Received: 23 April 2019 | Revised: 2 January 2020 | Accepted: 6 January 2020

DOI: 10.1111/sjos.12447

ORIGINAL ARTICLE

Scandinavian Journal of Statistics

Non-Gaussian geostatistical modeling using (skew) t processes

Moreno Bevilacqua[1,2] | **Christian Caamaño-Carrillo**[3] | **Reinaldo B. Arellano-Valle**[4] | **Víctor Morales-Oñate**[5]

[1]Departamento de Estadística, Universidad de Valparaíso

[2]Millennium Nucleus Center for the Discovery of Structures in Complex Data

[3]Departamento de Estadística, Universidad del Bío-Bío

[4]Departamento de Estadística, Pontificia Universidad Católica de Chile

[5]Departamento de Desarrollo, Ambiente y Territorio, Facultad Latinoamericana de Ciencias Sociales

Correspondence
Moreno Bevilacqua, Departamento de Estadística, Universidad de Valparaíso, Valparaíso 2340000, Chile.
Email: moreno.bevilacqua@uv.cl

Abstract

We propose a new model for regression and dependence analysis when addressing spatial 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

Gaussian scale mixture, heavy-tailed processes, hypergeometric functions, multivariate skew-normal distribution, pairwise likelihood

1 | INTRODUCTION

The geostatistical approach models data coming from a limited number of monitoring stations as a partial realization from a spatial stochastic process (or random field) defined on the continuum space. Gaussian stochastic processes are among the most popular tools for analyzing spatial data because a mean structure and a valid covariance function completely characterized the associated finite dimensional distribution. In addition, optimal prediction at an unobserved site depends on the knowledge of the covariance function of the process. Unfortunately, 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.

The focus of this work is on non-Gaussian models for stochastic processes that vary continuously in the euclidean space, even if the proposed methodology can be easily extended to the space-time framework or to the spherical space. In particular, we aim to accommodate heavier tails than the ones induced by Gaussian processes and wish to allow possible asymmetry. 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 (Allcroft & Glasbey, 2003; De Oliveira, 2006; De Oliveira, Kedem, & Short, 1997). 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. However, this approach is restricted to the Matérn covariance model with integer smoothness parameter and its statistical properties are much less understood than those of the Gaussian process.

The copula framework (Joe, 2014) has been adapted in the spatial context in order to account for possible deviations from the Gaussian distribution. Even though which copula model to use for a given analysis is not generally known a priori, the copula based on the multivariate Gaussian distribution has gained a general consensus (Gräler, 2014; Kazianka & Pilz, 2010; Masarotto & Varin, 2012) since the definition of the multivariate dependence relies again on the specification of the correlation function. However, Gaussian copula could be too restrictive in some cases since it expresses a symmetrical and elliptical dependence.

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 marginal distributions of the skew-Gaussian type (Azzalini & Capitanio, 2014). Zareifard, Khaledi, Rivaz, and Vahidi-Asl (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, Dey, and Márcia (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) and Zareifard

and Khaledi (2013) proposed a (skew) Gaussian-Log-Gaussian scale mixing approach in order to accommodate the presence of possible outliers for spatial data.

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 (Arellano-Valle & Bolfarine, 1995; Arellano-Valle, Castro, & Gonzalez-Farias, 2012; Ferrari & Arellano-Valle, 1996; Fonseca, Ferreira, & Migon, 2008; Lange, Little, & Taylor, 1989). Stochastic processes with marginal t distributions have been introduced in Røislien and Omre (2006), Ma (2009, 2010), and DeBastiani, Cysneiros, Uribe-Opazo, and Galea (2015), but as outlined in Genton and Zhang (2012), these models are not be identifiable when only a single realization is available (which is typically the case for spatial data).

In this article, we propose a process with marginal t distributions obtained though 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. 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 & Ryzhik, 2007). In addition, the bivariate distribution is not of elliptical type.

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, Caro, Bevilacqua, Porcu, and Clarke (2017). We then propose a process with marginal distribution of the skew-t type (Azzalini & 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 & 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 some of 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 as the Matérn model (Matèrn, 1986) or the Generalized Wendland model (Bevilacqua, Faouzi, Furrer, & Porcu, 2019; Gneiting, 2002).

For the t process estimation, we propose the method of weighted pairwise likelihood (Bevilacqua & Gaetan, 2015; Lindsay, 1988; Varin, Reid, & Firth, 2011) exploiting the bivariate distribution given in Theorem 3. In an extensive simulation study, we investigate the performance of the weighted pairwise likelihood (*wpl*) method under different scenarios including when the degrees of freedom are supposed to be unknown. We also study the performance of the *wpl* estimation by assuming a Gaussian process in the estimation step with correlation function equal to the correlation function of the t process. It turns out that the Gaussian misspecified *wpl* (see Gouriéroux, Monfort, & Renault, 2017 with the references therein) leads to a less efficient estimator, as expected. However, the method has some computational benefits.

In addition, 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 dataset 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 article is implemented in the R package `GeoModels` (Bevilacqua & Morales-Oñate, 2018). The remainder of the article 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 (misspecified) *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 dataset of maximum temperature in Australia. Finally, in Section 6, we give some conclusions. All the proofs have been deferred to the Appendix.

2 | A STOCHASTIC PROCESS WITH T MARGINAL DISTRIBUTION

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

As outlined in Palacios and Steel (2006), given a positive process $M = \{M(\boldsymbol{s}), \boldsymbol{s} \in A\}$ and an independent standard Gaussian process $G^* = \{G^*(\boldsymbol{s}), \boldsymbol{s} \in A\}$, a general class of non-Gaussian processes with marginal heavy tails can be obtained as scale mixture of G^*, that is, $\mu(\boldsymbol{s}) + \sigma M(\boldsymbol{s})^{-\frac{1}{2}} G^*(\boldsymbol{s})$, where $\mu(\boldsymbol{s})$ is the location-dependent mean and $\sigma > 0$ is a scale parameter. A typical parametric specification for the mean is given by $\mu(\boldsymbol{s}) = X(\boldsymbol{s})^T \boldsymbol{\beta}$, where $X(\boldsymbol{s}) \in \mathbb{R}^k$ 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 $\rho(\boldsymbol{h}) := \rho_{G^*}(\boldsymbol{h})$ and $G := G^*$. Our proposal considers a mixing process $W_\nu = \{W_\nu(\boldsymbol{s}), \boldsymbol{s} \in A\}$ with marginal distribution $\Gamma(\nu/2, \nu/2)$ defined as $W_\nu(\boldsymbol{s}) := \sum_{i=1}^{\nu} G_i(\boldsymbol{s})^2/\nu$, where $G_i, i = 1, \dots \nu$ are independent copies of G with $\mathrm{E}(W_\nu(\boldsymbol{s})) = 1$, $\mathrm{Var}(W_\nu(\boldsymbol{s})) = 2/\nu$ and $\rho_{W_\nu}(\boldsymbol{h}) = \rho^2(\boldsymbol{h})$ (Bevilacqua, Caamaño, & Gaetan, 2018). If we consider a process $Y_\nu^* = \{Y_\nu^*(\boldsymbol{s}), \boldsymbol{s} \in A\}$ defined as

$$Y_\nu^*(\boldsymbol{s}) := W_\nu(\boldsymbol{s})^{-\frac{1}{2}} G(\boldsymbol{s}), \tag{1}$$

then, by construction, Y_ν^* has the marginal t distribution with ν degrees of freedom with pdf given by:

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

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

$$Y_\nu(\boldsymbol{s}) := \mu(\boldsymbol{s}) + \sigma Y_\nu^*(\boldsymbol{s}) \tag{3}$$

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

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 zero. This feature is intimately related to the infinite divisibility of the squared Gaussian process $G^2 = \{G^2(\boldsymbol{s}), \boldsymbol{s} \in A\}$ as shown in Krishnaiah and Rao (1961). Characterization of the infinite 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 $\rho(\boldsymbol{h})$. Specifically, $\nu > 0$ if and only if there exists a matrix S_n such that $S_n\Omega^{-1}S_n$ is an M-matrix (Plemmons, 1977), where S_n is a signature matrix, that is, a diagonal matrix of size n 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. The t process Y_ν^* inherits this feature with the additional restriction $\nu > 2$. This implies that Y_ν^* is well defined for $\nu = 3, 4, \ldots$ and for $\nu > 2$ under noninfinite divisibility of G^2.

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 $f_{\boldsymbol{W}_\nu}$ was first discussed by Krishnamoorthy and Parthasarathy (1951) and its properties have been studied by different authors (Marcus, 2014; Royen, 2004). In the bivariate case, Vere-Jones (1967) showed that the bivariate Gamma distribution is infinite divisible, that is, $\nu > 0$ in (A2), irrespective of the correlation function. Note that this is consistent with the characterization given in Bapat (1989) since, given an arbitrary bivariate correlation matrix Ω, there exists a matrix S_2 such that $S_2\Omega^{-1}S_2$ is a M-matrix. In Theorem 3, we provide the bivariate distribution of Y_ν^*.

Note that, both W_ν and G in (3) are obtained through independent copies of the "parent" Gaussian process with correlation $\rho(\boldsymbol{h})$. For this reason, henceforth, in some cases, we will call Y_ν^* a standard t process with underlying correlation $\rho(\boldsymbol{h})$.

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

$${}_2F_1(a, b, c; x) = \sum_{k=0}^{\infty} \frac{(a)_k (b)_k}{(c)_k} \frac{x^k}{k!}, \tag{4}$$

with $(s)_k = \Gamma(s+k)/\Gamma(s)$ for $k \in \mathbb{N} \cup \{0\}$ being the Pochhammer symbol and we consider the restrictions $a > 0, b > 0, c > 0$, and $x \geq 0$. If $c > a + b$, the radius of convergence of (4) is $0 \leq x \leq 1$ and, in particular (4) is convergent at $x = 1$ through the identity:

$${}_2F_1(a, b; c; 1) = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)}. \tag{5}$$

We also consider the Appell hypergeometric function of the fourth type (Gradshteyn & 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 $\rho_{Y_\nu^*}(\boldsymbol{h})$ in terms of the Gauss hypergeometric function.

Theorem 1. *Let Y_ν^* be a standardized t process with underlying correlation $\rho(\boldsymbol{h})$. Then:*

$$\rho_{Y_\nu^*}(\boldsymbol{h}) = \frac{(\nu-2)\Gamma^2\left(\frac{\nu-1}{2}\right)}{2\Gamma^2\left(\frac{\nu}{2}\right)}[{}_2F_1\left(\frac{1}{2},\frac{1}{2};\frac{\nu}{2};\rho^2(\boldsymbol{h})\right)\rho(\boldsymbol{h})]. \tag{6}$$

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. Furthermore, 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 & Schlather, 2004; Lim & Teo, 2009) and Dagum (Berg, Mateu, & Porcu, 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 $\nu \to \infty$.

Theorem 2. *Let Y_ν^*, $\nu > 2$ be a standardized t process with underlying correlation $\rho(\boldsymbol{h})$. 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) Let G m-times mean-square differentiable, for $m = 0, 1, \ldots$
- *If $\nu > 2(2m+1)$ then Y_ν^* is m-times mean-square differentiable;*
- *If $\nu \le 2(2m+1)$ then Y_ν^* is $(m-k)$-times mean-square differentiable if $2(2(m-k)+1) < \nu \le 2(2(m-k)+3)$, for $k = 1, \ldots, m$.*

(d) Y_ν^* *is a long-range-dependent process if and only if G is a long-range-dependent process*
(e) $\rho_{Y_\nu^*}(\boldsymbol{h}) \le \rho(\boldsymbol{h})$ *and* $\lim_{\nu\to\infty} \rho_{Y_\nu^*}(\boldsymbol{h}) = \rho(\boldsymbol{h})$.

One implication of Theorem 2 point (c) is that the process Y_ν^* inherits the mean square differentiability of G under the condition $\nu > 2(2m+1)$. Otherwise, the mean square differentiability depends on ν. For instance, if G is one time mean square differentiable, then Y_ν^* can be zero or one time differentiable depending if $\nu > 6$ or not.

Remark 3. A simplified version of the t process in Equation (1), can be obtained assuming $W_\nu(\boldsymbol{s}_i) \perp W_\nu(\boldsymbol{s}_j)$, $i \neq j$. Under this assumption, Y_ν^* is still a process with t marginal distribution but, in this case, the geometrical properties are not inherited from the "parent" Gaussian process G. In particular, it can be shown that the resulting correlation function exhibits a discontinuity at the origin and, as a consequence, the process is not mean-square continuous. A not mean-square continuous version of the t process in Equation (1) can be obtained by introducing a nugget effect, that is, a discontinuity of $\rho_{Y_\nu^*}(\boldsymbol{h})$ at the origin. This can be easily achieved by replacing $\rho(\boldsymbol{h})$ in (6) with $\rho^*(\boldsymbol{h}) = 1$ if $\boldsymbol{h} = \boldsymbol{0}$ and $\rho^*(\boldsymbol{h}) = (1-\tau^2)\rho(\boldsymbol{h})$ otherwise, where $0 \le \tau^2 < 1$ represents the underlying nugget effect.

Since the t process inherits some of the geometrical properties of the "parent" Gaussian process, the choice of the covariance function is crucial. Two flexible isotropic 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 (Matèrn, 1986)

$$\mathcal{M}_{\alpha,\psi}(\boldsymbol{h}) = \frac{2^{1-\psi}}{\Gamma(\psi)}(||\boldsymbol{h}||/\alpha)^{\psi}\mathcal{K}_{\psi}(||\boldsymbol{h}||/\alpha), \tag{7}$$

where $\mathcal{K}_{\psi}$ 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:

$$\mathcal{GW}_{\alpha,\psi,\delta}(\boldsymbol{h}) := \begin{cases} \frac{\int_{||\boldsymbol{h}||/\alpha}^{1} u\left(u^2-(||\boldsymbol{h}||/\alpha)^2\right)^{\psi-1}(1-u)^{\delta}du}{\mathrm{B}(2\psi,\delta+1)} & ||\boldsymbol{h}|| < \alpha \\ 0 & \text{otherwise} \end{cases}, \tag{8}$$

and for $\psi = 0$ as:

$$\mathcal{GW}_{\alpha,0,\delta}(\boldsymbol{h}) := \begin{cases} (1-||\boldsymbol{h}||/\alpha)^{\delta} & ||\boldsymbol{h}|| < \alpha \\ 0 & \text{otherwise} \end{cases}. \tag{9}$$

Here $\mathrm{B}(\cdot,\cdot)$ is the Beta function and $\alpha > 0$, $\delta \geq (d+1)/2 + \psi$ 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 et al., 2019). In addition, the Generalized Wendland correlation is compactly supported, an interesting feature from computational point of view (Furrer, Genton, & Nychka, 2013), which is inherited by the t process since $\rho(\boldsymbol{h}) = 0$ implies $\rho_{Y_{\nu}^*}(\boldsymbol{h}) = 0$.

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 $\rho_{Y_{\nu}^*}(\boldsymbol{h})$ when $\nu = 3, 7$ with the correlation of the "parent" Gaussian process $\rho(\boldsymbol{h}) = \mathcal{M}_{1.5,\alpha^*}(\boldsymbol{h})$, where α^* is chosen such that the practical range is 0.2. It is apparent that when increasing the degrees of freedom $\rho_{Y_{\nu}^*}(\boldsymbol{h})$ approaches $\rho(\boldsymbol{h})$ and that the smoothness at the origin of $\rho_{Y_{\nu}^*}(\boldsymbol{h})$ is inherited by the smoothness of the Gaussian correlation $\rho(\boldsymbol{h})$ when $\nu = 7$ and if $\nu = 3$ then $\rho_{Y_{\nu}^*}(\boldsymbol{h})$ is not differentiable at the origin. On the right side of Figure 1, we compare a kernel nonparametric density estimation of a realization of G and a realization of Y_7^* (approximately 10 000 location sites in the unit square) using $\rho(\boldsymbol{h}) = \mathcal{M}_{1.5,\alpha^*}(\boldsymbol{h})$.

In Figure 2a,b, we compare, from left to right, two realizations of G with $\rho(\boldsymbol{h}) = \mathcal{M}_{0.5,\alpha^*}(\boldsymbol{h})$ and $\rho(\boldsymbol{h}) = \mathcal{M}_{1.5,\alpha^*}(\boldsymbol{h})$, 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 2c,d, it can be appreciated that this feature is inherited by the associated realizations of Y_7^*.

We now consider the bivariate random vector associated with Y_{ν}^* defined by:

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

where $\circ$ denotes the Schur product vector. The following theorem gives the pdf of $\boldsymbol{Y}_{\nu;ij}^*$ 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).

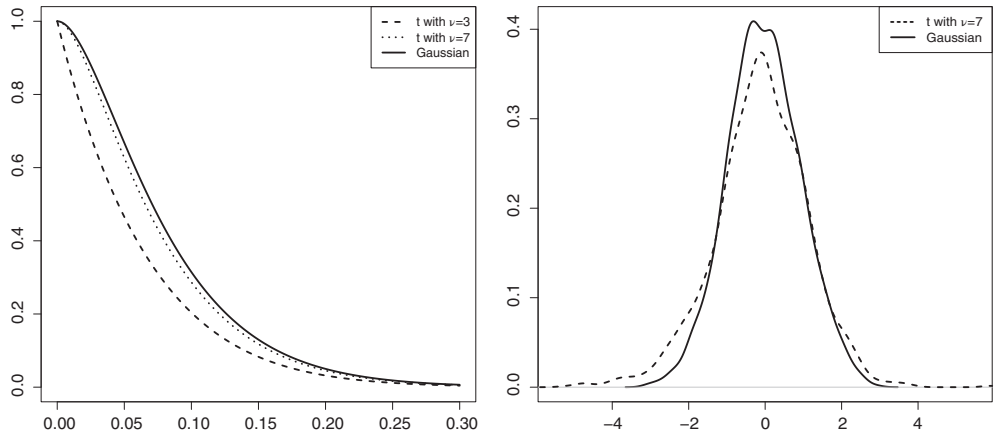

FIGURE 1 Left part: comparison of $\rho_{Y_\nu^*}(\boldsymbol{h})$, $\nu = 3, 7$ with the correlation $\rho(\boldsymbol{h})$ of the "parent" Gaussian process G when $\rho(\boldsymbol{h}) = \mathcal{M}_{1.5,\alpha^*}(\boldsymbol{h})$ with α^* such that the practical range is 0.2. Right part: a comparison of a nonparametric kernel density estimation of realizations from G and from the t process Y_7^*

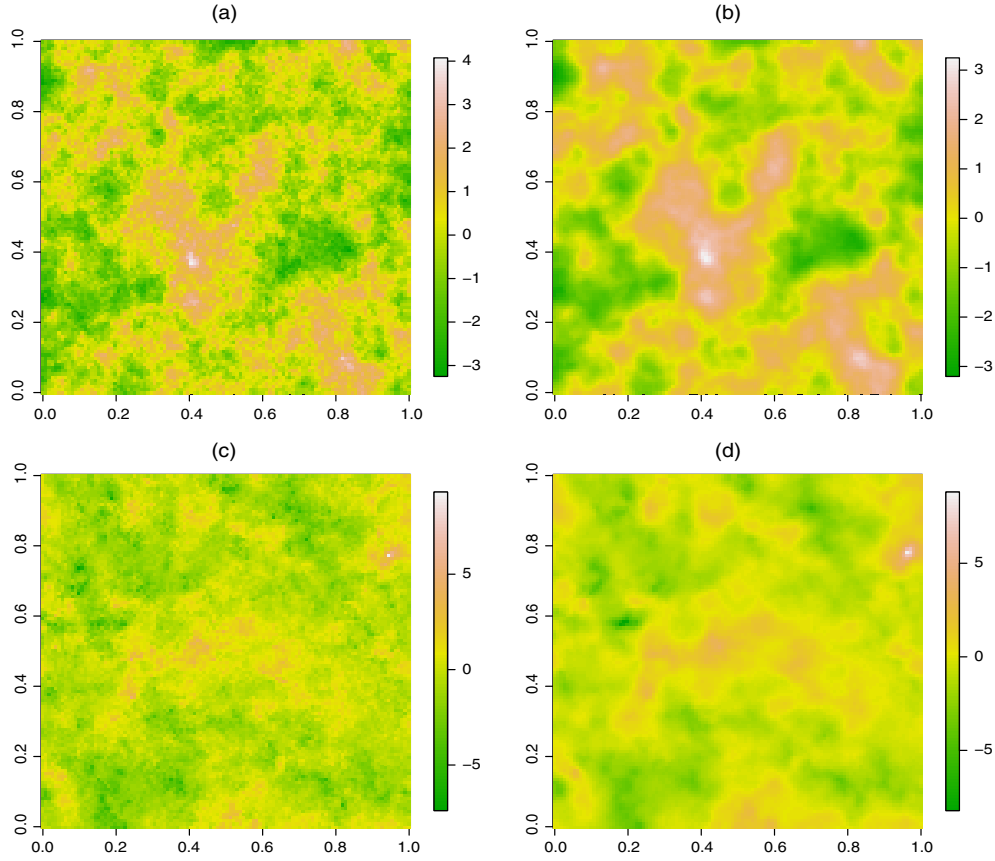

FIGURE 2 Upper part: two realizations of the "parent" Gaussian process G on $[0,1]^2$ with (a) $\rho(\boldsymbol{h}) = \mathcal{M}_{0.5,\alpha^*}(\boldsymbol{h})$ and (b) $\rho(\boldsymbol{h}) = \mathcal{M}_{1.5,\alpha^*}(\boldsymbol{h})$ (from left to right) with α^* such that the practical range is approximatively 0.2. Bottom part: (c) and (d) associated realizations of the t process Y_7^* [Color figure can be viewed at wileyonlinelibrary.com]

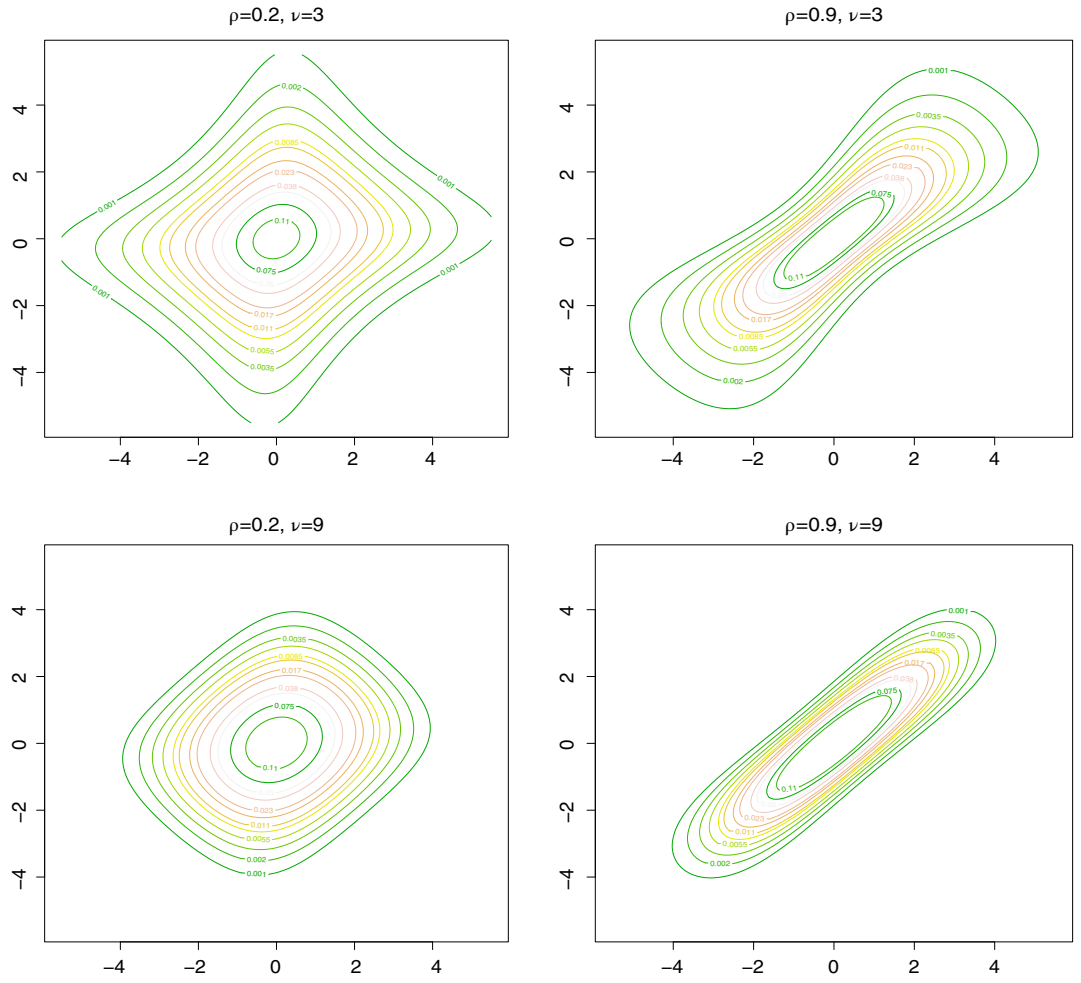

FIGURE 3 Contour plots of the bivariate t distribution (10) when $\rho(\boldsymbol{h}) = 0.2, 0.9$ and $\nu = 3, 9$ [Color figure can be viewed at wileyonlinelibrary.com]

Theorem 3. *Let Y_ν^*, $\nu > 2$ be a standard t process with underlying correlation $\rho(\boldsymbol{h})$. Then:*

$$f_{\boldsymbol{Y}^*_{\nu;ij}}(y_i, y_j) = \frac{\nu^\nu l_{ij}^{-\frac{(\nu+1)}{2}} \Gamma^2\left(\frac{\nu+1}{2}\right)}{\pi\Gamma^2\left(\frac{\nu}{2}\right)(1-\rho^2(\boldsymbol{h}))^{-(\nu+1)/2}} F_4\left(\frac{\nu+1}{2}, \frac{\nu+1}{2}, \frac{1}{2}, \frac{\nu}{2}; \frac{\rho^2(\boldsymbol{h}) y_i^2 y_j^2}{l_{ij}}, \frac{\nu^2\rho^2(\boldsymbol{h})}{l_{ij}}\right)$$
$$+ \frac{\rho(\boldsymbol{h}) y_i y_j \nu^{\nu+2} l_{ij}^{-\frac{\nu}{2}-1}}{2\pi(1-\rho^2(\boldsymbol{h}))^{-\frac{(\nu+1)}{2}}} F_4\left(\frac{\nu}{2}+1, \frac{\nu}{2}+1, \frac{3}{2}, \frac{\nu}{2}; \frac{\rho^2(\boldsymbol{h}) y_i^2 y_j^2}{l_{ij}}, \frac{\nu^2\rho^2(\boldsymbol{h})}{l_{ij}}\right), \tag{10}$$

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

Remark 4. Note that $f_{\boldsymbol{Y}^*_{\nu;ij}}(y_i, y_j)$ is defined for $\nu > 2$ irrespectively of the correlation function since it is obtained from a bivariate Gamma distribution (see Remark 2). Moreover, when $\rho(\boldsymbol{h}) = 0$, according to (21) 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, $f_{\boldsymbol{Y}^*_{\nu;ij}}(y_i, y_j)$ 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. Figure 3 shows the contour plots of (10) when $\nu = 3, 9$ and $\rho(\boldsymbol{h}) = 0.2, 0.9$. It turns out that the bivariate t distribution is not elliptical and when increasing ν, the contour plots tend toward an elliptical form. Finally, the bivariate density of the process Y_ν is easily obtained from (3):

$$f_{\boldsymbol{Y}_{\nu;ij}}(y_i, y_j) = \frac{1}{\sigma^2} f_{\boldsymbol{Y}^*_{\nu;ij}}\left(\frac{y_i - \mu_i}{\sigma}, \frac{y_j - \mu_j}{\sigma}\right). \tag{11}$$

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 & Capitanio, 2014).

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

$$U_\eta(\boldsymbol{s}) := g(\boldsymbol{s}) + \eta|X_1(\boldsymbol{s})| + \omega X_2(\boldsymbol{s}), \quad \boldsymbol{s} \in A \subset \mathbb{R}^d \tag{12}$$

where $\eta \in \mathbb{R}$, $\omega > 0$ and X_i $i = 1, 2$ are two independents copies of a process $X = \{X(\boldsymbol{s}), \boldsymbol{s} \in A\}$ 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 skew Gaussian marginal distributions (Azzalini & Capitanio, 2014) with pdf given by:

$$f_{U_\eta(\boldsymbol{s})}(u) = \frac{2}{(\eta^2 + \omega^2)^{1/2}} \phi\left(\frac{(u - g(\boldsymbol{s}))}{(\eta^2 + \omega^2)^{1/2}}\right) \Phi\left(\frac{\eta(u - g(\boldsymbol{s}))}{\omega(\eta^2 + \omega^2)^{1/2}}\right) \tag{13}$$

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

$$\rho_{U_\eta}(\boldsymbol{h}) = \frac{2\eta^2}{\pi\omega^2 + \eta^2(\pi - 2)} \left((1 - \rho^2(\boldsymbol{h}))^{1/2} + \rho(\boldsymbol{h}) \arcsin(\rho(\boldsymbol{h})) - 1\right) + \frac{\omega^2 \rho(\boldsymbol{h})}{\omega^2 + \eta^2(1 - 2/\pi)}. \tag{14}$$

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 $\boldsymbol{U}_\eta$.

Theorem 4. *Let $U_\eta(\boldsymbol{s}) = g(\boldsymbol{s}) + \eta|X_1(\boldsymbol{s})| + \omega X_2(\boldsymbol{s})$ where X_i $i = 1, 2$ are two independent copies of G the "parent" Gaussian process. Then:*

$$f_{\boldsymbol{U}_\eta}(\boldsymbol{u}) = 2 \sum_{l=1}^{2^{n-1}} \phi_n(\boldsymbol{u} - \boldsymbol{\alpha}; \boldsymbol{A}_l) \Phi_n(\boldsymbol{c}_l; \boldsymbol{0}, \boldsymbol{B}_l) \tag{15}$$

where

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

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

Some comments are in order. First, note that $f_{\boldsymbol{U}}$ can be viewed as a generalization of the multivariate skew-Gaussian distribution proposed in Azzalini and Dalla-Valle (1996). Second, using Theorem 4, 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 & Capitanio, 2014), we replace the process G in (3) with the process U_η. Specifically, we consider a process $S_{\nu,\eta} = \{S_{\nu,\eta}(\boldsymbol{s}), \boldsymbol{s}\in A\}$ defined as

$$S_{\nu,\eta}(\boldsymbol{s}) := \mu(\boldsymbol{s}) + \sigma W_\nu(\boldsymbol{s})^{-\frac{1}{2}} U_\eta(\boldsymbol{s}), \tag{16}$$

where W_ν and U_η are supposed to be independent. In (12), we assume $g(\boldsymbol{s}) = 0$ and $\eta^2 + \omega^2 = 1$. The pdf of the marginal distribution of $S^*_{\nu,\eta}$ is given by:

$$f_{S^*_{\nu,\eta}(\boldsymbol{s})}(g) = 2f_{Y^*_\nu(\boldsymbol{s})}(g;\nu)F_{Y^*_\nu(\boldsymbol{s})}\left(\eta g\sqrt{\frac{\nu+1}{\nu+g^2}};\nu+1\right) \tag{17}$$

with $\mathrm{E}(S^*_{\nu,\eta}(\boldsymbol{s})) = \dfrac{\sqrt{\nu}\Gamma\left(\frac{\nu-1}{2}\right)\eta}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}$, and $\mathrm{Var}(S^*_{\nu,\eta}(\boldsymbol{s})) = \left[\dfrac{\nu}{\nu-2} - \dfrac{\nu\Gamma^2\left(\frac{\nu-1}{2}\right)\eta^2}{\pi\Gamma^2\left(\frac{\nu}{2}\right)}\right]$.

If $\eta = 0$, (17) reduces to a marginal t density given in (2) and if $\nu\to\infty$, (17) converges to a skew-normal distribution. Moreover, coupling (6) and (14) the correlation function of the skew-t process is given by:

$$\rho_{S^*_{\nu,\eta}}(\boldsymbol{h}) = a(\nu,\eta)[{}_2F_1\left(\frac{1}{2},\frac{1}{2};\frac{\nu}{2};\rho^2(\boldsymbol{h})\right)\left\{(1+\eta^2(1-\frac{2}{\pi}))\rho_{U_\eta}(\boldsymbol{h}) + \frac{2\eta^2}{\pi}\right\} - \frac{2\eta^2}{\pi}], \tag{18}$$

where $a(\nu,\eta) = \dfrac{\pi(\nu-2)\Gamma^2\left(\frac{\nu-1}{2}\right)}{2\left[\pi\Gamma^2\left(\frac{\nu}{2}\right)(1+\eta^2)-\eta^2(\nu-2)\Gamma^2\left(\frac{\nu-1}{2}\right)\right]}$. Note that $\rho_{S^*_{\nu,\eta}}(\boldsymbol{h}) = \rho_{S^*_{\nu,-\eta}}(\boldsymbol{h})$, 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 point (e), it can be shown that $\lim_{\nu\to\infty}\rho_{S^*_{\nu,\eta}}(\boldsymbol{h}) = \rho_{U_\eta}(\boldsymbol{h})$.

Finally, following the steps of the proof of Theorem 2, it can be shown that properties (a), (b), (c), and (d) in Theorem 2 are true for the skew-t process $S^*_{\nu,\eta}$.

Figure 4, left part, compares $\rho_{S^*_{7,0.9}}(\boldsymbol{h})$ and $\rho_{S^*_{7,0}}(\boldsymbol{h}) = \rho_{Y^*_7}(\boldsymbol{h})$ with the underlying correlation $\rho(\boldsymbol{h}) = \mathcal{GW}_{0.3,1,5}(\boldsymbol{h})$. The right part shows a realization of $S^*_{7,0.9}$.

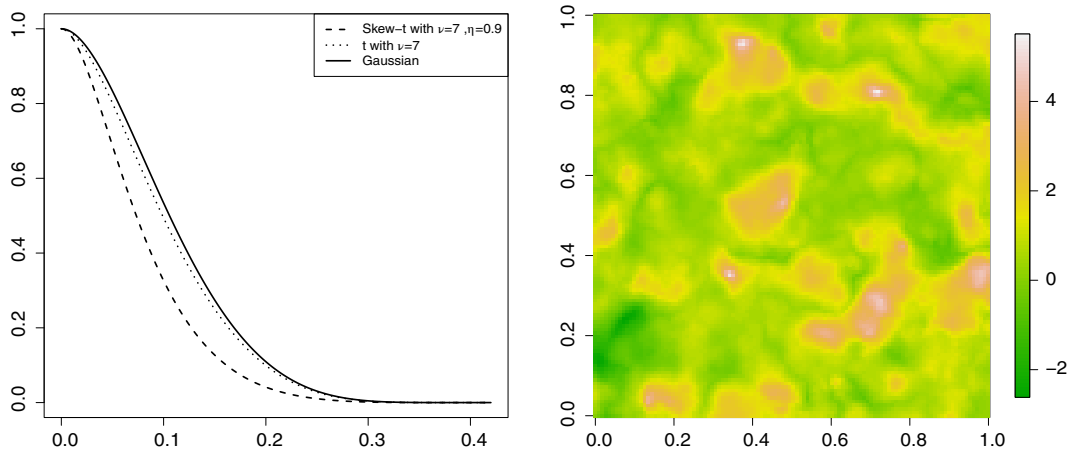

FIGURE 4 From left to right: (a) comparison between $\rho_{S^*_{7,0.9}}(\boldsymbol{h})$, $\rho_{S^*_{7,0}}(\boldsymbol{h}) = \rho_{Y^*_7}(\boldsymbol{h})$ and the underlying correlation $\rho(\boldsymbol{h}) = \mathcal{GW}_{0.3,1,5}(\boldsymbol{h})$ and (b) a realization from $S^*_{7,0.9}(\boldsymbol{s})$ [Color figure can be viewed at wileyonlinelibrary.com]

4 | NUMERICAL EXAMPLES

In this section, 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 $\nu > 2$ and $\nu = 3, 4, \ldots$. 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.

We also compare the performance of the *wpl* using the bivariate t distribution (10) with a misspecified Gaussian standard and weighted pairwise likelihood. Finally, we compare the performance of the optimal linear predictor of the t process using (6) versus the optimal predictor of the Gaussian process.

4.1 | Weighted pairwise likelihood estimation

Let $(y_1, \ldots, y_n)^T$ be a realization of the t random process Y_ν defined in Equation (3) observed at distinct spatial locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$, $\boldsymbol{s}_i \in A$ 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 underlying correlation model. 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

$$\mathrm{pl}(\boldsymbol{\theta}) := \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \log(f_{\boldsymbol{Y}_{\nu;ij}}(y_i, y_j; \boldsymbol{\theta}))c_{ij}, \tag{19}$$

where $f_{\boldsymbol{Y}_{\nu;ij}}(y_i, y_j; \boldsymbol{\theta})$ is the bivariate density in (11) and c_{ij} is a nonnegative suitable weight. The choice of cutoff weights, namely,

$$c_{ij} = \begin{cases} 1 & \|\boldsymbol{s}_i - \boldsymbol{s}_j\| \leq d_{ij} \\ 0 & \text{otherwise} \end{cases}, \tag{20}$$

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 & Gaetan, 2015; Joe & Lee, 2009). The maximum *wpl* estimator is given by

$$\hat{\boldsymbol{\theta}} := \arg\max_{\boldsymbol{\theta}} \text{pl}(\boldsymbol{\theta})$$

and, arguing as in Bevilacqua, Gaetan, Mateu, and Porcu (2012) and Bevilacqua and Gaetan (2015), under some mixing conditions of the t process, 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 $\mathcal{H}_n(\boldsymbol{\theta}) := \text{E}[-\nabla^2 \text{pl}(\boldsymbol{\theta})]$ 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 (Gao & Song, 2010; Varin & Vidoni, 2005). 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, Kang, & Song, 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 associated with the original parametrization. Here we consider two possible scenarios, that is, a t process observed on a subset of $\mathbb{R}$ and $\mathbb{R}^2$.

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, the t process is well defined for $0 < \lambda < 1/2$ and in this specific case, all the parameters 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 cutoff 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 $\boldsymbol{s}_i \in A = [0, 1]^2$, $i = 1, \dots, N$. Specifically, we simulate, using Cholesky decomposition, 500 realizations of a t process observed at $N = 500$ spatial location sites

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)

	1/3		1/6		1/9	
λ	**Bias**	**MSE**	**Bias**	**MSE**	**Bias**	**MSE**
$\hat{\lambda}$	−0.01022	0.00321	−0.01033	0.00215	−0.00924	0.00172
$\hat{\beta}_0$	−0.01466	0.06585	−0.00559	0.06154	0.00028	0.06532
$\hat{\beta}_1$	−0.00099	0.00062	−0.00232	0.00049	−0.00193	0.00049
$\hat{\alpha}$	−0.00213	0.0006	−0.00278	0.0007	−0.00189	0.0008
$\hat{\sigma}^2$	−0.01064	0.07493	−0.05502	0.06902	−0.02677	0.07052

Abbreviation: MSE, mean square error.

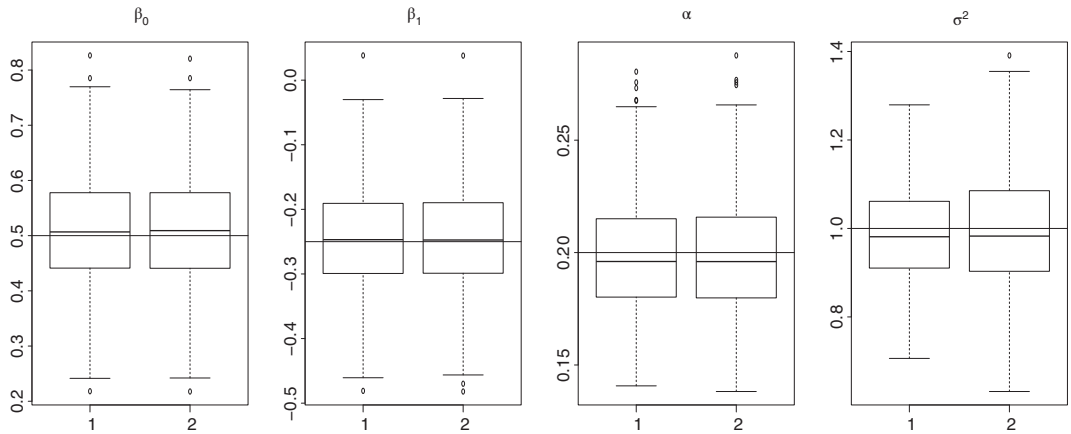

FIGURE 5 Boxplots of *wpl* estimates for $\beta_0 = 0.5$, $\beta_1 = -0.25$, $\alpha = 0.2$, $\sigma^2 = 1$ (from left to right) under Scenario 2 when estimating a t process with $\lambda = 1/\nu$, $\nu = 6$ when (1) ν is assumed known, (2) ν is assumed unknown and it is fixed to a positive integer through a two-step estimation

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, $\rho(\boldsymbol{h}) = \mathcal{GW}_{\alpha,0,4}(\boldsymbol{h})$ with $\alpha = 0.2$ is considered. In the *wpl* estimation, we consider a cut-off weight function with $d_{ij} = 0.05$ 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 freedom are supposed to be unknown. Recall that from Remark 1, ν must be fixed to a positive integer $\nu = 3, 4, \ldots$ in order to guarantee the existence of the t process. A brute force approach considers different *wpl* estimates using a fixed $\lambda = 1/\nu$, $\nu = 3, 4, \ldots$. 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, we estimate all the parameters including $0 < \lambda < 1/2$ maximizing the *wpl* function. This is possible since the bivariate t distribution is well defined for $0 < \lambda < 1/2$ (see Remark 4). 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 (1) known and

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)

	1/3				1/6				1/9			
	1		2		1		2		1		2	
λ	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\hat{\beta}_0$	−0.00024	0.01261	−0.00037	0.01271	0.00712	0.01139	0.00712	0.01135	0.00638	0.01153	0.00646	0.0115
$\hat{\beta}_1$	−0.00702	0.00874	−0.00721	0.00881	0.00419	0.00697	0.00419	0.00696	0.00089	0.00685	−0.00098	0.00694
$\hat{\alpha}$	−0.00233	0.00058	−0.00352	0.00058	−0.00115	0.00059	−0.00143	0.00061	−0.00425	0.00056	−0.0048	0.00058
$\hat{\sigma}^2$	−0.00971	0.01341	0.01768	0.01745	−0.01056	0.01142	−0.00461	0.01694	−0.01049	0.01148	−0.00015	0.01657

Abbreviation: MSE, mean square error.

fixed and (2) unknown and fixed using a two-step estimation and 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 Table 2, 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, that is, when the degrees of freedom are supposed to be known.

4.3 | Performance of the misspecified (pairwise) Gaussian likelihood estimation

Weighted pairwise likelihood estimation requires the evaluation of the bivariate distribution (10), that is, 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 & Saad, 2017):

$$F_4(a,b;c,c';w,z) = \sum_{k=0}^{\infty} \frac{(a)_k (b)_k z^k}{k!(c')_k} {}_2F_1(a+k, b+k; c; w), \quad |\sqrt{w}| + |\sqrt{z}| < 1, \tag{21}$$

truncating the series when the kth 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, Olver, & Porter, 2017). Evaluation of the F_4 function can be time consuming depending of the speed of convergence of (21) and, as a consequence, if the number of location sites is large, the computation of the *wpl* estimator can be computationally demanding.

An estimator that require smaller computational burden can be obtained by considering a misspecified *wpl*. Specifically, if in the estimation procedure we assume a Gaussian process with mean equal to $\mu(\boldsymbol{s})$, variance equal to $\sigma^2\nu/(\nu-2)$ and correlation $\rho_{Y^*_\nu}(\boldsymbol{h})$, then a Gaussian *wpl* only requires the computation of the Gaussian bivariate distribution and of the Gauss hypergeometric function in (6). Note that the misspecified Gaussian process matches mean, variance, and correlation function of the t process. To avoid identifiability problems, we need a reparametrization of the variance, that is, $\sigma^2_* := \sigma^2\nu/(\nu-2)$. Then, maximization of the Gaussian *wpl* function leads to the estimation of $\mu(\boldsymbol{s})$, σ^2_*, ν and the parameters of the underlying correlation model $\rho(\boldsymbol{h})$.

To investigate the performance of this kind of estimator, we consider 676 points on a regular spatial grid $A = [0,1]^2$, that is $(x_i, x_j)^T$ for $i,j = 1,\dots,21$ with $x_1 = 0, x_2 = 0.04, \dots x_{26} = 1$ and we simulate, using Cholesky decomposition, 500 realizations of a t process setting $\mu(\boldsymbol{s}) = \mu = 0$, $\sigma^2 = 1$, $\nu = 3, 6, 9$ and underlying correlation function $\rho(\boldsymbol{h}) = \mathcal{GW}_{\alpha,0,4}(\boldsymbol{h})$ with $\alpha = 0.2$. Then we estimate the parameters μ, σ^2_*, α (assuming ν known and fixed) with *wpl* using the bivariate t distribution (10) and with both misspecified Gaussian *wpl* and standard likelihood. In the *wpl* estimation, we consider a cutoff weight function with $d_{ij} = 0.05$.

TABLE 3 Bias and MSE associated with μ, α, and σ^2 for *wpl* with bivariate *t* distribution (WPL_T), standard misspecified Gaussian likelihood (L_G) and *wpl* with bivariate misspecified Gaussian distribution (WPL_G) when $\lambda = 1/\nu$, $\nu = 3, 6, 9$

		WPL_T		L_G		WPL_G	
λ	**Parameters**	**Bias**	**MSE**	**Bias**	**MSE**	**Bias**	**MSE**
	μ	0.0045	0.0088	0.0036	0.0154	0.0036	0.0158
$\lambda = 1/3$	α	−0.0016	0.0003	0.0084	0.0009	0.0129	0.0013
	σ^2	−0.0019	0.0102	−0.0577	0.0477	−0.0534	0.0478
	μ	−0.0057	0.0096	−0.0053	0.0106	−0.0052	0.0110
$\lambda = 1/6$	α	−0.0020	0.0003	−0.0007	0.0003	−0.0014	0.0003
	σ^2	−0.0105	0.0075	−0.0150	0.0095	−0.0168	0.0098
	μ	−0.0029	0.0091	−0.0038	0.0094	−0.0040	0.0096
$\lambda = 1/9$	α	−0.0019	0.0003	−0.0019	0.0003	−0.0022	0.0003
	σ^2	−0.0075	0.0081	−0.0089	0.0088	−0.0093	0.0088

Abbreviations: L_G, standard misspecified Gaussian likelihood; MSE, mean square error; WPL_G, weighted pairwise likelihood with bivariate Gaussian distribution; WPL_T, weighted pairwise likelihood with bivariate *t* distribution.

Table 3 shows the bias and MSE associated with μ, α, and σ^2 for the three methods of estimation. Note that, for comparison, the results of the variance parameter are reported in terms of the original parametrization. It is apparent that *wpl* with bivariate *t* distribution shows the best performance. In particular when $\lambda = 1/3$, the gains in terms of efficiency are considerable. However, when increasing the degrees of freedom, the gains tends to decrease and when $\nu = 9$ the efficiencies of the three estimators are quite similar (see boxplots in Figure 6).

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.

A practical and less efficient solution can be obtained using the optimal linear prediction. Assuming known mean, correlation, and the degrees of freedom of the *t* process, the predictor at an unknown location $\boldsymbol{s}_0$ is given by:

$$\widehat{y(\boldsymbol{s}_0)} = \mu(\boldsymbol{s}_0) + \boldsymbol{c}_\nu^T R_\nu^{-1}(\boldsymbol{Y} - \boldsymbol{\mu}), \tag{22}$$

where $\boldsymbol{\mu} = (\mu(\boldsymbol{s}_1), \dots, \mu(\boldsymbol{s}_n))^T$, $\boldsymbol{c}_\nu = [\rho_{Y_\nu^*}(\boldsymbol{s}_0 - \boldsymbol{s}_j)]_{i=1}^n$, and $R_\nu = [\rho_{Y_\nu^*}(\boldsymbol{s}_i - \boldsymbol{s}_j)]_{i,j=1}^n$, and the associated variance is given by:

$$\widehat{\text{Var}(y(\boldsymbol{s}_0))} = \sigma_*^2(1 - \boldsymbol{c}_\nu^T R_\nu^{-1}\boldsymbol{c}_\nu). \tag{23}$$

As an Associate Editor pointed out, this is equivalent to perform optimal Gaussian prediction with covariance function equal to $\sigma_*^2\rho_{Y_\nu^*}(\boldsymbol{h})$. Similarly, using (18), the optimal linear predictor of the skew-t process can be obtained.

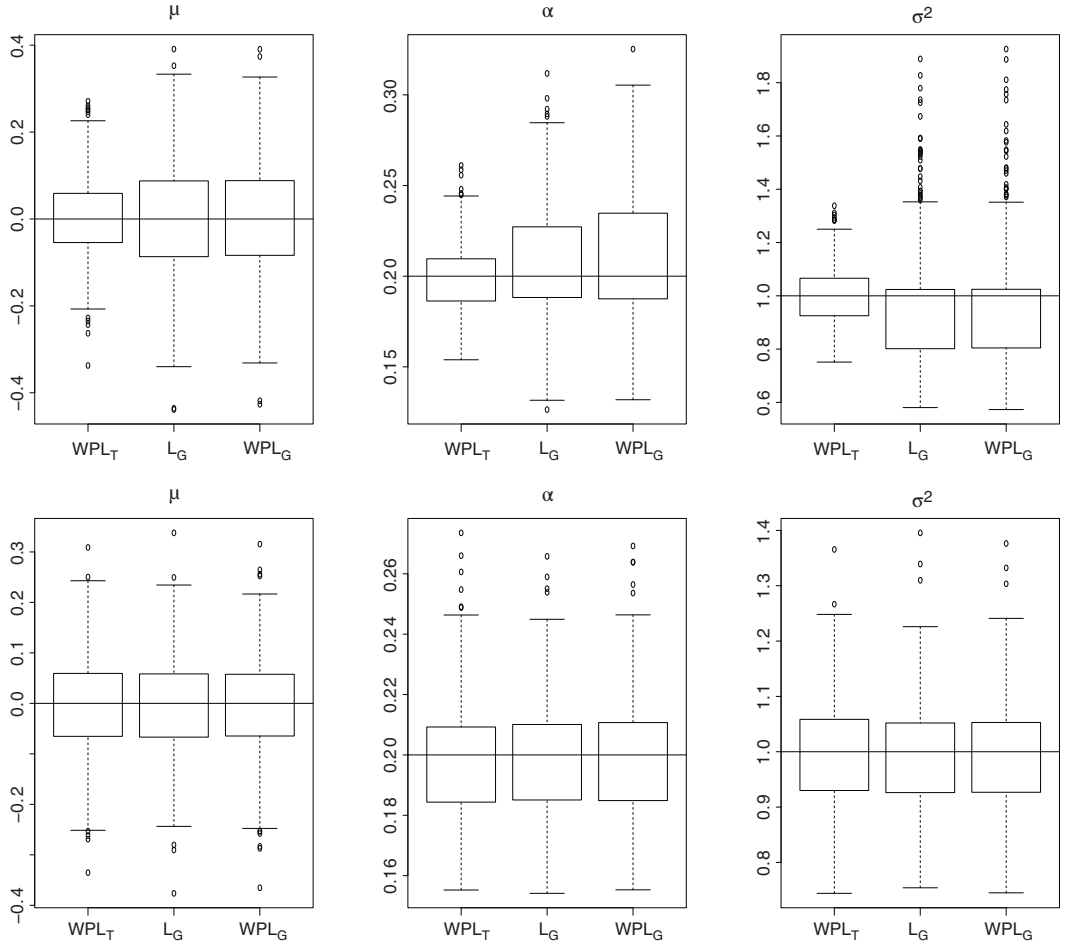

FIGURE 6 Upper part: boxplots of μ, α, σ^2 for *wpl* using t bivariate distribution (WPL_T), standard misspecified Gaussian likelihood (L_G) and *wpl* using Gaussian misspecified bivariate distribution (WPL_G) when $\nu = 3$. Bottom part: the same boxplots with $\nu = 9$.

We investigate the performance of (22) when compared with the Gaussian optimal predictor, assuming $\rho(\boldsymbol{h}) = \mathcal{GW}_{0.2,\psi,4}(\boldsymbol{h})$, $\psi = 0, 1, 2$ as underlying correlation function, using crossvalidation. With this goal in mind, we simulate 1000 realizations from a t process Y_ν^* with $\nu = 3, 7, 11$ and a Gaussian process under the settings of Section 4.3 and for each realization, we consider 80% of the data for prediction and leave 20% as validation dataset.

For each model and for each realization, we compute the root-mean-square errors (RMSEs), that is:

$$\text{RMSE}_l = \left(\frac{1}{n_l} \sum_{i=1}^{n_l} \left(\widehat{y(\boldsymbol{s}_{i,l})} - y(\boldsymbol{s}_{i,l}) \right) \right)^2 \Bigg)^{\frac{1}{2}},$$

where $y(\boldsymbol{s}_{i,l})$, $i = 1, \ldots, n_l$ are the observation in the lth validation set and n_l is the associated cardinality ($n_l = 135$ in our example).

TABLE 4 Relative mean root-mean-square errors prediction efficiency over 500 runs for a t process with $\nu = 3, 7, 11$ and a Gaussian process when predicting using $\rho_{Y_\nu^*}(\boldsymbol{h})$ for $\nu = 3, 7, 11$ and $\rho(\boldsymbol{h})$. The underlying correlation model is $\rho(\boldsymbol{h}) = \mathcal{GW}_{0.2,\psi,4}(\boldsymbol{h})$ with $\psi = 0, 1, 2$.

$\mathcal{GW}_{0.2,\psi,4}(\boldsymbol{h})$	Simulation from: / Prediction with	Y_3^*	Y_7^*	Y_{11}^*	Gaussian
$\psi = 0$	$\rho_{Y_3^*}(\boldsymbol{h})$	1	0.9896	0.9870	0.9838
	$\rho_{Y_7^*}(\boldsymbol{h})$	0.9904	1	0.9998	0.9987
	$\rho_{Y_{11}^*}(\boldsymbol{h})$	0.9877	0.9998	1	0.9996
	$\rho(\boldsymbol{h})$	0.9836	0.9988	0.9996	1
$\psi = 1$	$\rho_{Y_3^*}(\boldsymbol{h})$	1	0.9399	0.9180	0.8789
	$\rho_{Y_7^*}(\boldsymbol{h})$	0.9503	1	0.9984	0.9895
	$\rho_{Y_{11}^*}(\boldsymbol{h})$	0.9327	0.9984	1	0.9966
	$\rho(\boldsymbol{h})$	0.9095	0.9947	0.9969	1
$\psi = 2$	$\rho_{Y_3^*}(\boldsymbol{h})$	1	0.8827	0.8263	0.7109
	$\rho_{Y_7^*}(\boldsymbol{h})$	0.9223	1	0.9936	0.9511
	$\rho_{Y_{11}^*}(\boldsymbol{h})$	0.8834	0.9947	1	0.9817
	$\rho(\boldsymbol{h})$	0.8169	0.9643	0.9848	1

Abbreviation: $\mathcal{GW}_{0.2,\psi,4}(\boldsymbol{h})$, Generalized Wendland correlation function.

Finally, we compute the empirical mean of the 500 RMSEs when the prediction is performed with the optimal linear predictor (22), that is, using $\rho_{Y_\nu^*}(\boldsymbol{h})$ with $\nu = 3, 7, 11$ and the optimal Gaussian predictor with $\rho(\boldsymbol{h})$. Note that, from Theorem 2(e), the prediction using $\rho(\boldsymbol{h})$ can be viewed as the prediction using $\rho_{Y_\nu^*}(\boldsymbol{h})$ when $\nu \to \infty$.

In Table 4, we report the simulation results in terms of relative efficiency , that is, for a given process and a given $\psi = 0, 1, 2$, the ratio between the mean RMSE of the best predictor and the mean RMSE associated with a competitive predictor. This implies that relative efficiency prediction is lower than 1 and it is equal to 1 in the best case. From Table 4, it can be appreciated that under the t process Y_ν^*, the prediction with $\rho_{Y_\nu^*}(\boldsymbol{h})$, for $\nu = 3, 7, 11$, performs overall better than the optimal Gaussian prediction using $\rho(\boldsymbol{h})$. As expected, the gain is more apparent when decreasing the degrees of freedom and increasing ψ. For instance, if $\nu = 3$ and $\psi = 2$, the loss of efficiency predicting with $\rho(\boldsymbol{h})$ is 19% approximatively. It can also be noted that if Y_7^*, Y_{11}^*, or Gaussian are one or two mean squared differentiable ($\psi = 1, 2$), then the prediction using $\rho_{Y_3^*}(\boldsymbol{h})$ can be very inefficient. This is not surprising since from Theorem 2(c), Y_3^* is not mean square differentiable. Resuming this numerical experiment study suggests that when predicting data exhibiting heavy tails, the use of the correlation function $\rho_{Y_\nu^*}(\boldsymbol{h})$ should be preferred to the use of $\rho(\boldsymbol{h})$.

5 | APPLICATION TO MAXIMUM TEMPERATURE DATA

In this section, we apply the proposed t process to a dataset of maximum temperature data observed in Australia. Specifically, we consider a subset of a global dataset 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

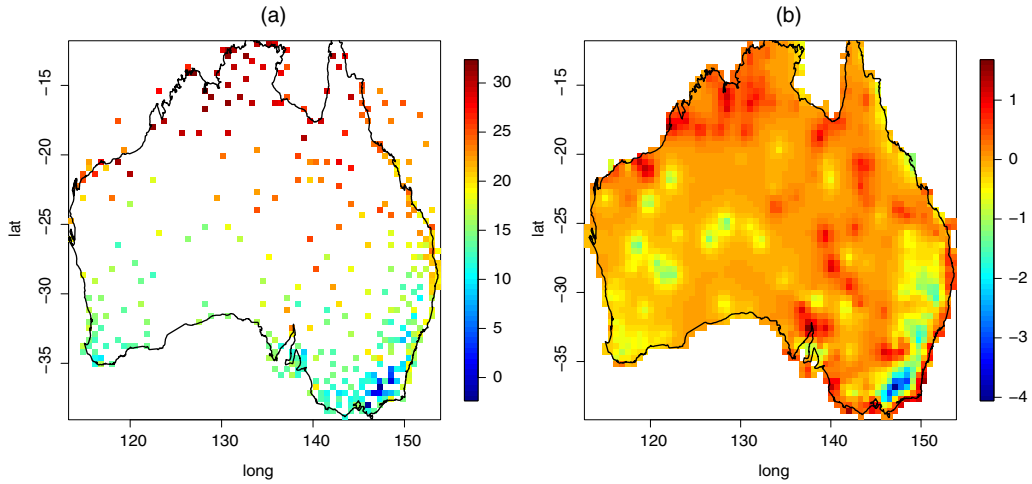

FIGURE 7 From left to right: (a) spatial locations of maximum temperature in Australia in July 2011 and (b) prediction of residuals of the estimated t process [Color figure can be viewed at wileyonlinelibrary.com]

we consider is depicted in Figure 7a and consists of the maximum temperature observed on July 5 in 446 location sites, $y(\boldsymbol{s}_i)$, $i = 1, \dots, 446$, 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 approximation $\mathbb{S}^2 = \{\boldsymbol{s} \in \mathbb{R}^3 : ||\boldsymbol{s}|| = 6371\}$. The first process we use to model this dataset is a t process:

$$Y_\nu(\boldsymbol{s}) = \beta_0 + \beta_1 X(\boldsymbol{s}) + \sigma Y^*_\nu(\boldsymbol{s}), \quad \boldsymbol{s} \in \mathbb{S}^2, \tag{24}$$

where Y^*_ν is a standard t process. Here, $X(\boldsymbol{s})$ 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). As a comparison, we also consider a Gaussian process:

$$G(\boldsymbol{s}) = \beta_0 + \beta_1 X(\boldsymbol{s}) + \sigma G^*(\boldsymbol{s}), \quad \boldsymbol{s} \in \mathbb{S}^2, \tag{25}$$

where G^* is a standard Gaussian process. We assume that the underlying geodesically isotropic correlation function (Gneiting, 2013; Porcu, Bevilacqua, & Genton, 2016) is of the Matérn and generalized Wendland type. A preliminary estimation of the t and Gaussian processes, including the smoothness parameters, highlights a multimodality of the (pairwise) likelihood surface, for both correlation models and a not mean-square differentiability of the process. For this reason, we fix the smoothness parameters and we consider the underlying correlation models $\mathcal{M}_{\alpha,0.5}(d_{GC}) = e^{-d_{GC}/\alpha}$ and $\mathcal{GW}_{\alpha,0,5}(d_{GC}) = (1 - d_{GC}/\alpha)^5_+$ where, given two spherical points $\boldsymbol{s}_i = (\text{lon}_i, \text{lat}_i)$ and $\boldsymbol{s}_j = (\text{lon}_j, \text{lat}_j)$, $d_{GC}(\boldsymbol{s}_i, \boldsymbol{s}_j) = 6371\theta_{ij}$, 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 $a_i = (\text{lat}_i)\pi/180$, $a_j = (\text{lat}_j)\pi/180$, $b_i = (\text{lon}_i)\pi/180$, $b_j = (\text{lon}_j)\pi/180$.

For the t process, the parameters were estimated using *wpl* using the bivariate t distribution with the two-step method described in Section 4 and using the weight function (20) with $d_{ij} = 150$ km. It turns out that the estimation at the first step leads to fix $\nu = 4$ in the second step, irrespective of the correlation model. We also consider a Gaussian misspecified standard

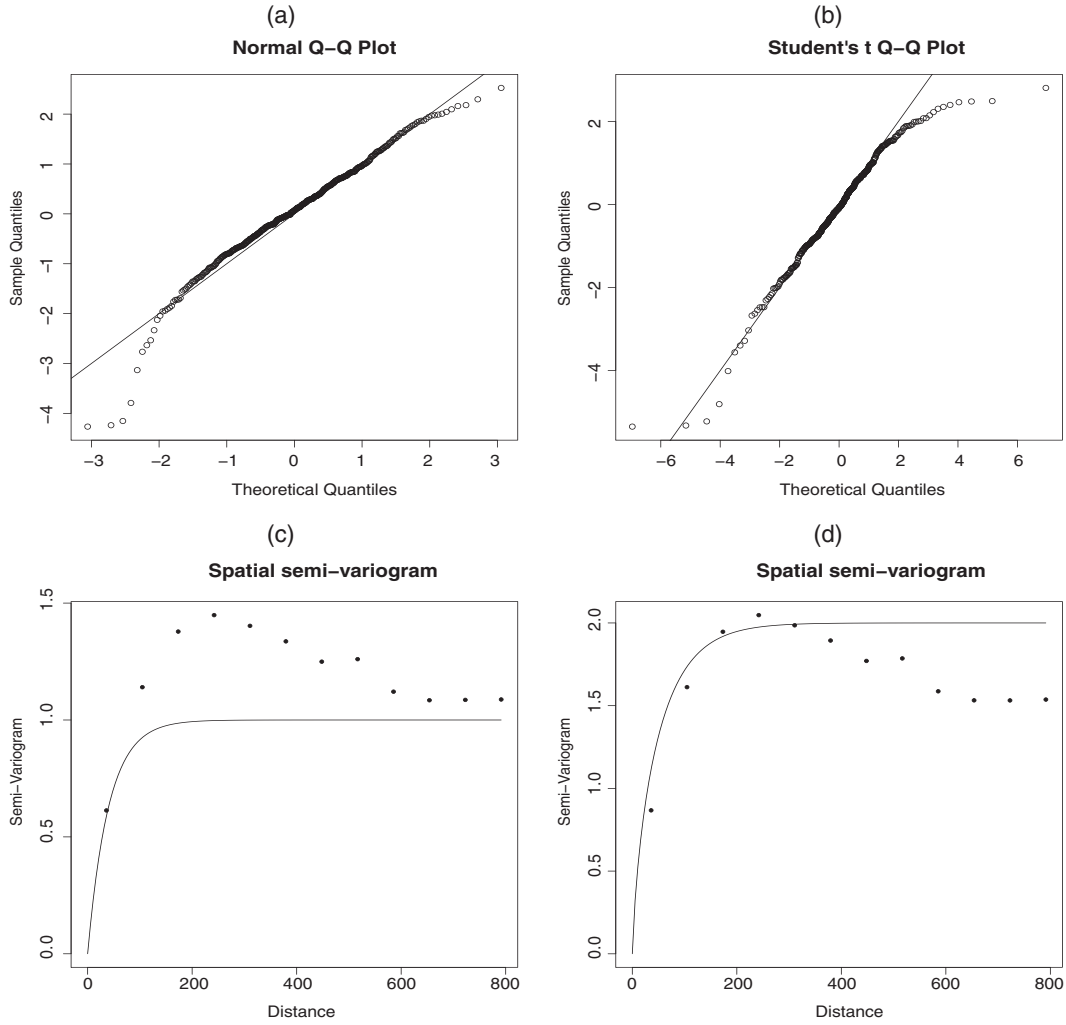

FIGURE 8 Upper part: Q-Q plot of the residuals versus the estimated quantiles in the Gaussian and t models (a and b, respectively). Bottom part: empirical semivariogram (dotted points) of the residuals versus the estimated semivariogram (solid line) in the Gaussian and t models (c and d, respectively). Distances are expressed in kilometers

likelihood and *wpl* estimation as described in Section 4.3, that is, we estimate using the Gaussian process (25) with the t correlation model 6 fixing $\nu = 4$.

In addition, we compute the standard error estimation, PLIC and BLIC values through parametric bootstrap estimation of the inverse of the Godambe information matrix (Bai et al., 2014). For standard maximum likelihood, we compute the standard errors as the square root of diagonal elements of the inverse of Fisher Information matrix (Mardia & Marshall, 1984). The results are summarized in Table 5. Note that the regression parameters estimates are quite similar for the t and Gaussian processes, irrespective of the correlation model. Furthermore, we note that the standard Gaussian process assigns lower spatial dependence and stronger variance compared with the other cases. Finally, for each correlation model, both the (pairwise) likelihood information criterion PLIC and BLIC select for the pairwise case the t model and for the standard case the Gaussian model with t correlation function.

TABLE 5 Estimates for the t process using *wpl* with bivariate t distribution and misspecified (full and weighted pairwise) Gaussian likelihood and for the Gaussian process using standard likelihood with associated standard error (in parenthesis) and PLIC and BLIC values, when estimating the Australian maximum temperature dataset using two correlation models: $\mathcal{M}_{\alpha,0.5}$ and $\mathcal{GW}_{\alpha,0,5}$

		Matérn								
	Method	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\alpha}$	$\hat{\sigma}^2$	**PLIC**	**BLIC**	**RMSE**	**MAE**	**CRPS**
t	Pairwise	6.652	0.994	58.582	7.589	23906	24689	2.775	2.138	1.807
		(1.192)	(0.107)	(13.492)	(1.451)					
Missp-Gaussian	Pairwise	5.656	1.064	72.435	7.062	24680	26172	2.743	2.112	1.809
		(1.514)	(0.130)	(18.303)	(2.024)					
Missp-Gaussian	Standard	5.673	1.050	63.437	5.385	2228	2244	2.755	2.119	1.813
		(0.660)	(0.045)	(10.41)	(0.432)					
Gaussian	Standard	5.508	1.060	40.484	10.762	2240	2257	2.812	2.155	1.815
		(0.558)	(0.039)	(5.346)	(0.813)					
		Wendland								
	Method	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\alpha}$	$\hat{\sigma}^2$	**PLIC**	**BLIC**	**RMSE**	**MAE**	**CRPS**
t	Pairwise	6.6375	0.995	331.70	7.622	23890	24624	2.796	2.155	1.807
		(1.134)	(0.103)	(69.03)	(1.482)					
Missp-Gaussian	Pairwise	5.647	1.065	404.81	7.080	24657	26100	2.760	2.127	1.810
		(1.446)	(0.124)	(92.99)	(2.032)					
Missp-Gaussian	Standard	5.579	1.056	349.90	5.485	2232	2249	2.774	2.134	1.815
		(0.618)	(0.042)	(52.030)	(0.434)					
Gaussian	Standard	5.400	1.066	225.89	11.081	2250	2266	2.846	2.177	1.820
		(0.530)	(0.038)	(27.72)	(0.835)					

Note: Last three columns: associated empirical mean of RMSEs, MAEs, and CRPSs.
Abbreviations: BLIC, bayesian likelihood information criterion; CRPS, continuous ranked probability score; MAE, mean absolute error; PLIC, pseudolikelihood information criterion; RMSE, root-mean-square error.

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

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

can be viewed as a realization of the process Y_4^*. Similarly we can compute the Gaussian residuals. For the t process, we use the *wpl* estimates obtained with the bivariate t distribution. 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 8a, Q-Q plot of the residuals of the Gaussian and t processes (from left to right) is depicted for the Matérn case. 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, the graphical comparison between the empirical and fitted semivariogram of the residuals (bottom part of Figure 8) highlights an apparent better fitting of the t model.

We want to further evaluate the predictive performances of Gaussian and t processes using RMSE and mean absolute error (MAE) as in Section 4.4. Specifically, we use the following resampling approach: we randomly choose 80% of the spatial locations and we use the remaining 20% as data for the predictions. Then, we use the estimates in Table 5 in order to compute RMSE and MAE. We repeat the approach for 2000 times and record all RMSEs and MAEs. Specifically, for each jth left-out sample $(y_j^L(\boldsymbol{s}_1), \dots, \dots, y_j^L(\boldsymbol{s}_K))$, we compute

$$\overline{\text{RMSE}}_j = \left[\frac{1}{K}\sum_{i=1}^{K}\left(y_j^L(\boldsymbol{s}_i) - \hat{Y}_j^L(\boldsymbol{s}_i)\right)^2\right]^{\frac{1}{2}}$$

and

$$\overline{\text{MAE}}_j = \frac{1}{K}\sum_{i=1}^{K}|y_j^L(\boldsymbol{s}_i) - \hat{Y}_j^L(\boldsymbol{s}_i)|,$$

where $\hat{Y}_j^L(\boldsymbol{s}_i)$ is the optimal linear predictor for the t process. Similarly, we can compute the optimal predictor for the Gaussian process. Finally, we compute the overall mean for both Gaussian and t processes and for both correlation models, that is, $\text{RMSE} = \sum_{j=1}^{2000}\overline{\text{RMSE}}_j/2000$ and $\text{MAE} = \sum_{j=1}^{2000}\overline{\text{MAE}}_j/2000$.

In addition, to evaluate the marginal predictive distribution performance, we also consider, for each sample, the continuous ranked probability score (CRPS) (Gneiting & Raftery, 2007). For a single predictive cdf F and a verifying observation y, it is defined as:

$$\text{CRPS}(F, y) = \int_{-\infty}^{\infty}(F(t) - \mathbf{1}_{[y,\infty]}(t))^2\, dt.$$

Specifically, for each jth left-out sample, we consider the averaged CRPS for the Gaussian and t distributions as:

$$\overline{\text{CRPS}}_j = \frac{1}{K}\sum_{i=1}^{K}\text{CRPS}(F, y_j^L(\boldsymbol{s}_i)) \quad F = F_G, F_{Y_4}, \tag{26}$$

for $j = 1, \dots, 2000$. In particular in the Gaussian case

$$\begin{aligned}\text{CRPS}(F_G, y_j^L(\boldsymbol{s}_i)) = \sigma\left(\frac{y_j^L(\boldsymbol{s}_i) - \mu(\boldsymbol{s})}{\sigma}\right)\left[2F_{G^*(\boldsymbol{s})}\left(\frac{y_j^L(\boldsymbol{s}_i) - \mu(\boldsymbol{s})}{\sigma}\right) - 1\right] \\ + 2\sigma f_{G^*(\boldsymbol{s})}\left(\frac{y_j^L(\boldsymbol{s}_i) - \mu(\boldsymbol{s})}{\sigma}\right) - \frac{\sigma}{\sqrt{\pi}},\end{aligned} \tag{27}$$

and in the t case with four degrees of freedom:

$$\begin{aligned}\text{CRPS}(F_{Y_4}, y_j^L(\boldsymbol{s}_i)) = \sigma\left(\frac{y_j^L(\boldsymbol{s}_i) - \mu(\boldsymbol{s})}{\sigma}\right)\left[2F_{Y_4^*(\boldsymbol{s})}\left(\frac{y_j^L(\boldsymbol{s}_i) - \mu(\boldsymbol{s})}{\sigma}\right) - 1\right] \\ + 2\left[\frac{4\sigma^2 + (y_j^L(\boldsymbol{s}_i) - \mu(\boldsymbol{s}))^2}{3\sigma}\right] f_{Y_4^*(\boldsymbol{s})}\left(\frac{y_j^L(\boldsymbol{s}_i) - \mu(\boldsymbol{s})}{\sigma}\right) - \frac{4\sigma B\left(\frac{1}{2}, \frac{7}{2}\right)}{3B^2\left(\frac{1}{2}, 2\right)},\end{aligned} \tag{28}$$

where $\mu(\boldsymbol{s}) = \beta_0 + \beta_1 X(\boldsymbol{s})$. We compute the CRPS in (27) and (28) plugging-in the estimates of the pairwise and standard (misspecified) likelihood estimation methods of β_0, β_1, and σ using the R package `scoringRules` (Jordan, Krueger, & Lerch, 2019). Finally, we compute the overall mean for both Gaussian and t processes and for both correlation models, that is, $\text{CRPS} = \sum_{j=1}^{2000} \overline{\text{CRPS}}_j / 2000$.

Table 5 reports the estimated RMSE, MAE, and CRPS. As a general remark, the t process outperforms the Gaussian process for the three measures of prediction performance irrespective of the method of estimation and for both correlation models. We point out that RMSE and MAE are computed using the optimal predictor in the Gaussian case and the linear optimal in the t case. However, the RMSE and MAE results highlight a better performance for the t process. In addition, a better RMSE and MAE for the Matérn correlation model with respect to the Wendland is apparent, irrespective of the type of process. The proposed t process also leads to a clear better performance of the CRPS with respect to the Gaussian case. In this specific example, the use of the misspecified Gaussian *wpl* estimates leads to the best results in terms of RMSE and MAE. On the other hand, the best CRPS results are achieved by using the *wpl* estimates using the proposed t bivariate distribution.

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 7b, we plot a high-resolution map of the predicted residuals using the t process with underlying Matérn correlation model estimated with *wpl*.

6 | CONCLUDING REMARKS

We have introduced a new stochastic process with t marginal distributions for regression and dependence analysis when addressing spatial 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. In addition, the proposed t process partially inherits some 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 an inferential approach based on *wpl*, using the bivariate t distribution given in Theorem 3, could be an effective solution for estimating the unknown parameters. An alternative less efficient solution that requires smaller computational burden can be obtained by considering a misspecified Gaussian *wpl* using the correlation function of the t process. In the prediction case, our numerical experiments show that the optimal linear predictor of the t process performs better than the optimal Gaussian predictor when working with spatial data with heavy tails.

Another possible drawback concerns the restriction of the degrees of freedom of the t process to $\nu = 3, 4, \ldots$. under noninfinite 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 article since the bivariate distribution in this case is quite complicated. In principle, after a suitable parametrization, Gaussian misspecified *wpl* can be performed using (18) to estimate the parameters of the skew-t process. In this case, an additional issue is the behavior of the information matrix when $\eta = 0$ (Arellano-Valle & Azzalini, 2013). 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 Rosco, Jones, and Pewsey (2011) or under the two-piece distribution framework (Arellano-Valle, Gómez, & Quintana, 2005) and this will be studied in future work.

ACKNOWLEDGEMENTS

This work is partially supported by FONDECYT grant 1200068, Chile and by Millennium Science Initiative of the Ministry of Economy, Development, and Tourism, grant "Millenium Nucleus Center for the Discovery of Structures in Complex Data" and by regional MATH-AmSud program, grant number 20-MATH-03 for Moreno Bevilacqua and by Proyecto de Iniciación Interno DIUBB 173408 2/I de la Universidad del Bío-Bío for Christian Caamaño.

ORCID

Moreno Bevilacqua https://orcid.org/0000-0001-8384-840X

REFERENCES

Alegría, A., Caro, S., Bevilacqua, M., Porcu, E., & Clarke, J. (2017). Estimating covariance functions of multivariate skew-Gaussian random fields on the sphere. *Spatial Statistics*, *22*, 388–402.

Allcroft, D., & 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., & 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., & Bolfarine, H. (1995). On some characterizations of the t distribution. *Statistics and Probability Letters*, *25*, 79–85.

Arellano-Valle, R. B., Castro, L., & 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., & 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., & Capitanio, A. (2014). *The skew-normal and related families*. New York, NY: United States of America by Cambridge University Press.

Azzalini, A., & Dalla-Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, *83*(4), 715–726.

Bai, Y., Kang, J., & 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: The Indian Journal of Statistics, Series A*, *51*, 73–78.

Berg, C., Mateu, J., & Porcu, E. (2008). The Dagum family of isotropic correlation functions. *Bernoulli*, *14*(4), 1134–1149.

Bevilacqua, M., Caamaño, C., & Gaetan, C. (2018). On modelling positive continuous data with spatio-temporal dependence. arXiv preprint arXiv:1808.03829.

Bevilacqua, M., Faouzi, T., Furrer, R., & Porcu, E. (2019). Estimation and prediction using generalized Wendland functions under fixed domain asymptotics. *The Annals of Statistics*, *47*, 828–856.

Bevilacqua, M., & 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., & 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., & Morales-Oñate, V. (2018). GeoModels: A package for geostatistical gaussian and non gaussian data analysis. R package version 1.0.3-4. https://vmoprojs.github.io/GeoModels-page/.

Brychkov, Y. A., & Saad, N. (2017). On some formulas for the Appell function $F_4(a, b, b'; c, c'; w; z)$. *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., & 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., & Galea, M. (2015). Influence diagnostics in elliptical spatial linear models. *Test*, *24*, 322–340.

DiCiccio, T. J., & Monti, A. C. (2011). Inferential aspects of the skew tdistribution. *Quaderni di Statistica*, *13*, 1–21.

Eisenbaum, N., & Kaspi, H. (2006). A characterization of the infinitely divisible squared gaussian processes. *The Annals of Probability*, *34*(2), 728–742.

Ferrari, S. L. P., & 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., & Seneta, E. (2006). Stationary-increment student and variancegamma processes. *Journal of Applied Probability*, *43*(2), 441–453.

Fonseca, T. C. O., Ferreira, M. A. R., & Migon, H. S. (2008). Objective Bayesian analysis for the student-t regression model. *Biometrika*, *95*(2), 325–333.

Furrer, R., Genton, M. G., & Nychka, D. (2013). Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*, *15*, 502–523.

Gao, X., & 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., & 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., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*, 359–378.

Gneiting, T., & Schlather, M. (2004). Stochastic models that separate fractal dimension and the Hurst effect. *SIAM Review*, *46*, 269–282.

Gouriéroux, C., Monfort, A., & Renault, E. (2017). Consistent pseudo-maximum likelihood estimators. *Annals of Economics and Statistics*, (125/126), 187–218. https://www.jstor.org/stable/10.15609/annaeconstat2009.125-126.0187#metadata_info_tab_contents.

Gradshteyn, I., & Ryzhik, I. (2007). *Table of integrals, series, and products* (7th ed.). New York, NY: 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: The Indian Journal of Statistics, Series A*, *32*, 393–404.

Heagerty, P., & Lele, S. (1998). A composite likelihood approach to binary spatial data. *Journal of the American Statistical Association*, *93*, 1099–1111.

Heyde, C. C., & Leonenko, N. N. (2005). Student processes. *Adv. In Appl. Probab.*, *37*(2), 342–365.

Joe, H. (2014). *Dependence modeling with copulas*. Boca Raton, FL: Chapman and Hall/CRC Press.

Joe, H., & Lee, Y. (2009). On weighting of bivariate margins in pairwise likelihood. *Journal of Multivariate Analysis*, *100*, 670–685.

Jordan, A., Krueger, F., & Lerch, S. (2019). Evaluating probabilistic forecasts with scoringRules. *Journal of Statistical Software 90*(12), 1–37.

Kazianka, H., & 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., Percec Tadic, M., & 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., & Rao, M. M. (1961). Remarks on a multivariate gamma distribution. *The American Mathematical Monthly*, *68*(4), 342–346.

Krishnamoorthy, A. S., & Parthasarathy, M. (1951). A multivariate gamma-type distribution. *The Annals of Mathematical Statistics*, *22*(4), 549–557.

Lange, K. L., Little, R. J. A., & Taylor, J. M. G. (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association*, *84*(408), 881–896.

Lim, S., & 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. (2010). 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. *Electronic Communications in Probability*, *19*, 10 pp.

Mardia, K. V., & Marshall, J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika*, *71*, 135–146.

Masarotto, G., & Varin, C. (2012). Gaussian copula marginal regression. *Electronic Journal of Statistics*, *6*, 1517–1549.

Matèrn, B. (1986). *Spatial variation: Stochastic models and their applications to some problems in forest surveys and other sampling investigations* (2nd ed.). Heidelberg, Germany: Springer.

Miller, K. S. (1968). Some multivariate t-distributions. *The Annals of Mathematical Statistics*, *39*(5), 1605–1609.

Palacios, M. B., & Steel, M. F. J. (2006). Non-Gaussian Bayesian geostatistical modeling. *Journal of the American Statistical Association*, *101*, 604–618.

Pearson, J. W., Olver, S., & 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., & 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., & Omre, H. (2006). T-distributed random fields: A parametric model for heavy-tailed well-log data. *Mathematical Geology*, *38*, 821–849.

Rosco, J. F., Jones., M., & Pewsey, A. (2011). Skew t distributions via the sinh-arcsinh transformation. *Test*, *30*, 630–652.

Royen, T. (2004). *Multivariate gamma distributions II*. In *Encyclopedia of statistical sciences* (pp. 419–425). New York, NY: John Wiley & Sons.

Sahu, K. S., Dey, D. K., & 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, NY: Springer-Verlag.

Varin, C., Reid, N., & Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, *21*, 5–42.

Varin, C., & 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 Journal of Mathematics*, *26*, 125–149.

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

Xua, G., & Genton, M. G. (2017). Tukey g-and-h random fields. *Journal of the American Statistical Association*, *112*, 1236–1249.

Yeo, G., & Milne, R. (1991). On characterizations of beta and gamma distributions. *Statistics and Probability Letters*, *11*(3), 239–242.

Zareifard, H., & Khaledi, M. J. (2013). Non-Gaussian modeling of spatial data using scale mixing of a unified skew Gaussian process. *Journal of Multivariate Analysis*, *114*, 16–28.

Zareifard, H., Khaledi, M. J., Rivaz, F., & Vahidi-Asl, M. Q. (2018). Modeling skewed spatial data using a convolution of Gaussian and log-Gaussian processes. *Bayesian Analysis*, *13*(2), 531–557.

Zhang, H., & El-Shaarawi, A. (2010). On spatial skew-Gaussian processes and applications. *Environmetrics*, *21*(1), 33–47.

How to cite this article: Bevilacqua M, Caamaño-Carrillo C, Arellano-Valle RB, Morales-Oñate V. Non-Gaussian geostatistical modeling using (skew) t processes. *Scand J Statist*. 2020;1–34. https://doi.org/10.1111/sjos.12447

APPENDIX A

A.1 Proof of Theorem 1

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

$$\rho_{Y_\nu^*}(\boldsymbol{h}) = \left(\frac{\nu-2}{\nu}\right)\left(\mathrm{E}(R_\nu(\boldsymbol{s}_i)R_\nu(\boldsymbol{s}_j))\rho(\boldsymbol{h})\right). \tag{A1}$$

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

$$f_{\boldsymbol{W}_{\nu;ij}}(w_i,w_j) = \frac{2^{-\nu}\nu^\nu(w_iw_j)^{\nu/2-1}e^{-\frac{\nu(w_i+w_j)}{2(1-\rho^2(\boldsymbol{h}))}}}{\Gamma\left(\frac{\nu}{2}\right)(1-\rho^2(\boldsymbol{h}))^{\nu/2}}\left(\frac{\nu\sqrt{\rho^2(\boldsymbol{h})w_iw_j}}{2(1-\rho^2(\boldsymbol{h}))}\right)^{1-\nu/2} I_{\nu/2-1}\left(\frac{\nu\sqrt{\rho^2(\boldsymbol{h})w_iw_j}}{(1-\rho^2(\boldsymbol{h}))}\right), \tag{A2}$$

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

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

$$f_{\boldsymbol{R}_{\nu;ij}}(\boldsymbol{r}_{ij}) = \frac{2^{-\nu+2}\nu^\nu(r_ir_j)^{-\nu-1}e^{-\frac{\nu}{2(1-\rho^2(\boldsymbol{h}))}\left(\frac{1}{r_i^2}+\frac{1}{r_j^2}\right)}}{\Gamma\left(\frac{\nu}{2}\right)(1-\rho^2(\boldsymbol{h}))^{\nu/2}}\left(\frac{\nu\rho(\boldsymbol{h})}{2(1-\rho^2(\boldsymbol{h}))r_ir_j}\right)^{1-\frac{\nu}{2}} I_{\frac{\nu}{2}-1}\left(\frac{\nu\rho(\boldsymbol{h})}{(1-\rho^2(\boldsymbol{h}))r_ir_j}\right) \tag{A3}$$

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 (A3) we have

$$\mathrm{E}(R^a(\boldsymbol{s}_i)R^b(\boldsymbol{s}_j)) = \frac{2^{-\nu+2}\nu^\nu}{\Gamma^2\left(\frac{\nu}{2}\right)(1-\rho^2(\boldsymbol{h}))^{\nu/2}}\int_{\mathbb{R}_+^2} r_i^{-\nu+a-1}r_j^{-\nu+b-1}e^{-\frac{\nu}{2(1-\rho^2(\boldsymbol{h}))r_i^2}}e^{-\frac{\nu}{2(1-\rho^2(\boldsymbol{h}))r_j^2}}$$

$$
\begin{aligned}
&\times {}_0F_1\left(\frac{\nu}{2}; \frac{\nu^2\rho^2(\boldsymbol{h})}{4(1-\rho^2(\boldsymbol{h}))^2 r_i^2 r_j^2}\right) d\boldsymbol{r}_{ij} \\
&= \frac{2^{-\nu+2}\nu^\nu}{\Gamma^2\left(\frac{\nu}{2}\right)(1-\rho^2(\boldsymbol{h}))^{\nu/2}} \sum_{k=0}^{\infty} \int_{\mathbb{R}_+^2} r_i^{-\nu+a-2k-1} r_j^{-\nu+b-2k-1} e^{-\frac{\nu}{2(1-\rho^2(\boldsymbol{h}))r_i^2}} e^{-\frac{\nu}{2(1-\rho^2(\boldsymbol{h}))r_j^2}} \\
&\quad \times \frac{1}{k!\left(\frac{\nu}{2}\right)_k} \left(\frac{\rho^2(\boldsymbol{h})\nu^2}{4(1-\rho^2(\boldsymbol{h}))}\right)^k d\boldsymbol{r}_{ij} \\
&= \frac{2^{-\nu+2}\nu^\nu}{\Gamma^2\left(\frac{\nu}{2}\right)(1-\rho^2(\boldsymbol{h}))^{\nu/2}} \sum_{k=0}^{\infty} \frac{I(k)}{k!\left(\frac{\nu}{2}\right)_k} \left(\frac{\rho^2(\boldsymbol{h})\nu^2}{4(1-\rho^2(\boldsymbol{h}))}\right)^k, \qquad \text{(A4)}
\end{aligned}
$$

where, using Fubini's theorem

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

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

$$
I(k) = \Gamma\left(\frac{\nu-a}{2}+k\right)\Gamma\left(\frac{\nu-b}{2}+k\right) 2^{\frac{\nu-a}{2}+k-1} 2^{\frac{\nu-b}{2}+k-1} \left(\frac{(1-\rho^2(\boldsymbol{h}))}{\nu}\right)^{\frac{\nu-a}{2}+k} \left(\frac{(1-\rho^2(\boldsymbol{h}))}{\nu}\right)^{\frac{\nu-b}{2}+k} \qquad \text{(A5)}
$$

and combining Equations (A5) and (A4), we obtain

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

Then, using the Euler transformation, we obtain

$$
\mathrm{E}(R_\nu^a(\boldsymbol{s}_i)R_\nu^b(\boldsymbol{s}_j)) = \frac{2^{-(a+b)/2}\nu^{(a+b)/2}}{\Gamma^2\left(\frac{\nu}{2}\right)}\Gamma\left(\frac{\nu-a}{2}\right)\Gamma\left(\frac{\nu-b}{2}\right) {}_2F_1\left(\frac{a}{2}, \frac{b}{2}; \frac{\nu}{2}; \rho^2(\boldsymbol{h})\right) \qquad \text{(A6)}
$$

for $\nu > a$ and $\nu > b$. Finally, setting $a = b = 1$ in (A6) and using it in (A1) we obtain (6). ∎

A.2 Proof of Theorem 2

Proof. If G is a weakly stationary Gaussian process with correlation $\rho(\boldsymbol{h})$, then from (6) 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, the mean-square continuity and the m-times mean-square differentiability of Y_ν^* are equivalent to the continuity and $2m$-times differentiability of $\rho_{Y_\nu^*}(\boldsymbol{h})$ at $\boldsymbol{h}=\mathbf{0}$, respectively (Stein, 1999). Recall that the correlation function of Y_ν^* is given by:

$$\rho_{Y_\nu^*}(\boldsymbol{h}) = a(\nu)[{}_2F_1\left(\frac{1}{2},\frac{1}{2};\frac{\nu}{2};\rho^2(\boldsymbol{h})\right)\rho(\boldsymbol{h})]. \tag{A7}$$

with $a(\nu)=\frac{(\nu-2)\Gamma^2\left(\frac{\nu-1}{2}\right)}{2\Gamma^2\left(\frac{\nu}{2}\right)}$. Using (5), it can be easily seen that $\rho(\mathbf{0})=1$ if and only if $\rho_{Y_\nu^*}(\mathbf{0})=1$. Then Y_ν^* is mean-square continuous if and only if G is mean-square continuous.

For the mean square differentiability, let G m-times mean square differentiable. Using iteratively the nth derivative of the ${}_2F_1$ function with respect to x:

$${}_2F_1^{(n)}(a,b,c,x) = \frac{(a)_n(b)_n}{(c)_n}{}_2F_1(a+n,b+n,c+n,x), \quad n=1,2,\dots \tag{A8}$$

and applying the convergence condition $c>b+a$ of identity (5), it can be shown that $\rho_{Y_\nu^*}^{(2m)}(\boldsymbol{h})|_{\boldsymbol{h}=\mathbf{0}}<\infty$ if $\nu>2(2m+1)$ and, as a consequence, Y_ν^* is m-times mean square differentiable under this condition. On the other hand, if $\nu\le 2(2m+1)$ then Y_ν^* is $(m-k)$-times mean-square differentiable if $2(2(m-k)+1)<\nu\le 2(2(m-k)+3)$, for $k=1,\dots,m$.

For instance, let assume that G is 1-times mean square differentiable. This implies that $\rho^{(i)}(\boldsymbol{h})|_{\boldsymbol{h}=\mathbf{0}}<\infty$, $i=1,2$. Applying (A8) to (A7), the second derivative of $\rho_{Y_\nu^*}(\boldsymbol{h})$ is given by:

$$\begin{aligned}\rho_{Y_\nu^*}^{(2)}(\boldsymbol{h}) = \frac{a(\nu)}{\nu(\nu+2)}\Big[&(6+3\nu){}_2F_1\left(1.5,1.5;1+\frac{\nu}{2};\rho^2(\boldsymbol{h})\right)\rho(\boldsymbol{h})\{\rho^{(1)}(\boldsymbol{h})\}^2\\ &+9{}_2F_1\left(2.5,2.5;2+\frac{\nu}{2};\rho^2(\boldsymbol{h})\right)\rho^3(\boldsymbol{h})\{\rho^{(1)}(\boldsymbol{h})\}^2\\ &+\nu(2+\nu){}_2F_1\left(0.5,0.5;\frac{\nu}{2};\rho^2(\boldsymbol{h})\right)\rho^{(2)}(\boldsymbol{h})\\ &+(2+\nu){}_2F_1\left(1.5,1.5;1+\frac{\nu}{2};\rho^2(\boldsymbol{h})\right)\rho^2(\boldsymbol{h})\rho^{(2)}(\boldsymbol{h})\Big].\end{aligned}$$

Then, applying the convergence condition of identity (5), $\rho_{Y_\nu^*}^{(2)}(\boldsymbol{h})|_{\boldsymbol{h}=\mathbf{0}}<\infty$ if $2+\nu/2>5$, that is, $\nu>6$. Therefore, Y_ν^* is 1-times mean square differentiable if $\nu>6$ and 0-times mean square differentiable if $2<\nu\le 6$. Point (d) can be shown recalling that a process F is long-range dependent if the correlation of F is such that $\int_{\mathbb{R}_+^n}|\rho_F(\boldsymbol{h})|d^n\boldsymbol{h}=\infty$ (Lim & Teo, 2009). Direct inspection, using series expansion of the hypergeometric function, shows that $\int_{\mathbb{R}_+^n}|\rho_{Y_\nu^*}(\boldsymbol{h})|d^n\boldsymbol{h}=\infty$ if and only if $\int_{\mathbb{R}_+^n}|\rho(\boldsymbol{h})|d^n\boldsymbol{h}=\infty$ and, as a consequence, Y_ν^* has long-range dependence if and only if G has long-range dependence. Finally, note that if $\nu>2$ then $a(\nu){}_2F_1\left(\frac{1}{2},\frac{1}{2};\frac{\nu}{2};0\right)=a(\nu)$, $a(\nu){}_2F_1\left(\frac{1}{2},\frac{1}{2};\frac{\nu}{2};1\right)=1$, and $a(\nu){}_2F_1\left(\frac{1}{2},\frac{1}{2};\frac{\nu}{2};x^2\right)$ is not decreasing in $0\le x\le 1$. This implies $a(\nu){}_2F_1\left(\frac{1}{2},\frac{1}{2};\frac{\nu}{2};x^2\right)\le 1$, that is, $\rho_{Y_\nu^*}(\boldsymbol{h})\le\rho(\boldsymbol{h})$. Moreover, $\lim_{\nu\to\infty}a(\nu)=1$ and using series expansion of the hypergeometric function:

$$\lim_{\nu\to\infty} {}_2F_1\left(\frac{1}{2},\frac{1}{2};\frac{\nu}{2};\rho(\boldsymbol{h})^2\right) = \lim_{\nu\to\infty}\left[1+\frac{\left(\frac{1}{2}\right)_1^2\rho(\boldsymbol{h})^2}{1\left(\frac{\nu}{2}\right)_1}+\frac{\left(\frac{1}{2}\right)_2^2\rho(\boldsymbol{h})^4}{1\left(\frac{\nu}{2}\right)_2}+\ldots+\frac{\left(\frac{1}{2}\right)_k^2\rho(\boldsymbol{h})^{2k}}{1\left(\frac{\nu}{2}\right)_k}+\ldots\right]$$

$$= \lim_{\nu\to\infty}\left[1+\frac{2\rho(\boldsymbol{h})^2}{\nu}+\frac{\left(\frac{1}{2}\right)^2\left(\frac{1}{2}+1\right)^2\rho(\boldsymbol{h})^2}{2!\left(\frac{\nu}{2}\right)\left(\frac{\nu}{2}+1\right)}\right.$$

$$\left.+\ldots+\frac{\left(\frac{1}{2}\right)^2\left(\frac{1}{2}+1\right)^2\ldots\left(\frac{1}{2}+k-1\right)^2\rho(\boldsymbol{h})^{2k}}{k!\left(\frac{\nu}{2}\right)\left(\frac{\nu}{2}+1\right)\ldots\left(\frac{\nu}{2}+k-1\right)}+\ldots\right]$$

$$= 1.$$

This implies $\lim_{\nu\to\infty}\rho_{Y_\nu^*}(\boldsymbol{h}) = \rho(\boldsymbol{h})$. ∎

A.3 Proof of Theorem 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 $J((g_i,g_j)\to(y_i,y_j)) = (w_iw_j)^{1/2}$, we have:

$$f_{\boldsymbol{Y}_{ij}^*}(\boldsymbol{y}_{ij}) = \int_{\mathbb{R}_+^2} f_{\boldsymbol{G}_{ij}|\boldsymbol{W}_{ij}}(\boldsymbol{g}_{ij}|\boldsymbol{w}_{ij})f_{\boldsymbol{W}_{ij}}(\boldsymbol{w}_{ij})J\,d\boldsymbol{w}_{ij}$$

$$= \frac{2^{-\nu}\nu^\nu}{2\pi\Gamma^2\left(\frac{\nu}{2}\right)(1-\rho^2(\boldsymbol{h}))^{(\nu+1)/2}}\int_{\mathbb{R}_+^2}(w_iw_j)^{(\nu+1)/2-1}e^{-\frac{1}{2(1-\rho^2(\boldsymbol{h}))}\left[w_iy_i^2+w_jy_j^2-2\rho(\boldsymbol{h})\sqrt{w_iw_j}y_iy_j\right]}$$

$$\times e^{-\frac{\nu(w_i+w_j)}{2(1-\rho^2(\boldsymbol{h}))}}{}_0F_1\left(\frac{\nu}{2};\frac{\nu^2\rho^2(\boldsymbol{h})w_iw_j}{4(1-\rho^2(\boldsymbol{h}))^2}\right)d\boldsymbol{w}_{ij}$$

$$= \frac{2^{-\nu}\nu^\nu}{2\pi\Gamma^2\left(\frac{\nu}{2}\right)(1-\rho^2(\boldsymbol{h}))^{(\nu+1)/2}}\int_{\mathbb{R}_+^2}(w_iw_j)^{(\nu+1)/2-1}e^{-\frac{1}{2(1-\rho^2(\boldsymbol{h}))}\left[y_i^2-2\rho(\boldsymbol{h})\sqrt{\frac{w_j}{w_i}}y_iy_j+\nu\right]w_i}e^{-\frac{(y_j^2+\nu)w_j}{2(1-\rho^2(\boldsymbol{h}))}}$$

$$\times\sum_{k=0}^{\infty}\frac{1}{k!\left(\frac{\nu}{2}\right)_k}\left(\frac{\nu^2\rho^2(\boldsymbol{h})w_iw_j}{4(1-\rho^2(\boldsymbol{h}))^2}\right)^k d\boldsymbol{w}_{ij}$$

$$= \frac{2^{-\nu}\nu^\nu}{2\pi\Gamma^2\left(\frac{\nu}{2}\right)(1-\rho^2(\boldsymbol{h}))^{(\nu+1)/2}}\sum_{k=0}^{\infty}\frac{I(k)}{k!\left(\frac{\nu}{2}\right)_k}\left(\frac{\nu^2\rho^2(\boldsymbol{h})}{4(1-\rho^2(\boldsymbol{h}))^2}\right)^k \tag{A9}$$

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

$$I(k) = \int_{\mathbb{R}_+} w_j^{(\nu+1)/2+k-1}e^{-\frac{(y_j^2+\nu)w_j}{2(1-\rho^2(\boldsymbol{h}))}}\left[\int_{\mathbb{R}_+} w_i^{(\nu+1)/2+k-1}e^{\left[-\frac{(y_i^2+\nu)}{2(1-\rho^2(\boldsymbol{h}))}w_i-\frac{\rho(\boldsymbol{h})\sqrt{w_j}y_iy_j}{(\rho^2(\boldsymbol{h})-1)}\sqrt{w_i}\right]}dw_i\right]dw_j$$

$$= 2\left(\frac{y_i^2+\nu}{(1-\rho^2(\boldsymbol{h}))}\right)^{-\left(\frac{\nu+1}{2}+k\right)}\Gamma(\nu+1+2k)\int_{\mathbb{R}_+} w_j^{(\nu+1)/2+k-1}e^{\left[\frac{\rho^2(\boldsymbol{h})y_i^2y_j^2}{4(1-\rho^2(\boldsymbol{h}))(y_i^2+\nu)}-\frac{(y_j^2+\nu)}{2(1-\rho^2(\boldsymbol{h}))}\right]w_j}$$

$$\times D_{-(\nu+1+2k)}\left(-\frac{\rho(\boldsymbol{h})y_iy_j\sqrt{w_j}}{\sqrt{(1-\rho^2(\boldsymbol{h}))(y_i^2+\nu)}}\right)dw_j$$

$$=2\left(\frac{y_i^2+\nu}{(1-\rho^2(\boldsymbol{h}))}\right)^{-\left(\frac{\nu+1}{2}+k\right)}\Gamma(\nu+1+2k)A(k), \tag{A10}$$

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

$$D_{-(\nu+1+2k)}\left(-\frac{\rho(\boldsymbol{h})y_iy_j\sqrt{w_j}}{\sqrt{(1-\rho^2(\boldsymbol{h}))(y_i^2+\nu)}}\right)=b_1e^{-\frac{\rho^2(\boldsymbol{h})y_i^2y_j^2w_j}{4(1-\rho^2(\boldsymbol{h}))(y_i^2+\nu)}}$$

$$\times{}_1F_1\left(\frac{\nu+1}{2}+k;\frac{1}{2};\frac{\rho^2(\boldsymbol{h})y_i^2y_j^2w_j}{2(1-\rho^2(\boldsymbol{h}))(y_i^2+\nu)}\right)$$

$$+b_2\sqrt{w_j}e^{-\frac{\rho^2(\boldsymbol{h})y_i^2y_j^2w_j}{4(1-\rho^2(\boldsymbol{h}))(y_i^2+\nu)}}$$

$$\times{}_1F_1\left(\frac{\nu}{2}+k+1;\frac{3}{2};\frac{\rho^2(\boldsymbol{h})y_i^2y_j^2w_j}{2(1-\rho^2(\boldsymbol{h}))(y_i^2+\nu)}\right), \tag{A11}$$

where $b_1=\frac{2^{-(\nu+1)/2+k}\sqrt{\pi}}{\Gamma\left(\frac{\nu}{2}+k+1\right)}$ and $b_2=\frac{2^{-\nu/2-k}\sqrt{\pi}\rho(\boldsymbol{h})y_iy_j}{\Gamma\left(\frac{\nu+1}{2}+k\right)\sqrt{(1-\rho^2(\boldsymbol{h}))(y_i^2+\nu)}}$. Replacing Equations (A11) in (A10) and using (7.621.4) of Gradshteyn and Ryzhik (2007), we obtain

$$A(k)=b_1\int_{\mathbb{R}_+}w_j^{(\nu+1)/2+k-1}e^{-\frac{(y_j^2+\nu)}{2(1-\rho^2(\boldsymbol{h}))}w_j}{}_1F_1\left(\frac{\nu+1}{2}+k;\frac{1}{2};\frac{\rho^2(\boldsymbol{h})y_i^2y_j^2w_j}{2(1-\rho^2(\boldsymbol{h}))(y_i^2+\nu)}\right)dw_j$$

$$+b_2\int_{\mathbb{R}_+}w_j^{\nu/2+k+1-1}e^{-\frac{(y_j^2+\nu)}{2(1-\rho^2(\boldsymbol{h}))}w_j}{}_1F_1\left(\frac{\nu}{2}+k+1;\frac{3}{2};\frac{\rho^2(\boldsymbol{h})y_i^2y_j^2w_j}{2(1-\rho^2(\boldsymbol{h}))(y_i^2+\nu)}\right)dw_j$$

$$=b_1\Gamma\left(\frac{\nu+1}{2}+k\right)\left(\frac{y_j^2+\nu}{2(1-\rho^2(\boldsymbol{h}))}\right)^{-\frac{(\nu+1)}{2}-k}{}_2F_1\left(\frac{\nu+1}{2}+k,\frac{\nu+1}{2}+k;\frac{1}{2};\frac{\rho^2(\boldsymbol{h})y_i^2y_j^2}{(y_i^2+\nu)(y_j^2+\nu)}\right)$$

$$+b_2\Gamma\left(\frac{\nu}{2}+k+1\right)\left(\frac{y_j^2+\nu}{2(1-\rho^2(\boldsymbol{h}))}\right)^{-\frac{\nu}{2}-k-1}{}_2F_1\left(\frac{\nu}{2}+k+1,\frac{\nu}{2}+k+1;\frac{3}{2};\frac{\rho^2(\boldsymbol{h})y_i^2y_j^2}{(y_i^2+\nu)(y_j^2+\nu)}\right) \tag{A12}$$

finally, combining Equations (A12), (A10), and (A9), we obtain

$$f_{\mathbf{Y}_{ij}^*}(\boldsymbol{y}_{ij})=\frac{\nu^\nu[(y_i^2+\nu)(y_j^2+\nu)]^{-(\nu+1)/2}\Gamma^2\left(\frac{\nu+1}{2}\right)}{\pi\Gamma^2\left(\frac{\nu}{2}\right)(1-\rho^2(\boldsymbol{h}))^{-(\nu+1)/2}}\sum_{k=0}^{\infty}\frac{\left(\frac{\nu+1}{2}\right)_k^2}{k!\left(\frac{\nu}{2}\right)_k}\left(\frac{\nu^2\rho^2(\boldsymbol{h})}{(y_i^2+\nu)(y_j^2+\nu)}\right)^k$$

$$\times {}_2F_1\left(\frac{\nu+1}{2}+k, \frac{\nu+1}{2}+k; \frac{1}{2}; \frac{\rho^2(\boldsymbol{h})y_i^2y_j^2}{(y_i^2+\nu)(y_j^2+\nu)}\right)$$

$$+\frac{\rho(\boldsymbol{h})y_iy_j\nu^{\nu+2}[(y_i^2+\nu)(y_j^2+\nu)]^{-\nu/2-1}}{2\pi(1-\rho^2(\boldsymbol{h}))^{-(\nu+1)/2}}\sum_{k=0}^{\infty}\frac{\left(\frac{\nu}{2}+1\right)_k^2}{k!\left(\frac{\nu}{2}\right)_k}\left(\frac{\nu^2\rho^2(\boldsymbol{h})}{(y_i^2+\nu)(y_j^2+\nu)}\right)^k$$

$$\times {}_2F_1\left(\frac{\nu}{2}+k+1, \frac{\nu}{2}+k+1; \frac{3}{2}; \frac{\rho^2(\boldsymbol{h})y_i^2y_j^2}{(y_i^2+\nu)(y_j^2+\nu)}\right)$$

■

and using (21) we obtain Theorem 3.

A.4 Proof of Theorem 1

Proof. Consider $\boldsymbol{U} = (U(\boldsymbol{s}_1), \dots, U(\boldsymbol{s}_n))^T$, $\boldsymbol{V} = (|X_1(\boldsymbol{s}_1)|, \dots, |X_1(\boldsymbol{s}_n)|)^T$, $\boldsymbol{Q} = (X_2(\boldsymbol{s}_1), \dots, X_2(\boldsymbol{s}_n))^T$ where $\boldsymbol{X}_k = (X_k(\boldsymbol{s}_1), \dots, X_k(\boldsymbol{s}_n))^T \sim N_n(\boldsymbol{0}, \Omega)$, for $k = 1, 2$, which are assumed to be independent. By definition of the skew-Gaussian process in (12) we have:

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

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

$$f_U(\boldsymbol{u}) = \int_{\mathbb{R}^n} \phi_n(\boldsymbol{u}; \boldsymbol{\alpha} + \eta\boldsymbol{v}, \omega^2\Omega) f_V(\boldsymbol{v})\, d\boldsymbol{v}.$$

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

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

Hence, by using that

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

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

$$\begin{aligned} f_{\boldsymbol{V}}(\boldsymbol{v}) &= \sum_{\boldsymbol{l} \in \{-1,+1\}^n} [\det\{\boldsymbol{D}(\boldsymbol{l})\}]^2 \phi_n(\boldsymbol{D}(\boldsymbol{l})\boldsymbol{v}; \boldsymbol{0}, \Omega) \\ &= \sum_{\boldsymbol{l} \in \{-1,+1\}^n} \phi_n(\boldsymbol{D}(\boldsymbol{l})\boldsymbol{v}; \boldsymbol{0}, \Omega), \quad ([\det\{\boldsymbol{D}(\boldsymbol{l})\}]^2 = 1) \end{aligned}$$

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

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

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

since

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

and both sets produce the same correlation matrices. The joint density of $\boldsymbol{U}$ is thus given by

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

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