

On Spatial (Skew) t Processes and Applications

Moreno Bevilacqua¹, Christian Caamaño², Reinaldo B. Arellano Valle³
and Víctor Morales-Oñate⁴

¹Departamento de Estadística, Universidad of Valparaíso, Chile

¹Millennium Nucleus Center for the Discovery of Structures in Complex Data , Chile,
`moreno.bevilacqua@uv.cl`

²Departamento de Estadística, Universidad del Bío-Bío, Concepción, Chile,
`chcaaman@ubiobio.cl`

³Departamento de Estadística, Pontificia Universidad Católica de Chile, Santiago, Chile,
`reivalle@mat.uc.cl`

⁴Departamento de Desarrollo, Ambiente y Territorio, Facultad Latinoamericana de
Ciencias Sociales, Quito, Ecuador `victor.morales@uv.cl`

December 18, 2018

Abstract

We propose a new model for regression and dependence analysis when addressing spatial or spatiotemporal data with possibly heavy tails and an asymmetric marginal distribution. We first propose a stationary process with t marginals obtained through scale mixing of a Gaussian process with an inverse square root process with Gamma marginals. We then generalize this construction by considering a skew-Gaussian process, thus obtaining a process with skew- t marginal distributions. For the proposed (skew) t process we study the second-order and geometrical properties and in the t case, we provide analytic expressions for the bivariate distribution. In an extensive simulation study, we investigate the use of the weighted pairwise likelihood as a method of estimation for the t process. Moreover we compare the performance of the optimal linear

predictor of the t process versus the optimal Gaussian predictor. Finally, the effectiveness of our methodology is illustrated by analyzing a georeferenced dataset on maximum temperatures in Australia.

Keywords: Heavy-tailed processes, Multivariate skew-normal distribution, Non-Gaussian processes, Pairwise likelihood.

1 Introduction

The focus of this work is on non-Gaussian models for stochastic processes (or random fields) that vary continuously in space or space-time. In particular, we aim to accommodate heavier tails than the ones induced by Gaussian processes and wish to allow possible asymmetry.

Indeed, in many geostatistical applications, including climatology, oceanography, the environment and the study of natural resources, the Gaussian framework is unrealistic because the observed data have specific features such as negative or positive asymmetry and/or heavy tails.

In recent years, different approaches have been proposed in order to analyze these kind of data. Transformation of Gaussian (trans-Gaussian) processes is a general method to model non-Gaussian spatial data obtained by applying some nonlinear transformations to the original data (De Oliveira et al., 1997; Allcroft and Glasbey, 2003; De Oliveira, 2006). Then statistical analyses can be carried out on the transformed data using any techniques available for Gaussian processes. However, it can be difficult to find an adequate nonlinear transformation and some appealing properties of the latent Gaussian process may not be inherited by the transformed process. A flexible trans-Gaussian process based on the Tukey $g-h$ distribution has been proposed in Xua and Genton (2017).

Wallin and Bolin (2015) proposed non-Gaussian processes derived from stochastic partial differential equations to model non-Gaussian spatial data. Nevertheless this approach is restricted to the Matérn covariance model and its statistical properties are much less understood than those of the Gaussian process.

The copula framework has been adapted in the spatial context in order to account for possible deviations from the Gaussian distribution, for instance in Kazianka and Pilz (2010), Masarotto and Varin (2012) and Gräler (2014).

Convolution of Gaussian and non-Gaussian processes is an appealing strategy for modeling spatial data with skewness. For instance, Zhang and El-Shaarawi (2010) proposed a Gaussian-Half Gaussian convolution in order to construct a process with a marginal distribution of the skew-Gaussian type (Azzalini and Capitanio, 2014). Zareifard et al. (2018) developed bayesian inference for the estimation of a process with asymmetric marginal distributions obtained through convolution of Gaussian and Log-Gaussian processes. Mahmoudian (2017) proposed a skew-Gaussian process using the skew-model proposed in Sahu et al. (2003). The resulting process is not mean square continuous and as a consequence it is not a suitable model for data exhibiting smooth behavior of the realization.

On the other hand, mixing of Gaussian and non-Gaussian processes is a useful strategy for modeling spatial data with heavy tails. For instance, Palacios and Steel (2006) proposed a Gaussian-Log-Gaussian scale mixing approach in order to accommodate the presence of possible outliers for spatial data and focused on Bayesian inference.

The t distribution is a parametric model that is able to accommodate flexible tail behavior, thus providing robust estimates against extreme data and it has been studied extensively in recent years (Lange et al., 1989; Fonseca et al., 2008; Ferrari and Arellano-Valle, 1996; Arellano-Valle et al., 2012). Stochastic processes with marginal t distributions have been introduced in Røislien and Omre (2006), Ma (2009), Ma (2010a) and DeBastiani et al. (2015), but as outlined in Genton and Zhang (2012), these models are not identifiable when only a single realization is available (which is typically the case for spatial and spatiotemporal data).

In this paper, we propose a process with marginal t distributions obtained through scale mixing of a standard Gaussian process with an inverse square root process with Gamma marginals. The latter is obtained through a rescaled sum of independent copies of a standard squared Gaussian process (Bevilacqua, Caamaño and Gaetan, 2018).

Although this can be viewed as a natural way to define a t process, the associated second-order, geometrical properties and bivariate distribution are somewhat unknown to the best of our knowledge. Some results can be found in Heyde and Leonenko (2005) and Finlay and Seneta (2006). We study the second-order and geometrical properties of the t process and we provide analytic expressions for the correlation and the bivariate distribution. It turns out that both depend on special functions, particularly the Gauss hypergeometric and Appell function of the fourth type (Gradshteyn and Ryzhik, 2007).

We then focus on processes with asymmetric marginal distributions and heavy tails. We first review the skew Gaussian process proposed in Zhang and El-Shaarawi (2010). For this process we provide an explicit expression of the finite dimensional distribution generalizing previous results in Alegría et al. (2017). We then propose a process with marginal distribution of the skew- t type (Azzalini and Capitanio, 2014) obtained through scale mixing of a skew-Gaussian with an inverse square root process with Gamma marginals.

Our proposals for the t and skew- t processes have two main features. First, they allow removal of any problem of identifiability (Genton and Zhang, 2012), and as a consequence, all the parameters can be estimated using one realization of the process. Second, the t and skew- t processes inherit the geometrical properties of the underlying Gaussian process. This implies that the mean square continuity and differentiability of the t and skew- t processes can be modeled using suitable parametric correlation models. From this point of view, two flexible correlation models are the Matérn model (Matérn, 1986) and the Generalized Wendland model (Gneiting, 2002; Bevilacqua, Faouzi, Furrer and Porcu, 2018).

For the t process estimation we propose the method of weighted pairwise likelihood (Lindsay, 1988; Varin et al., 2011; Bevilacqua and Gaetan, 2015) exploiting the bivariate distribution given in Theorem 2.3. In an extensive simulation study we investigate the performance of the weighted pairwise likelihood (*wpl*) method under different scenarios (temporal, spatial, and spatiotemporal) including when the degrees of freedom are supposed to be unknown. Moreover, we compare the performance of the optimal linear predictor of the t process with the optimal predictor of

the Gaussian process. Finally we apply the proposed methodology by analyzing a real data set of maximum temperature in Australia where, in this case, we consider a t process defined on a portion of the sphere (used as an approximation of the planet Earth) and use a correlation model depending on the great-circle distance (Gneiting, 2013).

The methodology considered in this paper is implemented in an upcoming version of the R package `GeoModels` (Bevilacqua and Morales-Oñate, 2018). In particular, the *wpl* estimation method has been implemented using the Open Computing Language (OpenCL) in order to reduce the computational costs associated with the Appell function evaluation.

The remainder of the paper is organized as follows. In Section 2 we introduce the t process, study the second-order and geometrical properties and provide an analytic expression for the bivariate distribution. In Section 3, we first study the finite dimensional distribution of the skew Gaussian process, and then we study the second-order properties of the skew- t process. In Section 4, we present a simulation study in order to investigate the performance of the *wpl* method when estimating the t process and the performance of the associated optimal linear predictor versus the optimal Gaussian predictor. In Section 5, we analyze a real data set of maximum temperature in Australia. Finally, in Section 6, we give some conclusions.

2 A stochastic process with t marginal distribution

For the rest of the paper, given a weakly stationary process $Q = \{Q(\mathbf{s}), \mathbf{s} \in A\}$ with $E(Q(\mathbf{s})) = \mu(\mathbf{s})$ and $Var(Q(\mathbf{s})) = \sigma^2$, we denote by $\rho_Q(\mathbf{h}) = Corr(Q(\mathbf{s}_i), Q(\mathbf{s}_j))$ its correlation function, where $\mathbf{h} = \mathbf{s}_i - \mathbf{s}_j \in A$ is the lag separation vector. For any set of distinct points $(\mathbf{s}_1, \dots, \mathbf{s}_n)^T$, $n \in \mathbb{N}$, we denote by $\mathbf{Q}_{ij} = (Q(\mathbf{s}_i), Q(\mathbf{s}_j))^T$, $i \neq j$, the bivariate random vector and by $\mathbf{Q} = (Q(\mathbf{s}_1), \dots, Q(\mathbf{s}_n))^T$ the multivariate random vector. Moreover, we denote with $f_{Q(\mathbf{s})}$ and $F_{Q(\mathbf{s})}$ the marginal probability density function (pdf) and cumulative distribution function (cdf) of $Q(\mathbf{s})$ respectively, with $f_{\mathbf{Q}_{ij}}$ the pdf of \mathbf{Q}_{ij} and with $f_{\mathbf{Q}}$ the pdf of \mathbf{Q} . Finally, we denote with Q^* the standardized process, *i.e.*, $Q^*(\mathbf{s}) := (Q(\mathbf{s}) - \mu(\mathbf{s}))/\sigma$.

For simplicity of presentation, we restrict the treatment to the spatial Euclidean setting $A \subseteq \mathbb{R}^d$. Nevertheless, the results presented in this paper can be applied to the spatiotemporal $A \subseteq \mathbb{R}^d \times \mathbb{R}$ or spherical $A \subseteq \mathbb{S}^d = \{\mathbf{s} \in \mathbb{R}^{d+1} : \|\mathbf{s}\| = R\}$, $R > 0$ setting (see Section 5).

As outlined in Palacios and Steel (2006), a general class of non-Gaussian processes with marginal heavy tails $T = \{T(\mathbf{s}), \mathbf{s} \in A\}$ can be obtained as scale mixture of $G^* = \{G^*(\mathbf{s}), \mathbf{s} \in A\}$, a standard Gaussian process:

$$T(\mathbf{s}) := \mu(\mathbf{s}) + \sigma M(\mathbf{s})^{-\frac{1}{2}} G^*(\mathbf{s}) \quad (2.1)$$

where $M = \{M(\mathbf{s}), \mathbf{s} \in A\}$ is a positive process independent of G^* , $\mu(\mathbf{s})$ is the location dependent mean and $\sigma > 0$ is a scale parameter. A typical parametric specification for the mean is given by $\mu(\mathbf{s}) = X(\mathbf{s})^T \boldsymbol{\beta}$ where $X(\mathbf{s}) \in \mathbb{R}^k$ is a vector of covariates and $\boldsymbol{\beta} \in \mathbb{R}^k$ but other types of parametric or nonparametric functions can be considered.

Henceforth, we call G^* the ‘parent’ process and with some abuse of notation we set $\rho(\mathbf{h}) := \rho_{G^*}(\mathbf{h})$ and $G := G^*$. Our proposal considers a mixing process $W_\nu = \{W_\nu(\mathbf{s}), \mathbf{s} \in A\}$ with marginal distribution $\Gamma(\nu/2, \nu/2)$ defined as $W_\nu(\mathbf{s}) := \sum_{i=1}^\nu G_i(\mathbf{s})^2 / \nu$ where $G_i, i = 1, \dots, \nu$ are independent copies of G with $E(W_\nu(\mathbf{s})) = 1$, $Var(W_\nu(\mathbf{s})) = 2/\nu$ and $\rho_{W_\nu}(\mathbf{h}) = \rho^2(\mathbf{h})$ (Bevilacqua, Caamaño and Gaetan, 2018).

If we consider a process $Y_\nu^* = \{Y_\nu^*(\mathbf{s}), \mathbf{s} \in A\}$ defined as $Y_\nu^*(\mathbf{s}) := W_\nu(\mathbf{s})^{-\frac{1}{2}} G(\mathbf{s})$, then, by construction, Y_ν^* has the marginal t distribution with ν degrees of freedom with pdf given by:

$$f_{Y_\nu^*(\mathbf{s})}(y; \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{y^2}{\nu}\right)^{-\frac{(\nu+1)}{2}}. \quad (2.2)$$

Then, we define the location-scale transformation process $Y_\nu = \{Y_\nu(\mathbf{s}), \mathbf{s} \in A\}$ as:

$$Y_\nu(\mathbf{s}) := \mu(\mathbf{s}) + \sigma Y_\nu^*(\mathbf{s}) \quad (2.3)$$

with $E(Y_\nu(\mathbf{s})) = \mu(\mathbf{s})$ and $Var(Y_\nu(\mathbf{s})) = \sigma^2 \nu / (\nu - 2)$, $\nu > 2$.

Remark 1: A possible drawback for the Gamma process W_ν is that it is a limited model due to the restrictions to the half-integers for the shape parameter. Actually,

in some special cases, it can assume any positive value greater than two. This feature is intimately related to the infinite divisibility of the squared Gaussian process $G^2 = \{G^2(\mathbf{s}), \mathbf{s} \in A\}$ as shown in Krishnaiah and Rao (1961). Characterization of the infinity divisibility of G^2 has been studied in Vere-Jones (1997), Bapat (1989), Griffiths (1970) and Eisenbaum and Kaspi (2006). In particular Bapat (1989) provides a characterization based on Ω , the correlation matrix associated with $\rho(\mathbf{h})$. Specifically, $\nu > 2$ if and only if $S\Omega^{-1}S$ is an M -matrix (Plemmons, 1977), where S is a signature matrix, *i.e.*, a diagonal matrix with entries either 1 or -1 . This condition is satisfied, for instance, by a stationary Gaussian random process G defined on $A = \mathbb{R}$ with an exponential correlation function.

Remark 2: The finite dimensional distribution of Y_ν^* is unknown to the best of our knowledge, but in principle, it can be derived by mixing the multivariate density associated with $W_\nu^{-\frac{1}{2}}$ with the multivariate standard Gaussian density. The multivariate Gamma density $f_{\mathbf{W}_\nu}$ was first discussed by Krishnamoorthy and Parthasarathy (1951) and its properties have been studied by different authors (Royen, 2004; Marcus, 2014). In the bivariate case, Vere-Jones (1967) showed that the bivariate Gamma distribution is infinite divisible, *i.e.* $\nu > 2$ in (A.2), irrespective of the correlation function. Note that this is consistent with the characterization given in Bapat (1989) since, given Ω_{ij} an arbitrary bivariate correlation matrix, $S\Omega_{ij}^{-1}S$ is an M -matrix. In Theorem 2.3 we provide the bivariate distribution of Y_ν^*

Note that, both W_ν and G in (2.3) are obtained through independent copies of the ‘parent’ Gaussian process with correlation $\rho(\mathbf{h})$. For this reason, henceforth, in some cases, we will call Y_ν^* a standard t process with underlying correlation $\rho(\mathbf{h})$.

In what follows, we make use of the Gauss hypergeometric function defined by (Gradshteyn and Ryzhik, 2007):

$${}_2F_1(a, b, c; x) = \sum_{k=0}^{\infty} \frac{(a)_k (b)_k}{(c)_k} \frac{x^k}{k!}, \quad |x| < 1$$

with $(s)_k = \Gamma(s+k)/\Gamma(s)$ for $k \in \mathbb{N} \cup \{0\}$ being the Pochhammer symbol. We also consider the Appell hypergeometric function of the fourth type (Gradshteyn and

Ryzhik, 2007) defined as:

$$F_4(a, b; c, c'; w, z) = \sum_{k=0}^{\infty} \sum_{m=0}^{\infty} \frac{(a)_{k+m} (b)_{k+m} w^k z^m}{k! m! (c)_k (c')_m}, \quad |\sqrt{w}| + |\sqrt{z}| < 1.$$

The following Theorem gives an analytic expression for $\rho_{Y_\nu^*}(\mathbf{h})$ in terms of the Gauss hypergeometric function. The proof can be found in the Appendix.

Theorem 2.1. *Let Y_ν^* be a standardized t process with underlying correlation $\rho(\mathbf{h})$. Then:*

$$\rho_{Y_\nu^*}(\mathbf{h}) = \frac{(\nu - 2)\Gamma^2\left(\frac{\nu-1}{2}\right)}{2\Gamma^2\left(\frac{\nu}{2}\right)} \left[{}_2F_1\left(\frac{1}{2}, \frac{1}{2}; \frac{\nu}{2}; \rho^2(\mathbf{h})\right) \rho(\mathbf{h}) \right]. \quad (2.4)$$

The following Theorem depicts some features of the t process. It turns out that nice properties such as stationarity, mean-square continuity and degrees of mean-square differentiability can be inherited from the ‘parent’ Gaussian process G . Further, the t process has long-range dependence when the ‘parent’ Gaussian process has long-range dependence and this can be achieved when the correlation has some specific features. For instance, the generalized Cauchy (Gneiting and Schlather, 2004; Lim and Teo, 2009) and Dagum (Berg et al., 2008) correlation models can lead to a Gaussian process with long range dependence.

Finally, an appealing and intuitive feature is that the correlation of Y_ν^* approaches the correlation of G when $\nu \rightarrow \infty$.

Theorem 2.2. *Let Y_ν^* , $\nu > 2$ be a standardized t process with underlying correlation $\rho(\mathbf{h})$. Then:*

- a) Y_ν^* is also weakly stationary;
- b) Y_ν^* is mean-square continuous if and only if G is mean-square continuous;
- c) Y_ν^* is m -times mean-square differentiable if G is m -times mean-square differentiable;
- d) Y_ν^* is a long-range dependent process if and only if G is a long-range dependent process
- e) $\rho_{Y_\nu^*}(\mathbf{h}) \leq \rho(\mathbf{h})$ and $\lim_{\nu \rightarrow \infty} \rho_{Y_\nu^*}(\mathbf{h}) = \rho(\mathbf{h})$.

Proof. If G is a weakly stationary Gaussian process with correlation $\rho(\mathbf{h})$ then from (2.4) it is straightforward to see that Y_ν^* is also weakly stationary. Points b) and c) can be shown using the relations between the geometrical properties of a stationary process and the associated correlation. Specifically, by Stein (1999), the mean-square continuity and the m -times mean-square differentiability of Y_ν^* are equivalent to the continuity and $2m$ -times differentiability of $\rho_{Y_\nu^*}(\mathbf{h})$ at $\mathbf{h} = \mathbf{0}$, respectively. Define the function

$$g(x) = a(\nu) \left[{}_2F_1 \left(\frac{1}{2}, \frac{1}{2}; \frac{\nu}{2}; x^2 \right) x \right]$$

where $a(\nu) = \frac{(\nu-2)\Gamma^2(\frac{\nu-1}{2})}{2\Gamma^2(\frac{\nu}{2})}$. Then, $\rho_{Y_\nu^*}(\mathbf{h}) = g\{\rho(\mathbf{h})\}$ and for any $-1 \leq x \leq 1$, $g(x)$ is a continuous and infinitely differentiable function. Hence, $\rho_{Y_\nu^*}(\mathbf{h})$ is continuous at $\mathbf{h} = \mathbf{0}$ if and only if $\rho(\mathbf{h})$ is continuous at $\mathbf{h} = \mathbf{0}$, which implies that Y_ν^* is mean-square continuous if and only if G is mean-square continuous. Furthermore, the $2m$ -th derivative of $\rho_{Y_\nu^*}(\mathbf{h})$ at $\mathbf{h} = \mathbf{0}$ (i.e., $\rho_{Y_\nu^*}^{2m}(\mathbf{0})$) exists if $\rho^{2m}(\mathbf{0})$ exists, which implies that Y_ν^* is m -times mean-square differentiable if G is m -times mean-square differentiable.

Point d) can be easily shown recalling that a process F is long-range dependent if the correlation of F is such that $\int_{\mathbb{R}_+^n} |\rho_F(\mathbf{h})| d^n \mathbf{h} = \infty$ (Lim and Teo, 2009). Direct inspection, using series expansion of the hypergeometric function, shows that $\int_{\mathbb{R}_+^n} |\rho_{Y_\nu^*}(\mathbf{h})| d^n \mathbf{h} = \infty$ if and only if $\int_{\mathbb{R}_+^n} |\rho(\mathbf{h})| d^n \mathbf{h} = \infty$ and, as a consequence, Y_ν^* has long-range dependence if and only if G has long-range dependence.

Finally, since $0 < a(\nu) \leq 1$ for $\nu > 2$, then $\rho_{Y_\nu^*}(\mathbf{h}) \leq \rho(\mathbf{h})$. Moreover, $\lim_{\nu \rightarrow \infty} a(\nu) = 1$ and using series expansion of the hypergeometric function, it can be shown that $\lim_{\nu \rightarrow \infty} {}_2F_1 \left(\frac{1}{2}, \frac{1}{2}; \frac{\nu}{2}; \rho^2(\mathbf{h}) \right) = 1$. This implies $\lim_{\nu \rightarrow \infty} \rho_{Y_\nu^*}(\mathbf{h}) = \rho(\mathbf{h})$. \square

A process with t marginals that is not mean-square continuous can be obtained by introducing a nugget effect, i.e., a discontinuity of $\rho_{Y_\nu^*}(\mathbf{h})$ at the origin. This can be easily achieved by replacing $\rho(\mathbf{h})$ in (2.4) with $\rho^*(\mathbf{h}) = 1$ if $\mathbf{h} = \mathbf{0}$ and $\rho^*(\mathbf{h}) = (1 - \tau^2)\rho(\mathbf{h})$ otherwise, where $0 \leq \tau^2 < 1$ represents the underlying nugget effect.

Since the t process inherits the geometrical properties of the ‘parent’ Gaussian process, the choice of the covariance function is crucial. Two flexible models that

allow parametrizing in a continuous fashion the mean square differentiability of a Gaussian process and its sample paths are as follows:

1. the Matérn correlation function (Matérn, 1986)

$$\mathcal{M}_{\alpha,\psi}(\mathbf{h}) = \frac{2^{1-\psi}}{\Gamma(\psi)} (||\mathbf{h}||/\alpha)^\psi \mathcal{K}_\psi (||\mathbf{h}||/\alpha), \quad ||\mathbf{h}|| \geq 0. \quad (2.5)$$

where \mathcal{K}_ψ is a modified Bessel function of the second kind of order ψ . Here, $\alpha > 0$ and $\psi > 0$ guarantee the positive definiteness of the model in any dimension.

2. the Generalized Wendland correlation function (Gneiting, 2002), defined for $\psi > 0$ as:

$$\mathcal{GW}_{\alpha,\psi,\delta}(\mathbf{h}) := \begin{cases} \frac{\int_{||\mathbf{h}||/\alpha}^1 u(u^2 - (||\mathbf{h}||/\alpha)^2)^{\psi-1} (1-u)^\delta du}{B(2\psi,\delta+1)} & ||\mathbf{h}|| < \alpha \\ 0 & \text{otherwise} \end{cases}, \quad (2.6)$$

and for $\psi = 0$ as:

$$\mathcal{GW}_{\alpha,0,\delta}(\mathbf{h}) := \begin{cases} (1 - ||\mathbf{h}||/\alpha)^\delta & ||\mathbf{h}|| < \alpha \\ 0 & \text{otherwise} \end{cases}. \quad (2.7)$$

Here $B(\cdot, \cdot)$ is the Beta function and $\alpha > 0$, $\psi > 0$, and $\delta \geq (d+1)/2 + \psi$ guarantee the positive definiteness of the model in \mathbb{R}^d .

In particular for a positive integer k , the sample paths of a Gaussian process are k times differentiable if and only if $\psi > k$ in the Matérn case (Stein, 1999) and if and only if $\psi > k - 1/2$ in the Generalized Wendland case (Bevilacqua, Faouzi, Furrer and Porcu, 2018). Additionally, the Generalized Wendland correlation is compactly supported, an interesting feature from computational point of view (Furrer et al., 2013), which is inherited by the t process since $\rho(\mathbf{h}) = 0$ implies $\rho_{Y_\nu^*}(\mathbf{h}) = 0$.

In order to illustrate some geometric features of the t process, we first compare the correlation functions of the Gaussian and t processes using an underlying Matérn model.

In Figure 1 (left part) we compare $\rho_{Y_\nu^*}(\mathbf{h})$ when $\nu = 5, 10$ with the correlation of the ‘parent’ Gaussian process $\rho(\mathbf{h}) = \mathcal{M}_{1.5, \alpha^*}(\mathbf{h})$ where α^* is chosen such that the practical range is 0.2. It is apparent that when increasing the degrees of freedom $\rho_{Y_\nu^*}(\mathbf{h})$ approaches $\rho(\mathbf{h})$ and that the smoothness at the origin of $\rho_{Y_\nu^*}(\mathbf{h})$ is inherited by the smoothness of the Gaussian correlation $\rho(\mathbf{h})$, as depicted in Theorem 2.2. On the right side of Figure (1) we compare a kernel nonparametric density estimation of a realization of G and a realization of Y_6^* (approximately 10000 location sites in the unit square) using $\rho(\mathbf{h}) = \mathcal{M}_{1.5, \alpha^*}(\mathbf{h})$.

In Figure 2 (a) and (b), we compare, from left to right, two realizations of G with $\rho(\mathbf{h}) = \mathcal{M}_{0.5, \alpha^*}(\mathbf{h})$ and $\rho(\mathbf{h}) = \mathcal{M}_{1.5, \alpha^*}(\mathbf{h})$ where α^* is chosen such that the practical range is 0.2. In this case, the sample paths of G are zero and one times differentiable. From the bottom part of Figure 2 (c) and (d) it can be appreciated that this feature is inherited by the associated realizations of Y_6^* .

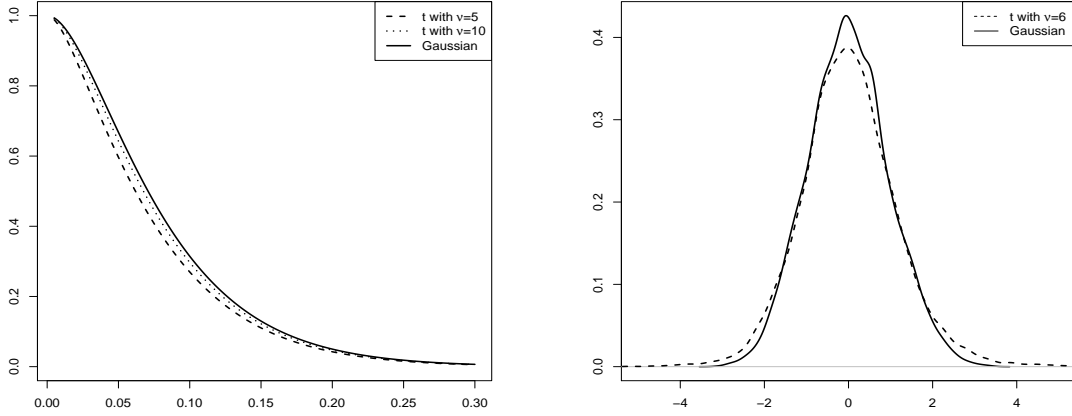


Figure 1: Left part: comparison of $\rho_{Y_\nu^*}(\mathbf{h})$, $\nu = 5, 10$ with the correlation $\rho(\mathbf{h})$ of the ‘parent’ Gaussian process G when $\rho(\mathbf{h}) = \mathcal{M}_{1.5, \alpha^*}(\mathbf{h})$ with α^* such that the practical range is 0.2. Right part: a comparison of a nonparametric kernel density estimation of realizations from G and from the t process Y_6^* .

We now consider the bivariate random vector associated with Y_ν^* defined by:

$$\mathbf{Y}_{\nu;ij}^* = \mathbf{W}_{\nu;ij}^{-\frac{1}{2}} \circ \mathbf{G}_{ij}$$

where \circ denotes the Schur product vector.

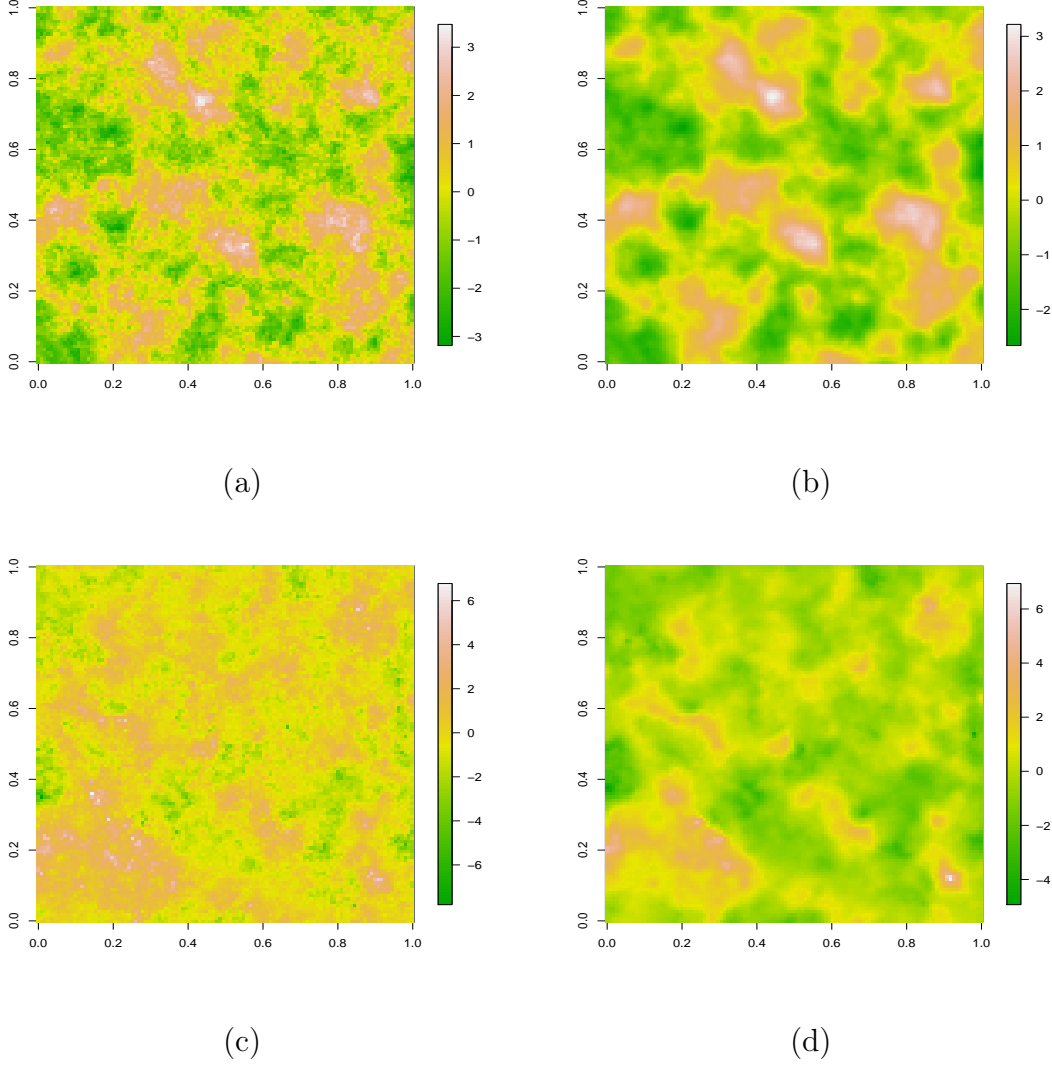


Figure 2: Upper part: two realizations of the ‘parent’ Gaussian process G on $[0, 1]^2$ with (a) $\rho(\mathbf{h}) = \mathcal{M}_{0.5, \alpha^*}(\mathbf{h})$ and (b) $\rho(\mathbf{h}) = \mathcal{M}_{1.5, \alpha^*}(\mathbf{h})$ (from left to right) with α^* such that the practical range is approximatively 0.2. Bottom part: (c) and (d) associated realizations of the t process Y_6^* .

The following Theorem gives the pdf of $\mathbf{Y}_{\nu; ij}^*$ in terms of the Appell function F_4 . It can be viewed as a generalization of the generalized bivariate t distribution proposed in Miller (1968). The proof has been deferred to the Appendix.

Theorem 2.3. *Let Y_ν^* , $\nu > 2$ be a standard t process with underlying correlation*

$\rho(\mathbf{h})$. Then:

$$f_{\mathbf{Y}_{\nu;ij}^*}(y_i, y_j) = \frac{\nu^\nu l_{ij}^{-\frac{(\nu+1)}{2}} \Gamma^2\left(\frac{\nu+1}{2}\right)}{\pi \Gamma^2\left(\frac{\nu}{2}\right) (1 - \rho^2(\mathbf{h}))^{-(\nu+1)/2}} F_4\left(\frac{\nu+1}{2}, \frac{\nu+1}{2}, \frac{1}{2}, \frac{\nu}{2}; \frac{\rho^2(\mathbf{h}) y_i^2 y_j^2}{l_{ij}}, \frac{\nu^2 \rho^2(\mathbf{h})}{l_{ij}}\right) \\ + \frac{\rho(\mathbf{h}) y_i y_j \nu^{\nu+2} l_{ij}^{-\frac{\nu}{2}-1}}{2\pi (1 - \rho^2(\mathbf{h}))^{-\frac{(\nu+1)}{2}}} F_4\left(\frac{\nu}{2} + 1, \frac{\nu}{2} + 1, \frac{3}{2}, \frac{\nu}{2}; \frac{\rho^2(\mathbf{h}) y_i^2 y_j^2}{l_{ij}}, \frac{\nu^2 \rho^2(\mathbf{h})}{l_{ij}}\right) \quad (2.8)$$

where $l_{ij} = [(y_i^2 + \nu)(y_j^2 + \nu)]$.

Remark 3: Note that $f_{\mathbf{Y}_{\nu;ij}^*}(y_i, y_j)$ is defined for $\nu > 2$ irrespectively of the correlation function since it is obtained from a bivariate Gamma distribution (see Remark 2). Moreover, when $\rho(\mathbf{h}) = 0$, according to (4.4) and using the identity ${}_2F_1(a, b; c'; 0) = 1$, we obtain $F_4(a, b; c, c'; 0, 0) = 1$, and as a consequence, $f_{\mathbf{Y}_{\nu;ij}^*}(y_i, y_j)$ can be written as the product of two independent t random variables with ν degrees of freedom. Thus, zero pairwise correlation implies pairwise independence, as in the Gaussian case.

Finally, the bivariate density of the process Y_ν is easily obtained from (4.3):

$$f_{\mathbf{Y}_{\nu;ij}}(y_i, y_j) = \frac{1}{\sigma^2} f_{\mathbf{Y}_{\nu;ij}^*}\left(\frac{y_i - \mu_i}{\sigma}, \frac{y_j - \mu_j}{\sigma}\right). \quad (2.9)$$

3 A stochastic process with skew-t marginal distribution

In this section we first review the skew-Gaussian process proposed in Zhang and El-Shaarawi (2010). For this process, we provide an explicit expression for the finite dimensional distribution generalizing previous results in Alegría et al. (2017). Then, using this skew-Gaussian process, we propose a generalization of the t process Y_ν obtaining a new process with marginal distribution of the skew-t type (Azzalini and Capitanio, 2014).

Following Zhang and El-Shaarawi (2010) a general construction for a process with asymmetric marginal distribution is given by:

$$U_\eta(\mathbf{s}) := g(\mathbf{s}) + \eta |X_1(\mathbf{s})| + \omega X_2(\mathbf{s}), \quad \mathbf{s} \in A \subset \mathbb{R}^d \quad (3.1)$$

where $\eta \in \mathbb{R}$ and X_i $i = 1, 2$ are two independents copies of a process $X = \{X(\mathbf{s}), \mathbf{s} \in$

$A\}$ with symmetric marginals. The parameters η and ω allow modeling the asymmetry and variance of the process simultaneously.

Zhang and El-Shaarawi (2010) studied the second-order properties of U_η when $X \equiv G$. In this case, U_η has marginal distribution of the skew Gaussian type given by (Azzalini and Capitanio, 2014):

$$f_{U_\eta(\mathbf{s})}(u) = \frac{2}{(\eta^2 + \omega^2)^{1/2}} \phi\left(\frac{(u - g(\mathbf{s}))}{(\eta^2 + \omega^2)^{1/2}}\right) \Phi\left(\frac{\eta(u - g(\mathbf{s}))}{\omega(\eta^2 + \omega^2)^{1/2}}\right) \quad (3.2)$$

with $E(U_\eta(\mathbf{s})) = g(\mathbf{s}) + \eta(2/\pi)^{1/2}$, $Var(U_\eta(\mathbf{s})) = \omega^2 + \eta^2(1 - 2/\pi)$ and with correlation function given:

$$\rho_{U_\eta}(\mathbf{h}) = \frac{2\eta^2}{\pi\omega^2 + \eta^2(\pi - 2)} \left((1 - \rho^2(\mathbf{h}))^{1/2} + \rho(\mathbf{h}) \arcsin(\rho(\mathbf{h})) - 1 \right) + \frac{\omega^2 \rho(\mathbf{h})}{\omega^2 + \eta^2(1 - 2/\pi)}. \quad (3.3)$$

The following theorem generalizes the results in Alegría et al. (2017) and gives an explicit closed-form expression for the pdf of the random vector \mathbf{U}_η . The proof can be found in the appendix.

Theorem 3.1. *Let $U_\eta(\mathbf{s}) = g(\mathbf{s}) + \eta|X_1(\mathbf{s})| + \omega X_2(\mathbf{s})$ where X_i $i = 1, 2$ are two independent copies of G the ‘parent’ Gaussian process. Then:*

$$f_{\mathbf{U}_\eta}(\mathbf{u}) = 2 \sum_{l=1}^{2^{n-1}} \phi_n(\mathbf{u} - \boldsymbol{\alpha}; \mathbf{A}_l) \Phi_n(\mathbf{c}_l; \mathbf{0}, \mathbf{B}_l) \quad (3.4)$$

where

$$\begin{aligned} \mathbf{A}_l &= \omega^2 \Omega + \eta^2 \Omega_l \\ \mathbf{c}_l &= \eta \Omega_l (\omega^2 \Omega + \eta^2 \Omega_l)^{-1} (\mathbf{u} - \boldsymbol{\alpha}) \\ \mathbf{B}_l &= \Omega_l - \eta^2 \Omega_l (\omega^2 \Omega + \eta^2 \Omega_l)^{-1} \Omega_l \\ \boldsymbol{\alpha} &= [g(\mathbf{s}_i)]_{i=1}^n \end{aligned}$$

and the Ω_l ’s are correlation matrices that depend on the correlation matrix Ω .

Some comments are in order. First, note that $f_{\mathbf{U}}$ can be viewed as a generalization of the multivariate skew-Gaussian distribution proposed in Azzalini and Dalla-Valle (1996). Second, using Theorem (3.1), it can be easily shown that the

consistency conditions given in Mahmoudian (2018) are satisfied. Third, it is apparent that likelihood-based methods for the skew-Gaussian process are impractical from computational point of view even for a relatively small dataset.

To obtain a process with skew- t marginal distributions (Azzalini and Capitanio, 2014), we replace the process G in (2.3) with the process U_η . Specifically, we consider a process $S_{\nu,\eta} = \{S_{\nu,\eta}(\mathbf{s}), \mathbf{s} \in A\}$ defined as

$$S_{\nu,\eta}(\mathbf{s}) := \mu(\mathbf{s}) + \sigma W_\nu(\mathbf{s})^{-\frac{1}{2}} U_\eta(\mathbf{s}) \quad (3.5)$$

where W_ν and U_η are supposed to be independent. In (3.1) we assume $g(\mathbf{s}) = 0$ and $\eta^2 + \omega^2 = 1$. The marginal distribution of $S_{\nu,\eta}^*$ has the skew- t marginal given by (Azzalini and Capitanio, 2014):

$$f_{S_{\nu,\eta}^*}(g) = 2f_{Y_\nu^*}(\mathbf{s})(g; \nu) F_{Y_\nu^*}(\mathbf{s}) \left(\eta g \sqrt{\frac{\nu+1}{\nu+g^2}}; \nu+1 \right) \quad (3.6)$$

with $E(S_{\nu,\eta}^*(\mathbf{s})) = \frac{\sqrt{\nu}\Gamma(\frac{\nu-1}{2})\eta}{\sqrt{\pi}\Gamma(\frac{\nu}{2})}$, and $Var(S_{\nu,\eta}^*(\mathbf{s})) = \left[\frac{\nu}{\nu-2}(1+\eta^2) - \frac{\nu\Gamma^2(\frac{\nu-1}{2})\eta^2}{\pi\Gamma^2(\frac{\nu}{2})} \right]$.

If $\eta = 0$, (3.6) reduces to a marginal t density given in (2.2) and if $\nu \rightarrow \infty$, (3.6) converges to a skew-normal distribution. Moreover, combining (2.4) and (3.3) the correlation function of the skew- t process is given by:

$$\begin{aligned} \rho_{S_{\nu,\eta}^*}(\mathbf{h}) &= \frac{\pi(\nu-2)\Gamma^2(\frac{\nu-1}{2})}{2[\pi\Gamma^2(\frac{\nu}{2})(1+\eta^2) - \eta^2(\nu-2)\Gamma^2(\frac{\nu-1}{2})]} \\ &\times \left[{}_2F_1\left(\frac{1}{2}, \frac{1}{2}; \frac{\nu}{2}; \rho^2(\mathbf{h})\right) \left\{ (1+\eta^2(1-\frac{2}{\pi}))\rho_{U_\eta}(\mathbf{h}) + \frac{2\eta^2}{\pi} \right\} - \frac{2\eta^2}{\pi} \right]. \end{aligned} \quad (3.7)$$

Note that $\rho_{S_{\nu,\eta}^*}(\mathbf{h}) = \rho_{S_{\nu,-\eta}^*}(\mathbf{h})$ that is, as in the skew-Gaussian process U_η , the correlation is invariant with respect to positive or negative asymmetry and using similar arguments of Theorem 2.2 point d), it can be shown that $\lim_{\nu \rightarrow \infty} \rho_{S_{\nu,\eta}^*}(\mathbf{h}) = \rho_{U_\eta}(\mathbf{h})$.

Finally, following the steps of the proof of Theorem 2.2, it can be shown that nice properties such as stationarity, mean-square continuity, degrees of mean-square differentiability and long-range dependence can be inherited by the skew or skew- t process from the ‘parent’ Gaussian process G .

Figure 3, left part, compares $\rho_{S_{6,0.9}^*}(\mathbf{h})$ and $\rho_{S_{6,0}^*}(\mathbf{h}) = \rho_{Y_6^*}(\mathbf{h})$ with the underlying correlation $\rho(\mathbf{h}) = \mathcal{GW}_{0.3,1,5}(\mathbf{h})$. The right part shows a realization of $S_{6,0.9}^*$.

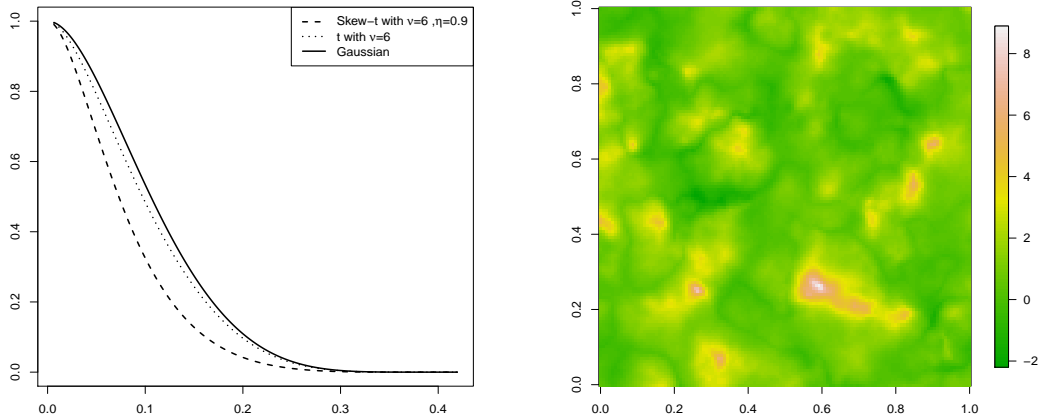


Figure 3: From left to right: a) comparison between $\rho_{S_{6,0.9}^*}(\mathbf{h})$, $\rho_{S_{6,0}^*}(\mathbf{h}) = \rho_{Y_6^*}(\mathbf{h})$ and the underlying correlation $\rho(\mathbf{h}) = \mathcal{GW}_{0.3,1,5}(\mathbf{h})$; b) a realization from $S_{6,0.9}^*(\mathbf{s})$.

4 Numerical examples

The main goals of this section are twofold: on the one hand, we analyze the performance of the *wpl* method when estimating the t process assuming ν known or unknown. Following Remark 1 in Section 2, we consider the cases when $\nu > 2$ or $\nu = 3, 4, \dots$. In the latter case, we give a practical solution for fixing the degrees of freedom parameter to a positive integer value through a two-step estimation.

On the other hand, we compare the performance of the optimal linear predictor of the t process using (2.4) versus the optimal predictor of the Gaussian process.

4.1 Weighted pairwise likelihood estimation

Let $\mathbf{Y} = (y_1, \dots, y_n)^T$ be a realization of the t random process Y_ν defined in equation (2.3) observed at distinct spatial locations $\mathbf{s}_1, \dots, \mathbf{s}_n$, $\mathbf{s}_i \in A$ and let $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \nu, \sigma^2, \boldsymbol{\alpha}^T)$ be the vector of unknown parameters where $\boldsymbol{\alpha}$ is the vector parameter associated with the correlation model of the ‘parent’ Gaussian process. The method of *wpl*, (Lindsay, 1988; Varin et al., 2011) combines the bivariate distributions of all possible distinct pairs of observations. The pairwise likelihood function

is given by

$$\text{pl}(\boldsymbol{\theta}) := \sum_{i=1}^{n-1} \sum_{j=i+1}^n \log(f_{\mathbf{Y}_{\nu;ij}}(y_i, y_j; \boldsymbol{\theta})) c_{ij} \quad (4.1)$$

where $f_{\mathbf{Y}_{\nu;ij}}(y_i, y_j; \boldsymbol{\theta})$ is the bivariate density in (2.9) and c_{ij} is a nonnegative suitable weight. The choice of cut-off weights, namely

$$c_{ij} = \begin{cases} 1 & \|\mathbf{s}_i - \mathbf{s}_j\| \leq d_{ij} \\ 0 & \text{otherwise} \end{cases}, \quad (4.2)$$

for a positive value of d_{ij} , can be motivated by its simplicity and by observing that the dependence between observations that are distant is weak. Therefore, the use of all pairs may skew the information confined in pairs of near observations (Bevilacqua and Gaetan, 2015; Joe and Lee, 2009). The maximum *wpl* estimator is given by

$$\hat{\boldsymbol{\theta}} := \operatorname{argmax}_{\boldsymbol{\theta}} \text{pl}(\boldsymbol{\theta})$$

and, arguing as in Bevilacqua et al. (2012) and Bevilacqua and Gaetan (2015), it can be shown that, under increasing domain asymptotics, $\hat{\boldsymbol{\theta}}$ is consistent and asymptotically Gaussian with the asymptotic covariance matrix given by $\mathcal{G}_n^{-1}(\boldsymbol{\theta})$ the inverse of the Godambe information $\mathcal{G}_n(\boldsymbol{\theta}) := \mathcal{H}_n(\boldsymbol{\theta}) \mathcal{J}_n(\boldsymbol{\theta})^{-1} \mathcal{H}_n(\boldsymbol{\theta})$, where $\mathcal{H}_n(\boldsymbol{\theta}) := \mathbb{E}[-\nabla^2 \text{pl}(\boldsymbol{\theta})]$ and $\mathcal{J}_n(\boldsymbol{\theta}) := \text{Var}[\nabla \text{pl}(\boldsymbol{\theta})]$. Standard error estimation can be obtained considering the square root diagonal elements of $\mathcal{G}_n^{-1}(\hat{\boldsymbol{\theta}})$. Moreover, model selection can be performed by considering two information criterion, defined as

$$\text{PLIC} := -2 \text{pl}(\hat{\boldsymbol{\theta}}) + 2 \text{tr}(\mathcal{H}_n(\hat{\boldsymbol{\theta}}) \mathcal{G}_n^{-1}(\hat{\boldsymbol{\theta}})), \quad \text{BLIC} := -2 \text{pl}(\hat{\boldsymbol{\theta}}) + \log(n) \text{tr}(\mathcal{H}_n(\hat{\boldsymbol{\theta}}) \mathcal{G}_n^{-1}(\hat{\boldsymbol{\theta}}))$$

which are composite likelihood version of the Akaike information criterion (AIC) and Bayesian information criterion (BIC) respectively (Varin and Vidoni, 2005; Gao and Song, 2010). Note that, the computation of standard errors, PLIC and BLIC require evaluation of the matrices $\mathcal{H}_n(\hat{\boldsymbol{\theta}})$ and $\mathcal{J}_n(\hat{\boldsymbol{\theta}})$. However, the evaluation of $\mathcal{J}_n(\hat{\boldsymbol{\theta}})$ is computationally unfeasible for large datasets and in this case subsampling techniques can be used in order to estimate $\mathcal{J}_n(\boldsymbol{\theta})$ as in Bevilacqua et al. (2012) and Heagerty and Lele (1998). A straightforward and more robust alternative is parametric bootstrap estimation of $\mathcal{G}_n^{-1}(\boldsymbol{\theta})$ (Bai et al., 2014). We adopt the second strategy in Section 5.

4.2 Performance of the weighted pairwise likelihood estimation

Following DiCiccio and Monti (2011) and Arellano-Valle and Azzalini (2013) we consider a reparametrization for the t process by using the inverse of degrees of freedom, $\lambda = 1/\nu$. In the standard i.i.d case this kind of parametrization has proven effective for solving some problems associated with the singularity of the Fisher information matrix. In our simulation study we consider three possible scenarios i.e. a t process observed on a subset of \mathbb{R} , \mathbb{R}^2 and $\mathbb{R}^2 \times \mathbb{R}$:

1. We consider points $s_i \in A = [0, 1]$, $i = 1, \dots, N$ and an exponential correlation function for the ‘parent’ Gaussian process. Then, according to Remark 1 in section 2, in this specific case all the parameters (including $0 < \lambda < 1/2$) can be jointly estimated. We simulate, using Cholesky decomposition, 500 realizations of a t process observed on a regular transect $s_1 = 0, s_2 = 0.002, \dots, s_{501} = 1$. We consider two mean regression parameters, that is, $\mu(s_i) = \beta_0 + \beta_1 u(s_i)$ with $\beta_0 = 0.5$, $\beta_1 = -0.25$ where $u(s_i)$ is a realization from a $U(0, 1)$. Then we set $\lambda = 1/\nu$, $\nu = 3, 6, 9$ and $\sigma^2 = 1$.

As correlation model we consider $\rho(h) = \mathcal{M}_{\alpha, 0.5}(h) = e^{-|h|/\alpha}$ with $\alpha = 0.1/3$ and in the *wpl* estimation we consider a cut-off weight function with $d_{ij} = 0.002$. Table 1 shows the bias and mean square error (MSE) associated with λ , β_0 , β_1 , α and σ^2 .

2. We consider points $\mathbf{s}_i \in A = [0, 1]^2$, $i = 1, \dots, N$. Specifically, we simulate, using Cholesky decomposition, 500 realizations of a t process observed at $N = 1000$ spatial location sites uniformly distributed in the unit square. Regression, variance and (inverse of) degrees of freedom parameters have been set as in the first scenario. As an isotropic parametric correlation model, $\rho(\mathbf{h}) = \mathcal{GW}_{\alpha, 0.4}(\mathbf{h})$ with $\alpha = 0.2$ is considered. In the *wpl* estimation we consider a cut-off weight function with $d_{ij} = 0.04$ and for each simulation we estimate with *wpl*, assuming the degrees of freedom are fixed and known.

We also consider the more realistic case when the (inverse of) degrees of free-

dom are supposed to be unknown. Recall that from Remark 1, ν must be fixed to a positive integer $\nu = 3, 4, \dots$. A brute force approach considers different wpl estimates using a fixed $\lambda = 1/\nu$, $\nu = 3, 4, \dots$ and then simply keeps the estimate with the best *PLIC* or *BLIC*. We propose a computationally easier approach by considering a two-step method. In the first step, an estimation for $0 < \lambda < 1/2$ can be obtained maximizing the wpl function. This is possible since the bivariate t distribution is well defined for $0 < \lambda < 1/2$ (see Remark 3). In the second step ν is fixed equal to the rounded value of $1/\hat{\lambda}_1$ where $\hat{\lambda}_1$ is the estimation at first step. (If at the first step, the estimation of $1/\hat{\lambda}_1$ is lower than 2.5, then it is rounded to 3).

Table 2 shows the bias and MSE associated with β_0 , β_1 , α and σ^2 when estimating with wpl , assuming (the inverse of) degrees of freedom are 1) known and fixed, and 2) unknown and fixed using a two-step estimation and Figure 4 shows the boxplots of the wpl estimates for the case 1) and 2).

3. We consider points $(\mathbf{s}_i, t_k) \in A = [0, 1]^2 \times [0.5, 10]$, $i = 1, \dots, N$, $k = 1, \dots, T$. Specifically, we simulate, using Cholesky decomposition, 500 realizations of a t process observed at $N = 80$ spatial location sites uniformly distributed in the unit square and $T = 20$ temporal instants $t_1 = 0.5$, $t_2 = 1$, \dots , $t_{20} = 10$. As spatially isotropic and temporally symmetric correlation model we consider a special case of the nonseparable class proposed in Porcu et al. (2018)

$$\rho(\mathbf{h}, u) = \frac{1}{\gamma(u/\alpha_t)^{2.5}} \mathcal{GW}_{\alpha_s \gamma(u/\alpha_t)^{0.5}, 0.4}(\mathbf{h}) \quad (4.3)$$

where $\gamma(u) = (1 + |u|)$, $\alpha_t > 0$ is a temporal scale parameter and $\alpha_s > 0$ is a spatial compact support. Regression, variance and (inverse of) degrees of freedom parameters have been set as in the first two settings, and additionally, we set $\alpha_s = 0.3$ and $\alpha_t = 0.5$. In the wpl estimation we consider a spacetime cut off weight function with $d_{ij} = 0.05$ and $d_{lk} = 0.5$. Table 3 shows the bias and MSE associated with β_0 , β_1 , α_s , α_t and σ^2 when estimating with wpl assuming (the inverse of) degrees of freedom 1) known and fixed, 2) unknown and fixed using the two-step estimation depicted in the second scenario. Finally, Figure 5 shows the boxplots of the wpl estimates for the case 1) and 2).

As a general comment, the distribution of the estimates are quite symmetric, numerically stable and with very few outliers for the three scenarios. In Scenario 1, the MSE of $\lambda = 1/\nu$ slightly decreases when increasing ν . Moreover, in Tables 2 and 3, it can be appreciated that only the estimation of σ^2 is affected when considering a two step estimation. Specifically, the MSE of σ^2 slightly increases with respect to the one-step estimation, *i.e.*, when the degrees of freedom are supposed to be known.

λ	1/3		1/6		1/9	
	Bias	MSE	Bias	MSE	Bias	MSE
$\hat{\lambda}$	-0.01027	0.00326	-0.01040	0.00215	-0.00803	0.00160
$\hat{\beta}_0$	-0.00150	0.06463	-0.00621	0.06180	0.00029	0.06523
$\hat{\beta}_1$	-0.00067	0.00065	-0.00231	0.00049	-0.00197	0.00049
$\hat{\alpha}$	-0.00218	0.00006	-0.00286	0.00006	-0.00190	0.00007
$\hat{\sigma}^2$	-0.02668	0.07401	-0.05659	0.06838	-0.03073	0.06923

Table 1: Bias and MSE when estimating with *wpl* the t process with $\lambda = 1/\nu$, $\nu = 3, 6, 9$ and exponential correlation function (Scenario 1).

λ	1/3				1/6				1/9			
	1		2		1		2		1		2	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\hat{\beta}_0$	0.00519	0.00971	0.00519	0.00974	0.00448	0.00943	0.0042	0.00943	0.00673	0.00947	0.00686	0.00947
$\hat{\beta}_1$	0.00119	0.00332	0.00112	0.00333	0.00211	0.00264	0.00205	0.00266	0.00194	0.00249	0.00204	0.00249
$\hat{\alpha}$	-0.00164	0.00033	-0.00216	0.00033	-0.00187	0.00033	-0.00206	0.00035	-0.00204	0.00037	-0.00234	0.00038
$\hat{\sigma}^2$	-0.00331	0.00937	0.00689	0.01066	-0.00253	0.00872	-0.00036	0.01108	-0.00481	0.00802	-0.00145	0.01003

Table 2: Bias and MSE when estimating with *wpl* the t process when the (inverse of) degrees of freedom ($\lambda = 1/\nu$, $\nu = 3, 6, 9$) are: 1) fixed and known, 2) unknown and fixed through a two-step estimation (Scenario 2).

4.3 Computational details

Weighted pairwise likelihood estimation requires the evaluation of the bivariate distribution *i.e.* the computation of the Appell F_4 function. Standard statistical software libraries for the computation of the F_4 function are unavailable to the best of our knowledge. In our implementation we exploit the following relation with the Gaussian hypergeometric function (Brychkov and Saad, 2017):

$$F_4(a, b; c, c'; w, z) = \sum_{k=0}^{\infty} \frac{(a)_k (b)_k z^k}{k! (c')_k} {}_2F_1(a+k, b+k; c; w), \quad |\sqrt{w}| + |\sqrt{z}| < 1, \quad (4.4)$$

λ	1/3				1/6				1/9			
	1		2		1		2		1		2	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\hat{\beta}_0$	0.00159	0.00341	0.00160	0.00342	-0.00173	0.00381	-0.00180	0.00381	-0.00055	0.00331	-0.00057	0.00331
$\hat{\beta}_1$	0.00416	0.00348	0.00409	0.00348	-0.00120	0.00329	-0.00133	0.00330	0.00005	0.00306	0.00005	0.00307
$\hat{\alpha}_s$	-0.00171	0.00139	-0.00255	0.00139	-0.00117	0.00119	-0.00146	0.00120	-0.00050	0.00125	-0.00075	0.00127
$\hat{\alpha}_t$	-0.00077	0.00738	-0.00167	0.00734	-0.00262	0.00721	-0.00313	0.00715	-0.00334	0.00691	-0.00392	0.00685
$\hat{\sigma}^2$	0.00020	0.00537	0.01270	0.00731	-0.00016	0.00412	0.00370	0.00665	-0.00100	0.00377	0.00224	0.00585

Table 3: Bias and MSE when estimating with wpl the space-time t process when the (inverse of) degrees of freedom ($\lambda = 1/\nu$, $\nu = 3, 6, 9$) are: 1) fixed and known, 2) unknown and fixed through a two-step estimation (Scenario 3):

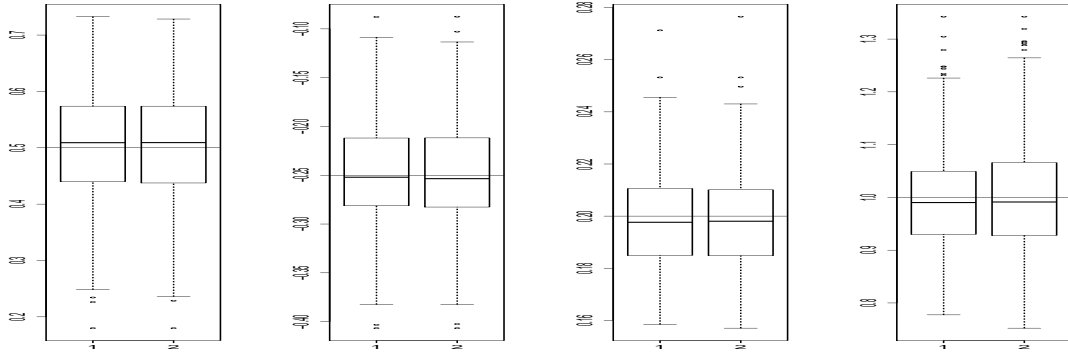


Figure 4: Boxplots of wpl estimates for $\beta_0 = 0.5$, $\beta_1 = -0.25$, $\alpha = 0.2$, $\sigma^2 = 1$ (from left to right) under Scenario 2 when estimating a t process with $\lambda = 1/\nu$, $\nu = 6$ when 1) ν is assumed known, 2) ν is assumed unknown and it is fixed to a positive integer through a two-step estimation .

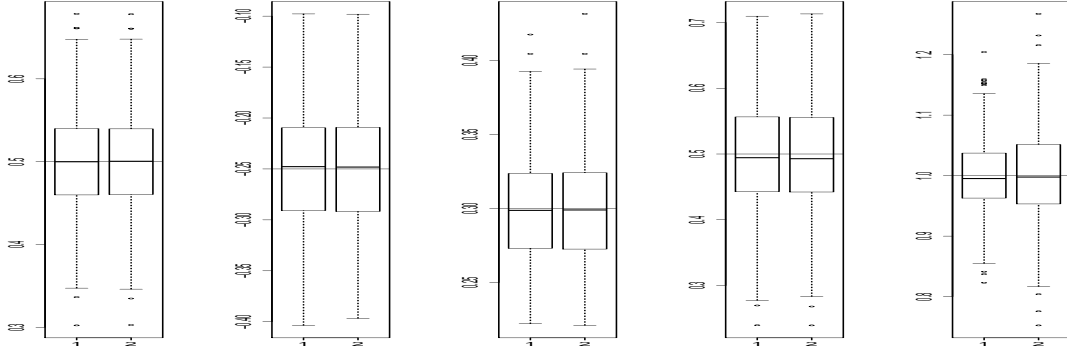


Figure 5: Boxplots of wpl estimates for $\beta_0 = 0.5$, $\beta_1 = -0.25$, $\alpha_s = 0.3$, $\alpha_t = 0.5$, $\sigma^2 = 1$ (from left to right) under Scenario 3 when estimating a space-time t process with $\lambda = 1/\nu$, $\nu = 9$ when 1) ν is assumed known, 2) ν is assumed unknown and it is fixed to a positive integer through a two-step estimation.

truncating the series when the k -th generic element of the series is smaller than a fixed ϵ and where standard libraries for the computation of the ${}_2F_1$ function can be used (Pearson et al., 2017). Evaluation of the F_4 function can be time consuming and in order to speed up the computation, the implementation in the **GeoModels** package (Bevilacqua and Morales-Oñate, 2018) uses the OpenCL framework for parallel computing. We skip technical details on the OpenCL implementation but we highlight that the computational savings with respect to a classical sequential C language implementation can reach up to 10 times. For instance, evaluation of the pairwise t function (4.3) in scenario 2, using a machine with 16 GB of memory and with two devices (a 2.6 GHz CPU and a GPU device) requires approximately 0.7 and 0.1 (in terms of R elapsed time in seconds) for the classical and the OpenCL implementation respectively.

4.4 t optimal linear prediction versus Gaussian optimal prediction

One of the primary goals of geostatistical modeling is to make predictions at spatial locations without observations. The optimal predictor for the t process, with respect

to the mean squared error criterion, is nonlinear and difficult to evaluate explicitly since it requires the knowledge of the finite dimensional distribution. Monte Carlo methods can be an appealing option in this case. However, from a computational point of view, Monte Carlo samples are difficult to produce efficiently and such a method can be unfeasible for large data sets (Zhang and El-Shaarawi, 2010). A more practical and less efficient solution can be obtained using the optimal linear prediction. Assuming known mean and covariance of the t process, the predictor at an unknown location \mathbf{s}_0 is given by:

$$\widehat{y(\mathbf{s}_0)} = \mu(\mathbf{s}_0) + \mathbf{c}_\nu R_\nu^{-1}(\mathbf{Y} - \boldsymbol{\mu}) \quad (4.5)$$

where $\boldsymbol{\mu} = (\mu(\mathbf{s}_1), \dots, \mu(\mathbf{s}_n))^T$, $\mathbf{c}_\nu = [\rho_{Y_\nu^*}(\mathbf{s}_0 - \mathbf{s}_j)]_{j=1}^n$ and $R_\nu = [\rho_{Y_\nu^*}(\mathbf{s}_i - \mathbf{s}_j)]_{i,j=1}^n$. Note that, from Theorem 2.2 (d), optimal Gaussian predictor can be viewed as a special case of (4.5) when $\nu \rightarrow \infty$.

We investigate the performance of (4.5) when compared with the Gaussian optimal predictor under the t process using 2-fold cross-validation. With this goal in mind, we consider a zero mean t process and the same settings of Scenario 2. We simulate 500 realizations and for each realization, we consider 80% of the data for estimation and leave 20% as validation dataset. We use *wpl* (using $d_{ij} = 0.04$) for both t and Gaussian processes and standard likelihood for the Gaussian process and then we use the estimated parameters for computing the associated optimal (linear) predictors. The use of a Generalized Wendland model induces sparsity in the covariance matrix of the t process and this allows speeding up the computation of the optimal linear predictor using specific sparse matrix algorithms (Furrer et al., 2013; Bevilacqua, Faouzi, Furrer and Porcu, 2018).

We then compute the root mean square errors (RMSEs) and median absolute errors (MAEs) for the t process that is:

$$RMSE_l = \left(\frac{1}{n_l} \sum_{i=1}^{n_l} \left(\widehat{y(\mathbf{s}_{i,l})} - y(\mathbf{s}_{i,l}) \right)^2 \right)^{\frac{1}{2}}, \quad MAE_l = \frac{1}{n_l} \sum_{i=1}^{n_l} |\widehat{y(\mathbf{s}_{i,l})} - y(\mathbf{s}_{i,l})|, \quad l = 1, \dots, 500$$

where $y(\mathbf{s}_{i,l})$, $i = 1, \dots, n_l$ are the observation in the l -th validation set and n_l is the associated cardinality ($n_l = 200$ in our example). Similarly we compute $RMSE_l$ and MAE_l , $l = 1, \dots, 500$ in the Gaussian case.

In Table 4 we report the empirical mean of the 500 RMSEs and MAEs for both the Gaussian and t cases. It can be appreciated from Table 4 that (4.5) performs overall better than the optimal Gaussian predictor even when the standard likelihood is used as method of estimation and, as expected, it approaches the optimal Gaussian predictor when increasing the degrees of freedom.

ν	Method	3	6	9	12	20
RMSE						
t	Pairwise	1.44206	0.91531	0.81538	0.78129	0.74684
Gaussian	Pairwise	1.45118	0.91666	0.81575	0.78144	0.74685
Gaussian	Likelihood	1.45104	0.91656	0.81568	0.78131	0.74677
MAE						
t	Pairwise	0.87721	0.67872	0.62534	0.60639	0.58619
Gaussian	Pairwise	0.88605	0.68020	0.62594	0.60660	0.58624
Gaussian	Likelihood	0.88747	0.68043	0.62600	0.60659	0.58622

Table 4: Root mean square error (RMSE) and mean absolute error (MAE) for the Gaussian process (using wpl and standard likelihood) and t process (using wpl) when increasing the degrees of freedom ν .

This simulation study suggests that, using an optimal Gaussian predictor when addressing spatial symmetric data with heavy tails can lead to a loss of performance with respect to a simple optimal linear t predictor.

5 Application to Maximum Temperature Data

In this section, we apply the proposed t process to a data set of maximum temperature data observed in Australia. Specifically, we consider a subset of a global data set of merged maximum daily temperature measurements from the Global Surface Summary of Day data (GSOD) with European Climate Assessment & Dataset (ECA&D) data in July 2011. The dataset is described in detail in Kilibarda et al. (2014) and it is available in the R package `meteo`. The subset we consider is depicted in Figure 6 (a) and consists of the maximum temperature observed on July 5 in 446 location sites in the region with longitude $[110, 154]$ and latitude $[-39, -12]$.

Spatial coordinates are given in longitude and latitude expressed as decimal degrees and we consider the proposed t process defined on the planet Earth sphere

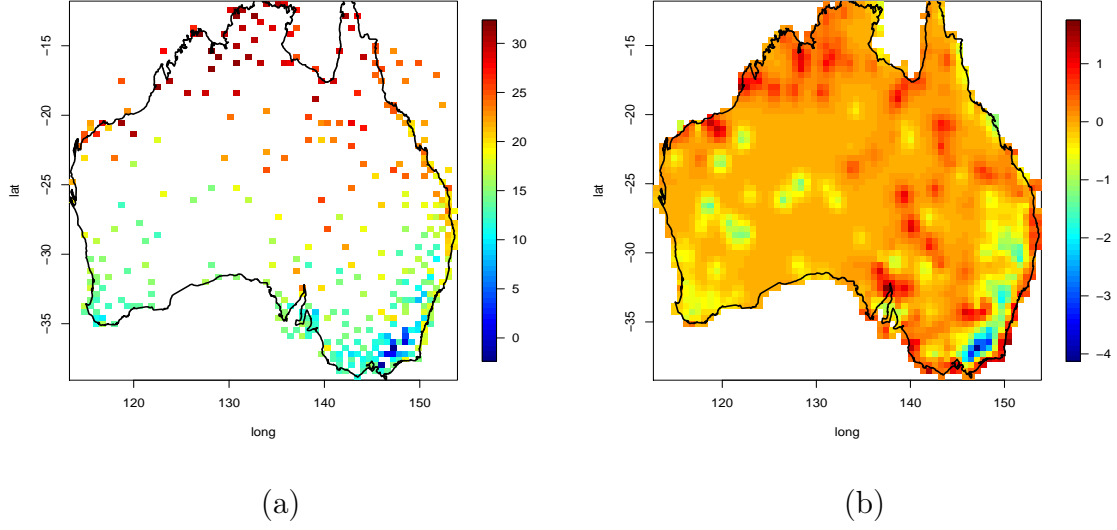


Figure 6: From left to right: a) spatial locations of maximum temperature in Australia in July 2011 and b) prediction of residuals of the estimated t process

approximation $\mathbb{S}^2 = \{\mathbf{s} \in \mathbb{R}^3 : \|\mathbf{s}\| = 6371\}$. The first process we use to model this dataset is a t process:

$$Y_\nu(\mathbf{s}) = \beta_0 + \beta_1 X(\mathbf{s}) + \sigma Y_\nu^*(\mathbf{s}), \quad \mathbf{s} \in \mathbb{S}^2 \quad (5.1)$$

where Y_ν^* is a standardized t process and $X(\mathbf{s})$ is a covariate called *geometric temperature* which represents the geometric position of a particular location on Earth and the day of the year (Kilibarda et al., 2014).

We assume that the underlying geodesically isotropic correlation function (Gneiting, 2013; Porcu et al., 2016) is $\rho(d_{GC}) = \mathcal{GW}_{\alpha,0,4}(d_{GC})$ where, given two spherical points $\mathbf{s}_i = (\text{lon}_i, \text{lat}_i)$ and $\mathbf{s}_j = (\text{lon}_j, \text{lat}_j)$, $d_{GC}(\mathbf{s}_i, \mathbf{s}_j) = 6371\theta_{ij}$, is the great circle distance. Here $\theta_{ij} = \arccos\{\sin a_i \sin a_j + \cos a_i \cos a_j \cos(b_i - b_j)\}$ is the great circle distance on the unit sphere with $a_i = (\text{lat}_i)\pi/180$, $a_j = (\text{lat}_j)\pi/180$, $b_i = (\text{lon}_i)\pi/180$, $b_j = (\text{lon}_j)\pi/180$.

As a comparison, we also consider a Gaussian process:

$$Y(\mathbf{s}) = \beta_0 + \beta_1 X(\mathbf{s}) + \sigma G(\mathbf{s}), \quad \mathbf{s} \in \mathbb{S}^2 \quad (5.2)$$

where G is a standardized Gaussian process with correlation $\mathcal{GW}_{\alpha,0,4}(d_{GC})$. For the t process the parameters were estimated using *wpl* using the two-step method de-

scribed in Section 4 and using the weight function (4.2) with $d_{ij} = 150$ Km. It turns out that the estimation at the first step leads to fix $\nu = 4$ in the second step. For the Gaussian model we also consider pairwise likelihood estimation using the same weight function. In addition, we compute the standard error estimation and PLIC and BLIC values through parametric bootstrap estimation of the inverse of the Godambe information matrix. The results are summarized in Table 5. Note that the regression parameters estimations are quite similar and the t model assigns more spatial dependence than the Gaussian one. More importantly, both the pairwise likelihood information criterion PLIC and BLIC select the t model.

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\alpha}$	$\hat{\sigma}^2$	PLIC	BLIC	RMSE	MAE
Gaussian	5.622 (1.2300)	1.067 (0.1010)	260.894 (47.2179)	14.275 (1.9241)	24294	24955	2.83298	2.17895
t	6.6336 (1.1262)	0.9958 (0.1037)	273.8178 (56.8324)	7.6321 (1.3991)	23888	24614	2.81506	2.17576

Table 5: Weighted pairwise likelihood estimates with associated standard error (in parenthesis) and PLIC and BLIC values for the Gaussian process and t process Y_4 when estimating the Australian maximum temperature dataset. The empirical mean of RMSEs and MAEs is obtained in order to assess the prediction performance of the two processes through cross-validation.

Given the estimation of the mean regression and variance parameters, the estimated residuals

$$\hat{Y}_4^*(\mathbf{s}_i) = \frac{Y_4(\mathbf{s}_i) - (\hat{\beta}_0 + \hat{\beta}_1 X(\mathbf{s}_i))}{(\hat{\sigma}^2)^{\frac{1}{2}}} \quad i = 1, \dots, N$$

can be viewed as a realization of the process Y_4^* . Similarly we can compute the Gaussian residuals. Both residuals can be useful in order to check the model assumptions, in particular the marginal and dependence assumptions. In the top part of Figure 7 a qq -plot of the residuals in the Gaussian and t cases (from left to right) is depicted. It can be appreciated that the t model overall fits better with respect the Gaussian model even if it seems to fail to model properly the right tail behavior. Moreover, a graphical comparison between the empirical and fitted semivariogram of the residuals is shown in the bottom part of Figure 7.

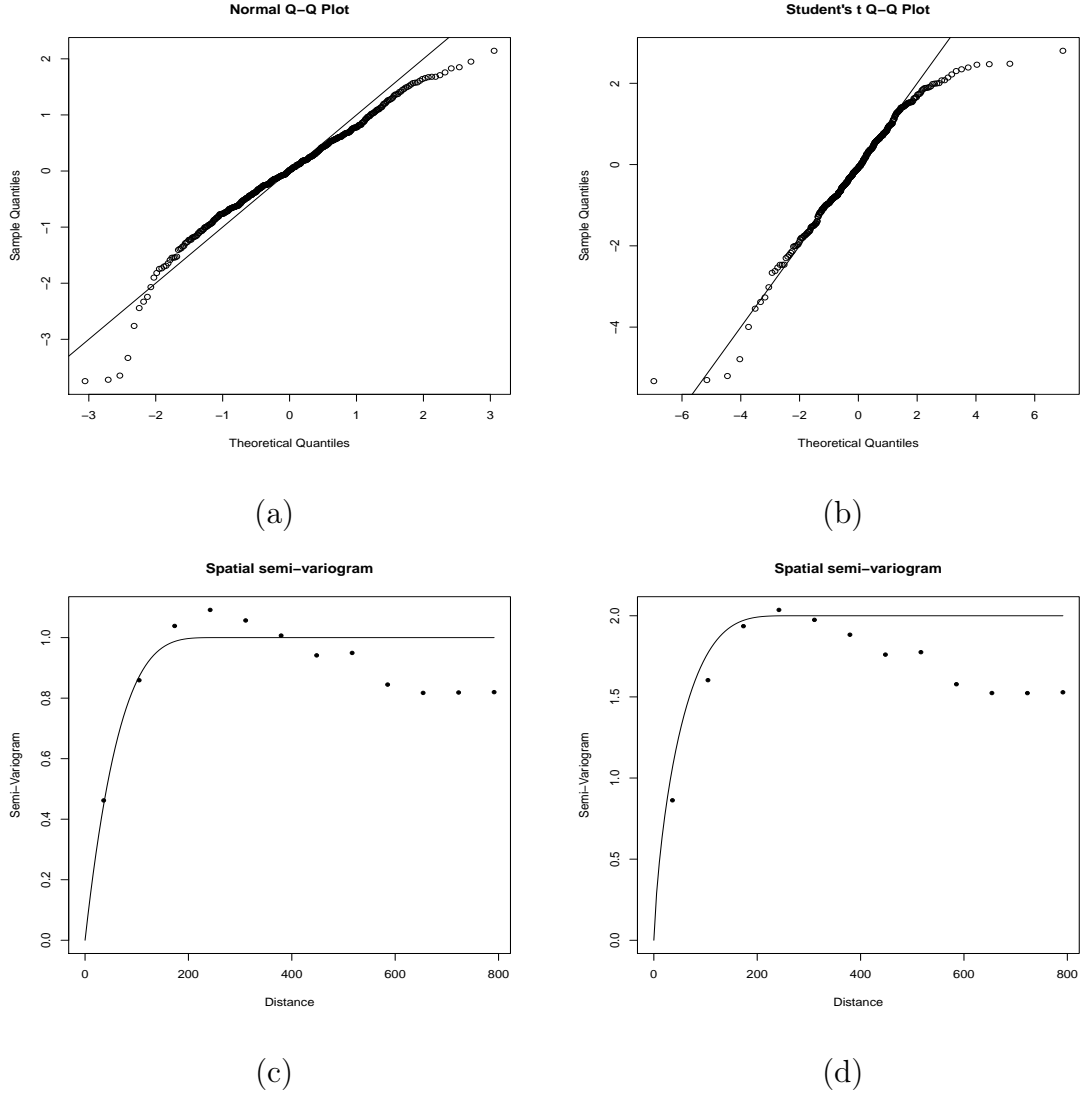


Figure 7: Upper part: Q-Q plot of the residuals versus the estimated quantiles in the Gaussian and t models ((a) and (b) respectively). Bottom part: Empirical semi-variogram (dotted points) of the residuals versus the estimated semivariogram (solid line) in the Gaussian and t models ((c) and (d) respectively). Distances are expressed in Km.

We want to further evaluate the predictive performances of Gaussian and t processes using RMSE and MAE as in Section 4.4. Specifically, we use the following resampling approach: we randomly choose 80% of the data to estimate and we use the estimates in order to compute, for each model, RMSE and MAE values at the remaining 20% of the spatial locations. We repeat the approach for 500 times and

record all RMSEs and MAEs. Table 5 reports the empirical mean of the five hundred RMSE and MAEs for the Gaussian and t cases. We can see that the t process outperforms the Gaussian process for both measures of prediction performance.

Finally, one important goal in spatial modeling of temperature data is to create a high resolution map in a spatial region using the observed data. In Figure 6 (b), we plot a high resolution map of the predicted residuals using the fitted t process.

6 Concluding remarks

We have introduced a new stochastic process with t marginal distributions for regression and dependence analysis when addressing spatial or spatiotemporal data with heavy tails. Our proposal allows overcoming any problem of identifiability associated with previously proposed spatial models with t marginals and, as a consequence, the model parameters can be estimated with just one realization of the process. Moreover the proposed t process inherits the geometrical properties of the ‘parent’ Gaussian process, an appealing feature from a data analysis point of view. We have also proposed a possible generalization, obtaining a new process with the marginal distribution of the skew- t type using the skew-Gaussian process proposed in Zhang and El-Shaarawi (2010).

In our proposal, a possible limitation is the lack of amenable expressions of the associated multivariate distributions. This prevents an inference approach based on the full likelihood and the computation of the optimal predictor. In the first case, our simulation study shows that, for the t process, an inferential approach based on weighted pairwise likelihood, using the bivariate t distribution given in Theorem 2.3, could be an effective solution for estimating the unknown parameters. In the second case, our numerical experiments show that the optimal linear predictor of the t process performs better than the optimal Gaussian prediction when working with spatial data with heavy tails.

Another possible drawback concerns the restriction of the degrees of freedom of the t process to $\nu = 3, 4, \dots$ under noninfinity divisibility of the associated Gamma process. This problem could be solved by considering a Gamma process obtained by

mixing the proposed Gamma process with a process with beta marginals and using the results in Yeo and Milne (1991); however the mathematics involved with this approach are much more challenging.

The estimation of the skew- t process has not been addressed in this paper since the bivariate distribution in this case is quite complicated. We believe this issue can be resolved under a suitable Bayesian framework. Finally, a t process with asymmetric marginal distribution can also be obtained by considering some specific transformations of the proposed standard t process as in J. F. Rosco and Pewsey (2011) or under the two-piece distribution framework (Arellano-Valle et al., 2005) and this will be studied in future work.

Acknowledgements

Partial support was provided by FONDECYT grant 1160280, Chile and by Millennium Science Initiative of the Ministry of Economy, Development, and Tourism, grant "Millennium Nucleus Center for the Discovery of Structures in Complex Data" for Moreno Bevilacqua and by Proyecto de Iniciación Interno DIUBB 173408 2/I de la Universidad del Bío-Bío for Christian Caamaño.

A Appendix

A.1 Proof Theorem 2.1

Proof. Set $R_\nu \equiv W_\nu^{-\frac{1}{2}}$. Then the correlation function of Y_ν^* is given by

$$\rho_{Y_\nu^*}(\mathbf{h}) = \left(\frac{\nu - 2}{\nu} \right) (E(R_\nu(\mathbf{s}_i)R_\nu(\mathbf{s}_j))\rho(\mathbf{h})). \quad (\text{A.1})$$

To find a closed form for $E(R_\nu(\mathbf{s}_i)R_\nu(\mathbf{s}_j))$, we need the bivariate distribution of $\mathbf{R}_{\nu;ij}$ that can be easily obtained from density of the bivariate random vector $\mathbf{W}_{\nu;ij}$ given by (Bevilacqua, Caamaño and Gaetan, 2018):

$$f_{\mathbf{W}_{\nu;ij}}(w_i, w_j) = \frac{2^{-\nu} \nu^\nu (w_i w_j)^{\nu/2-1} e^{-\frac{\nu(w_i+w_j)}{2(1-\rho^2(\mathbf{h}))}}}{\Gamma(\frac{\nu}{2}) (1-\rho^2(\mathbf{h}))^{\nu/2}} \left(\frac{\nu \sqrt{\rho^2(\mathbf{h}) w_i w_j}}{2(1-\rho^2(\mathbf{h}))} \right)^{1-\nu/2} I_{\nu/2-1} \left(\frac{\nu \sqrt{\rho^2(\mathbf{h}) w_i w_j}}{(1-\rho^2(\mathbf{h}))} \right) \quad (\text{A.2})$$

where $I_\alpha(\cdot)$ denotes the modified Bessel function of the first kind of order α . Vere-Jones (1967) show the infinite divisibility of $\mathbf{W}_{\nu;ij}$.

Then, for each $\nu > 2$, the bivariate distribution of $\mathbf{R}_{\nu;ij}$ is given by:

$$f_{\mathbf{R}_{\nu;ij}}(\mathbf{r}_{ij}) = \frac{2^{-\nu+2}\nu^\nu (r_i r_j)^{-\nu-1} e^{-\frac{\nu}{2(1-\rho^2(\mathbf{h}))}\left(\frac{1}{r_i^2} + \frac{1}{r_j^2}\right)}}{\Gamma\left(\frac{\nu}{2}\right) (1-\rho^2(\mathbf{h}))^{\nu/2}} \left(\frac{\nu\rho(\mathbf{h})}{2(1-\rho^2(\mathbf{h}))r_i r_j}\right)^{1-\frac{\nu}{2}} I_{\frac{\nu}{2}-1}\left(\frac{\nu\rho(\mathbf{h})}{(1-\rho^2(\mathbf{h}))r_i r_j}\right) \quad (\text{A.3})$$

Using the identity ${}_0F_1(;b;x) = \Gamma(b)x^{(1-b)/2}I_{b-1}(2\sqrt{x})$ and the series expansion of hypergeometric function ${}_0F_1$ in (A.3) we have

$$\begin{aligned} E(R^a(\mathbf{s}_i)R^b(\mathbf{s}_j)) &= \frac{2^{-\nu+2}\nu^\nu}{\Gamma^2\left(\frac{\nu}{2}\right) (1-\rho^2(\mathbf{h}))^{\nu/2}} \int_{\mathbb{R}_+^2} r_i^{-\nu+a-1} r_j^{-\nu+b-1} e^{-\frac{\nu}{2(1-\rho^2(\mathbf{h}))r_i^2}} e^{-\frac{\nu}{2(1-\rho^2(\mathbf{h}))r_j^2}} \\ &\quad \times {}_0F_1\left(\frac{\nu}{2}; \frac{\nu^2\rho^2(\mathbf{h})}{4(1-\rho^2(\mathbf{h}))^2r_i^2r_j^2}\right) d\mathbf{r}_{ij} \\ &= \frac{2^{-\nu+2}\nu^\nu}{\Gamma^2\left(\frac{\nu}{2}\right) (1-\rho^2(\mathbf{h}))^{\nu/2}} \sum_{k=0}^{\infty} \int_{\mathbb{R}_+^2} r_i^{-\nu+a-2k-1} r_j^{-\nu+b-2k-1} e^{-\frac{\nu}{2(1-\rho^2(\mathbf{h}))r_i^2}} e^{-\frac{\nu}{2(1-\rho^2(\mathbf{h}))r_j^2}} \\ &\quad \times \frac{1}{k! \left(\frac{\nu}{2}\right)_k} \left(\frac{\rho^2(\mathbf{h})\nu^2}{4(1-\rho^2(\mathbf{h}))}\right)^k d\mathbf{r}_{ij} \\ &= \frac{2^{-\nu+2}\nu^\nu}{\Gamma^2\left(\frac{\nu}{2}\right) (1-\rho^2(\mathbf{h}))^{\nu/2}} \sum_{k=0}^{\infty} \frac{I(k)}{k! \left(\frac{\nu}{2}\right)_k} \left(\frac{\rho^2(\mathbf{h})\nu^2}{4(1-\rho^2(\mathbf{h}))}\right)^k \end{aligned} \quad (\text{A.4})$$

where, using Fubini's Theorem

$$I(k) = \int_{\mathbb{R}_+} r_i^{-\nu+a-2k-1} e^{-\frac{\nu}{2(1-\rho^2(\mathbf{h}))r_i^2}} dr_i \int_{\mathbb{R}_+} r_j^{\nu+b-2k-1} e^{-\frac{\nu}{2(1-\rho^2(\mathbf{h}))r_j^2}} dr_j$$

Using the univariate density $f_{R_\nu(\mathbf{s})}(r) = 2\left(\frac{\nu}{2}\right)^{\nu/2} r^{-\nu-1} e^{-\frac{\nu}{2r^2}} / \Gamma\left(\frac{\nu}{2}\right)$, we obtain

$$I(k) = \Gamma\left(\frac{\nu-a}{2} + k\right) \Gamma\left(\frac{\nu-b}{2} + k\right) 2^{\frac{\nu-a}{2}+k-1} 2^{\frac{\nu-b}{2}+k-1} \left(\frac{(1-\rho^2(\mathbf{h}))}{\nu}\right)^{\frac{\nu-a}{2}+k} \left(\frac{(1-\rho^2(\mathbf{h}))}{\nu}\right)^{\frac{\nu-b}{2}+k} \quad (\text{A.5})$$

and combining equations (A.5) and (A.4), we obtain

$$\begin{aligned} E(R^a(\mathbf{s}_i)R^b(\mathbf{s}_j)) &= \frac{2^{-(a+b)/2}\nu^{(a+b)/2}(1-\rho^2(\mathbf{h}))^{(\nu-a-b)/2}\Gamma\left(\frac{\nu-a}{2}\right)\Gamma\left(\frac{\nu-b}{2}\right)}{\Gamma^2\left(\frac{\nu}{2}\right)} \sum_{k=0}^{\infty} \frac{\left(\frac{\nu-a}{2}\right)_k \left(\frac{\nu-b}{2}\right)_k}{k! \Gamma\left(\frac{\nu}{2}\right)} \rho_G^{2k}(\mathbf{h}) \\ &= \frac{2^{-(a+b)/2}\nu^{(a+b)/2}(1-\rho^2(\mathbf{h}))^{(\nu-a-b)/2}\Gamma\left(\frac{\nu-a}{2}\right)\Gamma\left(\frac{\nu-b}{2}\right)}{\Gamma^2\left(\frac{\nu}{2}\right)} {}_2F_1\left(\frac{\nu-a}{2}, \frac{\nu-b}{2}; \frac{\nu}{2}; \rho^2(\mathbf{h})\right) \end{aligned}$$

Then, using the Euler transformation, we obtain

$$E(R_\nu^a(\mathbf{s}_i)R_\nu^b(\mathbf{s}_j)) = \frac{2^{-(a+b)/2}\nu^{(a+b)/2}}{\Gamma^2\left(\frac{\nu}{2}\right)} \Gamma\left(\frac{\nu-a}{2}\right) \Gamma\left(\frac{\nu-b}{2}\right) {}_2F_1\left(\frac{a}{2}, \frac{b}{2}; \frac{\nu}{2}; \rho^2(\mathbf{h})\right) \quad (\text{A.6})$$

for $\nu > a$ and $\nu > b$. Finally, setting $a = b = 1$ in (A.6) and using it in (A.1) we obtain (2.4). \square

A.2 Proof Theorem 2.3

Proof. Using the identity ${}_0F_1(; b; x) = \Gamma(b)x^{(1-b)/2}I_{b-1}(2\sqrt{x})$ and the series expansion of hypergeometric function ${}_0F_1$, then under the transformation $g_i = y_i\sqrt{w_i}$ and $g_j = y_j\sqrt{w_j}$ with Jacobian $J((g_i, g_j) \rightarrow (y_i, y_j)) = (w_i w_j)^{1/2}$, we have:

$$\begin{aligned}
f_{Y_{ij}^*}(y_{ij}) &= \int_{\mathbb{R}_+^2} f_{G_{ij}|\mathbf{W}_{ij}}(\mathbf{g}_{ij}|\mathbf{w}_{ij}) f_{\mathbf{W}_{ij}}(\mathbf{w}_{ij}) J d\mathbf{w}_{ij} \\
&= \frac{2^{-\nu}\nu^\nu}{2\pi\Gamma^2\left(\frac{\nu}{2}\right)(1-\rho^2(\mathbf{h}))^{(\nu+1)/2}} \int_{\mathbb{R}_+^2} (w_i w_j)^{(\nu+1)/2-1} e^{-\frac{1}{2(1-\rho^2(\mathbf{h}))}[w_i y_i^2 + w_j y_j^2 - 2\rho(\mathbf{h})\sqrt{w_i w_j} y_i y_j]} \\
&\quad \times e^{-\frac{\nu(w_i + w_j)}{2(1-\rho^2(\mathbf{h}))}} {}_0F_1\left(\frac{\nu}{2}; \frac{\nu^2 \rho^2(\mathbf{h}) w_i w_j}{4(1-\rho^2(\mathbf{h}))^2}\right) d\mathbf{w}_{ij} \\
&= \frac{2^{-\nu}\nu^\nu}{2\pi\Gamma^2\left(\frac{\nu}{2}\right)(1-\rho^2(\mathbf{h}))^{(\nu+1)/2}} \int_{\mathbb{R}_+^2} (w_i w_j)^{(\nu+1)/2-1} e^{-\frac{1}{2(1-\rho^2(\mathbf{h}))}[y_i^2 - 2\rho(\mathbf{h})\sqrt{\frac{w_j}{w_i}} y_i y_j + \nu]} w_i e^{-\frac{(y_j^2 + \nu)w_j}{2(1-\rho^2(\mathbf{h}))}} \\
&\quad \times \sum_{k=0}^{\infty} \frac{1}{k! \left(\frac{\nu}{2}\right)_k} \left(\frac{\nu^2 \rho^2(\mathbf{h}) w_i w_j}{4(1-\rho^2(\mathbf{h}))^2}\right)^k d\mathbf{w}_{ij} \\
&= \frac{2^{-\nu}\nu^\nu}{2\pi\Gamma^2\left(\frac{\nu}{2}\right)(1-\rho^2(\mathbf{h}))^{(\nu+1)/2}} \sum_{k=0}^{\infty} \frac{I(k)}{k! \left(\frac{\nu}{2}\right)_k} \left(\frac{\nu^2 \rho^2(\mathbf{h})}{4(1-\rho^2(\mathbf{h}))^2}\right)^k \tag{A.7}
\end{aligned}$$

using (3.462.1) of Gradshteyn and Ryzhik (2007), we obtain

$$\begin{aligned}
I(k) &= \int_{\mathbb{R}_+} w_j^{(\nu+1)/2+k-1} e^{-\frac{(y_j^2 + \nu)w_j}{2(1-\rho^2(\mathbf{h}))}} \left[\int_{\mathbb{R}_+} w_i^{(\nu+1)/2+k-1} e^{-\frac{(y_i^2 + \nu)}{2(1-\rho^2(\mathbf{h}))} w_i - \frac{\rho(\mathbf{h})\sqrt{w_j} y_i y_j}{(\rho^2(\mathbf{h})-1)\sqrt{w_i}}} dw_i \right] dw_j \\
&= 2 \left(\frac{y_i^2 + \nu}{(1-\rho^2(\mathbf{h}))}\right)^{-\left(\frac{\nu+1}{2}+k\right)} \Gamma(\nu+1+2k) \int_{\mathbb{R}_+} w_j^{(\nu+1)/2+k-1} e^{-\left[\frac{\rho^2(\mathbf{h}) y_i^2 y_j^2}{4(1-\rho^2(\mathbf{h}))(y_i^2 + \nu)} - \frac{(y_j^2 + \nu)}{2(1-\rho^2(\mathbf{h}))}\right]} w_j \\
&\quad \times D_{-(\nu+1+2k)}\left(-\frac{\rho(\mathbf{h}) y_i y_j \sqrt{w_j}}{\sqrt{(1-\rho^2(\mathbf{h}))(y_i^2 + \nu)}}\right) dw_j \\
&= 2 \left(\frac{y_i^2 + \nu}{(1-\rho^2(\mathbf{h}))}\right)^{-\left(\frac{\nu+1}{2}+k\right)} \Gamma(\nu+1+2k) A(k) \tag{A.8}
\end{aligned}$$

where $D_n(x)$ is the parabolic cylinder function. Now, considering (9.240) of Gradshteyn and Ryzhik (2007):

$$\begin{aligned}
D_{-(\nu+1+2k)}\left(-\frac{\rho(\mathbf{h}) y_i y_j \sqrt{w_j}}{\sqrt{(1-\rho^2(\mathbf{h}))(y_i^2 + \nu)}}\right) &= \frac{2^{-(\nu+1)/2+k} \sqrt{\pi}}{\Gamma\left(\frac{\nu}{2} + k + 1\right)} e^{-\frac{\rho^2(\mathbf{h}) y_i^2 y_j^2 w_j}{4(1-\rho^2(\mathbf{h}))(y_i^2 + \nu)}} \\
&\quad \times {}_1F_1\left(\frac{\nu+1}{2} + k; \frac{1}{2}; \frac{\rho^2(\mathbf{h}) y_i^2 y_j^2 w_j}{2(1-\rho^2(\mathbf{h}))(y_i^2 + \nu)}\right) \\
&\quad + \frac{2^{-\nu/2-k} \sqrt{\pi} \rho(\mathbf{h}) y_i y_j \sqrt{w_j}}{\Gamma\left(\frac{\nu+1}{2} + k\right) \sqrt{(1-\rho^2(\mathbf{h}))(y_i^2 + \nu)}} e^{-\frac{\rho^2(\mathbf{h}) y_i^2 y_j^2 w_j}{4(1-\rho^2(\mathbf{h}))(y_i^2 + \nu)}} \\
&\quad \times {}_1F_1\left(\frac{\nu}{2} + k + 1; \frac{3}{2}; \frac{\rho^2(\mathbf{h}) y_i^2 y_j^2 w_j}{2(1-\rho^2(\mathbf{h}))(y_i^2 + \nu)}\right) \tag{A.9}
\end{aligned}$$

combining equations (A.9) and the integral of (A.8) and using (7.621.4) of Gradshteyn and Ryzhik (2007), we obtain

$$\begin{aligned}
A(k) &= \int_{\mathbb{R}_+} w_j^{(\nu+1)/2+k-1} e^{-\frac{(y_j^2+\nu)}{2(1-\rho^2(\mathbf{h}))} w_j} {}_1F_1 \left(\frac{\nu+1}{2} + k; \frac{1}{2}; \frac{\rho^2(\mathbf{h}) y_i^2 y_j^2 w_j}{2(1-\rho^2(\mathbf{h}))(y_i^2+\nu)} \right) dw_j \\
&+ \int_{\mathbb{R}_+} w_j^{\nu/2+k+1-1} e^{-\frac{(y_j^2+\nu)}{2(1-\rho^2(\mathbf{h}))} w_j} {}_1F_1 \left(\frac{\nu}{2} + k + 1; \frac{3}{2}; \frac{\rho^2(\mathbf{h}) y_i^2 y_j^2 w_j}{2(1-\rho^2(\mathbf{h}))(y_i^2+\nu)} \right) dw_j \\
&= \Gamma \left(\frac{\nu+1}{2} + k \right) \left(\frac{y_j^2 + \nu}{2(1-\rho^2(\mathbf{h}))} \right)^{-\frac{(\nu+1)}{2}-k} {}_2F_1 \left(\frac{\nu+1}{2} + k, \frac{\nu+1}{2} + k; \frac{1}{2}; \frac{\rho^2(\mathbf{h}) y_i^2 y_j^2}{(y_i^2 + \nu)(y_j^2 + \nu)} \right) \\
&+ \Gamma \left(\frac{\nu}{2} + k + 1 \right) \left(\frac{y_j^2 + \nu}{2(1-\rho^2(\mathbf{h}))} \right)^{-\frac{\nu}{2}-k-1} {}_2F_1 \left(\frac{\nu}{2} + k + 1, \frac{\nu}{2} + k + 1; \frac{3}{2}; \frac{\rho^2(\mathbf{h}) y_i^2 y_j^2}{(y_i^2 + \nu)(y_j^2 + \nu)} \right)
\end{aligned} \tag{A.10}$$

finally, combining equations (A.10), (A.8) and (A.7), we obtain

$$\begin{aligned}
f_{\mathbf{Y}_{ij}^*}(\mathbf{y}_{ij}) &= \frac{\nu^\nu [(y_i^2 + \nu)(y_j^2 + \nu)]^{-(\nu+1)/2} \Gamma^2 \left(\frac{\nu+1}{2} \right)}{\pi \Gamma^2 \left(\frac{\nu}{2} \right) (1 - \rho^2(\mathbf{h}))^{-(\nu+1)/2}} \sum_{k=0}^{\infty} \frac{\left(\frac{\nu+1}{2} \right)_k^2}{k! \left(\frac{\nu}{2} \right)_k} \left(\frac{\nu^2 \rho^2(\mathbf{h})}{(y_i^2 + \nu)(y_j^2 + \nu)} \right)^k \\
&\times {}_2F_1 \left(\frac{\nu+1}{2} + k, \frac{\nu+1}{2} + k; \frac{1}{2}; \frac{\rho^2(\mathbf{h}) y_i^2 y_j^2}{(y_i^2 + \nu)(y_j^2 + \nu)} \right) \\
&+ \frac{\rho(\mathbf{h}) y_i y_j \nu^{\nu+2} [(y_i^2 + \nu)(y_j^2 + \nu)]^{-\nu/2-1}}{2\pi (1 - \rho^2(\mathbf{h}))^{-(\nu+1)/2}} \sum_{k=0}^{\infty} \frac{\left(\frac{\nu}{2} + 1 \right)_k^2}{k! \left(\frac{\nu}{2} \right)_k} \left(\frac{\nu^2 \rho^2(\mathbf{h})}{(y_i^2 + \nu)(y_j^2 + \nu)} \right)^k \\
&\times {}_2F_1 \left(\frac{\nu}{2} + k + 1, \frac{\nu}{2} + k + 1; \frac{3}{2}; \frac{\rho^2(\mathbf{h}) y_i^2 y_j^2}{(y_i^2 + \nu)(y_j^2 + \nu)} \right)
\end{aligned}$$

and using (4.4) we obtain theorem 2.3. \square

A.3 Proof Theorem 3.1

Proof. Consider $\mathbf{U} = (U(\mathbf{s}_1), \dots, U(\mathbf{s}_n))^T$, $\mathbf{V} = (|X_1(\mathbf{s}_1)|, \dots, |X_1(\mathbf{s}_n)|)^T$, $\mathbf{Q} = (X_2(\mathbf{s}_1), \dots, X_2(\mathbf{s}_n))^T$ where $\mathbf{X}_k = (X_k(\mathbf{s}_1), \dots, X_k(\mathbf{s}_n))^T \sim N_n(\mathbf{0}, \Omega)$, for $k = 1, 2$, which are assumed to be independent. By definition of the skew-Gaussian process in (3.1) we have:

$$\mathbf{U} = \boldsymbol{\alpha} + \eta \mathbf{V} + \omega \mathbf{Q}$$

where, by assumption \mathbf{V} and \mathbf{Q} are independent. Thus, by conditioning on $\mathbf{V} = \mathbf{v}$, we have $\mathbf{U} | \mathbf{V} = \mathbf{v} \sim N_n(\boldsymbol{\alpha} + \eta \mathbf{v}, \omega^2 \Omega)$, from which we obtain

$$f_{\mathbf{U}}(\mathbf{u}) = \int_{\mathbb{R}^n} \phi_n(\mathbf{u}; \boldsymbol{\alpha} + \eta \mathbf{v}, \omega^2 \Omega) f_{\mathbf{V}}(\mathbf{v}) d\mathbf{v}$$

To solve this integral we need $f_{\mathbf{V}}(\mathbf{v})$, *i.e.*, the joint density of $\mathbf{V} = (|X_1(\mathbf{s}_1)|, \dots, |X_1(\mathbf{s}_n)|)^T$. Let $\mathbf{X}_k = (X_1, \dots, X_n)^T = (X_1(\mathbf{s}_1), \dots, X_1(\mathbf{s}_n))^T$ and $\mathbf{V} = (|X_1|, \dots, |X_n|)^T$. Additionally, consider the diagonal matrices $\mathbf{D}(\mathbf{l}) = \text{diag}\{l_1, \dots, l_n\}$, with $\mathbf{l} = (l_1, \dots, l_n) \in \{-1, +1\}^n$, which are such that $\mathbf{D}(\mathbf{l})^2 = \mathbf{D}(\mathbf{l})$ for all $\mathbf{l} \in \{-1, +1\}^n$. Since $\mathbf{l} \circ \mathbf{v} = \mathbf{D}(\mathbf{l})\mathbf{v}$ (the componentwise product) and $\mathbf{X} \sim N_n(\mathbf{0}, \Omega)$, we then have

$$\begin{aligned} F_{\mathbf{V}}(\mathbf{v}) &= Pr(\mathbf{V} \leq \mathbf{v}) = Pr(|\mathbf{X}| \leq \mathbf{v}) = Pr(-\mathbf{v} \leq \mathbf{X} \leq \mathbf{v}) \\ &= \sum_{\mathbf{l} \in \{-1, +1\}^n} (-1)^{N_-} \Phi_n(\mathbf{D}(\mathbf{l})\mathbf{v}; \mathbf{0}, \Omega), \quad (N_- = \sum_{i=1}^n I_{l_i=-1} \det\{\mathbf{D}(\mathbf{l})\}) \\ &= \sum_{\mathbf{l} \in \{-1, +1\}^n} \det\{\mathbf{D}(\mathbf{l})\} \Phi_n(\mathbf{D}(\mathbf{l})\mathbf{v}; \mathbf{0}, \Omega) \end{aligned}$$

Hence, by using that

$$\frac{\partial^n \Phi_n(\mathbf{D}(\mathbf{l})\mathbf{v}; \mathbf{0}, \Omega)}{\partial v_1 \cdots \partial v_n} = \det\{\mathbf{D}(\mathbf{l})\} \Phi_n(\mathbf{D}(\mathbf{l})\mathbf{v}; \mathbf{0}, \Omega)$$

we find that the joint density of \mathbf{V} is

$$\begin{aligned} f_{\mathbf{V}}(\mathbf{v}) &= \sum_{\mathbf{l} \in \{-1, +1\}^n} [\det\{\mathbf{D}(\mathbf{l})\}]^2 \phi_n(\mathbf{D}(\mathbf{l})\mathbf{v}; \mathbf{0}, \Omega) \\ &= \sum_{\mathbf{l} \in \{-1, +1\}^n} \phi_n(\mathbf{D}(\mathbf{l})\mathbf{v}; \mathbf{0}, \Omega), \quad ([\det\{\mathbf{D}(\mathbf{l})\}]^2 = 1) \\ &= \sum_{\mathbf{l} \in \{-1, +1\}^n} |\det\{\mathbf{D}(\mathbf{l})\}| \phi_n(\mathbf{v}; \mathbf{0}, \Omega_{\mathbf{l}}), \quad (\Omega_{\mathbf{l}} = \mathbf{D}(\mathbf{l})\Omega\mathbf{D}(\mathbf{l}) = (l_i l_j \rho_{ij})) \\ &= \sum_{\mathbf{l} \in \{-1, +1\}^n} \phi_n(\mathbf{v}; \mathbf{0}, \Omega_{\mathbf{l}}), \quad (|\det\{\mathbf{D}(\mathbf{l})\}| = 1) \\ &= 2 \sum_{\mathbf{l} \in \{-1, +1\}^n: \mathbf{l} \neq -\mathbf{l}} \phi_n(\mathbf{v}; \mathbf{0}, \Omega_{\mathbf{l}}) \end{aligned}$$

where the last identity is due to $\Omega_{-\mathbf{l}} = \mathbf{D}(-\mathbf{l})\Omega\mathbf{D}(-\mathbf{l}) = \mathbf{D}(\mathbf{l})\Omega\mathbf{D}(\mathbf{l}) = \Omega_{\mathbf{l}}$ for all $\mathbf{l} \in \{-1, +1\}^n$; e.g. for $n = 3$ the sum must be performed on

$$\mathbf{l} \in \{(+1, +1, +1), (+1, +1, -1), (+1, -1, +1), (-1, +1, +1)\}$$

since

$$-\mathbf{l} \in \{(-1, -1, -1), (-1, -1, +1), (-1, +1, -1), (+1, -1, -1)\}$$

and both sets produce the same correlation matrices. The joint density of \mathbf{U} is thus

given by

$$\begin{aligned}
f_U(\mathbf{u}) &= 2 \sum_{w \in \{-1, +1\}^n: w \neq -w} \int_{\mathbb{R}_+^n} \phi_n(\mathbf{u}; \boldsymbol{\alpha} + \eta \mathbf{v}, \omega^2 \Omega) \phi_n(\mathbf{v}; \mathbf{0}, \Omega_l) d\mathbf{v} \\
&= 2 \sum_{w \in \{-1, +1\}^n: w \neq -w} \phi_n(\mathbf{u}; \boldsymbol{\alpha}, \mathbf{A}_l) \int_{\mathbb{R}_+^n} \phi_n(\mathbf{v}; \mathbf{c}_l, \mathbf{B}_l) d\mathbf{v} \\
&= 2 \sum_{w \in \{-1, +1\}^n: w \neq -w} \phi_n(\mathbf{u}; \boldsymbol{\alpha}, \mathbf{A}_l) \Phi_n(\mathbf{c}_l; \mathbf{0}, \mathbf{B}_l)
\end{aligned}$$

where $\mathbf{A}_l = \omega^2 \Omega + \eta^2 \Omega_l$, $\mathbf{c}_l = \eta \Omega_l \mathbf{A}_l^{-1}(\mathbf{u} - \boldsymbol{\alpha})$, $\mathbf{B}_l = \Omega_l - \eta^2 \Omega_l \mathbf{A}_l^{-1} \Omega_l$, and we have used the identity $\phi_n(\mathbf{u}; \boldsymbol{\alpha} + \eta \mathbf{v}, \omega^2 \Omega) \phi_n(\mathbf{v}; \mathbf{0}, \Omega_l) = \phi_n(\mathbf{u}; \boldsymbol{\alpha}, \mathbf{A}_l) \phi_n(\mathbf{v}; \mathbf{c}_l, \mathbf{B}_l)$ which follows straightforwardly from the standard marginal-conditional factorizations of the underlying multivariate normal joint density. \square

References

- Alegría, A., Caro, S., Bevilacqua, M., Porcu, E., and Clarke, J. (2017), “Estimating covariance functions of multivariate skew-Gaussian random fields on the sphere,” *Spatial Statistics*, 22, 388 – 402.
- Allcroft, D., and Glasbey, C. (2003), “A latent Gaussian Markov random field model for spatio-temporal rainfall disaggregation,” *Journal of the Royal Statistical Society, C*, 52, 487–498.
- Arellano-Valle, R. B., and Azzalini, A. (2013), “The centred parameterization and related quantities of the skew-t distribution,” *Journal of Multivariate Analysis*, 113, 73 – 90. Special Issue on Multivariate Distribution Theory in Memory of Samuel Kotz.
- Arellano-Valle, R. B., Castro, L., and Gonzalez-Farias, G. (2012), “Student-t censored regression model: properties and inference,” *Statistical Methods and Applications*, 21(4), 453–473.
- Arellano-Valle, R. B., Gómez, H. W., and Quintana, F. A. (2005), “Statistical inference for a general class of asymmetric distributions,” *Journal of Statistical Planning and Inference*, 128(2), 427 – 443.

- Azzalini, A., and Capitanio, A. (2014), *The Skew-Normal and Related Families*, New York: United States of America by Cambridge University Press.
- Azzalini, A., and Dalla-Valle, A. (1996), “The Multivariate Skew-Normal Distribution,” *Biometrika*, 83(4), 715–726.
- Bai, Y., Kang, J., and Song, P. (2014), “Efficient Pairwise Composite Likelihood Estimation for Spatial-Clustered Data,” *Biometrics*, 7(3), 661–670.
- Bapat, R. B. (1989), “Infinite divisibility of multivariate gamma distributions and M-matrices,” *Sankhyā A*, 51, 73–78.
- Berg, C., Mateu, J., and Porcu, E. (2008), “The Dagum family of isotropic correlation functions,” *Bernoulli*, 14(4), 1134–1149.
- Bevilacqua, M., Caamaño, C., and Gaetan, C. (2018), “On modelling positive continuous data with spatio-temporal dependence,” *ArXiv e-prints*, .
- Bevilacqua, M., Faouzi, T., Furrer, R., and Porcu, E. (2018), “Estimation and Prediction using generalized Wendland functions under fixed domain asymptotics,” *The Annals of Statistics*, . Forthcoming.
- Bevilacqua, M., and Gaetan, C. (2015), “Comparing composite likelihood methods based on pairs for spatial Gaussian random fields,” *Statistics and Computing*, 25, 877–892.
- Bevilacqua, M., Gaetan, C., Mateu, J., and Porcu, E. (2012), “Estimating space and space-time covariance functions for large data sets: a weighted composite likelihood approach,” *Journal of the American Statistical Association*, 107, 268–280.
- Bevilacqua, M., and Morales-Oñate, V. (2018), *GeoModels: A Package for Geostatistical Gaussian and non Gaussian Data Analysis*. R package version 1.0.3-4.
URL: <https://vmoprojs.github.io/GeoModels-page/>

- Brychkov, Y. A., and Saad, N. (2017), “On some formulas for the Appell function $F_2(a, b, b'; c, c'; w; z)$,” *Journal Integral Transforms and Special Functions*, 25, 1465–1483.
- De Oliveira, V. (2006), “On optimal point and block prediction in log-gaussian random fields,” *Scandinavian Journal of Statistics*, 33, 523–540.
- De Oliveira, V., Kedem, B., and Short, D. A. (1997), “Bayesian prediction of transformed Gaussian random fields,” *Journal of the American Statistical Association*, 92, 1422–1433.
- DeBastiani, F., Cysneiros, A., Uribe-Opazo, M., and Galea, M. (2015), “Influence diagnostics in elliptical spatial linear models,” *TEST*, 24, 322–340.
- DiCiccio, T. J., and Monti, A. C. (2011), “Inferential aspects of the skew tdistribution,” *Quaderni di Statistica*, 13, 1–21.
- Eisenbaum, N., and Kaspi, H. (2006), “A characterization of the infinitely divisible squared gaussian processes,” *The Annals of Probability*, pp. 728–742.
- Ferrari, S. L. P., and Arellano-Valle, R. B. (1996), “Modified likelihood ratio and score tests in linear regression models using the t distribution,” *Brazilian Journal of Probability and Statistics*, 10(1), 15–33.
- Finlay, R., and Seneta, E. (2006), “Stationary-Increment Student and Variance-Gamma Processes,” *Journal of Applied Probability*, 43(2), 441–453.
- Fonseca, T. C. O., Ferreira, M. A. R., and Migon, H. S. (2008), “Objective Bayesian analysis for the Student-t regression model,” *Biometrika*, 95(2), 325–333.
- Furrer, R., Genton, M. G., and Nychka, D. (2013), “Covariance tapering for interpolation of large spatial datasets,” *Journal of Computational and Graphical Statistics*, 15, 502–523.
- Gao, X., and Song, P. X.-K. (2010), “Composite Likelihood Bayesian Information Criteria for Model Selection in High-Dimensional Data,” *Journal of the American Statistical Association*, 105(492), 1531–1540.

- Genton, M. G., and Zhang, H. (2012), “Identifiability problems in some non-Gaussian spatial random fields,” *Chilean Journal of Statistics*, 3(171–179).
- Gneiting, T. (2002), “Compactly supported correlation functions,” *Journal of Multivariate Analysis*, 83, 493–508.
- Gneiting, T. (2013), “Strictly and non-strictly positive definite functions on spheres,” *Bernoulli*, 19(4), 1327–1349.
- Gneiting, T., and Schlather, M. (2004), “Stochastic models that separate fractal dimension and the Hurst effect,” *SIAM Rev.*, 46, 269–282.
- Gradshteyn, I., and Ryzhik, I. (2007), *Table of Integrals, Series, and Products*, 7 edn, New York: Academic Press.
- Gräler, B. (2014), “Modelling Skewed Spatial Random Fields Through the Spatial Vine Copula,” *Spatial Statistics*, 10, 87 –102.
- Griffiths, R. C. (1970), “Infinitely divisible multivariate gamma distributions,” *Sankhyā Ser. A*, 32, 393–404.
- Heagerty, P., and Lele, S. (1998), “A Composite Likelihood Approach to Binary Spatial Data,” *Journal of the American Statistical Association*, 93, 1099 –1111.
- Heyde, C. C., and Leonenko, N. N. (2005), “Student processes,” *Adv. in Appl. Probab.*, 37(2), 342–365.
- J. F. Rosco, M. C. J., and Pewsey, A. (2011), “Skew t distributions via the sinh-arcsinh transformation,” *TEST*, 30, 630 – 652.
- Joe, H., and Lee, Y. (2009), “On weighting of bivariate margins in pairwise likelihood,” *Journal of Multivariate Analysis*, 100, 670–685.
- Kazianka, H., and Pilz, J. (2010), “Copula-based geostatistical modeling of continuous and discrete data including covariates,” *Stochastic Environmental Research and Risk Assessment*, 24, 661–673.

- Kilibarda, M., Hengl, T., Heuvelink, G. B. M., Gräler, B., Pebesma, E., Perčec Tadić, M., and Bajat, B. (2014), “Spatio-temporal interpolation of daily temperatures for global land areas at 1km resolution,” *Journal of Geophysical Research: Atmospheres*, 119(5), 2294–2313.
- Krishnaiah, P. R., and Rao, M. M. (1961), “Remarks on a Multivariate Gamma Distribution,” *The American Mathematical Monthly*, 68(4), 342–346.
- Krishnamoorthy, A. S., and Parthasarathy, M. (1951), “A Multivariate Gamma-Type Distribution,” *Ann. Math. Statist.*, 22(4), 549–557.
- Lange, K. L., Little, R. J. A., and Taylor, J. M. G. (1989), “Robust Statistical Modeling Using the t Distribution,” *Journal of the American Statistical Association*, 84(408), 881–896.
- Lim, S., and Teo, L. (2009), “Gaussian fields and Gaussian sheets with Generalized Cauchy covariance structure,” *Stochastic Processes and Their Applications*, 119(4), 1325–1356.
- Lindsay, B. (1988), “Composite likelihood methods,” *Contemporary Mathematics*, 80, 221–239.
- Ma, C. (2009), “Construction of non-Gaussian random fields with any given correlation structure,” *Journal of Statistical Planning and Inference*, 139, 780–787.
- Ma, C. (2010a), “Elliptically contoured random fields in space and time,” *Journal of Physics A: Mathematical and Theoretical*, 43, 167–209.
- Mahmoudian, B. (2017), “A Skewed and Heavy-Tailed Latent Random Field Model for Spatial Extremes,” *Journal of Computational and Graphical Statistics*, 26(3), 658–670.
- Mahmoudian, B. (2018), “On the existence of some skew-Gaussian random field models,” *Statistics and Probability Letters*, 137, 331 – 335.

- Marcus, M. (2014), “Multivariate gamma distributions,” *Electron. Commun. Probab.*, 19, 10 pp.
URL: <https://doi.org/10.1214/ECP.v19-3794>
- Masarotto, G., and Varin, C. (2012), “Gaussian copula marginal regression,” *Electronic Journal of Statistics*, 6, 1517–1549.
- Matérn, B. (1986), *Spatial Variation: Stochastic Models and their Applications to Some Problems in Forest Surveys and Other Sampling Investigations*, 2nd edn, Heidelberg: Springer.
- Miller, K. S. (1968), “Some Multivariate t -Distributions,” *Ann. Math. Statist.*, 39(5), 1605–1609.
URL: <https://doi.org/10.1214/aoms/1177698141>
- Palacios, M. B., and Steel, M. F. J. (2006), “Non-Gaussian Bayesian Geostatistical Modeling,” *Journal of the American Statistical Association*, 101, 604–618.
- Pearson, J. W., S. Olver, and Porter, M. A. (2017), “Numerical methods for the computation of the confluent and Gauss hypergeometric functions,” *Numerical Algorithms*, 74(3), 821 – 866.
- Plemmons, R. (1977), “M-Matrix Characterizations. I – Nonsingular M-Matrices,” *Linear Algebra and its Applications*, 18(2), 175 –188.
- Porcu, E., Bevilacqua, M., and Genton, M. (2018), “Nonseparable, Space-Time Covariance Functions with Dynamical Compact Supports,” *Statistica Sinica*, to appear.
- Porcu, E., Bevilacqua, M., and Genton, M. G. (2016), “Spatio-Temporal Covariance and Cross-Covariance Functions of the Great Circle Distance on a Sphere,” *Journal of the American Statistical Association*, 111(514), 888–898.
- Røislien, J., and Omre, H. (2006), “T-Distributed Random Fields: A Parametric Model for Heavy-Tailed Well-Log Data,” *Mathematical Geology*, 38, 821–849.

- Royen, T. (2004), Multivariate Gamma distributions II,, in *Encyclopedia of Statistical Sciences*, New York: John Wiley & Sons, pp. 419–425.
- Sahu, K. S., Dey, D. K., and Márcia, M. D. B. (2003), “A new class of multivariate skew distributions with applications to bayesian regression models,” *Canadian Journal of Statistics*, 31(2), 129–150.
- Stein, M. (1999), *Interpolation of Spatial Data. Some Theory of Kriging*, New York: Springer-Verlag.
- Varin, C., Reid, N., and Firth, D. (2011), “An overview of composite likelihood methods,” *Statistica Sinica*, 21, 5–42.
- Varin, C., and Vidoni, P. (2005), “A note on composite likelihood inference and model selection,” *Biometrika*, 52, 519–528.
- Vere-Jones, D. (1967), “The infinite divisibility of a bivariate gamma distribution,” *Sankhyā: The Indian Journal of Statistics*, 29, 421–422.
- Vere-Jones, D. (1997), “Alpha-permanents and their applications to multivariate gamma, negative binomial and ordinary binomial distributions,” *New Zealand J. Math*, 26, 125–149.
- Wallin, J., and Bolin, D. (2015), “Geostatistical Modelling using Non-Gaussian Matérn Fields,” *Scandinavian Journal of Statistics*, 42, 872 –890.
- Xua, G., and Genton, M. G. (2017), “Tukey g-and-h Random Fields,” *Journal of the American Statistical Association*, 112, 1236 –1249.
- Yeo, G., and Milne, R. (1991), “On characterizations of beta and gamma distributions,” *Statistics and Probability Letters*, 11(3), 239 – 242.
- Zareifard, H., Khaledi, M. J., Rivaz, F., and Vahidi-Asl, M. Q. (2018), “Modeling Skewed Spatial Data Using a Convolution of Gaussian and Log-Gaussian Processes,” *Bayesian Anal.*, 13(2), 531–557.
- Zhang, H., and El-Shaarawi, A. (2010), “On spatial skew-gaussian processes and applications,” *Environmetrics*, 21(1), 33–47.