S}_0^* = \frac{1}{n_0^*}(\sum_{t=1}^T(\beta_t - \mathbf{G}_t\beta_{t-1})(\beta_t - \mathbf{G}_t\beta_{t-1})' + n_0\mathbf{S}_0)$;

- $\sigma^2 \mid \mathbf{y}, \theta_{-\sigma^2}, \mathbf{U} \sim GI(a_\sigma, b_\sigma)$,
  with $a_\sigma = nT/2 + a_\varepsilon + a_\xi$ and $b_\sigma = \frac{1}{2}\sum_{i=1}^T(\mathbf{y}_t - \mu_t)'(U_t^{-1}\Delta)^{-1}(\mathbf{y}_t - \mu_t) + \frac{b_\xi}{1-\kappa} + \frac{b_\varepsilon}{\kappa}$, where
  $\Delta = (1-\kappa)\mathbf{R}(\phi) + \kappa\mathbf{I}_n$;

- $U_t \mid \mathbf{y}, \theta \sim G(a_U, b_U)$,
  with $a_U = \frac{nT+\nu}{2}$ and $b_U = \frac{1}{2}[(\mathbf{y}_t - \mu_t)'\Sigma^{-1}(\mathbf{y}_t - \mu_t) + \nu]$, with $\Sigma_t = U_t^{-1}\Sigma$, for $t = 1, \ldots, T$.

The full conditional distributions for the parameters κ, ϕ, ν, and $\mathbf{d}^G$ do not have a closed form and they are sampled through Metropolis-Hastings steps. Their full conditionals are given below:

$$\pi(\kappa \mid \mathbf{y}, \theta_{-\kappa}, \mathbf{U}) \propto \mid \Delta \mid^{-T/2} \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}(\mathbf{y}_t-\mu_t)'\Sigma_t^{-1}(\mathbf{y}_t-\mu_t) - (\frac{b_\varepsilon}{1-\kappa}+\frac{b_\varepsilon}{\kappa})\sigma^2\right\} \frac{\kappa^{a_\xi-1}(1-\kappa)^{a_\varepsilon-1}}{[b_\xi\kappa+b_\varepsilon(1-\kappa)]^{a_\xi+a_\varepsilon}};$$

$$\pi(\phi \mid \mathbf{y}, \theta_{-\phi}, \mathbf{U}) \propto \mid \Delta \mid^{-T/2} \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}(\mathbf{y}_t-\mu_t)'\Sigma_t^{-1}(\mathbf{y}_t-\mu_t) - b_\phi\phi\right\}\phi^{a_\phi-1};$$

$$\pi(\nu \mid \mathbf{y}, \theta_{-\nu}, \mathbf{U}) \propto \left[\frac{(\frac{\nu}{2})^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})}\right]^T \exp\left\{-\left(\frac{\sum_{t=1}^{T}U_t}{2}+\lambda\right)\nu\right\}\prod_{t=1}^{T}U_t^{\frac{\nu}{2}};$$

$$\pi(\mathbf{d}^G \mid \mathbf{y}, \theta_{-\mathbf{d}^G}, \mathbf{U}) \propto \exp\left\{\frac{1}{2}tr[(\mathbf{d}^G-\mathbf{S})'(\sigma^2_{\mathbf{d}})^{-1}(\mathbf{d}^G-\mathbf{S})\mathbf{R}_d^{-1}]\right\}\prod_{t=1}^{T}\mid \Sigma_t \mid^{-1/2}\exp\left\{-\frac{1}{2}\sum_{t=1}^{T}(\mathbf{y}_t-\mu_t)'\Sigma_t^{-1}(\mathbf{y}_t-\mu_t)\right\}.$$

Table 1 shows the transition functions and acceptance probability used in the Metropolis-Hastings algorithm to obtain samples of the full conditional distributions listed above. Details about approximation for the posterior distribution of **d** can be see in Morales and Vicini (2020).

Table 1 Proposed function and acceptance probability considered in the Metropolis-Hastings steps.

	Proposed function $\pi^*(\cdot \mid \theta_i)$	$\alpha(\theta^{(j)}, \theta^{prop})$
β	$N(\cdot \mid \beta, u\mathbf{I})$	$\frac{\pi(\beta^{(prop)}\mid\theta_{-\beta},\mathbf{y})}{\pi(\beta^{(j)}\mid\theta_{-\beta},\mathbf{y})}$
σ^2	$G(\cdot \mid \sigma^2 \times u, u)$	$\frac{\pi(\sigma^{2^{prop}}\mid\theta_{-\sigma^2},\mathbf{y})\pi^*_G(\sigma^{2^{prop}}\mid\sigma^{2^{(j)}}\times u,u)}{\pi(\sigma^{2^{(j)}}\mid\theta_{-\sigma^2},\mathbf{y})\pi^*_G(\sigma^{2^{(j)}}\mid\sigma^{2^{prop}}\times u,u)}$
ϕ	$G(\cdot \mid \phi \times u, u)$	$\frac{\pi(\phi^{prop}\mid\theta_{-\phi},\mathbf{y})\pi^*_G(\phi^{prop}\mid\phi^{(j)}\times u,u)}{\pi(\phi^{(j)}\mid\theta_{-\phi},\mathbf{y})\pi^*_G(\phi^{(j)}\mid\phi^{prop}\times u,u)}$
κ	$U(\cdot \mid \max\{0, \kappa-u\}, \min\{\kappa+u, 1\})$	$\frac{\pi(\kappa^{prop}\mid\theta_{-\kappa},\mathbf{y})\pi^*_U(\kappa^{prop}\mid\max\{0,\kappa^{(j)}-u\},\min\{\kappa^{(j)}+u,1\})}{\pi(\kappa^{(j)}\mid\theta_{-\kappa},\mathbf{y})\pi^*_U(\kappa^{(j)}\mid\max\{0,\kappa^{prop}-u\},\min\{\kappa^{prop}+u,1\})}$
ν	$G(\cdot \mid \nu \times u, u)$	$\frac{\pi(\nu^{prop}\mid\theta_{-\nu},\mathbf{y})\pi^*_G(\nu^{prop}\mid\nu^{(j)}\times u,u)}{\pi(\nu^{(j)}\mid\theta_{-\nu},\mathbf{y})\pi^*_G(\nu^{(j)}\mid\nu^{prop}\times u,u)}$
λ	$U(\cdot \mid \max\{c, \lambda-u\}, \min\{\lambda+u, d\})$	$\frac{\pi(\lambda^{prop}\mid\theta_{-\lambda},\mathbf{y})\pi^*_U(\lambda^{prop}\mid\max\{c,\lambda^{(j)}-u\},\min\{\lambda^{(j)}+u,d\})}{\pi(\lambda^{(j)}\mid\theta_{-\lambda},\mathbf{y})\pi^*_U(\lambda^{(j)}\mid\max\{c,\lambda^{prop}-u\},\min\{\lambda^{prop}+u,d\})}$

It is important to point out that one of the main goals in Geostatistics is to be able to make predictions of the process of interest at any set of points in the region of study. After obtaining samples from the posterior distribution of the unknown model parameters, extra steps can be added to the algorithm to sample from predictive distributions, as presented in Section 2.4.

2.4 Interpolation

One of the main goals in geostatistics is to allow predictions of the process of interest anywhere in the region of study after observing the process at n fixed locations in space. This section presents the interpolation method proposed here, which consists of two main steps: first mapping the new locations to the D-space and then, given these new locations, interpolating the observations in the original space.

Let $\mathbf{S}^{NO} = (\mathbf{s}_{n+1}, \ldots, \mathbf{s}_{n+m})$ be a matrix of dimension $2 \times m$, representing m ungauged sites of interest in G space. Given these locations and the unknown parameters of the model, we can obtain a distribution for $\mathbf{d}^{NO} = (\mathbf{d}_{n+1}, \ldots, \mathbf{d}_{n+m})$, which are the corresponding m ungauged sites in the deformed D space, based on properties of the matrix-variate normal distribution. Firstly, note that

$$\begin{pmatrix} \mathbf{d}^{G} \\ \mathbf{d}^{N0} \end{pmatrix} \mid \mathbf{R} \sim N\left[\begin{pmatrix} \mathbf{S} \\ \mathbf{S}^{NO} \end{pmatrix}, \sigma_d^2, \mathbf{R}\right],$$

where $N(\mathbf{a}, \mathbf{A}, \mathbf{B})$ denotes the matrix-variate normal distribution with mean $\mathbf{a}$, row covariance matrix $\mathbf{A}$ and column covariance matrix $\mathbf{B}$. $\mathbf{R} = \begin{pmatrix} \mathbf{R_{A_1}} & \mathbf{R'_{A_{12}}} \\ \mathbf{R_{A_{12}}} & \mathbf{R_{A_2}} \end{pmatrix}$, in which $\mathbf{R_{A_1}}$ represents the correlation matrix of $\mathbf{d}^G$, $\mathbf{R_{A_2}}$ represents the correlation matrix of $\mathbf{d}^{NO}$ and $\mathbf{R_{A_{12}}}$ represents the correlation matrix between $\mathbf{d}^G$ and $\mathbf{d}^{NO}$. Then, it can be shown that:

$$\mathrm{vec}(\mathbf{d}^{NO}) \mid \mathrm{vec}(\mathbf{d}^{G}) \sim N(\mathbf{A}_{\mathbf{2}}^{*}, \Sigma_{A_2}^{*}), \tag{6}$$

where $\mathbf{A}_{\mathbf{2}}^{*} = \mathrm{vec}(\mathbf{S}^{NO}) + (\mathbf{I}_2 \otimes \mathbf{R_{A_{12}}}\mathbf{R_{A_1}^{-1}})(\mathrm{vec}(\mathbf{d}^G) - \mathrm{vec}(\mathbf{S}))$ and $\Sigma_{A_2}^{*} = \sigma_{\mathbf{d}}^{\mathbf{2}} \otimes (\mathbf{I_m} - \mathbf{R'_{A_{12}}}\mathbf{R_{A_1}^{-1}}\mathbf{R_{A_{12}}})$, with $\mathrm{vec}(\mathbf{A})$ denoting a vectorized version of matrix $\mathbf{A}$ and $\otimes$ denoting the Kronecker product.

Following equation (6), samples from the conditional distribution of $\mathbf{d}^{NO}$ can be obtained, adding an additional step to the proposed MCMC algorithm.

Now let $\mathbf{Y}_t^{Total} = \left(\mathbf{Y}_t, \mathbf{Y}_t^{NO}\right)'$ be a vector such that $\mathbf{Y}_t^{Total} \mid \theta \sim t_\nu(\mu_t^{Total}, \Sigma^{Total})$ where $\mathbf{Y}_t^{NO} = (Y_t(\mathbf{s}_{n+1}), \ldots, Y_t(\mathbf{s}_{n+m}))'$ is a vector of non-observed values of the process $Y(\cdot, t)$ at the ungauged sites $\mathbf{s}_{n+1}, \ldots, \mathbf{s}_{n+m} \in G$, $\mu_t^{Total} = \left(\mu_t, \mu_t^{NO}\right)'$ where μ_t^{NO} corresponds to the mean of the process at $(\mathbf{s}_{n+1}, \ldots, \mathbf{s}_{n+m})$ at time t. We also have that $\Sigma^{Total} = \begin{pmatrix} \Sigma & \Sigma^{*} \\ \Sigma^{*\prime} & \Sigma_{NO}, \end{pmatrix}$ where Σ^{NO} represents the covariance matrix of $\mathbf{Y}_t^{NO}$ and Σ^{*}, the covariance matrix between $\mathbf{Y}_t$ and $\mathbf{Y}_t^{NO}$. Note that given $\mathbf{d}^{NO}$ and the collection of parameters θ, it is easy to evaluate the corresponding Σ^{*} and Σ^{NO}.

Through properties of the multivariate t distribution (Ding 2016) we obtain:

$$\mathbf{Y}_t^{NO} \mid \mathbf{Y}_t, \theta \sim t_{\nu+n}\left(\mu_{NO|*}, \frac{\nu + d_1}{\nu + n}\Sigma_{NO|*}\right), \tag{7}$$

where

$$\begin{aligned} \mu_{NO|*} &= \mu_t^{NO} + \Sigma^{*\prime}\Sigma^{-1}(\mathbf{Y}_t - \mu_t), \\ \Sigma_{NO|*} &= \Sigma_{NO} - \Sigma^{*\prime}\Sigma^{-1}\Sigma^{*}, \end{aligned}$$

and d_1 is the squared Mahalanobis distance of $\mathbf{Y}_t$ from μ_t with scale matrix Σ, given by:

$$d_1 = (\mathbf{Y}_t - \mu_t)'\Sigma^{-1}(\mathbf{Y}_t - \mu_t).$$

That way, to sample from $\mathbf{Y}_t^{NO}|\mathbf{Y}_t, \theta$ we add an extra step on the MCMC algorithm after sampling from the conditional distribution of $\mathbf{d}^{NO}$. The samples obtained from $\mathbf{Y}_t^{NO}|\mathbf{Y}_t, \theta$ are used to compute an approximate predictive distribution of $\mathbf{Y}_t^{NO}|\mathbf{Y}_t$, and then to compute the posterior mean of $\mathbf{Y}_t^{NO}|\mathbf{Y}_t$ through Monte Carlo.

3 Simulation exercise

In this section, we present a simulation exercise aiming firstly to validate the algorithm used to sample from the posterior distribution of the model parameters and then quantify the proposed approach's performance in terms of accuracy of parameters estimation compared to simpler models. We also aim to compare the proposed and alternative models through their interpolation performance.

For the simulation, we assume that observations are made at 100 periods of time, and at 16 points in space, located in a 4×4 regular grid in a region of interest G. The data are generated by the following model:

$$\mathbf{Y}_t \sim t_\nu(\mathbf{F}_t\beta_t, \Sigma), \tag{8}$$

where

$$\beta_t = \beta_{t-1} + \mathbf{h}_t, \mathbf{h}_t \sim N(\mathbf{0}, \mathbf{H}), \qquad t = 1, \ldots, 100$$

with $\beta_0 = (2, 3, 0.5)$; $\mathbf{F}_t$ is a matrix 16×3 with elements $\mathbf{F}_{ti} = (1, s_{i,1}, s_{i,2})$, $i = 1, \ldots, 16$, where $\mathbf{s}_{i,j}, j = 1, 2$ denotes the j^{th} coordinate of location $\mathbf{s}_i$; $\Sigma = \sigma^2[(1-\kappa)\mathbf{R}(\phi) + \kappa\mathbf{I}]$, with $\sigma^2 = 0.2$, $\kappa = 0.005$; and $\mathbf{R}(\phi)$ is a matrix with elements $R_{ij} = \exp\{-\phi|\mathbf{d}_i - \mathbf{d}_j|\}$, where $\phi = 0.3$, $\mathbf{d}_i \sim GP(\mathbf{s}_i, \sigma_d^2\rho_\gamma)$, $\rho_\gamma(\mathbf{s}_i, \mathbf{s}_j) = exp(-\gamma|\mathbf{s}_i - \mathbf{s}_j|^2)$, $\sigma_d^2 = \text{diag}(0.45, 0.45)$ and $\gamma = 3$.

We simulate from the proposed model considering four different scenarios, varying the number of degrees of freedom in the Student's-t process: $\nu = 3$ (Scenario 1), $\nu = 10$ (Scenario 2), $\nu = 30$ (Scenario 3), and $\nu \to \infty$ which corresponds to $\mathbf{Y}_t \sim N(\mathbf{F}_t\beta_t, \Sigma)$ (Scenario 4). The results presented in this section were obtained considering the exponential correlation function for the elements of $\mathbf{R}(\phi)$. Similar results were obtained when specifying the wave correlation function instead, and these results are presented in the supplementary material.

We generate 100 data-sets for each of the above scenarios, and after removing the data from location $\mathbf{s}_6$, the following models were fit to each of them:

- Model A: Multivariate Student's-t distribution model with anisotropic spatial correlation function (Proposed model).

– Model B: Multivariate Student's-t distribution model with isotropic spatial correlation function.

– Model C: Multivariate normal distribution model with anisotropic spatial correlation function.

– Model D: Multivariate normal distribution model with isotropic spatial correlation function.

Note that the data from location $\mathbf{s}_6$ is considered missing, and their predictive distribution is obtained under each of the above models for comparison.

To complete the models, we chose the following hyperparameters for the prior distributions in 2.1:

– $\sigma^2_{d_i} \sim GI(1002, 1002), i = 1,2.$;

– $a_\lambda = 0.01$ and $b_\lambda = 1$, such that $\lambda \sim U(0.01, 1)$;

– $a_\phi = -2\log(0.05)/max(\mid \mathbf{s}_i - \mathbf{s}_j \mid)$ and $\eta = 1$, such that $\phi \sim G(3, 1)$;

– $a_\xi = a_\varepsilon = 2.01$ and $b_\xi = b_\varepsilon = 1.005$;

– $\mathbf{m}_0 = \mathbf{0}$ and $\mathbf{C}_0 = 1000\mathbf{I}$, such that $\beta_0 \sim N(\mathbf{0}, 1000\mathbf{I})$;

– $n_0 = 2$ and $\mathbf{S}_0 = \mathbf{I}_{15}$, such that $\mathbf{H} \sim WI^{-1}_{n_0}(\mathbf{S}_0)$.

Note that we use informative prior distributions for the σ^2_d and ϕ parameters proposed by Schmidt and Gelfand (2003); Morales et al. (2013). On the other hand, we use non-informative prior distributions for the β_0 and $\mathbf{H}$ parameters.

We sampled from the posterior distribution of the parameters through the MCMC algorithm described in Section 2.3, considering a sample of $10,000$ iterations obtained after a burn-in of $50,000$ iterations. Analysing the results from the MCMC, for the most part the real values of the model parameters were included in the 95% credibility intervals of their posterior distribution, showing that the estimation through MCMC was satisfactory.

After obtaining samples from the posterior distribution, we used the mean squared error to assess the accuracy of the estimation of the model parameters. Table 2 presents the estimated mean squared error for the parameters ϕ, κ, σ^2, and ν obtained under models A, B, C, and D for each of the four scenarios. We observe that in most scenarios, the estimated mean squared error for the parameters ϕ and κ obtained under the anisotropic models (A and C) were lower than those obtained under the isotropic models (B and D). This result was obtained due to the inclusion of the spatial deformation in the definition of the spatial correlation function, which helped explain the spatial variability in the data.

The estimated mean squared error for parameters σ^2 and ν were smaller under Model B than under Model A, possibly due to the uncertainty included in estimating these parameters when including spatial deformation in the model. This result is aggravated by the low number of locations in space, as increasing the number of monitoring stations not only reduces the MSE of ϕ and κ but also leads to a reduction in the discrepancy between the MSE of σ^2 and ν between the t models (results obtained under Scenario 1 and a larger number of locations in space can be seen in the supplementary material). However, as it can be seen in Table 3, on average, the parameter ν was better estimated under model A. Table 3 shows the quantiles of the posterior means of ν estimated for the 100 simulated data-sets under models A and B.

Table 2 Estimated mean square error for the parameters ϕ, κ, σ^2, and ν obtained under models A, B, C, and D for four different scenarios.

	Scenario 1 ($\nu = 3$)				Scenario 2 ($\nu = 10$)			
Model	ϕ	κ	σ^2	ν	ϕ	κ	σ^2	ν
A	0.02	0.0004	0.31	0.31	0.04	0.0003	0.33	18.42
B	0.04	0.0088	0.09	0.22	0.05	0.0053	0.12	5.98
C	0.02	0.0009	0.34		0.03	0.0003	0.32	
D	0.33	0.0490	0.12		0.04	0.0091	0.11	
	Scenario 3 ($\nu = 30$)				Scenario 4 ($\nu \to \infty$)			
Model	ϕ	κ	σ^2	ν	ϕ	κ	σ^2	ν
A	0.04	0.0002	0.34	561.85	0.04	0.0002	0.34	
B	0.05	0.0058	0.12	237.38	0.05	0.0059	0.12	
C	0.04	0.0002	0.34		0.04	0.0002	0.33	
D	0.05	0.0077	0.11		0.05	0.0068	0.11	

Table 3 Q_1, Q_2, and Q_3 are quantiles of the 100 estimates of the ν parameter obtained under models A and B. We used the mean posterior of parameter ν as how to point estimator of parameter ν.

	Scenario 1 ($\nu = 3$)			Scenario 2 ($\nu = 10$)		
Model	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3
A	2.8	3.0	3.4	10.3	12.7	14.2
B	2.7	2.9	3.2	7.2	8.5	9.3
	Scenario 3 ($\nu = 30$)			Scenario 4 ($\nu \to \infty$)		
Model	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3
A	29.7	43.7	58.3	73.7	86.1	94.7
B	11.9	16.0	18.4	18.6	24.9	35.8

We also performed tests to verify the goodness of fit of the models. To assess the goodness of fit for the multivariate normal distribution models (C and D) we utilized a few tests that were implemented in the mvnTest package in R by Pya et al. (2016), which were: Anderson-Darling (AD) (Paulson et al. 1987; Henze and Zirkler 1990), Cramer-von Mises (CM) (Koziol 1982; Henze and Zirkler 1990), Doornik-Hansen (DH) (Doornik and Hansen 2008), and Henze-Zirkler (HZ) (Henze and Zirkler 1990). These tests verify the assumption that the residuals obtained after fitting the model follow a multivariate normal distribution (null hypothesis). It is important to notice that as the data were simulated from a Student's-t distribution under scenarios 1, 2, and 3, it would be desirable to detect a lack of fit for these cases.

Table 4 shows how many times (in a total of 100) the null hypothesis was rejected when using each of the tests to assess the goodness of fit of models C and D. Under scenario 1, which generated data further from the normal, it was clear the lack of adequacy of both models as the tests rejected the null hypothesis 100% of the time. Under scenario 2, the AD, CM, and HD tests were still able to reject the null hypothesis most of the time, while the HZ test detected it 50% of the time under model C and 47% under model D. Under scenario 3, generated with

$\nu = 30$, the residuals behave mostly accordingly to a normal distribution, with the percentage of rejection being just slightly over the percentage of rejection under scenario 4.

Table 4 Number of times (in a total of 100) that the null hypothesis was rejected when using the Anderson-Darling (AD), Cramer-von Mises (CM), Doornik-Hansen (DH), and Henze-Zirkler (HZ) tests to assess the goodness of fit of models C and D.

Scenarios	Model	AD	CM	DH	HZ
1 ($\nu = 3$)	C	100	100	100	100
	D	100	100	100	100
2 ($\nu = 10$)	C	87	74	53	50
	D	91	73	58	47
3 ($\nu = 30$)	C	8	4	12	14
	D	6	3	17	11
4 ($\nu \to \infty$)	C	7	6	4	7
	D	8	7	6	7

To assess the goodness of fit for the multivariate Student's-t distribution models (A and B), we used the test proposed by McAssey (2013). This test verifies the assumption that the residuals obtained after fitting the model follow a multivariate Student's-t distribution. Table 5 shows how many times (in a total of 100) the null hypothesis was rejected when using this test. The number of tests which had the null hypothesis rejected was relatively low under scenarios 1, 2, and 3, with a higher percentage of rejection under scenario 4.

Table 5 Number of times (in a total of 100) that the null hypothesis was rejected when using the test proposed by McAssey (2013) to assess the goodness of fit of models A and B.

	Scenarios			
Model	1 ($\nu = 3$)	2 ($\nu = 10$)	3 ($\nu = 30$)	4 ($\nu \to \infty$)
A	3	2	4	14
B	3	3	10	17

To directly compare the models and verify which one performs better under different scenarios, we used the Deviance information criterion (DIC) (Spiegelhalter et al. 2002) and the Log Pseudo-Marginal Likelihood (LPML) criteria (Gelfand and Dey 1994). We used the Mean Absolute Deviation (MAD) and Measures of Predictive Precision (MSE) criteria to measure the quality of the interpolation produced by the models comparing to the data observed in $\mathbf{s}_6$. The MAD and MSE criteria are defined as follows:

$$\text{MAD} = \frac{1}{T}\sum_{t=1}^{T} |y_t(\mathbf{s}_6) - \hat{y}_t(\mathbf{s}_6)|,$$

and

$$\text{MSE} = \frac{1}{T}\sum_{t=1}^{T} (y_t(\mathbf{s}_6) - \hat{y}_t(\mathbf{s}_6))^2,$$

where $\hat{y}_t(\mathbf{s}_6)$ is the estimate produced by the interpolation method (predictive mean).

Table 6 presents the quartiles of the DIC, LPML, MAD, and MSE criteria obtained from the 100 fits of the models. When we analyze scenarios 1 and 2, we observe that using DIC and LPML criteria, the best model is Model A, with smaller values under the DIC and larger values under the LPML. The quality of prediction, however, is very similar amongst the different models. The fact that the proposed model did not significantly improve the prediction performance is probably due to the fact that the number of grid points used in the G region (15 locations) is large enough for the predictive distribution at location $\mathbf{s}_6$ be well approximated by a normal distribution.

Table 6 Quantiles Q_i, $i = 1,2,3$, of the DIC, LPML, MAS, and MSE criteria were obtained with the 100 fits for each of the models analyzed for each of the studied scenarios.

Scenario 1 ($\nu = 3$)												
	DIC			LPML			MAD			MSE		
Model	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3
A	558	775	1000	-7.1	-5.9	-4.8	0.16	0.20	0.24	0.10	0.15	0.24
B	1081	1274	1461	-8.1	-7.4	-6.4	0.16	0.18	0.21	0.10	0.13	0.17
C	1069	1324	1598	-11.6	-10.2	-8.6	0.17	0.20	0.24	0.12	0.16	0.23
D	1681	1917	2121	-Inf	-16.5	-11.7	0.16	0.18	0.20	0.09	0.13	0.17
Scenario 2 ($\nu = 10$)												
	DIC			LPML			MAD			MSE		
Model	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3
A	152	348	519	-4.6	-3.8	-2.8	0.14	0.17	0.20	0.05	0.08	0.11
B	635	796	914	-6.5	-5.6	-4.7	0.14	0.16	0.18	0.05	0.06	0.08
C	190	398	569	-5.5	-4.8	-3.7	0.14	0.18	0.20	0.05	0.08	0.11
D	790	947	1133	-35.3	-7.7	-6.4	0.14	0.16	0.18	0.05	0.06	0.08
Scenario 3 ($\nu = 30$)												
	DIC			LPML			MAD			MSE		
Model	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3
A	23	167	299	-3.4	-2.8	-2.1	0.14	0.16	0.20	0.04	0.07	0.09
B	468	590	710	-6.9	-5.3	-4.5	0.13	0.15	0.18	0.05	0.05	0.07
C	48	162	295	-4.0	-3.2	-2.5	0.14	0.16	0.20	0.04	0.06	0.10
D	586	732	847	-65.7	-10.8	-5.3	0.13	0.16	0.18	0.05	0.05	0.07
Scenario 4 ($\nu \to \infty$)												
	DIC			LPML			MAD			MSE		
Model	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3
A	-11	130	254	-3.4	-2.7	-2.1	0.13	0.17	0.19	0.04	0.06	0.08
B	397	540	630	-9.1	-5.4	-4.4	0.13	0.16	0.17	0.04	0.05	0.07
C	-31	141	255	-3.7	-2.9	-2.2	0.13	0.16	0.20	0.04	0.07	0.09
D	536	663	769	-47.0	-10.5	-4.9	0.13	0.16	0.17	0.04	0.05	0.07

4 Application to evaporation data in Rio Grande do Sul, Brazil

In this section, we are interested in modeling the space-time variability of evaporation data from a Piche evaporimeter measured daily at 11 stations distributed in the state of Rio Grande do Sul, Brazil, in the years 2017 and 2018, for a total T =730 observations per monitoring station. The map with the monitoring stations is in Figure 1. This dataset is collected by the National Institute of Meteorology and can be accessed at the website `https://portal.inmet.gov.br/`.

Exploratory analysis of the evaporation time series (see Figure 8 in Appendix A) show some atypical values, which can be an indication that the process under investigation is generated by a distribution with heavy tails. We also compared the variance of the observations for each period of time (representing the variability in space) and

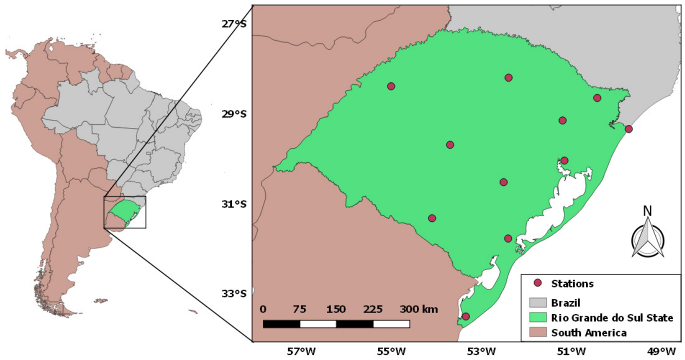

Fig. 1 Map of the state of Rio Grande do Sul (G space), located in the south of Brazil, and the location of the 11 monitoring stations in G.

the variance of the observations for each location in space (representing the variability in time). Figure 2 shows the boxplot of the variances over time and space. It is essential to notice that even though the variability is usually higher in time, there are a few atypically high space variabilities. Such high variability in space indicates the need to consider a heavy tailed distribution to handle the spatial variability in the data.

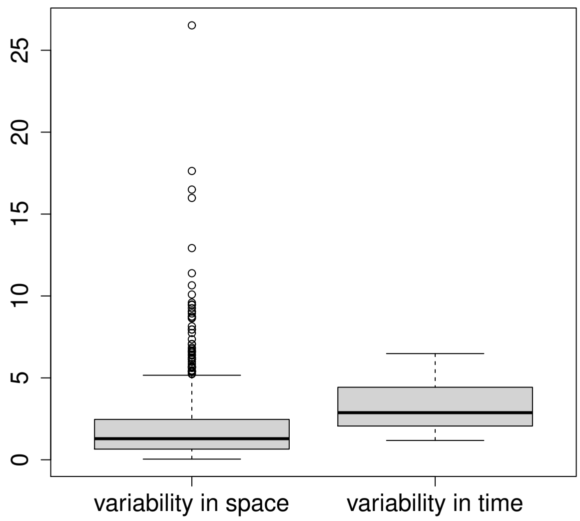

Fig. 2 Boxplot of the variances calculated for each period of time - space variability (left) and boxplot of the variances calculated for each location in space - time variability (right) .

The heterogenic topography of the region under study, as well as the climatic variations produced by the geographical location (south latitude) and the cold fronts coming from the south pole (Reboita et al. 2010) are an indication that the hypothesis of isotropy would not be reasonable for the evaporation process. Figure 3 presents the directional sample variogram at $0^o, 45^o, 90^o$ and 135^o of the residuals after removing a temporal trend. This

Figure shows that even though the nugget effect is similar at all directions (it is high for all), the range vary (at 135^o is smaller than at 0^o and 90^o) and the sill also vary (at 90^o being larger and at 45^o being smaller). It is interesting to point out that at 45^o is it not possible to notice spatial dependence.

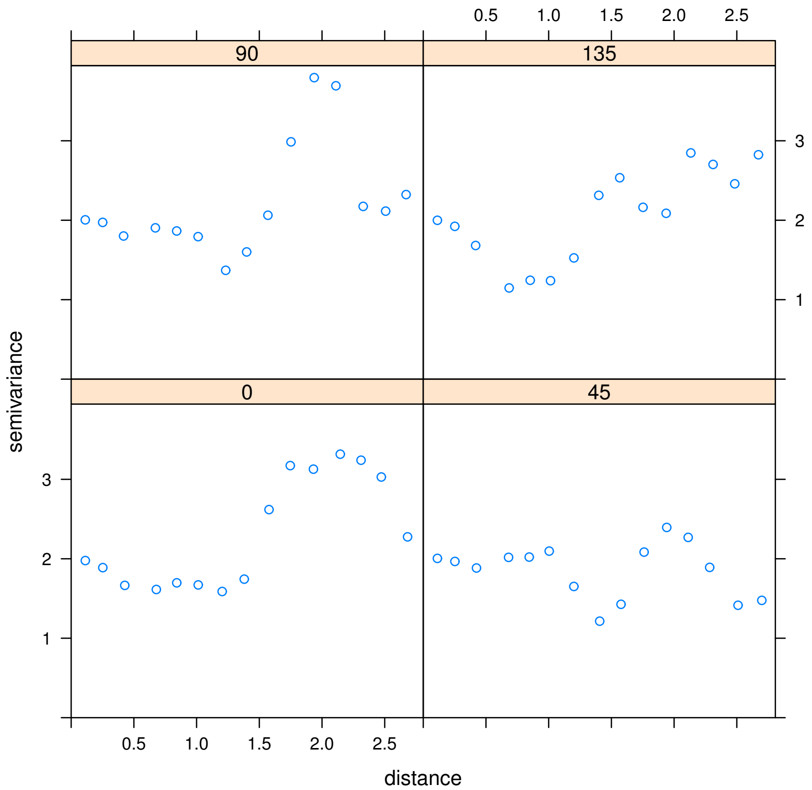

Fig. 3 Directional sample variogram at $0^o, 45^o, 90^o$ and 135^o of the residuals after removing a temporal trend.

This analysis confirms that the hypothesis of isotropy does not seem appropriate to handle the spatial dependency in the observations. This way, we model this dataset through the spatial-temporal model proposed in Section 2, considering the following covariates to explain the mean of the process: Latitude (denoted by x_i); Longitude (denoted by y_i); Altitude (denoted by z_i) and the iterations between x_i, y_i, and z_i. We also incorporate an annual and semiannual seasonal in the specification of $\mathbf{G}_t$. That is, the state space model is defined with the covariate vector $\mathbf{F}_{\mathbf{t}i} = (1\ x_i\ y_i\ z_i\ x_iy_i\ x_iz_i\ y_iz_i\ x_iy_iz_i\ 1\ 0\ 1\ 0)$, for $i = 1, \ldots, 11$, and the transition matrix $\mathbf{G}_t = \begin{pmatrix} \mathbf{I}_8 & \mathbf{0}_{8\times 4} \\ \mathbf{0}_{4\times 8} & \mathbf{G}_3 \end{pmatrix}$, where $\mathbf{G}_1 = \begin{pmatrix} \cos(\omega) & \sin(\omega) \\ -\sin(\omega) & \cos(\omega) \end{pmatrix}$, $\mathbf{G}_2 = \begin{pmatrix} \cos(2\omega) & \sin(2\omega) \\ -\sin(2\omega) & \cos(2\omega) \end{pmatrix}$, $\mathbf{G}_3 = \mathrm{diag}(\mathbf{G}_1, \mathbf{G}_2)$, $\omega = 2\pi/365$. Note that $\mathbf{G}_1$ and $\mathbf{G}_2$ incorporate sines and cosines in the evolution matrix $\mathbf{G}_t$ to model cycles attributed to the seasons of the year and semiannual rainfall regime (summer and winter rains), respectively.

Besides the proposed model in Section 2, which we will refer to as Model A, we also estimated for comparison an isotropic version of this model (Model B), and Gaussian versions either considering spatial deformation (Model

C) or isotropy (Model D). These models are summarized bellow:

$$
\begin{aligned}
\text{Model A} &: \mathbf{W}_t \sim t_n(\mathbf{0}, \Sigma), \Sigma_{ij} = \sigma^2[(1-\kappa)\frac{\phi \sin(\mid \mathbf{d}_i - \mathbf{d}_j \mid / \phi)}{\mid \mathbf{d}_i - \mathbf{d}_j \mid} + \kappa],\\
\text{Model B} &: \mathbf{W}_t \sim t_n(\mathbf{0}, \Sigma), \Sigma_{ij} = \sigma^2[(1-\kappa)\frac{\phi \sin(\mid \mathbf{s}_i - \mathbf{s}_j \mid / \phi)}{\mid \mathbf{s}_i - \mathbf{s}_j \mid} + \kappa],\\
\text{Model C} &: \mathbf{W}_t \sim N_n(\mathbf{0}, \Sigma), \Sigma_{ij} = \sigma^2[(1-\kappa)\frac{\phi \sin(\mid \mathbf{d}_i - \mathbf{d}_j \mid / \phi)}{\mid \mathbf{d}_i - \mathbf{d}_j \mid} + \kappa],\\
\text{Model D} &: \mathbf{W}_t \sim N_n(\mathbf{0}, \Sigma), \Sigma_{ij} = \sigma^2[(1-\kappa)\frac{\phi \sin(\mid \mathbf{s}_i - \mathbf{s}_j \mid / \phi)}{\mid \mathbf{s}_i - \mathbf{s}_j \mid} + \kappa].
\end{aligned}
$$

As can be seen, we opted to work with the wave correlation function for this application by the specifications above. Similar results were obtained when working with the exponential correlation function instead, and these results we present in the supplementary material.

We assume a priori that $\phi \sim G(a_\phi, \eta)$ with $a_\phi = -2\log(0.05)/max(\mid \mathbf{s}_i - \mathbf{s}_j \mid) = 0.19$ and $\eta = 1$, so that the mean is equal to a_ϕ and the spatial correlation function has a smooth decay Schmidt and Gelfand (2003); Morales et al. (2013). As we expect a soft deformation of the original space a priori, we define an informative prior distribution for the parameter σ^2_d, that is, $\sigma^2_{d_i} \sim GI(1000, 100)$, $i = 1, 2$ (Morales et al. 2013). For λ we specified $\lambda \sim U(\frac{1}{50}, \frac{1}{3})$ so that $3 < E(\nu \mid \lambda) < 50$ a priori. Finally, we present the following uninformative prior distributions for the β_0 and $\mathbf{H}$ parameters: $\beta_0 \sim N(\mathbf{0}, 1000\mathbf{I})$ and $\mathbf{H} \sim WI_2^{-1}(\mathbf{I})$.

We sampled from the posterior distribution of the parameters through the MCMC algorithm described in Section 2.3, considering a sample of $10,000$ iterations obtained after a burn-in of $100,000$ iterations. To verify the convergence of the MCMC we run two chains starting from different initial values, for each model. We used the Gelman and Rubin's criteria (Gelman and Rubin 1992; Plummer et al. 2006) to verify convergence.

Table 7 presents descriptive statistics of the posterior samples from the parameters σ^2, σ^2_ε, σ^2_ξ, ϕ, κ, and ν under each one of the adjusted models. The results show that Model A is the best when the models are compared through the DIC criteria. Through the LPML criteria, Models A, B and C presented similar results, with Model C showing a slight advantage. If we compare the t models (Models A and B) we observed that there is no significant difference in the estimation of parameters σ^2 and ν. However, the estimates of ϕ and κ produced by Model A are smaller when compared to Model B. In this case, the value of $\kappa = 0.44$ estimated under Model A, shows that 44% of the variability of the covariance structure is explained by the nugget effect against 81% under Model B. This result suggests that the improvement obtained with the increase in the percentage of the explanation of the variability of the structure of the covariance of to the spatial component is due to spatial deformation (see Figure 4). On the other hand, when comparing the Gaussian models (Models C and D) we observe significant difference in the estimated values for κ and ϕ. In this case, the variability of the covariance structure explained by the nugget effect is 28% under Model C against 72% under Model D. This indicates that the inclusion of spatial deformation increases significantly the percentage of explanation of the variability of the covariance of to the spatial component. To compare the goodness of fit of the models we performed residual analysis and construct Q-q plots with 95% confidence intervals. Figure 5 shows the residuals estimated under Models A and C for the monitoring station s_1 as well as the Q-q plots of the Student's-t and Normal distributions (respectively) with simulated confidence envelopes (95%) for the estimated residuals under these models. Similar results were obtained with other monitoring stations. This analysis shows that Model A have a better fit than Model C, with the residuals under model A following inside the Q-q plot confidence intervals. Similar results were obtained under models B and D, meaning the Student's-t models present a better fit then the Gaussian models.

In Figure 6, we present the estimated surface (posterior mean) and the amplitude of the 95% credibility intervals at time $t = 160$ under model A (proposed model). The Figure suggests that, on average, higher Piche evaporation levels are obtained in the north-east of the studied region, and, as one would expect, the credibility intervals have smaller amplitude in the center of the map, where the observations were made.

Table 7 Posterior mode, mean, median, and 95 % credibility interval for the parameters σ^2, σ^2_ε, σ^2_ξ, ϕ, κ, and ν of the models adjusted for evaporation data in the states of Rio Grande do Sul in Brazil.

	Model A					Model B				
Parameter	Mean	Median	2.5%	97.5%	SD	Mean	Median	2.5%	97.5%	SD
σ^2	0.63	0.63	0.58	0.69	0.03	0.62	0.62	0.57	0.69	0.03
σ^2_ε	0.27	0.27	0.25	0.30	0.01	0.50	0.50	0.47	0.54	0.02
σ^2_ξ	0.35	0.35	0.31	0.40	0.02	0.12	0.12	0.08	0.15	0.02
ϕ	0.63	0.63	0.60	0.66	0.03	4.95	4.82	2.99	7.88	1.24
κ	0.44	0.44	0.40	0.47	0.02	0.81	0.81	0.76	0.86	0.03
ν	4.07	4.06	3.56	4.65	0.28	4.04	4.04	3.47	4.69	0.31
DIC	2809					3237				
LPML	-18					-19				
	Model C					Model D				
Parameter	Mean	Median	2.5%	97.5%	SD	Mean	Median	2.5%	97.5%	SD
σ^2	1.25	1.25	1.17	1.32	0.04	1.31	1.28	1.12	1.64	0.14
σ^2_ε	0.35	0.34	0.31	0.40	0.02	0.94	0.94	0.90	0.99	0.02
σ^2_ξ	0.90	0.91	0.80	0.98	0.05	0.37	0.34	0.27	0.70	0.13
ϕ	0.56	0.56	0.54	0.60	0.01	6.18	5.97	3.38	10.02	1.70
κ	0.28	0.27	0.24	0.33	0.02	0.72	0.74	0.58	0.84	0.07
DIC	10357					12245				
LPML	-17					-24				

Finally, to study the predictive performance of our models, they were fit removing one of the stations at a time, and then the observations in the station that was removed are predicted through time and compared to the real observations. Table 8 shows the percentage of coverage of the observations of the station left-out, in the 95% confidence envelopes under models A, B, C and D. Figure 7 shows the observations made at monitoring station s_6 through time and the simulated 95% confidence envelopes under the proposed model (model A). Similar results were obtained for the other monitoring stations. The Figure shows that the predictions seem reasonable under the proposed model with most of the real values following inside the credibility intervals.

In conclusion, the results suggest a significant gain in modeling of evaporation when we relax the assumption that the observations come from an isotropic Gaussian process, and assume an anisotropic Student's-t distribution for the process of interest.

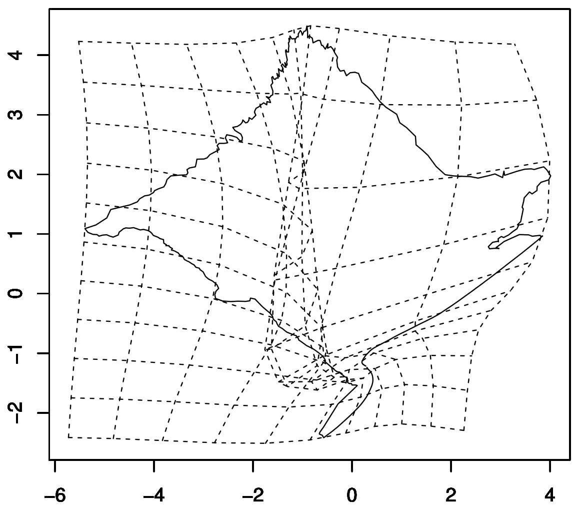

Fig. 4 Estimated deformed map of the state of Rio Grande do Sul (D space), under model A (proposed model)

Table 8 Percentage of coverage of the observations of the station left-out in the 95% confidence envelopes under models A, B, C and D.

Station left-out	Model A	Model B	Model C	Model D
s1	0.97	0.97	0.96	0.97
s2	0.54	0.68	0.50	0.74
s3	0.65	0.61	0.56	0.65
s4	0.67	0.69	0.64	0.67
s5	0.68	0.71	0.68	0.73
s6	0.97	0.92	0.89	0.51
s7	0.79	0.95	0.97	0.94
s8	0.90	0.90	0.89	0.90
s9	0.91	0.76	0.80	0.75
s10	0.98	0.90	0.88	0.88
s11	0.93	0.87	0.93	0.88

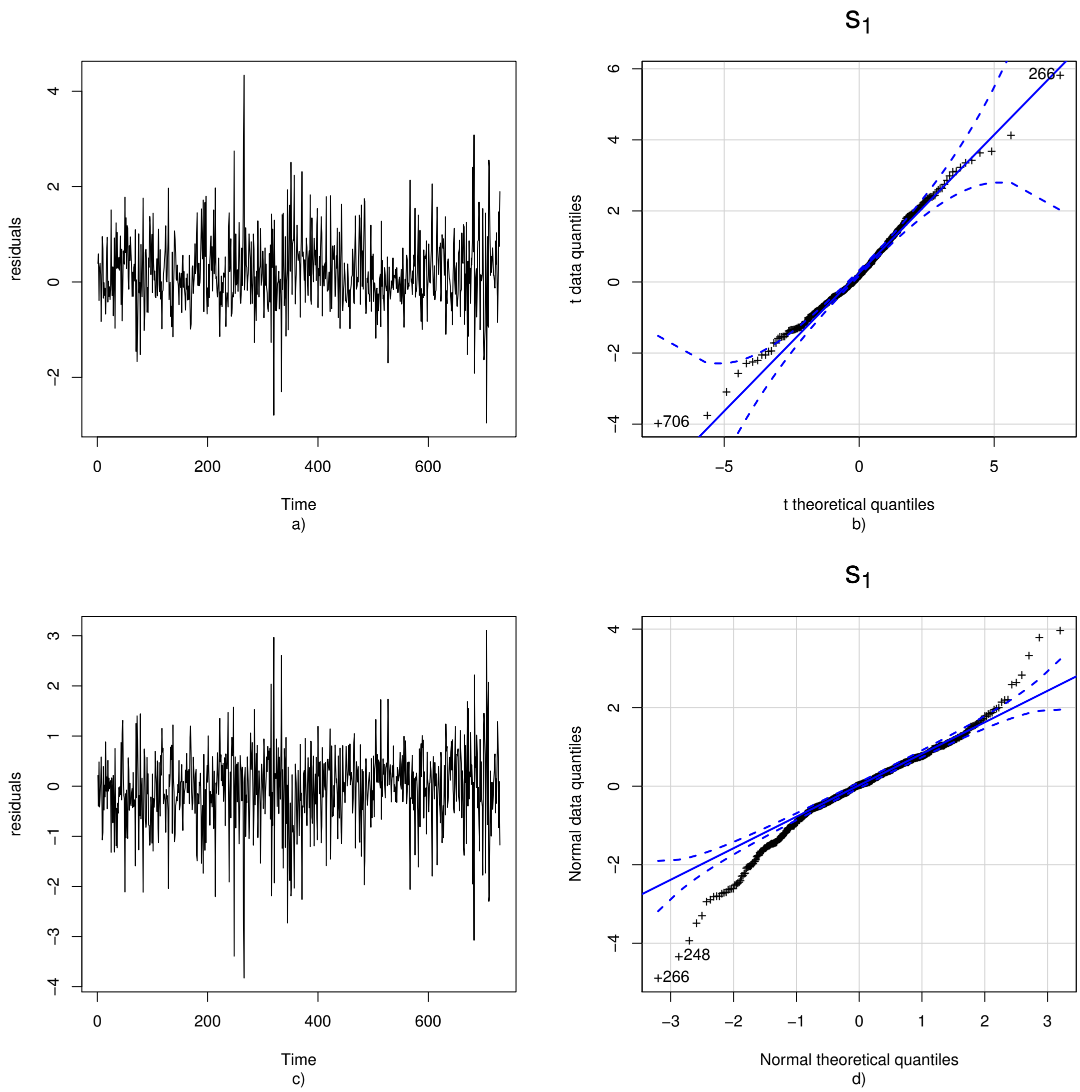

Fig. 5 a) Residuals estimated under Model A for the monitoring station s_1. b) Q-q plot of the Student's-t distribution with simulated confidence envelopes (95%) for the estimated residual under Model A. c) Residuals estimated under Model C for the monitoring station s_1. d) Q-q plot of the normal distribution with simulated confidence envelopes (95%) for the estimated residual under Model C.

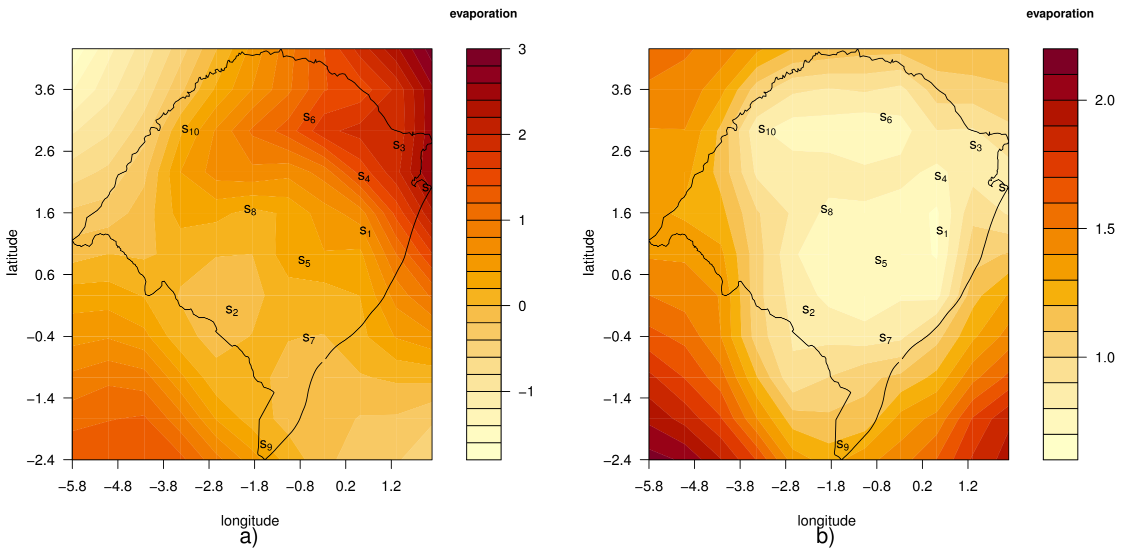

Fig. 6 a) Estimated surface (posterior mean) at t=160, b) Amplitude of the 95 % credibility intervals for evaporation in the state of Rio Grande do Sul at t = 160, under model A (proposed model)

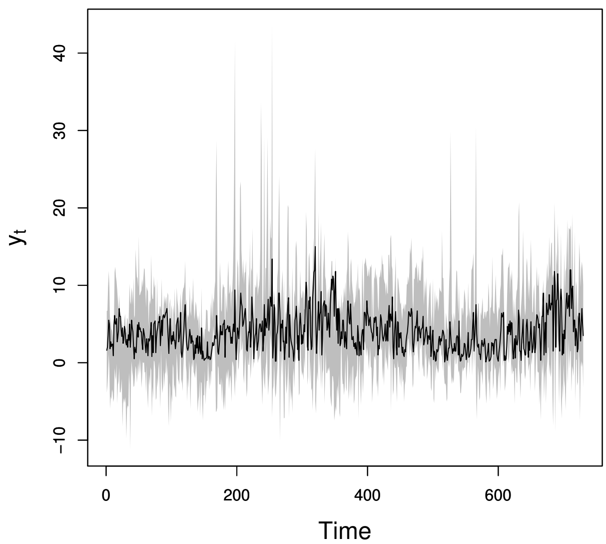

Fig. 7 Observations made at s_6 through time (black line) and simulated 95% confidence envelopes (gray areas) under model A (proposed model).

5 Conclusions

In this paper, we propose a spatiotemporal model to deal with environmental data that contain outliers. Our model also allows us to treat situations in which it is not reasonable to assume that the covariance structure is isotropic. Data of this kind are not uncommon when observing environmental processes. We propose a Student's-t spatial model in which the mean process incorporates time variation through a state space approach (West and Harrison 1997) and the spatial correlation function incorporates anisotropy via spatial deformation (Sampson and Guttorp 1992). The estimation of the parameters of the proposed model is performed from a Bayesian perspective through MCMC methods.

The algorithm is validated through a simulation exercise. In this simulated study we also compare the performance of the proposed Student's-t anisotropic model (Model A), with three alternative models: a Student's-t isotropic model (Model B), a Gaussian anisotropic model (Model C) and a Gaussian isotropic model (Model D). The proposed model (Model A) performed better than the others according to the DIC and LPML criteria, specially when simulating from a Student's-t process with a small number of degrees of freedom.

The proposed model is applied to an evaporation dataset collected at 11 stations distributed in the state of Rio Grande do Sul, Brazil. For comparison we also fit Models B, C and D to this dataset. We used the DIC and LPML criteria to compare the models and concluded that overall Model A is the one with the best performance.

To study the predictive performance of our models, models A, B, C and D were fit, removing one of the stations at a time, and then the observations in the station that was removed were predicted through time and compared to the real observations. All four models showed reasonable predictions, with the Student's-t models A and B presenting larger 95% credibility intervals. These larger intervals guarantee a higher percentage of the real observations inside the intervals, showing a better performance of the Student's-t models.

As a future line of work, we could extend the proposed model to handle temporal and spatial variations jointly instead of having independent spatial processes over time. A possible approach is to work with a sequence of intensity surfaces (varying in space) linked through time, as in Gelfand et al. (2005) and Reis et al. (2013).

We would also like to explore other approaches for heavy tailed distributions with atypical observations, such as the family of elliptical distributions proposed by De Bastiani et al. (2015). Recently, De Bastiani et al. (2015) proposed a family of elliptical distributions to model spatial covariance structures in order to minimize the influence of atypical data on maximum likelihood estimates. This family of elliptical distributions is also a great alternative to model heavy tailed processes. One important feature of this family is that it incorporates some interesting distributions, such as: Normal, Student's-t, generalized t, and contaminated normal, among others.

Another interesting comparison could be made between our approach to anisotropy and other approaches, such as the one proposed by Haskard et al. (2007), where the autors generalize the Mátern family to allow for anisotropy.

Acknowledgements

The first author acknowledges support by CNPq-Brazil.

A Appendix

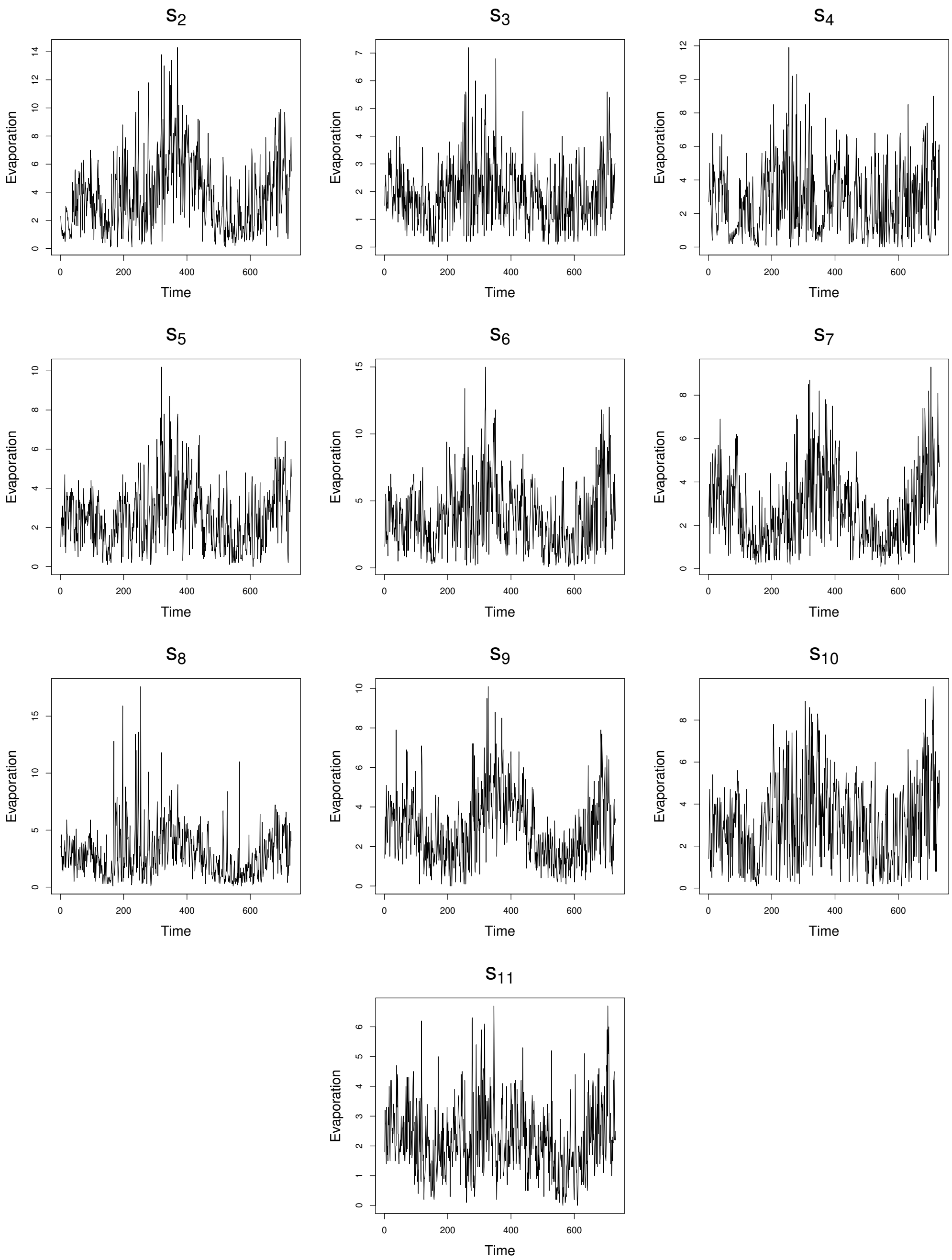

Fig. 8 Evaporation time series observed in each monitoring station.

References

Assumpcao RAB, Uribe-Opazo MA, Galea M (2011) Local influence for spatial analysis of soil physical properties and soybean yield student's t-distribuition. Revista Brasileira de Ciencia do Solo 35(5):1917–1926

Bruno F, Guttorp P, Sampson P, Cocchi D (2008) A simple non-separable, non-stationary spatiotemporal model for ozone. Environmental and Ecological Statistics 16:515–529, DOI 10.1007/s10651-008-0094-8

Cabral CRB, Lachos VH, Madruga MR (2012) Bayesian analysis of skew-normal independent linear mixed models with heterogeneity in the random-effects population. Journal of Statistical Planning and Inference 142(1):181–200, DOI 10.1016/j.jspi.2011.07.007, URL https://app.dimensions.ai/details/publication/pub.1052500334

Carlin BP, Polson NG, Stoffer DS (1992) A monte carlo approach to nonnormal and nonlinear state-space modeling. Journal of the American Statistical Association 87(418):493–500, URL http://www.jstor.org/stable/2290282

Chib S, Ramamurthy S (2014) Dsge models with student-t errors. Econometric Reviews 33(1-4):152–171, DOI 10.1080/07474938.2013.807152, URL https://doi.org/10.1080/07474938.2013.807152

Damian D, Sampson P, Guttorp P (2001) Bayesian estimation of semi-parametric non-stationary spatial covariance structure. Environmetrics 12:161 – 178, DOI 10.1002/1099-095X(200103)12:2⟨161::AID-ENV452⟩3.0.CO;2-G

De Bastiani F, de Aquino Cysneiros AM, Uribe-Opazo M, Galea M (2015) Influence diagnostics in elliptical spatial linear models. TEST: An Official Journal of the Spanish Society of Statistics and Operations Research 24(2):322–340, URL https://EconPapers.repec.org/RePEc:spr:testjl:v:24:y:2015:i:2:p:322-340

Ding P (2016) On the conditional distribution of the multivariate t distribution. The American Statistician 70(3):293–295, DOI 10.1080/00031305.2016.1164756, URL https://doi.org/10.1080/00031305.2016.1164756

Doornik JA, Hansen H (2008) An omnibus test for univariate and multivariate normality. Oxford Bulletin of Economics and Statistics 70(s1):927–939, DOI https://doi.org/10.1111/j.1468-0084.2008.00537.x, URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-0084.2008.00537.x

Fouedjio F, Desassis N, Romary T (2015) Estimation of space deformation model for non-stationary random functions. Spatial Statistics 13:45 – 61, DOI 10.1016/j.spasta.2015.05.001

Gelfand AE, Dey DK (1994) Bayesian model choice: Asymptotics and exact calculations. Journal of the Royal Statistical Society Series B (Methodological) 56(3):501–514, URL http://www.jstor.org/stable/2346123

Gelfand AE, Banerjee S, Gamerman D (2005) Spatial process modelling for univariate and multivariate dynamic spatial data. Environmetrics 16(5):465–479, DOI https://doi.org/10.1002/env.715, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/env.715

Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. Statistical Science 7(4):457 – 472, DOI 10.1214/ss/1177011136, URL https://doi.org/10.1214/ss/1177011136

Haskard KA, Cullis BR, Verbyla AP (2007) Anisotropic matérn correlation and spatial prediction using reml. Journal of Agricultural, Biological, and Environmental Statistics 12(2):147–160, URL http://www.jstor.org/stable/27595633

Henze N, Zirkler B (1990) A class of invariant consistent tests for multivariate normality. Communications in statistics-Theory and Methods 19(10):3595–3617

Kang EL, Cressie N (2011) Bayesian inference for the spatial random effects model. Journal of the American Statistical Association 106(495):972–983, URL http://www.jstor.org/stable/23427567

Koziol JA (1982) A class of invariant procedures for assessing multivariate normality. Biometrika 69(2):423–427, URL http://www.jstor.org/stable/2335417

Kuo L, Yang T (1996) Bayesian computation for nonhomogeneous poisson processes in software reliability. Journal of The American Statistical Association - J AMER STATIST ASSN 91:763 – 773, DOI 10.1080/01621459.1996.10476944

Marengo JA, Rodrigues R, Alves L (2017) Drought in northeast brazil—past, present, and future. Theoretical and Applied Climatology 129:1189 – 1200, DOI 10.1007/s00704-016-1840-8

McAssey MP (2013) An empirical goodness-of-fit test for multivariate distributions. Journal of Applied Statistics 40(5):1120–1131, DOI 10.1080/02664763.2013.780160, URL https://doi.org/10.1080/02664763.2013.780160

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of state calculations by fast computing machines. The journal of chemical physics 21(6):1087–1092

Morales F, Vicini L (2020) A non-homogeneous poisson process geostatistical model with spatial deformation. AStA Advances in Statistical Analysis URL https://doi.org/10.1007/s10182-020-00373-6

Morales F, Gamerman D, Paez M (2013) State space models with spatial deformation. Environmental and Ecological Statistics 20:191 – 214, DOI 10.1007/s10651-012-0215-2

Neal P, Kypraios T (2015) Exact bayesian inference via data augmentation. Statistics and Computing 25:pages333–347

Paulson AS, Roohan P, Sullo P (1987) Some empirical distribution function tests for multivariate normality. Journal of Statistical Computation and Simulation 28(1):15–30, DOI 10.1080/00949658708811005, URL https://doi.org/10.1080/00949658708811005

Plummer M, Best N, Cowles K, Vines K (2006) Coda: Convergence diagnosis and output analysis for mcmc. R News 6(1):7–11, URL https://journal.r-project.org/archive/

do Prado NV, Uribe-Opazo MA, Galea M, Assumpção RAB (2013) Influência local em um modelo espacial linear da produtividade da soja utilizando distribuição t-student. Engenharia Agrícola 33(5):1003–1016, URL https://dx.doi.org/10.1590/S0100-69162013000500012

Pya N, Voinov V, Makarov R, Voinov Y (2016) mvnTest: Goodness of Fit Tests for Multivariate Normality. URL https://CRAN.R-project.org/package=mvnTest, r package version 1.1-0

Reboita MS, Gan MA, Rocha RP, Ambrizzi T (2010) Regimes de precipitacao na america do sul: Uma revisao bibliografica. Revista Brasileira de Meteorologia 25(2):185–204

Reis EA, Gamerman D, Paez MS, Martins TG (2013) Bayesian dynamic models for space–time point processes. Computational Statistics & Data Analysis 60:146–156, DOI https://doi.org/10.1016/j.csda.2012.11.008, URL https://www.sciencedirect.com/science/article/pii/S0167947312003994

Ropelewski CF, Halpert MS (1987) Global and regional scale precipitation patterns associated with the el niño/southern oscillation. Monthly Weather Review 115(8):1606–1626, DOI 10.1175/1520-0493(1987)115⟨1606:GARSPP⟩2.0.CO;2, URL https://doi.org/10.1175/1520-0493(1987)115<1606:GARSPP>2.0.CO;2

Roth M (2013) On the multivariate t distribution. Tech. Rep. 3059, Automatic Control at Linköpings universitet

Sampson PD, Guttorp P (1992) Nonparametric estimation of nonstationary spatial covariance structure. Journal of the American Statistical Association 87(417):108 – 119, URL http://www.jstor.org/stable/2290458

Schemmer RC, Uribe-Opazo MA, Galea M, Assumpcao RAB (2017) Spatial variability of soybean yield through a reparametrized t-student model. Engenharia Agrícola 37(4):760–770, URL https://dx.doi.org/10.1590/1809-4430-eng.agric.v37n4p760-770/2017

Schmidt A, Gelfand A (2003) A bayesian coregionalization approach to multivariate pollutant data. Journal of Geophysical Research 108, DOI 10.1029/2002JD002905

Schmidt A, O'Hagan A (2003) Bayesian inference for non-stationary spatial covariance structure via spatial deformations. Journal of the Royal Statistical Society Series B 65:743 – 758, DOI 10.1111/1467-9868.00413

Schmidt A, Guttorp P, O'Hagan A (2011) Considering covariates in the covariance structure of spatial processes. Environmetrics 22:487 – 500, DOI 10.1002/env.1101

Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A (2002) Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64(4):583–639, DOI https://doi.org/10.1111/1467-9868.00353, URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/1467-9868.00353

West M, Harrison J (1997) Bayesian Forecasting and Dynamic Models (2Nd Ed.). Springer-Verlag, Berlin, Heidelberg

Yan J, Cowles MK, Wang S, Armstrong MP (2007) Parallelizing mcmc for bayesian spatiotemporal geostatistical models. Statistics and Computing 17(4):323–335, DOI 10.1007/s11222-007-9022-2, URL http://dx.doi.org/10.1007/s11222-007-9022-2