Estimating covariance functions of multivariate skew-Gaussian random fields on the sphere

Alegría A. [*a], Caro S.[b], Bevilacqua M.[c], Porcu E.[a], Clarke J.[a]

[a]*Departamento de Matemática, Universidad Técnica Federico Santa María, Valparaíso, Chile*

[b]*Departamento de Matemática y Computación, Universidad de Santiago, Santiago, Chile*

[c]*Instituto de Estadística, Universidad de Valparaíso, Valparaíso, Chile*

Abstract

This paper considers a multivariate spatial random field, with each component having univariate marginal distributions of the skew-Gaussian type. We assume that the field is defined spatially on the unit sphere embedded in $\mathbb{R}^3$, allowing for modeling data available over large portions of planet Earth. This model admits explicit expressions for the marginal and cross covariances. However, the n-dimensional distributions of the field are difficult to evaluate, because it requires the sum of 2^n terms involving the cumulative and probability density functions of a n-dimensional Gaussian distribution. Since in this case inference based on the full likelihood is computationally unfeasible, we propose a composite likelihood approach based on pairs of spatial observations. This last being possible thanks to the fact that we have a closed form expression for the bivariate distribution. We illustrate the effectiveness of the method through simulation experiments and the analysis of a real data set of minimum and maximum surface air temperatures.

Keywords: Composite likelihood, Geodesic distance, Global data

1. Introduction

The Gaussian assumption is an appealing option to model spatial data. First, the Gaussian distribution is completely characterized by its first two moments. Another interesting property is the tractability of the Gaussian distribution under linear combinations and conditioning. However, in many geostatistical applications, including oceanography, environment and the study of natural resources, the Gaussian framework is unrealistic, because the observed data have different features, such as, positivity, skewness or heavy tails, among others.

Transformations of Gaussian random fields (RFs) is the most common alternative to model non-Gaussian fields. Consider a spatial domain $\mathcal{D}$ and $\{Z(\boldsymbol{s}), \boldsymbol{s} \in \mathcal{D}\}$ defined as $Z(\boldsymbol{s}) = \varphi(Y(\boldsymbol{s}))$, where φ is a real-valued mapping and $\{Y(\boldsymbol{s}), \boldsymbol{s} \in \mathcal{D}\}$ is Gaussian. Apparently the finite-dimensional distributions of Z depend on the choice of φ. In some cases, such mapping is one-to-one and admits an inverse simplifying the analysis. In this class, we can highlight the log-Gaussian RFs,

---

[*]Corresponding author. Email: alfredo.alegria.jimenez@gmail.com

which are generated as a particular example of the Box-Cox transformation (see De Oliveira et al., 1997). However, a one-to-one transformation is not appropriate in general. For instance, Stein (1992) considers a truncated Gaussian RF, taking φ as an indicator function, in order to model data sets with a given percentage of zeros (for instance, precipitations). On the other hand, wrapped-Gaussian RFs have been introduced in the literature for modeling directional spatial data, arising in the study of wave and wind directions (see Jona-Lasinio et al., 2012). In addition, Xu and Genton (2016a) propose a flexible class of fields named the Tukey g-and-h RFs.

Another approach consists in taking advantage of the stochastic representations of random variables. For instance, Ma (2009) considers a general approach to construct elliptically contoured RFs through mixtures of Gaussian fields. These models have an explicit covariance structure and allow a wide range of finite-dimensional distributions. Other related works have been developed by Du et al. (2012), Ma (2013a) and Ma (2013b) including Hyperbolic, K and Student's t distributed fields. Moreover, Kim and Mallick (2004), Gualtierotti (2005) and Allard and Naveau (2007) have introduced skew-Gaussian RFs for modeling data with skewed distributions. However, Minozzo and Ferracuti (2012) and Genton and Zhang (2012) show that all these models are not valid because they cannot be identified with probability one using a single realization, i.e., in practice it is impossible to make inference on the basis of these models. Such results do not prevent the existence of RFs having univariate marginal distributions belonging to a given family.

In this paper, we consider multivariate stationary RFs, where each component has a univariate marginal distribution of the skew-Gaussian type. We follow the representation proposed in the univariate case by Zhang and El-Shaarawi (2010) and extend it to the multivariate case. This construction allows for modeling data with different degrees of skewness as well as explicit expressions for the covariance function. Estimation methods for this model are still unexplored. Maximum likelihood is certainly a useful tool, but it is impracticable, because the full likelihood does not have a simple form. Indeed, if n is the number of observations, it can be explicitly expressed as the sum of 2^n terms depending on the probability density function (pdf) and the cumulative distribution function (cdf) of the n-variate Gaussian distribution. Direct maximization of the likelihood seems intractable from a computational and analytical point of view. Zhang and El-Shaarawi (2010) consider the EM algorithm to perform inference on the skew-Gaussian model. However, the iterations of the EM algorithm are difficult to evaluate because each step requires Monte Carlo simulations of a non-trivial conditional expectation. On the other hand, composite likelihood (CL) is an estimation procedure (Lindsay, 1988; Varin et al., 2011; Cox and Reid, 2004) based on the likelihood of marginal or conditional events. CL methods are an attractive option when the full likelihood is difficult to write and/or when the data sets are large. This approach has been used in several spatial and space-time contexts, mainly in the Gaussian case (Vecchia, 1988; Curriero and Lele, 1999; Stein et al., 2004; Bevilacqua et al., 2012; Bevilacqua and Gaetan, 2015; Bevilacqua et al., 2016). Outside the Gaussian scenario, Heagerty and Lele (1998) propose CL inference for binary spatial data. Moreover, Padoan et al. (2010) and Sang and Genton (2014) have used CL methods for the estimation of max-stable fields, whereas Alegría et al. (2016) consider a truncated CL approach for wrapped-Gaussian fields.

The implementation of the CL method on multivariate skew-Gaussian fields is still unexplored. Our goal consists in developing a CL approach based on pairs of observations for a multivariate skew-Gaussian RF. Our contribution provides a fast and accurate tool to make inference on skewed data. The main ingredient of the pairwise CL method is the characterization of the bivariate

distributions of the RF, that is, we derive a closed form expression for the joint distribution between two correlated skew-Gaussian random variables (possibly with different means, variances and degrees of skewness).

In addition, in order to work with data collected over the whole planet Earth, we consider the spatial domain as the unit sphere $\mathcal{D} = \mathbb{S}^2 := \{\boldsymbol{s} \in \mathbb{R}^3 : \|\boldsymbol{s}\| = 1\}$, where $\|\cdot\|$ denotes the Euclidean distance. We refer the reader to Marinucci and Peccati (2011) for a more detailed study about RFs on spheres. An important implication is that the covariance function depends on a different metric, called geodesic distance. In general, covariance models valid on Euclidean spaces are not valid on the sphere, and we refer the reader to Gneiting (2013) for a overview of the problem.

The paper is organized as follows. In Section 2, the multivariate skew-Gaussian RF is introduced. The bivariate distributions of the skew-Gaussian field and the CL approach are discussed in Section 3. In section 4, simulation experiments are developed. Section 5 contains a real application for a bivariate data set of minimum and maximum surface air temperatures. Finally, Section 6 provides a brief discussion.

2. Skew-Gaussian RFs on the unit sphere

In this section, we introduce a skew-Gaussian model generated as a mixture of two latent Gaussian RFs. Such construction is based on the stochastic representation of skew-Gaussian random variables. Let X and Y be two independent standard Gaussian random variables and $-1 \leq \delta \leq 1$. Then, the distribution of

$$Z = \delta|X| + \sqrt{1 - \delta^2}Y \tag{1}$$

is called skew-Gaussian, with pdf $2\phi(z)\Phi(\alpha z)$, for $z \in \mathbb{R}$, where $\alpha = \delta/\sqrt{1 - \delta^2}$. Here, $\phi(\cdot)$ and $\Phi(\cdot)$ denote the pdf and cdf of the standard Gaussian distribution. If $\delta > 0$, we say that Z is right-skewed, whereas for $\delta < 0$, Z is left-skewed. Of course, for $\delta = 0$ we have the Gaussian case. For a detailed study of the skew-Gaussian distribution, we refer the reader to Azzalini (1985, 1986), Azzalini and Dalla Valle (1996), Azzalini and Capitanio (1999), Arellano-Valle and Azzalini (2006) and Azzalini (2013).

We work with the spatial counterpart of Equation (1). Let $\{\boldsymbol{X}(\boldsymbol{s}) = (X_1(\boldsymbol{s}), \ldots, X_m(\boldsymbol{s}))^\top : \boldsymbol{s} \in \mathbb{S}^2\}$ and $\{\boldsymbol{Y}(\boldsymbol{s}) = (Y_1(\boldsymbol{s}), \ldots, Y_m(\boldsymbol{s}))^\top : \boldsymbol{s} \in \mathbb{S}^2\}$ be two stationary multivariate Gaussian RFs defined on $\mathbb{S}^2$. Here, $m \in \mathbb{N}$ denotes the number of components of the fields and $\top$ is the transpose operator. In addition, we assume that $\boldsymbol{X}(\boldsymbol{s})$ and $\boldsymbol{Y}(\boldsymbol{s})$ are independent, with components having zero mean and unit variance. In the spherical framework, the covariances are given in terms of the geodesic distance, defined by

$$\theta := \theta(\boldsymbol{s}_1, \boldsymbol{s}_2) = \arccos(\boldsymbol{s}_1^\top \boldsymbol{s}_2) \in [0, \pi], \qquad \boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathbb{S}^2,$$

which is the most natural metric on the spherical surface. Therefore, we suppose that there exists two matrix-valued mappings $r^x, r^y : [0, \pi] \to \mathbb{R}^{m \times m}$ such that

$$\mathrm{cov}\{X_i(\boldsymbol{s}), X_j(\boldsymbol{s}')\} = r_{ij}^x(\theta(\boldsymbol{s}, \boldsymbol{s}')) \text{ and } \mathrm{cov}\{Y_i(\boldsymbol{s}), Y_j(\boldsymbol{s}')\} = r_{ij}^y(\theta(\boldsymbol{s}, \boldsymbol{s}')),$$

for all $\boldsymbol{s}, \boldsymbol{s}' \in \mathbb{S}^2$ and $i, j = 1, \ldots, m$. In such case, we say that the covariance function is spherically isotropic. In the univariate case ($m = 1$), Gneiting (2013) establishes that some classical covariances

3

such as the Cauchy, Matérn, Askey and Spherical models, given in the classical literature in terms of Euclidean metrics, can be coupled with the geodesic distance under specific constraints for the parameters. Furthermore, Porcu et al. (2016) propose several covariance models for space-time and multivariate RFs on spherical spatial domains.

Next, we define a multivariate spatial RF with each component being skew-Gaussian distributed, according to Equation (1). It is a multivariate extension of the univariate skew-Gaussian field proposed by Zhang and El-Shaarawi (2010). This model allows different means, variances and skewness in the components of the RF.

Definition 2.1. A multivariate field, $\{\boldsymbol{Z}(\boldsymbol{s}) = (Z_1(\boldsymbol{s}), \ldots, Z_m(\boldsymbol{s}))^\top : \boldsymbol{s} \in \mathbb{S}^2\}$, with components having skew-Gaussian marginal distributions, can be defined through

$$Z_i(\boldsymbol{s}) = \mu_i + \eta_i |X_i(\boldsymbol{s})| + \sigma_i Y_i(\boldsymbol{s}), \qquad \boldsymbol{s} \in \mathbb{S}^2, \qquad i = 1, \ldots, m, \tag{2}$$

where $\mu_i, \eta_i \in \mathbb{R}$ and $\sigma_i \in \mathbb{R}_+$. Note that $(Z_i(\boldsymbol{s}) - \mu_i)/\sqrt{\eta_i^2 + \sigma_i^2}$, for all $\boldsymbol{s} \in \mathbb{S}^2$, follows a skew-Gaussian distribution with pdf given by $f_{Z_i}(z) = 2\phi(z)\,\Phi\left((\eta_i/\sigma_i)z\right)$.

Remark 2.1. Recent literature considers the latent fields $X_i(\boldsymbol{s})$, $i = 1, \ldots, m$, as single random variables X_i, being constants along the spatial domain. However, this approach has apparent identifiability problems, since in practice we only work with one realization and there is no information about the variability of X_i. Thus, this approach only produces a shift effect in the model. These considerations are studied by Minozzo and Ferracuti (2012) and Genton and Zhang (2012).

Direct application of the results given by Zhang and El-Shaarawi (2010) provides the following proposition.

Proposition 2.1. The field $\boldsymbol{Z}(\boldsymbol{s})$ defined through Equation (2) is stationary with expectations

$$\mathbb{E}(Z_i(\boldsymbol{s})) = \mu_i + \eta_i \sqrt{\frac{2}{\pi}}, \qquad \boldsymbol{s} \in \mathbb{S}^2,$$

and covariances

$$C_{ij}(\theta) := \text{cov}\{Z_i(\boldsymbol{s}), Z_j(\boldsymbol{s}')\} = \frac{2\eta_i\eta_j}{\pi} g(r_{ij}^x(\theta)) + \sigma_i\sigma_j r_{ij}^y(\theta), \tag{3}$$

for all $i, j = 1, \ldots, m$, where $\theta = \theta(\boldsymbol{s}, \boldsymbol{s}')$ and $g(t) = \sqrt{1 - t^2} + t \arcsin(t) - 1$, for $|t| \leq 1$.

The proof is omitted because it is obtained by using the same arguments as in Zhang and El-Shaarawi (2010).

3. Composite likelihood estimation

3.1. General framework

We first introduce the CL approach from a general point of view. CL methods (Lindsay, 1988) are likelihood approximations for dealing with large data sets. In addition, in the last years, these techniques have been used to study data with intractable analytical expressions for the full likelihood. The objective function for CL methods is constructed through the likelihood of marginal or conditional events. Formally, let $f(\boldsymbol{z}; \boldsymbol{\lambda})$ be the pdf of a n-dimensional random vector, where

$\boldsymbol{\lambda} \in \Lambda \subset \mathbb{R}^p$ is an unknown parameter vector, and Λ is the parametric space. We denote by $\{\mathcal{A}_1, ..., \mathcal{A}_K\}$ a set of marginal or conditional events with associated likelihoods $\mathcal{L}_k(\boldsymbol{\lambda}; \boldsymbol{z})$. Then, the objective function for the composite likelihood method is defined as the weighted product

$$\mathcal{L}_C(\boldsymbol{\lambda}; \boldsymbol{z}) := \prod_{k=1}^{K} \mathcal{L}_k(\boldsymbol{\lambda}; \boldsymbol{z})^{w_k},$$

where the non-negative weights w_k must be chosen according to an appropriate criterion. In principle, the weights can improve the statistical and/or computational efficiency of the estimation. We use the notation $\ell_k(\boldsymbol{\lambda}; \boldsymbol{z}) = \log \mathcal{L}_k(\boldsymbol{\lambda}; \boldsymbol{z})$, thus the log composite likelihood is given by

$$\mathrm{CL}(\boldsymbol{\lambda}; \boldsymbol{z}) := \sum_{k=1}^{K} \ell_k(\boldsymbol{\lambda}; \boldsymbol{z}) w_k.$$

The maximum CL estimator is defined as

$$\widehat{\boldsymbol{\lambda}}_n = \mathrm{argmax}_{\boldsymbol{\lambda} \in \Lambda} \mathrm{CL}(\boldsymbol{\lambda}; \boldsymbol{z}).$$

By construction, the composite score $\nabla \mathrm{CL}(\boldsymbol{\lambda})$ is an unbiased estimating equation, i.e., $\mathbb{E}(\nabla \mathrm{CL}(\boldsymbol{\lambda})) = \boldsymbol{0} \in \mathbb{R}^p$. This is an appealing property of CL methods, since it is a first order likelihood property. On the other hand, the second order properties are related to the Godambe information matrix, defined as

$$G_n(\boldsymbol{\lambda}) = H_n(\boldsymbol{\lambda}) J_n(\boldsymbol{\lambda})^{-1} H_n(\boldsymbol{\lambda})^\top,$$

where $H_n(\boldsymbol{\lambda}) = -\mathbb{E}(\nabla^2 \mathrm{CL}(\boldsymbol{\lambda}; \boldsymbol{z}))$ and $J_n(\boldsymbol{\lambda}) = \mathbb{E}(\nabla \mathrm{CL}(\boldsymbol{\lambda}; \boldsymbol{z}) \nabla \mathrm{CL}(\boldsymbol{\lambda}; \boldsymbol{z})^\top)$. The inverse of $G_n(\boldsymbol{\lambda})$ is an approximation of the asymptotic variance of the CL estimator. Under increasing domain and regularity assumptions, CL estimates are consistent and asymptotically Gaussian.

3.2. Pairwise CL approach for the multivariate skew-Gaussian model

We now develop a CL method based on pairs of observations for the multivariate skew-Gaussian RF. We consider the m-variate field, $\{\boldsymbol{Z}(\boldsymbol{s}) = (Z_1(\boldsymbol{s}), \dots, Z_m(\boldsymbol{s}))^\top, \boldsymbol{s} \in \mathbb{S}^2\}$, defined in (2), and a realization of $\boldsymbol{Z}(\boldsymbol{s})$ at n spatial locations, namely, $(\boldsymbol{Z}(\boldsymbol{s}_1)^\top, \dots, \boldsymbol{Z}(\boldsymbol{s}_n)^\top)^\top$. Then, we define all possible pairs $\boldsymbol{Z}_{ijkl} = (Z_i(\boldsymbol{s}_k), Z_j(\boldsymbol{s}_l))^\top$ with associated log likelihood $\ell_{ijkl}(\boldsymbol{\lambda})$, where $\boldsymbol{\lambda} \in \mathbb{R}^p$ is the parameter vector. Therefore, the corresponding log composite likelihood equation is defined by (see Bevilacqua et al., 2016)

$$\mathrm{CL}(\boldsymbol{\lambda}) = \sum_{i=1}^{m} \sum_{k=1}^{n-1} \sum_{l=k+1}^{n} \omega_{iikl} \ell_{iikl}(\boldsymbol{\lambda}) + \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \sum_{k=1}^{n} \sum_{l=1}^{n} \omega_{ijkl} \ell_{ijkl}(\boldsymbol{\lambda}).$$

Note that the order of computation of the method is $\mathcal{O}\{mn(n-1)/2 + m(m-1)n^2/2\}$. Throughout, we consider a cut-off weight function (0/1 weights): $w_{ijkl} = 1$ if $\theta(\boldsymbol{s}_k, \boldsymbol{s}_l) \leq d_{ij}$, and 0 otherwise, for some cut-off distances d_{ij}. This choice has apparent computational advantages. Moreover, it can improve the efficiency as it has been shown in Joe and Lee (2009), Davis and Yau (2011) and Bevilacqua et al. (2012). The intuition behind this approach is that the correlations between pairs of distant observations are often nearly zero. Therefore, using all possible pairs can generate a loss of efficiency, since redundant pairs can produce bias in the results.

5

From now on, we use the notation

$$\Omega(r) = \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}. \tag{4}$$

The following result characterizes the pairwise distributions for the multivariate skew-Gaussian RF. We suppose that $\eta_i \neq 0$, for all $i = 1, \ldots, m$. The case with zero skewness is reduced to the Gaussian scenario.

Proposition 3.1. Consider two sites $s_k, s_l \in \mathbb{S}^2$ and $\theta = \theta(s_k, s_l)$. The log likelihood associated to the pair $\boldsymbol{Z}_{ijkl} = (Z_i(\boldsymbol{s}_k), Z_j(\boldsymbol{s}_l))^\top$ in the multivariate skew-Gaussian model (2) is given by

$$\ell_{ijkl}(\boldsymbol{\lambda}) = \log\left(2 \sum_{t=1}^{2} \phi_2(\boldsymbol{Z}_{ijkl} - \boldsymbol{\mu}; A_t)\Phi_2(L_t; B_t) \right) \tag{5}$$

where $\phi_2(\boldsymbol{y}; \Sigma)$ denotes the bivariate Gaussian density function with zero mean and covariance matrix Σ. Similarly, $\Phi_2(\boldsymbol{l}; \Sigma)$ denotes the corresponding Gaussian cdf. Here,

$$
\begin{aligned}
\boldsymbol{\mu} &= (\mu_i, \mu_j)^\top, \\
A_t &= \Omega_2 + \Upsilon^{-1}\Omega[(-1)^t r_{ij}^x(\theta)]\Upsilon^{-1}, \\
B_t &= ([\Upsilon\Omega_2\Upsilon]^{-1} + \Omega[(-1)^t r_{ij}^x(\theta)]^{-1})^{-1}, \\
L_t &= \left[I_2 + \Upsilon\Omega_2\Upsilon\Omega[(-1)^t r_{ij}^x(\theta)]^{-1} \right]^{-1} \Upsilon(\boldsymbol{Z}_{ijkl} - \boldsymbol{\mu}),
\end{aligned}
$$

where $\Upsilon = \mathrm{diag}\{1/\eta_i, 1/\eta_j\}$, I_2 is the identity matrix of order (2×2) and

$$\Omega_2 = \begin{pmatrix} \sigma_i^2 & \sigma_i\sigma_j \\ \sigma_i\sigma_j & \sigma_j^2 \end{pmatrix} \circ \Omega(r_{ij}^y(\theta)),$$

where $\circ$ denotes the Hadamard product.

We have deduced a closed form expression for the bivariate distributions of the field. Note that the correlation function r_{ij}^x alternates its sign in each element of the sum. Evaluation of Equation (5) requires the numerical calculation of the bivariate Gaussian cdf. The proof of Proposition 3.1 is deferred to Appendix A.

4. Simulation study

This section assesses through simulation experiments the statistical and computational performance of the pairwise CL method. We pay attention to bivariate $(m = 2)$ skew-Gaussian RFs on $\mathbb{S}^2$.

4.1. Parameterization

We believe that there are no strong arguments to consider different correlation structures $r^x(\cdot)$ and $r^y(\cdot)$ for the latent RFs. For example, the smoothness of the skew-Gaussian RF is the same as the smoothness of the roughest latent Gaussian field. Moreover, if both latent correlations are

compactly supported, thus the covariance generated has also compact support. We thus consider latent fields belonging to the same parametric family of correlation functions,

$$r_{ij}^x(\theta) = \rho_{ij}^x r(\theta; c_{ij}^x), \qquad r_{ij}^y(\theta) = \rho_{ij}^y r(\theta; c_{ij}^y), \qquad i, j = 1, 2, \qquad \theta \in [0, \pi],$$

where $\rho_{ii}^x = 1$, $\rho_{12}^x = \rho_{21}^x$, $|\rho_{12}^x| \leq 1$ and $c_{ij}^x > 0$ (and similar conditions for ρ_{ij}^y and c_{ij}^y), with mapping $\theta \mapsto r(\theta; c)$ being any univariate correlation function on the sphere (see Gneiting, 2013).

The particular choice of $r(\cdot; \cdot)$ produces additional restrictions on the parameters. Throughout, we work with

$$r(\theta; c) = \exp\left(-\frac{3\theta}{c}\right), \qquad \theta \in [0, \pi], \tag{6}$$

or

$$r(\theta; c) = \left(1 - \frac{\theta}{c}\right)_+^4, \qquad \theta \in [0, \pi], \tag{7}$$

where $c > 0$ is a scale parameter and $(a)_+ = \max\{0, a\}$. Mappings (6) and (7) are known as Exponential and Askey models, respectively. The former decreases exponentially to zero and it takes values less than 0.05 for $\theta > c$, whereas the second is compactly supported, that is, it is identically equal to zero beyond the cut-off distance c. Explicit parametric conditions for the validity of the bivariate Exponential model are provided by Porcu et al. (2016). On the other hand, Appendix B illustrates a construction principle that justify the use of bivariate Askey models on spheres.

An interesting property is that the collocated correlation coefficient between the components of a bivariate skew-Gaussian RF, $C_{12}(0)/\sqrt{C_{11}(0)C_{22}(0)}$, with $C_{12}(\cdot)$ defined in (3), depends on the majority of the model parameters. Figure 1 shows the behavior of this coefficient in terms of the latent correlation coefficients ρ_{ij}^x and ρ_{ij}^y, with $\sigma_1^2 = \sigma_2^2 = 1$ and under two different settings for the skewness parameters: $(\eta_1, \eta_2) = (1, 3)$ and $(\eta_1, \eta_2) = (1, -3)$. Note that the correlation between left-skewed and right-skewed fields has a more restrictive upper bound and this case admits strong negative correlations. The following studies are based on the parsimonious parameterizations $\rho_{12} := \rho_{12}^x = \rho_{12}^y$, $c_{ij} := c_{ij}^x = c_{ij}^y$ and $c_{12} = (c_{11} + c_{22})/2$. The parameter vector is given by $\boldsymbol{\lambda} = (\sigma_1^2, \sigma_2^2, \eta_1, \eta_2, c_{11}, c_{22}, \rho_{12}, \mu_1, \mu_2)^\top$. In addition, this parameterization avoids identifiability problems.

Figure 2 shows the covariance $C_{12}(\theta)$, in Equation (3), generated from the latent correlation functions (6) and (7). The skew-Gaussian RF preserves the correlation shape of the latent fields. Figure 3 depicts a bivariate realization of a skew-Gaussian RF, over 15000 spatial locations, with latent fields having Exponential correlations. We have simulated using Cholesky decomposition with $\sigma_1^2 = 0.1$, $\sigma_2^2 = 0.5$, $\eta_1 = 2$, $\eta_2 = 1$, $\mu_1 = \mu_2 = 0$, $\rho_{12} = 0.9$ and $c_{11} = c_{22} = 0.6$. The skewness of the simulated data is illustrated through the corresponding histograms.

4.2. Results

We illustrate the saving of the pairwise CL method in terms of computational burden. All experiments were carried out on a 2.7 GHz processor with 8 GB of memory and the estimation procedures were implemented coupling R functions and C routines. Table 1 provides the computational times (in seconds) in evaluating the weighted CL method, with cut-off distance equal
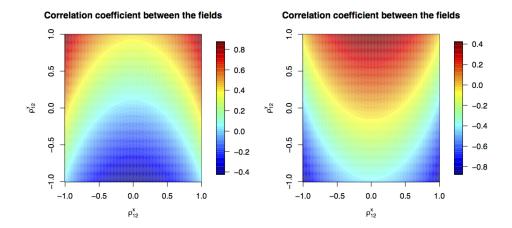
Figure 1: Collocated correlation coefficient between the components of a bivariate skew-Gaussian RF in terms of ρ_{ij}^x and ρ_{ij}^y. We consider $\sigma_1^2 = \sigma_2^2 = 1$ and two scenarios for the skewness parameters: $(\eta_1, \eta_2) = (1, 3)$ (left) and $(\eta_1, \eta_2) = (1, -3)$ (right).
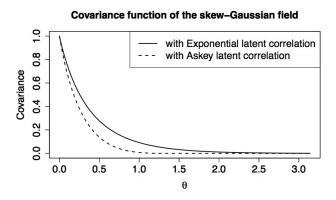

Figure 2: Covariance function associated to the skew-Gaussian RF, with latent correlations of Exponential (solid line) and Askey (dashed line) types.
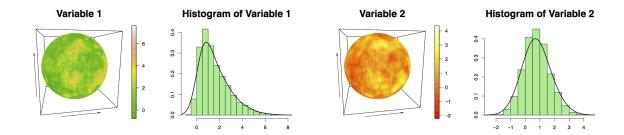

Figure 3: Bivariate simulation from the skew-Gaussian model with latent fields having Exponential correlation functions. Both variables are right-skewed and the empirical collocated correlation coefficient is approximately 0.7.

8

Table 1: Time (in seconds) in evaluating CL method, with 0/1 weights, considering different number of observations and cut-off distances $d_{ij} = 0.25, 0.5, 0.75, 1$ (in radians).

	Number of observations						
	250	500	1000	2000	4000	8000	16000
$d_{ij} = 0.25$	0.003	0.007	0.016	0.068	0.254	0.954	3.753
$d_{ij} = 0.5$	0.005	0.012	0.033	0.134	0.498	1.910	8.021
$d_{ij} = 0.75$	0.007	0.017	0.050	0.205	0.796	3.020	12.024
$d_{ij} = 1$	0.008	0.022	0.066	0.290	1.139	4.357	17.365

to $d_{ij} = 0.25, 0.5, 0.75, 1$ radians, for all $i, j = 1, 2$. These results show that CL has a moderate computational cost even for large data sets. Indeed, the most demanding part in the evaluation of the objective function is the repeated numerical calculation of the bivariate Gaussian cdf.

We now study the statistical efficiency of the estimation method. We consider 289 spatial sites in a grid on $\mathbb{S}^2$, which is generated with 17 equispaced longitude and latitude points. We use the latent correlation functions (6) and (7), with $\sigma_1^2 = \sigma_2^2 = 1$, $\mu_1 = \mu_2 = 0$, $c_{11} = 0.15$, $c_{22} = 0.25$ and $\rho_{12} = 0.5$, under the following choices for the skewness parameters:

(A) We set $\eta_1 = 1$ and $\eta_2 = 2$. In this case, both components are right-skewed and the collocated correlation coefficient between the components of the field is approximately 0.45.

(B) We set $\eta_1 = 1$ and $\eta_2 = -2$. The first component of the field is right-skewed, whereas the second one is left-skewed, and the collocated correlation coefficient between the components of the field is approximately 0.19.

In total, we have four scenarios:

- **Scenario (I)**. Exponential model under choice (A).

- **Scenario (II)**. Askey model under choice (A).

- **Scenario (III)**. Exponential model under choice (B).

- **Scenario (IV)**. Askey model under choice (B).

For each scenario, we simulate 500 independent realizations from the bivariate skew-Gaussian RF. Then, we estimate the parameters using the weighted pairwise CL method. We set $d_{ij} = 0.5$ radians for the 0/1 weights, for all $i, j = 1, 2$. Figure 4 reports the boxplots of the CL estimates. All studies show the effectiveness of our proposal. We have also applied CL estimation to other parametric models, such as, the Cauchy and Wendland correlation functions. For each case considered the pairwise CL performs well.

Finally, we assess the performance of the pairwise CL method with increasing scale parameters c_{ij} as well as increasing sample sizes n. For simplicity, all the subsequent experiments consider a single scale parameter $c_{ij} = c$, for all $i, j = 1, 2$. In this case, the parameter vector reduces to $\boldsymbol{\lambda} = (\sigma_1^2, \sigma_2^2, \eta_1, \eta_2, c, \rho_{12}, \mu_1, \mu_2)^\top$. We consider an Exponential latent correlation under the following parametric setting: $\sigma_1^2 = \sigma_2^2 = \eta_1 = 1$, $\eta_2 = 2$, $\mu_1 = \mu_2 = 0$ and $\rho_{12} = 0.5$. Figure 5 reports the centered boxplots of the CL estimates in three different cases: $c = 0.15, 0.45, 0.75$. The increase of the parameter c imply that the spatial dependence will be strengthened, and it produces
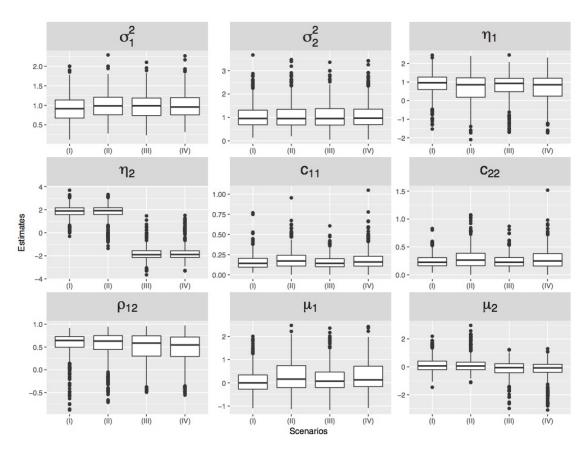
Figure 4: Boxplots of the CL estimates for the bivariate skew-Gaussian RF, under Scenarios (I)-(IV).
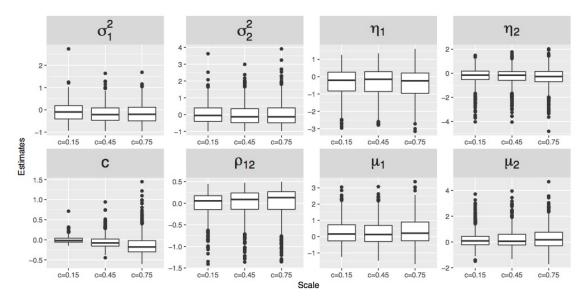

Figure 5: Centered boxplots of the CL estimates, for the bivariate skew-Gaussian RF, using an Exponential latent correlation and different scale parameters: $c = 0.15, 0.45, 0.75$.
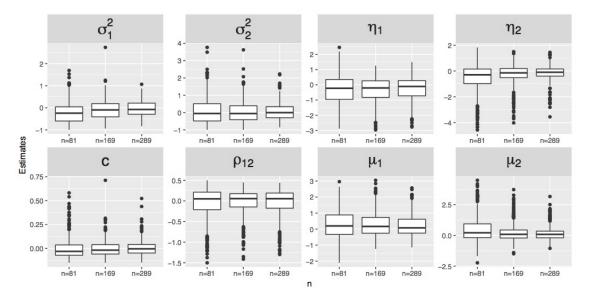
Figure 6: Centered boxplots of the CL estimates, for the bivariate skew-Gaussian RF, using an Exponential latent correlation and different sample sizes: $n = 81, 169, 289$.

biased estimates of c and ρ_{12}. Our findings are consistent and add more evidence to the results reported in the previous literature (Zhang, 2004; Bevilacqua et al., 2012; Xu and Genton, 2016b). On the other hand, we set $c = 0.15$, and we consider increasing sample sizes: $n = 81, 169, 289$, in grids generated with 9, 13 and 17 equispaced longitude and latitude points, respectively. As expected, more observations produce better estimations in terms of variability and bias.

5. A bivariate data set

We analyze a bivariate data set of Minimum (Variable 1) and Maximum (Variable 2) surface air temperatures. The spatial variability of temperatures is crucial for modeling hydrological and agricultural phenomena. These data outputs come from the Community Climate System Model (CCSM4.0) (see Gent et al., 2011) provided by NCAR (National Center for Atmospheric Research) located at Boulder, CO, USA.

We have monthly data over a grid of 2.5×2.5 degrees of spatial resolution. The unit for temperatures is Kelvin degrees. We focus on July of 2015 and we subtract the historical location-wise July average (considering the previous 50 years). Figure 7 depicts the resulting residuals for the global data set. In order to ensure spherical isotropy, we only consider locations with latitudes between -30 and 30 degrees. The final data set consists of 3456 observations per each variable. These variables are strongly correlated, since the empirical correlation is 0.68. The histogram of each variable reflects a certain degree of right skewness (see Figure 8 below). Thus, the residuals can be modeled approximately with our proposal, considering planet Earth as a sphere of radius 6378 kilometers.

We fit a bivariate skew-Gaussian RF with latent correlations of Exponential type. We use as benchmark a purely Gaussian model by taking Equation (2) with $\eta_i = 0$, for $i = 1, 2$. We have considered the parameterization introduced in the previous sections, so that, the skew-Gaussian
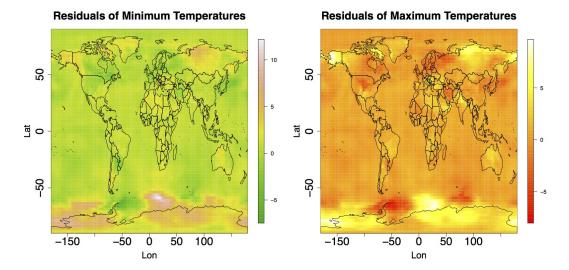
11

Figure 7: Residuals of the Minimum (left) and Maximum (right) surface air temperatures in July of 2015.

model has 9 parameters, whereas the Gaussian model has 7 parameters. The CL estimation is carried out using only pairs of observations whose spatial distances are less than 1592.75 kilometers (equivalent to 0.25 radians on the unit sphere). Table 2 reports the CL estimates for the skew-Gaussian and Gaussian models. The units of the scale parameters are kilometers.

The optimal values of the CL objective functions are given in Table 3. Note that the maximum CL value under the skew-Gaussian model is superior to the merely Gaussian model. It is clear that the incorporation of skewness produces improvements in goodness-of-fit. Figure 8 shows the histograms of each variable and the fitted skew-Gaussian and Gaussian density functions. In Figure 9, the marginal and cross empirical semi-variograms are compared to the theoretical models.

Finally, we compare both models in terms of their predictive performance. Since the covariance structure of the skew-Gaussian field is known explicitly, we use the classical best linear unbiased predictor (cokriging), which is optimal for the Gaussian model, in terms of mean squared error. However, it is not optimal for the skew-Gaussian RF. In spite of this, we will show that the skew-Gaussian model provides better predictive results. We use a drop-one prediction strategy and quantify the discrepancy between the real and predicted values through the root mean squared prediction error (RMSPE)

$$\text{RMSPE} = \sqrt{\frac{1}{2n} \sum_{i=1}^{2} \sum_{k=1}^{n} (Z_i(\boldsymbol{s}_k) - \widehat{Z}_i(\boldsymbol{s}_k))^2}$$

and the Log-score (LSCORE)

$$\text{LSCORE} = \frac{1}{2n} \sum_{i=1}^{2} \sum_{k=1}^{n} \left[\frac{\log(2\pi\widehat{\sigma}_i^2(\boldsymbol{s}_k))}{2} + \frac{(Z_i(\boldsymbol{s}_k) - \widehat{Z}_i(\boldsymbol{s}_k))^2}{2\widehat{\sigma}_i^2(\boldsymbol{s}_k)} \right],$$

where n is the number of spatial locations, $\widehat{Z}_i(\boldsymbol{s}_k)$ is the drop-one prediction of $Z_i(\boldsymbol{s}_k)$ at location $\boldsymbol{s}_k$ and $\widehat{\sigma}_i^2(\boldsymbol{s}_k)$ is the drop-one prediction variance (see Zhang and Wang, 2010). Note that the
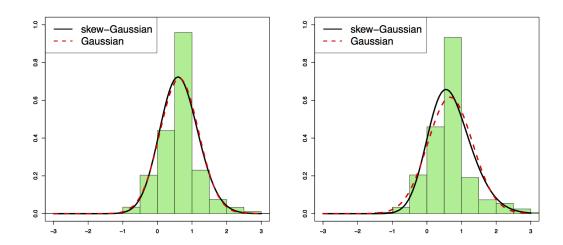
Figure 8: Histograms for the residuals of the Minimum (left) and Maximum (right) surface air temperatures, considering observations with latitudes between -30 and 30 degrees, and the fitted skew-Gaussian (solid line) and Gaussian (dashed line) probability density functions.

skew-Gaussian model generates better results since the mentioned indicators are smaller. In terms of RMSPE, the improvement in the prediction is approximately 3.1%.

6. Discussion

Building models for non-Gaussian RFs has become a major challenge and more efforts should be devoted to such constructions. In particular, it seems that the main difficulties arise when trying to build models that are statistically identifiable. Another major problem, on the other hand, comes when building the finite dimensional distributions, which are analytically intractable in most cases. This paper has provided an approach that allows to avoid the identifiability problem in multivariate skew-Gaussian RFs, and that permits to implement a CL approach.
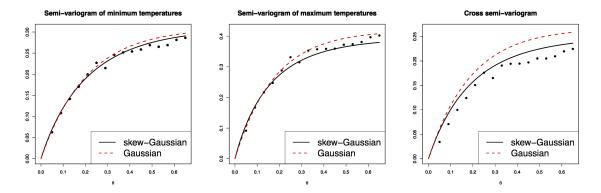

Figure 9: Empirical semi-variograms versus fitted semi-variograms, using Exponential latent correlations, for the skew-Gaussian (solid line) and Gaussian (dashed line) models.

13

Table 2: CL estimates for the skew-Gaussian and Gaussian RFs, using Exponential latent correlations. Scale parameters are given in kilometers.

	$\widehat{\sigma}_1^2$	$\widehat{\sigma_2}^2$	$\widehat{\eta}_1$	$\widehat{\eta}_2$	$\widehat{c}_{11}$	$\widehat{c}_{22}$	$\widehat{\rho}_{12}$	$\widehat{\mu}_1$	$\widehat{\mu}_2$
skew-Gaussian	0.223	0.179	0.487	0.769	4995.9	4867.1	0.819	0.250	0.079
Gaussian	0.310	0.419	-	-	3922.4	3393.1	0.743	0.635	0.674

Table 3: Prediction performance of the skew-Gaussian and Gaussian RFs, using Exponential latent correlations.

	# parameters	CL	RMSPE	LSCORE
skew-Gaussian	9	-1095912	0.219	-0.101
Gaussian	7	-1127334	0.226	0.028

We have shown that the pairwise CL method performs well under different correlation structures and parametric settings. We believe that CL approaches can be adapted to other families of non-Gaussian fields, such as, Student's t or Laplace RFs, among others. At the same time, the real data example illustrates that the incorporation of skewness can produce significant improvements in terms of prediction in comparison to a Gaussian model.

Since the optimal predictor for the skew-Gaussian model, with respect to a squared error criterion, is non-linear and difficult to evaluate explicitly, a relevant research direction is the search for methods that approximate this predictor. Indeed, Monte Carlo methods are an appealing option (Zhang and El-Shaarawi, 2010). However, from a computational point of view, Monte Carlo samples are difficult to produce efficiently and such a method can be unfeasible for large data sets.

Acknowledgments

Appendix A. Pairwise distributions for the multivariate skew-Gaussian RF

Before we state the proof of Proposition 3.1, we need the following property of quadratic forms.

Lemma Let A and B be two symmetric positive definite matrices of order $(n \times n)$ and $\boldsymbol{x}, \boldsymbol{a} \in \mathbb{R}^n$. Then, we have the following identity:

$$(\boldsymbol{a} - \boldsymbol{x})^\top A^{-1}(\boldsymbol{a} - \boldsymbol{x}) + \boldsymbol{x}^\top B^{-1}\boldsymbol{x} = (\boldsymbol{x} - \boldsymbol{c})^\top (A^{-1} + B^{-1})(\boldsymbol{x} - \boldsymbol{c}) + \boldsymbol{a}^\top (A + B)^{-1}\boldsymbol{a}, \qquad \text{(A.1)}$$

where $\boldsymbol{c} = (A^{-1} + B^{-1})^{-1} A^{-1}\boldsymbol{a}$.

Proof of Proposition 3.1 Consider $\boldsymbol{W} = (|X_1|, |X_2|)^\top$, $\boldsymbol{V} = (Y_1, Y_2)^\top$, $\boldsymbol{\mu} = (\mu_1, \mu_2)^\top$, $\boldsymbol{\eta} = (\eta_1, \eta_2)^\top$ and $\boldsymbol{\sigma} = (\sigma_1, \sigma_2)^\top$, where $(X_1, X_2)^\top \sim \mathcal{N}_2(\boldsymbol{0}, \Omega(r^x))$ and $(Y_1, Y_2)^\top \sim \mathcal{N}_2(\boldsymbol{0}, \Omega(r^y))$ are independent, with $\Omega(r)$ as defined in (4). Let $\boldsymbol{Z} = (Z_1, Z_2)^\top$ defined through

$$\boldsymbol{Z} = \boldsymbol{\mu} + \boldsymbol{\eta} \circ \boldsymbol{W} + \boldsymbol{\sigma} \circ \boldsymbol{V},$$

where $\circ$ denotes the Hadamard product. Therefore, the joint probability density function of $\boldsymbol{Z}$ is given by

$$f_{\boldsymbol{Z}}(\boldsymbol{z}) = \int_{\mathbb{R}^2_+} f_{\boldsymbol{Z}|(\boldsymbol{W}=\boldsymbol{w})}(\boldsymbol{z}|\boldsymbol{w}) f_{\boldsymbol{W}}(\boldsymbol{w}) \mathrm{d}\boldsymbol{w}, \tag{A.2}$$

Here, $f_{\boldsymbol{W}}(\boldsymbol{w})$ is the pdf of the random vector $\boldsymbol{W}$ and $\boldsymbol{w} = (w_1, w_2)^\top$. Note that the cdf of the random vector $\boldsymbol{W}$ can be written as

$$F_{\boldsymbol{W}}(w_1, w_2) = \Phi_2(w_1, w_2; \Omega(r^x)) - \Phi_2(-w_1, w_2; \Omega(r^x)) - \Phi_2(w_1, -w_2; \Omega(r^x)) + \Phi_2(-w_1, -w_2; \Omega(r^x))$$

Then, we can obtain the pdf of $\boldsymbol{W}$,

$$
\begin{aligned}
f_{\boldsymbol{W}}(w_1, w_2) &= 2\bigg(\phi_2(w_1, w_2; \Omega(r^x)) + \phi_2(-w_1, w_2; \Omega(r^x)) \bigg), \\
&= 2\bigg(\phi_2(w_1, w_2; \Omega(r^x)) + \phi_2(w_1, w_2; \Omega(-r^x)) \bigg).
\end{aligned}
$$

On the other hand, $f_{\boldsymbol{Z}|\boldsymbol{W}=\boldsymbol{w}}(\boldsymbol{z}|\boldsymbol{w})$ is the pdf of the random vector $\boldsymbol{Z}|(\boldsymbol{W} = \boldsymbol{w}) \sim \mathcal{N}_2(\boldsymbol{\mu} + \boldsymbol{\eta} \circ \boldsymbol{w}; \Omega_2)$, with

$$\Omega_2 = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2 \\ \sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \circ \Omega(r^y).$$

Therefore, evaluation of the integral (A.2) requires the characterization of an integral of the form

$$\mathcal{I} = \int_{\mathbb{R}^2_+} \phi_2(\boldsymbol{z} - \boldsymbol{\mu} - \boldsymbol{\eta} \circ \boldsymbol{w}; \Omega_2) \phi_2(\boldsymbol{w}; \Omega_1) \mathrm{d}\boldsymbol{w},$$

with Ω_1 being $\Omega(r^x)$ or $\Omega(-r^x)$. Moreover, $\mathcal{I}$ can be written as

$$\mathcal{I} = |\Upsilon| \int_{\mathbb{R}^2_+} \phi_2(\Upsilon(\boldsymbol{z} - \boldsymbol{\mu}) - \boldsymbol{w}; \Upsilon\Omega_2\Upsilon) \phi_2(\boldsymbol{w}; \Omega_1) \mathrm{d}\boldsymbol{w}$$

where $\Upsilon = \mathrm{diag}\{1/\eta_1, 1/\eta_2\}$. Thus, using Equation (A.1) we have

$$
\begin{aligned}
\mathcal{I} &= |\Upsilon| \phi_2\bigg(\Upsilon(\boldsymbol{z} - \boldsymbol{\mu}); \Upsilon\Omega_2\Upsilon + \Omega_1 \bigg) \int_{\mathbb{R}^2_+} \phi_2\bigg(\boldsymbol{w} - L; ([\Upsilon\Omega_2\Upsilon]^{-1} + \Omega_1^{-1})^{-1} \bigg) \mathrm{d}\boldsymbol{w}, \\
&= |\Upsilon| \phi_2\bigg(\Upsilon(\boldsymbol{z} - \boldsymbol{\mu}); \Upsilon\Omega_2\Upsilon + \Omega_1 \bigg) \Phi_2(L; ([\Upsilon\Omega_2\Upsilon]^{-1} + \Omega_1^{-1})^{-1}) \\
&= \phi_2\bigg(\boldsymbol{z} - \boldsymbol{\mu}; \Omega_2 + \Upsilon^{-1}\Omega_1\Upsilon^{-1} \bigg) \Phi_2(L; ([\Upsilon\Omega_2\Upsilon]^{-1} + \Omega_1^{-1})^{-1})
\end{aligned}
$$

where $L = \left[I_2 + \Upsilon\Omega_2\Upsilon\Omega_1^{-1} \right]^{-1} \Upsilon(\boldsymbol{z} - \boldsymbol{\mu})$.

Appendix B. Multivariate Askey model on the sphere

We consider a multivariate Askey model for the sphere $\mathbb{S}^2$, defined according to

$$\rho_{ij} r(\theta; c_{ij}) = \rho_{ij} \left(1 - \frac{\theta}{c_{ij}}\right)^4_+, \qquad \theta \in [0, \pi], \quad i, j = 1, \ldots, m,$$

which has been used through Sections 4 and 5, considering $c_{ij} = (c_{ii} + c_{jj})/2$. We claim that such a model is positive definite using the following mixture (see Daley et al., 2015)

$$r(\theta; c_{ij}) \propto \int_0^\infty r(\theta; \xi) \xi^2 r(\xi; c_{ij}) \mathrm{d}\xi, \qquad \theta \in [0, \pi].$$

The results of Gneiting (2013) for the univariate Askey model, coupled with the conditions developed by Daley et al. (2015) complete our assertion.

References

References

Alegría, A., Bevilacqua, M., Porcu, E., 2016. Likelihood-based inference for multivariate space-time wrapped-Gaussian fields. Journal of Statistical Computation and Simulation 86 (13), 2583–2597.

Allard, D., Naveau, P., 2007. A New Spatial Skew-Normal Random Field Model. Communications in Statistics - Theory and Methods 36 (9), 1821–1834.

Arellano-Valle, R., Azzalini, A., 2006. On the unification of families of skew-normal distributions. Scandinavian Journal of Statistics 33, 561–574.

Azzalini, A., 1985. A class of distributions which includes the normal ones. Scandinavian Journal of Statistics 12 (2), 171–178.

Azzalini, A., 1986. Further results on a class of distributions which includes the normal ones. Statistica 46 (2), 199–208.

Azzalini, A., 2013. The skew-normal and related families. Vol. 3. Cambridge University Press.

Azzalini, A., Capitanio, A., 1999. Statistical applications of the multivariate skew normal distribution. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 61 (3), 579–602.

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

Bevilacqua, M., Alegria, A., Velandia, D., Porcu, E., 2016. Composite likelihood inference for multivariate Gaussian random fields. Journal of Agricultural, Biological, and Environmental Statistics 21 (3), 448–469.

Bevilacqua, M., Gaetan, C., 2015. Comparing composite likelihood methods based on pairs for spatial Gaussian random fields. Statistics and Computing 25 (5), 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.

Cox, D., Reid, N., 2004. Miscellanea: A note on pseudolikelihood constructed from marginal densities. Biometrika 91, 729–37.

Curriero, F., Lele, S., 1999. A composite likelihood approach to semivariogram estimation. Journal of Agricultural, Biological and Environmental Statistics 4, 9–28.

Daley, D., Porcu, E., Bevilacqua, M., 2015. Classes of compactly supported covariance functions for multivariate random fields. Stochastic Environmental Research and Risk Assessment 29 (4), 1249–1263.

Davis, R., Yau, C.-Y., 2011. Comments on pairwise likelihood in time series models. Statistica Sinica 21, 255–277.

De Oliveira, V., Kedem, B., Short, D., 1997. Bayesian prediction of transformed Gaussian random fields. Journal of the American Statistical Association 92, 1422–1433.

Du, J., Leonenko, N., Ma, C., Shu, H., 2012. Hyperbolic vector random fields with hyperbolic direct and cross covariance functions. Stochastic Analysis and Applications 30 (4), 662–674.

Gent, P., Danabasoglu, G., Donner, L., Holland, M., Hunke, E., Jayne, S., Lawrence, D., Neale, R., Rasch, P., Vertenstein, M., Worley, P., Yang, Z.-L., Zhang, M., 2011. The Community Climate System Model Version 4. Journal of Climate 24 (19), 4973–4991.

Genton, M. G., Zhang, H., 2012. Identifiability problems in some non-Gaussian spatial random fields. Chilean Journal of Statistics 3 (2).

Gneiting, T., 2013. Strictly and non-strictly positive definite functions on spheres. Bernoulli 19 (4), 1327–1349.

Gualtierotti, A., 2005. Skew-normal processes as models for random signals corrupted by Gaussian noise. International Journal of Pure and Applied Mathematics 20, 109–142.

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

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

Jona-Lasinio, G., Gelfand, A., Jona-Lasinio, M., 12 2012. Spatial analysis of wave direction data using wrapped Gaussian processes. The Annals of Applied Statistics 6 (4), 1478–1498.

Kim, H., Mallick, B., 2004. A bayesian prediction using the skew Gaussian distribution. Journal of Statistical Planning and Inference 120 (1-2), 85–101.

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., 2013a. K-distributed vector random fields in space and time. Statistics & Probability Letters 83 (4), 1143–1150.

Ma, C., 2013b. Student's t vector random fields with power-law and log-law decaying direct and cross covariances. Stochastic Analysis and Applications 31 (1), 167–182.

Marinucci, D., Peccati, G., 2011. Random fields on the sphere: representation, limit theorems and cosmological applications. Vol. 389. Cambridge University Press.

Minozzo, M., Ferracuti, L., 2012. On the existence of some skew-normal stationary processes. Chilean Journal of Statistics 3, 157.

Padoan, S., Ribatet, M., Sisson, S., 2010. Likelihood-based inference for max-stable processes. Journal of the American Statistical Association 105, 263–277.

Porcu, E., Bevilacqua, M., Genton, M., 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.

Sang, H., Genton, M. G., 2014. Tapered composite likelihood for spatial max-stable models. Spatial Statistics 8, 86–103.

Stein, M., 1992. Prediction and Inference for Truncated Spatial Data. Journal of Computational and Graphical Statistics 1, 91–110.

Stein, M., Chi, Z., Welty, L., 2004. Approximating likelihoods for large spatial data sets. Journal of the Royal Statistical Society B 66, 275–296.

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

Vecchia, A., 1988. Estimation and model identification for continuous spatial processes. Journal of the Royal Statistical Society B 50, 297–312.

Xu, G., Genton, M. G., 2016a. Tukey g-and-h Random Fields. Journal of the American Statistical Association (To appear).

Xu, G., Genton, M. G., 2016b. Tukey max-stable processes for spatial extremes. Spatial Statistics 18, 431 – 443.

Zhang, H., 2004. Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. Journal of the American Statistical Association 99 (465), 250–261.

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

Zhang, H., Wang, Y., 2010. Kriging and cross-validation for massive spatial data. Environmetrics 21 (3-4), 290–304.