Modeling and estimation of some non Gaussian random fields

Christian Caamaño Carrillo

Supervisor: Prof. Moreno Bevilacqua

Advisor: Prof. Carlo Gaetan

Instituto de Estadística
Universidad de Valparaíso

This dissertation is submitted for the degree of
Doctor en Estadística

May 2018

Executive Summary

Climatology, Environmental sciences and Engineering, to name some fields, show an increasing interest in statistical analysis of temporal and/or spatial data. To model the inherent uncertainty of the data the Gaussian random processes play a fundamental role (see Cressie, 1993; Cressie and Wikle, 2011, for instance). Indeed they have to offer marginal and dependence modeling in terms of means and covariances, methods of inference well studied and scalable with large dataset (Heaton et al., 2017), optimality in the prediction (Stein, 1999).

However a bell shaped distribution of the data could be limiting, because the observed data have different features such negative or positive asymmetry, heavy tails or simply take only positive values.

In such circumstance a large part of the statistical literature have considered nonlinear transformations of a Gaussian random process. Examples of this transformed processes are the trans-Gaussian random processes and copula models (Cressie, 1993; Oliveira et al., 1997; Butler and Glasbey, 2008; Kazianka and Pilz, 2010; Masarotto and Varin, 2012).

A complementary approach specifies a hierarchical model where an hidden Gaussian random process regulates the parameters of the likelihood (Diggle et al., 1998; Banerjee et al., 2014). In such case the computational amount of the fitting procedure can be greatly reduced under a Markov assumption about the underlying process (Lindgren et al., 2011).

The aim of this thesis is to develop theoretical and methodological aspects for the construction and estimation of non-Gaussian random fields. Specifically we focus on random fields with marginal distributions of the Gamma, Weibull, t and asymmetric t type.

The thesis consists of four chapters in which different topics were studied according to the proposed objectives. In Chapter 1 and 2 we present the preliminary concepts associated with spatial and spatio-temporal modelling. Important concepts such as stationarity, covariance function, isotropy, separability, geometric properties and kriging are introduced.

In Chapter 3 we introduce two possibly non stationary random fields with Gamma and Weibull marginal distributions models for positive continuous values. The random fields are obtained transforming a rescaled sum of independent copies of squared Gaussian random fields. The non-stationarity is introduced by multiplying the stationary versions of the random fields by a positive functions. We derives the bivariate distributions and we illustrates the second order and geometrical properties in both cases.

A limitation of our modeling proposal is the difficulty of evaluating the multivariate density. For this reason we investigate the use of a weighted version of the pairwise likelihood (Lindsay, 1988; Varin et al., 2011) for estimating the marginal and dependence parameters. No effort is needed to separate the evaluation of the pairwise likelihood into a number of parallel tasks. Then the estimation has been implemented in OpenCL as a part of `GeoModels` an upcoming R package.

The effectiveness of our proposal is illustrated through a simulation study and are analysis of a well-known dataset, Ireland wind data (Haslett and Raftery, 1989). The analysis of wind speed is performed on the raw data without a preliminary transformation of the data, differently from the previous literature (see Gneiting, 2002b; Stein et al., 2004; Stein, 2005a; De Luna and Genton, 2005, for instance).

Finally, in Chapter 4 we propose a random field with t marginals obtained mixing a Gaussian random field with an inverse square root Gamma random field where the Gamma random field is introduced in Chapter 3.

Despite this can be viewed as a natural way to define a t random field, the associated second order and geometrical properties and the finite dimensional distribution are unknown. We study the second order and geometrical properties of the t random field and specifically we give an analytic closed form expression for the covariance and the bivariate distribution. We then consider two possible generalizations allowing for possible asymmetry. In the first approach, following Azzalini and Capitanio (2014), we obtain a skew-t random field mixing a skew Gaussian random field, as defined in Zhang and El-Shaarawi (2010) and Alegría et al. (2017), with an inverse square root Gamma random field. For the skew Gaussian random field, we find a closed form expression for the finite dimensional distribution generalizing previous results in Alegría et al. (2017).

In the second approach, inspired by Arellano-Valle et al. (2005), we consider a random field obtained mixing a discrete random fields with the half-t random field where the discrete random fields is obtained truncating a Gaussian random field. In this case a random field with marginals of the two piece t is obtained. We study in both cases the second order and geometrical properties and in the second approach we additionally give a closed form expression for the bivariate distribution.

As suggested in Genton and Zhang (2012), our definition of t or skew-t random fields allows to remove the problem of identifiability in Ma (2009), Ma (2010), Kim and Mallick (2004) or DeBastiani et al. (2015). The effectiveness of our proposal is illustrated through a simulation study estimating the parameters by the weighted pairwise composite likelihood method.

Acknowledgements

In these first lines, I would like to express my gratitude to all those people whose advice I am thankful, their presence and support have been by my side during the preparation of this thesis.

First of all, I would like to thank God for giving me the opportunity to improve myself every day, giving me the knowledge and guided my steps.

To my family, thank you for your support and unconditional understanding. Especially my mother Elisa Carrillo, my girlfriend Margarita Silva and those who have been important to me and today are not present but are always in my memory, my grandparents José Carrillo Fuentes and Nolfa Cisternas, and my uncle José Carrillo Cisternas.

I would like to express my gratitude to my advisor, Dr. Moreno Bevilacqua, for his encouragement, insightful guidance and endless support throughout my postgraduate experience. It has been a privilege, both from an academic and personal point of view, to work with him. I would also like to thank my co-advisors Dr. Carlo Gaetan and Dr. Reinaldo Arellano who guided me in the correct direction of this thesis.

I would also like to thank all the professors of the Institute of Statistics of the University of Valparaíso, who contributed to my training to reaching this goal.

To the University of Bío-Bío, especially to the Department of Statistics, who gave me the opportunity to enroll and finish this program.

To my colleagues and friends in the program, with whom I share a wonderful friendship, and without whom these years would have been an even more difficult path to follow.

To all those who once helped me in a disinterested way, secretaries, assistants and many people I met during these years at the University of Valparaíso.

Abstract

Spanish Version

En esta tesis, proponemos dos tipos de modelos para el análisis de regresión y dependencia de datos positivos y continuos, y de datos continuos con posible asimetría y/o cola pesada. Para el primer caso, proponemos dos campos aleatorios (posiblemente no estacionarios) con marginales de tipo Gamma y Weibull. Ambos campos aleatorios se obtienen transformando una suma rescalada de copias independientes de campos aleatorios normales al cuadrado. Para el segundo caso, proponemos un campo aleatorio con marginal t. Luego consideramos dos posibles generalizaciones que permiten una posible asimetría. En la primera aproximación obtenemos un campo aleatorio skew-t que mezcla un campo aleatorio skew-normal con la raíz cuadrada inversa de un campo aleatorio Gamma. En la segunda aproximación, obtenemos un campo aleatorio de dos piezas t que mezcla un campo aleatorio discreto binario específico con un campo aleatorio half-t.

Estudiamos las propiedades de segundo orden asociadas y en el caso estacionario, las propiedades geométricas. Dado que la estimación de máxima verosimilitud es computacionalmente inviable, incluso para un conjunto de datos relativamente pequeño, proponemos el uso de la verosimilitud por parejas. La eficacia de nuestra propuesta para los casos Gamma y Weibull, se ilustra mediante un estudio de simulación y un nuevo análisis de los datos de velocidad del viento de Irlanda (Haslett and Raftery, 1989) sin considerar ninguna transformación previa de los datos como en análisis estadísticos previos. Para los casos t y t asimétrico presentamos un estudio de simulación para mostrar el rendimiento de nuestro método.

English Version

In this thesis, we propose two types of models for the analysis of regression and dependence of positive and continuous spatio-temporal data, and of continuous spatio-temporal data with possible asymmetry and/or heavy tails. For the first case, we propose two (possibly non stationary) random fields with Gamma and Weibull marginals. Both random fields are obtained transforming a rescaled sum of independent copies of squared Gaussian random fields. For the second case, we propose a random field with t marginal distribution. We then consider two possible generalizations allowing for possible asymmetry. In the first approach we obtain a skew-t random field mixing a skew Gaussian random field with an inverse square root Gamma random field. In the second approach we obtain a two piece t random field mixing a specific binary discrete random field with half-t random field.

We study the associated second order properties and in the stationary case, the geometrical properties. Since maximum likelihood estimation is computationally unfeasible, even for relatively small data-set, we propose the use of the pairwise likelihood. The effectiveness of our proposal for the gamma and weibull cases, is illustrated through a simulation study and a re-analysis of the Irish Wind speed data (Haslett and Raftery, 1989) without considering any prior transformation of the data as in previous statistical analysis. For the t and asymmetric t cases we present a simulated study in order to show the performance of our method.

Table of contents

List of figures

List of tables

Chapter 1

Spatial modelling

1.1 Spatial processes

We review a few important basics of spatial processes. We denote as $Z(\boldsymbol{s})$, where $\boldsymbol{s}$ indexes location and $\boldsymbol{s} \in D \subset \mathbb{R}^d$, the quantity we are studying. For each $\boldsymbol{s}$, $Z(\boldsymbol{s})$ is a random variable. The collection, $Z(\boldsymbol{s})$, when $\boldsymbol{s}$ varies over all its possible values, is called a spatial process or random field *i.e.* $Z = \{Z(\boldsymbol{s}),\ \boldsymbol{s} \in D \subset \mathbb{R}^d\}$. In practice, $Z(\boldsymbol{s})$ is a random function indexed by the symbol $\boldsymbol{s}$ which belongs to some index set D. When $d = 1$ it is usually called random stochastic or random process, while when $d \geq 2$ it is defined as random field. Formally

Definition 1. *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and let $D \subset \mathbb{R}^d$ be an arbitrary set. For each $\boldsymbol{s} \in D$ the function $Z(\boldsymbol{s}, \cdot) : \Omega \to \mathbb{R},\ \omega \to (\boldsymbol{s}, \omega)$ is a random variable, and any collection of random variables $Z = \{Z(\boldsymbol{s}, \cdot),\ \boldsymbol{s} \in D \subset \mathbb{R}^d\}$ defined on $(\Omega, \mathscr{F}, \mathbb{P})$ is a stochastic process with index set D.*

Definition 2. *A sample path of spatial stochastic process is a mapping $Z : D \to \mathbb{R}$, $\boldsymbol{s} \to Z(\boldsymbol{s}, \omega)$ which to every event $\omega \in \Omega$ corresponds a sample path or trajectory of random field process Z.*

$Z(\boldsymbol{s})$ is simply a random variable for each $\boldsymbol{s}$ and its properties (e.g. mean and variance) can be described by its distribution function. More generally we are interested in studying the whole collection of random variables $\{Z(\boldsymbol{s}),\ \boldsymbol{s} \in D\}$ and its joint distribution function.

Let $Z(\boldsymbol{s})$ be a random filed and let us consider the set of points $(\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n)$. Then, the random field $Z(\boldsymbol{s})$ is characterized by its joint distribution function. The set of joint distribution functions for all values of n and all possible choices of $(\boldsymbol{s}_1,\ldots,\boldsymbol{s}_n)$ in the domain is called the spatial law of probability.

Definition 3. *For a given random field,* $Z(\boldsymbol{s})$*, the joint distribution function* $F_{Z(\boldsymbol{s}_1),Z(\boldsymbol{s}_2),\ldots,Z(\boldsymbol{s}_n)}(z_1,\ldots,z_n) : \mathbb{R}^d \to [0,1]$ *is defined as*

$$F_{Z(\boldsymbol{s}_1),Z(\boldsymbol{s}_2),\ldots,Z(\boldsymbol{s}_n)}(z_1,\ldots,z_n) = Pr(Z(\boldsymbol{s}_1) \le z_1,\ldots,Z(\boldsymbol{s}_n) \le z_n)$$

The Kolmogorov consistency theorem states that, under fairly general conditions, the probability structure of $Z(\boldsymbol{s})$ is fully specified if the joint distribution of $\{Z(\boldsymbol{s}_1),Z(\boldsymbol{s}_2),\ldots,Z(\boldsymbol{s}_n)\}$, *i.e.* the finite-dimensional distribution, is given for arbitrary choice of n and $\boldsymbol{s}_1,\ldots,\boldsymbol{s}_n$. Usually simplifying assumptions are considered on the probability structure.

1.2 Stationarity

When making inference on the probability structure of the spatial process based on what we observe (often just a single realization of the process) a simplifying assumption often made is the stationarity assumption. Stationarity in simple terms means that the random field looks similar in different parts of the domain. There are different kinds of stationarity. Now suppose $D = \mathbb{R}^d$

Definition 4. *A process* $Z(\boldsymbol{s})$ *is said strictly stationary if for all* $\boldsymbol{s}_1,\ldots,\boldsymbol{s}_n$ *and any* $\boldsymbol{h} \in \mathbb{R}^d$*, the joint distribution of* $Z(\boldsymbol{s}_1),\ldots,Z(\boldsymbol{s}_n)$ *is identical with the joint distribution of* $Z(\boldsymbol{s}_1+\boldsymbol{h}),\ldots,Z(\boldsymbol{s}_n+\boldsymbol{h})$*, i.e.,*

$$Pr(Z(\boldsymbol{s}_1) \le z_1,\ldots,Z(\boldsymbol{s}_n) \le z_n) = Pr(Z(\boldsymbol{s}_1+\boldsymbol{h}) \le z_1,\ldots,Z(\boldsymbol{s}_n+\boldsymbol{h}) \le z_n), \tag{1.1}$$

where $z_1,\ldots,z_n \in \mathbb{R}$.

That is the probability law of a strictly stationary process is invariant under a shift in space. A lighter type of stationarity is the weak stationarity:

Definition 5. *A process* $Z(\boldsymbol{s})$ *is weakly stationary (WS) if:*

$$E(Z(\boldsymbol{s})) = \mu$$

and

$$Cov(Z(\boldsymbol{s}_1),Z(\boldsymbol{s}_2)) = \mathrm{C}(\boldsymbol{s}_2-\boldsymbol{s}_1) = \mathrm{C}(\boldsymbol{h}).$$

Thus a spatial process which mean does not depend on the spatial location and which covariance is a function of the separation lag $\boldsymbol{h}$, is a WS process. $\mathrm{C}(\boldsymbol{h})$ is called the covariance function of $Z(\boldsymbol{s})$.

For a WS process $Z(\boldsymbol{s})$, the correlation between $Z(\boldsymbol{s}_1)$ and $Z(\boldsymbol{s}_2)$ is defined as:

$$Corr(Z(\boldsymbol{s}_1), Z(\boldsymbol{s}_2)) = \frac{\mathrm{C}(\boldsymbol{h})}{\mathrm{C}(\boldsymbol{0})} = \rho(\boldsymbol{h}).$$

Strict stationarity (if exist moments of second order) implies WS while the reverse is not true. A common hypothesis regarding the finite-dimensional distribution of the random fields $Z(\boldsymbol{s})$, is Gaussianity.

Definition 6. *$Z(\boldsymbol{s})$ is called Gaussian process if for all n and admissible $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$, the joint distribution of $Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_n)$ is multivariate normal.*

A multivariate normal distribution is characterized by its mean and covariance matrix, so the first two moments of a Gaussian process completely specify its probability structure. Thus for Gaussian processes, WS implies strict stationarity.

Last kind of stationarity regards the increments of the process.

Definition 7. *A process $Z(\boldsymbol{s})$ is intrinsic stationary (IS) if:*

$$E(Z(\boldsymbol{s})) = \mu$$

and

$$Var(Z(\boldsymbol{s}_1) - Z(\boldsymbol{s}_2)) = 2\gamma(\boldsymbol{s}_2 - \boldsymbol{s}_1) = 2\gamma(\boldsymbol{h}).$$

The function $2\gamma(\boldsymbol{h})$ is called variogram. IS is a weaker property than WS. If the process is WS, it is easy to verify that:

$$Var(Z(\boldsymbol{s}_1) - Z(\boldsymbol{s}_2)) = 2\mathrm{C}(\boldsymbol{0}) - 2\mathrm{C}(\boldsymbol{h})$$

and so $\gamma(\boldsymbol{h}) = \mathrm{C}(\boldsymbol{0}) - \mathrm{C}(\boldsymbol{h})$. Conversely, in general IS does not imply weak stationarity. For instance if $Z(\boldsymbol{s})$ is the standard Brownian motion in one dimension, the variogram function is $Var(Z(\boldsymbol{s}_1) - Z(\boldsymbol{s}_2)) = |\boldsymbol{s}_1 - \boldsymbol{s}_2|$. However, we can not recover the covariance function since the variogram is unbounded for $\boldsymbol{h}$ tends to infinity.

Gaussian processes play a central role in modeling spatial data. The advantages of the Gaussian process assumption are obvious: it allows convenient distribution theory (for instance, conditional distributions are easily obtained from the joint distributions). Gaussian processes have a rich, detailed and very well understood general theory. Furthermore, in most applications, we observe a single realization of the process at a finite set of locations. It is not easy to criticize a Gaussian assumption since we only have a sample size of one from a finite dimensional distribution. Nevertheless, there are situations in which it is more appropriate to use other processes to model spatial data.

1.3 Covariance functions and variograms properties

Valid (or permissible) covariance function or variogram means that they must respect some mathematical constraints. Indeed, one cannot define a spatial covariance or variogram function in a totally arbitrary way. The key property which has to satisfy is semi-positive definiteness.

Definition 8. *For a spatial process* $Z(\boldsymbol{s})$ *with finite second moments, its associated covariance matrix* $\mathrm{C} = \{\mathrm{C}(\boldsymbol{s}_i, \boldsymbol{s}_j)\}_{i,j}^{n}$, *is positive semi-definite if:*

$$\sum_{i}^{n}\sum_{j}^{n} a_i a_j \mathrm{C}(\boldsymbol{s}_i, \boldsymbol{s}_j) \geq 0 \tag{1.2}$$

for any set of $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$ *and all real* $a_1, \ldots, a_n$.

The positive semi-definite condition is necessary for the existence of a random field with finite second moments. This condition guarantees that the variance of spatial predictions is non-negative. This simply follows noting that (1.2) is $Var\left(\sum_i^n a_i Z(\boldsymbol{s}_i)\right)$.

On the other hand, if C is positive semi-definite, there exists a Gaussian random field with covariance matrix C and mean $E(Z(\boldsymbol{s})) = m < \infty$. Thus, positive definiteness is a necessary and sufficient condition for a covariance function.

If Z is a stationary process, $\mathrm{C}(\boldsymbol{s}_i, \boldsymbol{s}_j)$ depends only on $\boldsymbol{s}_i - \boldsymbol{s}_j$ (see definition 5). So we can use covariance function $\mathrm{C}(\boldsymbol{s}_i - \boldsymbol{s}_j)$ to describe the covariance structure of Z.

Bochner's theorem (1933) provides necessary and sufficient conditions for a covariance function $\mathrm{C}(\boldsymbol{h})$ of a WS process to be positive semi-definite.

Theorem 1. *(Bochner's Theorem). For a real-valued WS process on* $\mathbb{R}^d$, $\mathrm{C}(\boldsymbol{h})$ *is positive semi-definite if and only if it can be represented as:*

$$\mathrm{C}(\boldsymbol{h}) = \int e^{i\boldsymbol{\omega}^T\boldsymbol{h}} dF(\boldsymbol{\omega}) \tag{1.3}$$

where F is a positive, symmetric, and finite measure and is called the spectral measure of $\mathrm{C}(\boldsymbol{h})$. *If F is absolutely continuous with respect to Lebesgue measure, i.e., $dF(\boldsymbol{\omega}) = f(\boldsymbol{\omega})d\boldsymbol{\omega}$, $f(\boldsymbol{\omega})$ is called the spectral density.*

Analogously to the covariance function, the variogram must respect some conditions to be permissible. Specifically, for any set of $\boldsymbol{s}_1, \dots, \boldsymbol{s}_n$ and any set of real $a_1, \dots, a_n$ such that $\sum_i^n a_i = 0$,

$$\sum_i^n \sum_j^n a_i a_j \gamma(\boldsymbol{s}_i - \boldsymbol{s}_j) \leq 0. \tag{1.4}$$

This follows by noting:

$$\sum_i^n \sum_j^n a_i a_j \gamma(\boldsymbol{s}_i - \boldsymbol{s}_j) = -E\left(\sum_i^n (a_i Z(\boldsymbol{s}_i))^2\right) \leq 0.$$

The variogram and covariance functions are parameters of the spatial process and play a critical role in the geostatistical method of spatial data analysis. Both are important ingredients of the kriging methods for spatial prediction. Statisticians are more familiar with covariance functions, while geostatisticians prefer the variogram. Under WS, use $\mathrm{C}(\boldsymbol{h})$ or $\gamma(\boldsymbol{h})$ for statistical or prediction purpose is equivalent.

Under the hypothesis of second-order stationarity, the covariance function verifies the following theoretical properties:

(i) $\mathrm{C}(\boldsymbol{0}) \geq 0$;

(ii) $\mathrm{C}(\boldsymbol{h}) = \mathrm{C}(-\boldsymbol{h})$, i.e., C is an even function;

(iii) $\mathrm{C}(\boldsymbol{0}) \geq |\mathrm{C}(\boldsymbol{h})|$;

(iv) If $\mathrm{C}_j(\boldsymbol{h})$ are valid covariance function $j = 1, \dots, k$ then $\sum_{j=1}^k b_j \mathrm{C}_j(\boldsymbol{h})$ is a valid covariance function, if $b_j \geq 0, \quad \forall j$;

(v) If $\mathrm{C}_j(\boldsymbol{h})$ are valid covariance function $j = 1, \dots, k$ then $\prod_{j=1}^k \mathrm{C}_j(\boldsymbol{h})$ is a valid covariance function;

(vi) If $\mathrm{C}(\boldsymbol{h})$ is a valid covariance function in $\mathbb{R}^d$, then it is also a valid covariance function in $\mathbb{R}^p$, $p \leq d$.

Analogous properties of the variogram are:

(i) $\gamma(\mathbf{0}) = 0$;

(ii) $\gamma(\boldsymbol{h}) = \gamma(-\boldsymbol{h})$, i.e., γ is an even function;

(iii) $\gamma(\boldsymbol{h}) \geq 0$;

(iv) If $\gamma_j(\boldsymbol{h})$ are valid variograms $j = 1, \ldots, k$ then $\sum_{j=1}^{k} b_j \gamma_j(\boldsymbol{h})$ is a valid variogram,, if $b_j \geq 0, \quad \forall j$;

1.4 Isotropy and some parametric models

A WS random field is said isotropic if its covariance function $\mathrm{C}(\boldsymbol{h})$ only depends on $||\boldsymbol{h}||$, where $||\cdot||$ indicates the Euclidean distance. The isotropy property can be thought as an invariance property under rotations.

The class of all valid continuous covariance functions on $\mathbb{R}^d$ can be characterized by the Fourier transforms of all finite positive measures on $\mathbb{R}^d$ (see Theorem 1). There is an analogous characterization for isotropic covariance functions (see Yaglom, 1987, Section 22). Specifically,

Theorem 2. *For $d \geq 2$ a function $\mathrm{C}(\boldsymbol{h})$ is a continuous isotropic covariance function of a WS random field on $\mathbb{R}^d$ if and only if it can be represented as:*

$$\mathrm{C}(\boldsymbol{h}) = 2^{\frac{d-2}{2}} \Gamma\left(\frac{d}{2}\right) \int_0^\infty (\boldsymbol{\omega}||\boldsymbol{h}||)^{-\frac{d-2}{2}} J_{\frac{d-2}{2}}(\boldsymbol{\omega}||\boldsymbol{h}||) dG(\boldsymbol{\omega}) \tag{1.5}$$

Here J_k is the Bessel function of the first kind of order k (Abromowitz and Stegun, 1967) and the measure $G(\cdot)$ is nondecreasing bounded in $\mathbb{R}^+$ and $G(\mathbf{0}) = 0$.

A general form of an isotropic covariance function is:

$$\mathrm{C}(\boldsymbol{h}, \boldsymbol{\psi}) = \begin{cases} \sigma^2 \rho(||\boldsymbol{h}||, \boldsymbol{\psi}), & ||\boldsymbol{h}|| > 0 \\ \sigma^2 + \tau^2, & ||\boldsymbol{h}|| = 0 \end{cases} \tag{1.6}$$

where τ is the nugget parameter. This parameter describes the behavior of the covariance near the origin. A phenomenon quite common in applications is that the variogram at the origin does not attain the zero. This is due to microscale variability (variability of a spatial process operating at lag distances shorter than the smallest lag observed in the data) or/and measurement error. In geostatistical literature τ is the nugget, $\sigma^2 + \tau$ is the sill and σ^2 is the partial sill or variance.

In (1.6) $\rho(||\boldsymbol{h}||,\boldsymbol{\psi})$ is a parametric correlation function which depends on $\boldsymbol{\psi} \in \Psi \subset \mathbb{R}^p$. Typically parametric correlation models depend on few parameters . Following is a list of popular parametric isotropic correlation functions:

1. The (isotropic) Matérn correlation model (Matérn, 1986) is defined as:

$$\rho(||\boldsymbol{h}||,\upsilon,\alpha) = \mathscr{M}_{\upsilon,\alpha}(\boldsymbol{h}) = \frac{2^{1-\upsilon}}{\Gamma(\upsilon)}\left(\frac{||\boldsymbol{h}||}{\alpha}\right)^{\upsilon}\mathscr{K}_{\upsilon}\left(\frac{||\boldsymbol{h}||}{\alpha}\right), \qquad ||\boldsymbol{h}| \geq 0, \tag{1.7}$$

   Here, $\mathscr{K}_\upsilon$ is a modified Bessel function of the second kind of order $\upsilon > 0$ and $\alpha > 0$ a spatial scale parameter. The parameter υ characterizes the differentiability at the origin and, as a consequence, the mean square differentiability of the random field. In particular for a positive integer k, the sample paths of a Gaussian process are k times differentiable if and only if $\upsilon > k$.

2. The (isotropic) compactly supported Generalized Wendland correlation model (Gneiting, 2002a) is defined:

$$\mathscr{GW}_{\gamma,\delta,\alpha}(\boldsymbol{h}) = \begin{cases} \frac{1}{B(2\gamma,\delta+1)}\int_{||\boldsymbol{h}||/\alpha}^{1} u\left(u^2-\frac{||\boldsymbol{h}||^2}{\alpha^2}\right)^{\gamma-1}(1-u)^{\delta}\,du & ||\boldsymbol{h}|| < \alpha, \\ 0 & \text{otherwise} \end{cases} \tag{1.8}$$

   where $B(\cdot,\cdot)$ is the beta function. For $\gamma = 0$ the correlation resuces to $\mathscr{GW}_{0,\delta,\alpha}(\boldsymbol{h}) = (1-||\boldsymbol{h}||/\alpha)_+^{\delta}$ where $(\cdot)_+$ denotes the positive part, defined as Askey function.

   Here, $\gamma \geq 0$, $\delta > (d+1)/2+\gamma$ and $\alpha > 0$ is the spatial compact support. The parameter δ characterizes the differentiability at the origin and, as a consequence, the mean square differentiability of the random field. In particular for a positive integer k, the sample paths of a Gaussian process are k times differentiable if and only if $\gamma > k-1/2$. Closed form solution of the integral in Equation (1.8) can be obtained when γ is a positive integer.

Correlation models (1.8) and (1.7) allows a continuous parametrization of the smoothness of the associated Gaussian random field (Stein, 1999; Bevilacqua et al., 2018) and as a consequence they represent flexible parametric models for the correlation of spatial data. Generalized Wendland is additionally compactly supported, an interesting feature from computational point of view.

γ	$\mathcal{GW}_{\gamma,\delta,1}(\boldsymbol{h})$	υ	$\mathcal{M}_{\upsilon,1}(\boldsymbol{h})$	$SP(k)$
0	$(1-h)_+^{\delta}$	0.5	e^{-h}	0
1	$(1-h)_+^{\delta+1}(1+h(\delta+1))$	1.5	$e^{-h}(1+h)$	1
2	$(1-h)_+^{\delta+2}(1+h(\delta+2)+h^2(\delta^2+4\delta+3)\frac{1}{3})$	2.5	$e^{-h}(1+h+\frac{h^2}{3})$	2
3	$(1-h)_+^{\delta+3}(1+h(\delta+3)+h^2(2\delta^2+12\delta+15)\frac{1}{5}$ $+h^3(\delta^3+9\delta^2+23\delta+15)\frac{1}{15})$	3.5	$e^{-h}(1+\frac{h}{2}+h^2\frac{6}{15}+\frac{h^3}{15})$	3

Table 1.1 Generalized Wendland correlation $\mathcal{GW}_{\gamma,\delta,1}(\boldsymbol{h})$ and Matérn correlation $\mathcal{M}_{\upsilon,1}(\boldsymbol{h})$ with increasing smoothness parameters γ and υ. $SP(k)$ means that the sample paths of the associated Gaussian field are k times differentiable.

The model in (1.7) has the advantage to not depend on d. The flexibility in parameterizing the smoothness of the process by changing υ is the main reason why this family has been advocated as a default covariance model for most spatial applications ((Stein, 1999)).

Table 1.1 compares the Generalized Wendland correlation model $\mathcal{GW}_{\gamma,\delta,1}(\boldsymbol{h})$ for $\gamma = 0,1,2,3$ with the Matérn correlation model $\mathcal{M}_{\upsilon,1}(\boldsymbol{h})$ for $\upsilon = 0.5, 1.5, 2.5, 3.5$ with the associated degree of sample paths differentiability.

If the covariance function of a WS process is anisotropic, the spatial structure is directional dependent. Anisotropy is generally difficult to deal with but there are special cases that are tractable yet still interesting:

- Geometrical Anisotropy: variograms in two or more directions have different ranges, but the same sill value. That is, the variability of an observation is the same, but they are correlated over longer or shorter ranges, depending on the direction.

- Zonal Anisotropy: here the variogram has different sills and different ranges in two or more directions.

To correct for geometric anisotropy, a linear transformation of the coordinates is performed in order to reduce to an isotropic space. A linear transformation may correspond to rotation or stretching of the coordinate axes. Thus, in general if $\rho_0(\cdot,\boldsymbol{\psi})$ is an isotropic parametric covariance function, then

$$\rho(\boldsymbol{h},\boldsymbol{\psi}) = \rho_0(||\boldsymbol{A}\boldsymbol{h}||,\boldsymbol{\psi}),$$

is a geometrically anisotropic covariance function. Here $\boldsymbol{A}$ is a $d\times d$ matrix describing the linear transformation. There are different approach to face zonal anisotropy. Basically they are based on weighted linear combination of valid variograms (Rouhani and Hall, 1989). A special kind of anisotropy is attained considering variogram that are function of the sub-vectors $\boldsymbol{h} = (|h_1|, |h_2|)$, that is considering isotropy in the two main directions (Shapiro and Botha, 1991).

1.5 Geometric Properties

We are mainly interested in regularity properties such as continuity and differentiability of the sample paths of a random field. Nevertheless we shall also introduce the (weaker) concepts of mean square continuity and mean square differentiability as these are directly linked to the second-order structure and the smoothness properties of a random field.

1.5.1 Continuity

As there are different types to the convergence for random variables there are corresponding types of convergence for random fields. Three types of continuity will be considered based on almost sure convergence (*i.e.* convergence with probability one) and mean square convergence.

Definition 9. *Consider a* $D \subset \mathbb{R}^d$

(i) A random field $Z(\boldsymbol{s})$ *has continuous sample path with probability one in* D *if for every* $\boldsymbol{s}_0$ *for which* $||\boldsymbol{s}_0 - \boldsymbol{s}|| \to 0$ *as* $\boldsymbol{s} \to \boldsymbol{s}_0$*, then*

$$Pr(\omega : \ |Z(\boldsymbol{s}_0, \omega) - Z(\boldsymbol{s}, \omega)| \to 0 \quad as \quad \boldsymbol{s} \to \boldsymbol{s}_0 \quad for\ all\ \boldsymbol{s} \in D) = 1$$

(ii) A random field $Z(\boldsymbol{s})$ *is almost surely continuous in* D *if for every* $\boldsymbol{s}_0$ *for which* $||\boldsymbol{s}_0 - \boldsymbol{s}|| \to 0$ *as* $\boldsymbol{s} \to \boldsymbol{s}_0$*, then*

$$Pr(\omega : \ |Z(\boldsymbol{s}_0, \omega) - Z(\boldsymbol{s}, \omega)| \to 0 \quad as \quad \boldsymbol{s} \to \boldsymbol{s}_0) = 1 \quad for\ all\ \boldsymbol{s} \in D$$

(iii) A random field $Z(\boldsymbol{s})$ *is mean square continuous in* D *if for every* $\boldsymbol{s}_0$ *for which* $||\boldsymbol{s}_0 - \boldsymbol{s}|| \to 0$ *as* $\boldsymbol{s} \to \boldsymbol{s}_0$*, then*

$$E\{[Z(\boldsymbol{s}) - Z(\boldsymbol{s}_0)]^2\} \to 0 \quad as \quad \boldsymbol{s} \to \boldsymbol{s}_0 \quad for\ all\ \boldsymbol{s} \in D$$

Continuous sample paths with probability one means that there are, with probability one, no discontinuities within the whole domain D. Almost sure continuity however, allows discontinuities within D. There is no contradiction here; although a realization has discontinuities, the probability for finding a discontinuity at a particular position, say $\boldsymbol{s}$, is zero. Evidently, sample path continuity is a far stronger property than almost sure continuity. In general, mean square continuity is not implied by sample path continuity. Nor does mean square continuity imply sample path continuity. Sample path continuity is a far stronger condition relying on much more specific behavior of the covariance function.

For gaussian random fields however, mean square continuity is implied by the sufficient conditions for continuous sample paths given for the following theorems.

Theorem 3. *Let $Z(\boldsymbol{s})$ be a random field on $\mathbb{R}^d$. Then if, for some $c > 0$, some $\alpha > 0$, and some $\eta > \alpha$,*

$$E\{|Z(\boldsymbol{s}) - Z(\boldsymbol{s}+\boldsymbol{h})|^{\alpha}\} = \frac{c||\boldsymbol{h}||^{2d}}{|\log||\boldsymbol{h}|||^{1+\eta}}$$

then the random field $Z(\boldsymbol{s})$ will be continuous sample path with probability one over any compact set $D \in \mathbb{R}^d$.

For gaussian random fields this result can be sharpened. The following theorem is from Alder (1981).

Theorem 4. *Let $Z(\boldsymbol{s})$ be zero-mean, Gaussian random field with a continuous covariance function. Then if, for some $c < 0$ and some $\varepsilon > 0$,*

$$E\{|Z(\boldsymbol{s}) - Z(\boldsymbol{s}+\boldsymbol{h})|^{2}\} = \frac{c}{|\log||\boldsymbol{h}|||^{1+\varepsilon}}$$

then the random field $Z(\boldsymbol{s})$ will be continuous sample path with probability one.

Finally, if $Z(\boldsymbol{s})$ is a bounded process then almost sure continuity implies mean square continuity. It is easy to show that for a WS random field, mean square continuity at $\boldsymbol{s}$ implies that:

$$\lim_{\mathbf{h} \longrightarrow 0} E[Z(\boldsymbol{s}) - Z(\boldsymbol{s}+\boldsymbol{h})]^2 = 0$$

Thus, it is easily shown that for a WS random field mean square continuity is equivalent to the covariance function $\mathrm{C}(\boldsymbol{h})$ being continuous at 0. That is mean square continuity can be verified through the behavior of the covariance function near 0. As explained in section 1.4 some process appear to have a covariance for which $\mathrm{C}(\boldsymbol{h}) \longrightarrow \tau > 0$ as $\boldsymbol{h} \longrightarrow 0$, *i.e* the nugget effect.

Means square continuity by itself does not convey much about the smoothness of the process and how it is related to the covariance function. The smoothness concept is brought into focus by studying the partial derivatives of the random field and the introducing the mean square differentiability.

1.5.2 Differentiability

Consider Gaussian random fields $Z(\boldsymbol{s})$ on $\mathbb{R}^d$. Assume that Z has differentiable sample paths. Then the associated gradient field, $\dot{\boldsymbol{Z}}(\boldsymbol{s})$, is a space-vector in $\mathbb{R}^d$ defined by its components in a cartesian coordinate system:

$$\dot{Z}_i(\boldsymbol{s},\omega) = \frac{\partial Z(\boldsymbol{s})}{\partial \boldsymbol{s}_i} = \lim_{\Delta\to 0} \frac{Z(\boldsymbol{s}-\Delta\boldsymbol{e}_i,\omega) - Z(\boldsymbol{s},\omega)}{\Delta}$$

where ω is kept fixed and $\boldsymbol{e}_i$ is a unit vector in the i-th direction. A gradient field is commonly called a potential vector field and is characterized by zero curl. The components of $\dot{\boldsymbol{Z}}$ are also gaussian random fields since the different operator is lineal.

There are different forms of differentiability based on different forms the convergence.

Definition 10. *Consider a $D \subset \mathbb{R}^d$*

(i) A random field $Z(\boldsymbol{s})$ has differentiable sample path with probability one in D if for every $\boldsymbol{s}_0$ for which $||\boldsymbol{s}_0 - \boldsymbol{s}|| \to 0$ as $\boldsymbol{s} \to \boldsymbol{s}_0$, then

$$Pr(\omega:\ |\dot{Z}_i(\boldsymbol{s}_0,\omega) - \dot{Z}_i(\boldsymbol{s},\omega)| \to 0 \ \text{ as } \ \boldsymbol{s}\to\boldsymbol{s}_0 \ \text{ for all } \boldsymbol{s}\in D) = 1$$

(ii) A random field $Z(\boldsymbol{s})$ is almost surely differentiable in D if for every $\boldsymbol{s}_0$ for which $||\boldsymbol{s}_0 - \boldsymbol{s}|| \to 0$ as $\boldsymbol{s} \to \boldsymbol{s}_0$, then

$$Pr(\omega:\ |\dot{Z}_i(\boldsymbol{s}_0,\omega) - \dot{Z}_i(\boldsymbol{s},\omega)| \to 0 \ \text{ as } \ \boldsymbol{s}\to\boldsymbol{s}_0) = 1 \ \text{ for all } \boldsymbol{s}\in D$$

(iii) A random field $Z(\boldsymbol{s})$ is mean square differentiable in D if for every $\boldsymbol{s}_0$ for which $||\boldsymbol{s}_0 - \boldsymbol{s}|| \to 0$ as $\boldsymbol{s} \to \boldsymbol{s}_0$, then

$$E\{[\dot{Z}_i(\boldsymbol{s}) - \dot{Z}_i(\boldsymbol{s}_0)]^2\} \to 0 \ \text{ as } \ \boldsymbol{s}\to\boldsymbol{s}_0 \ \text{ for all } i=1,\ldots,n \ \text{ and } \ \boldsymbol{s}\in D$$

The discussion following definition 9 of continuity of random fields applies to differentiability simply by replacing continuity by differentiability.

Higher-order mean-square differentiability is then defined sequentially in the obvious way (that is $X(\boldsymbol{s})$ is twice mean-square differentiable if $X(\boldsymbol{s})$ is mean-square differentiable, and so on.)

As for mean square continuity, the significance of mean square differentiability is that it is a necessary condition for differentiable sample paths and it has a simple relation to the covariance function.

Theorem 5. *Consider a random field $Z(\boldsymbol{s})$ on $\mathbb{R}^d$ with covariance function C and differentiable expectation. If the derivative $\partial^2\mathrm{C}(\boldsymbol{s},\boldsymbol{t})/\partial \boldsymbol{s}_i \partial \boldsymbol{t}_i$ exists and is finite for all $i = 1,\ldots,n$ at the point $(\boldsymbol{s},\boldsymbol{s})$ then $Z(\boldsymbol{s})$ is mean square differentiability at $\boldsymbol{s}$. The covariance function the $\dot{Z}_i(\boldsymbol{s})$ is then given by $\partial^2\mathrm{C}(\boldsymbol{s},\boldsymbol{t})/\partial \boldsymbol{s}_i \partial \boldsymbol{t}_i$.*

For stationary random fields the theorem simplifies.

Corollary 1. *Consider a random field $Z(\boldsymbol{s})$ on $\mathbb{R}^d$ with covariance function C. If the derivative $\partial^2\mathrm{C}(\boldsymbol{h})/\partial h_i^2$ exists and is finite for all $i = 1,\ldots,n$ at the point $\mathbf{0}$ then $Z(\boldsymbol{s})$ is mean square differentiability at any $\boldsymbol{h}$. The covariance function the $\dot{Z}_i(\boldsymbol{s})$ is then given by $-\partial^2\mathrm{C}(\boldsymbol{h})/\partial h_i^2$.*

For isotropic covariance function this simplifies even further since all partial derivatives are equal; it is enough to consider $-d^2\mathrm{C}(\boldsymbol{h})/dh^2$.

Corollary 2. *Consider a random field $Z(\boldsymbol{s})$ on $\mathbb{R}^d$ with covariance function C and expectation possessing the necessary derivatives. If the derivative*

$$\frac{\partial^{2|k|}}{\partial s_1^{k_1}\cdots\partial s_n^{k_n}\partial t_1^{k_1}\cdots\partial t_n^{k_n}}\mathrm{C}(\boldsymbol{s},\boldsymbol{t}) \tag{1.9}$$

exists and is finite for all $i = 1,\ldots,n$ at the point $(\boldsymbol{s},\boldsymbol{s})$ then $Z(\boldsymbol{s})$ is $|k|$ times mean square differentiability at $\boldsymbol{s}$. The covariance function of

$$\frac{\partial^{|k|}Z(\boldsymbol{s})}{\partial s_1^{k_1}\cdots\partial s_n^{k_n}}$$

is the given by (1.9).

This means that the smoothness of the random field is related to the smoothness of the covariance function.

From this point of view it can be possible to make the selection of a particular correlation function based upon theoretical considerations. This possibility arises from the powerful fact that the choice of correlation function determines the smoothness of realizations from the spatial process. In this sense Stein (1999) and recommends the Matérn class as a general tool for building spatial models since the parameter υ can control the degree of smoothness.

1.6 Modeling spatial data

A common spatial process model is constructed as follows

$$Y(\boldsymbol{s}_i) = \mu(\boldsymbol{s}_i) + Z(\boldsymbol{s}_i) + \varepsilon(\boldsymbol{s}_i), \tag{1.10}$$

where $\mu(\boldsymbol{s}_i)$ is the mean of the response $Y(\boldsymbol{s}_i)$, typically of the form $X^T(\boldsymbol{s}_i)\boldsymbol{\beta}$. $X(\boldsymbol{s}_i)$ is a p-dimensional vector of explanatory variables at location $\boldsymbol{s}_i$ and $\boldsymbol{\beta}$ is a p-dimensional vector of parameters. $Z(\boldsymbol{s})$ is a zero mean spatial process ($Z(\boldsymbol{s})$ is often assumed to be a WS Gaussian process with a parametric covariance function and $\varepsilon(\boldsymbol{s})$ is a pure error process with mean 0 and variance nugget). Thus, the spatial signal is decomposed into a determinist trend, the pure spatial variability explained by the covariance function and a pure error process explained by the nugget.

The model (1.10) can be viewed as a hierarchical model with a conditionally independent first stage given $Z(\boldsymbol{s})$ and $\mu(\boldsymbol{s})$. In the second stage, usually we assume $Z(\boldsymbol{s})$ to be a Gaussian random field with mean zero and certain parametric covariance structure.

In some situations, the response variable $Y(\boldsymbol{s})$ (even after transformation) is not appropriate to be treated as a normal random variable. For instance, $Y(\boldsymbol{s})$ might be a binary variable or a count variable. It is natural to consider an extension of the model (1.10) analogous to the generalized linear model and considering at first stage the distribution of $Y(\boldsymbol{s}_i)$ conditionally $\boldsymbol{\beta}$ and $W(\boldsymbol{s}_i)$ belonging to the exponential family and consider $W(\boldsymbol{s}_i)$ as a WS process at second stage (Diggle et al., 1998).

1.7 Kriging

The main goal in spatial statistics is often interpolation. There exists different kinds of interpolators but kriging presents relevant advantages. For instance it provides some measure of the accuracy of the prediction with respect to deterministic interpolator such as splines or inverse distance method.

Kriging is a geostatistical interpolation technique that considers both the distance and the degree of variation between known data points when estimating values in unknown location points. A kriged estimate is a weighted linear combination of the known sample values around the point to be estimated. Applied properly, kriging allows the user to derive weights that result in optimal and unbiased estimates. Let us consider a random field $Z(\boldsymbol{s})$, $\boldsymbol{s} \in D \subset \mathbb{R}^d$ and the linear model

$$Z(\boldsymbol{s}) = \mu(\boldsymbol{s}) + \varepsilon(\boldsymbol{s}) \quad \boldsymbol{s} \in D, \tag{1.11}$$

where $\mu(\boldsymbol{s})$ is a deterministic function and $\varepsilon(\boldsymbol{s})$ is a WS random process. We observe the process at n different locations, $\boldsymbol{Z} = \{Z(\boldsymbol{s}_1), \dots Z, (\boldsymbol{s}_n)\}$, and wish to predict the process Z at an unobserved location $\boldsymbol{s}_0$. Let us denote with $p(\boldsymbol{s}_0, \boldsymbol{Z})$ the kriging interpolator at $\boldsymbol{s}_0$. The main properties of this object are:

- $p(\boldsymbol{s}_0, \boldsymbol{Z}) = \sum_i^n \lambda_i Z(\mathbf{s}_i)$.
- $E(p(\boldsymbol{s}_0, \boldsymbol{Z})) = \mu(\mathbf{s}_0)$.
- $E\{[Z(\boldsymbol{s}_0) - p(\boldsymbol{s}_0, \boldsymbol{Z})]^2\}$ is minimum.

That is, kriging interpolator is optimal in the class of linear interpolator, *i.e* it is unbiased and with minimum variance error. Relaxing first condition, it is easy to show that the best interpolator is the conditional expectation of $Z(\boldsymbol{s}_0)$ given the observed data:

$$p(\boldsymbol{s}_0, \boldsymbol{Z}) = E(Z(\boldsymbol{s}_0)|\boldsymbol{Z}) \tag{1.12}$$

However, in general $p(\boldsymbol{s}_0, \boldsymbol{Z})$ is not a linear function of the data and establishing the statistical properties of the best predictor under squared-error loss can be difficult. Fortunately, if the random field is Gaussian the best linear interpolator is also the best interpolator. Kriging appears in many forms and flavors, distinguished by whether the mean is known or not, what the distribution of is, whether predictions are made for points or areas and so forth. Here we describe classical ordinary kriging.

Let be $\mu(\boldsymbol{s}) = \mu$ and $\varepsilon(\boldsymbol{s})$ a zero mean random field with associated covariance matrix C, where μ is a unknown constant and C is known. We consider linear predictor of the form $p(\boldsymbol{s}_0, \boldsymbol{Z}) = \lambda_0 + \boldsymbol{\lambda}^T \boldsymbol{Z}$, where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N)$. Since we are looking for unbiased interpolator it is easy to show that it is equivalent to set $\lambda_0 = 0$ and $\sum_i^N \lambda_i = 1$.

Thus the problem now is to choose the $\boldsymbol{\lambda}$ weights that minimize:

$$E((\boldsymbol{\lambda}^T \boldsymbol{Z} - Z(\boldsymbol{s}_0))^2) \qquad \text{subject to} \quad \sum_i^N \lambda_i = 1.$$

This can be accomplished as an unconstrained minimization problem introducing the Lagrange multiplier m:

$$\underset{\boldsymbol{\lambda}}{\operatorname{argmin}} E((\boldsymbol{\lambda}^T \boldsymbol{Z} - Z(\boldsymbol{s}_0))^2) - 2m(\sum_i^N \lambda_i - 1). \tag{1.13}$$

It can be shown that (Cressie, 1993) the solution to this problem is:

$$\boldsymbol{\lambda}_{OK}^{opt} = \mathrm{C}^{-1}\left(\boldsymbol{c} + \mathbf{1}\frac{(1 - \mathbf{1}^T \mathrm{C}^{-1}\boldsymbol{c})}{\mathbf{1}^T \mathrm{C}^{-1}\mathbf{1}}\right)$$

$$m_{OK}^{opt} = \frac{1 - \mathbf{1}^T \mathrm{C}^{-1}\boldsymbol{c}}{\mathbf{1}^T \mathrm{C}^{-1}\mathbf{1}}.$$

Thus the optimal linear predictor is,

$$p(\boldsymbol{s}_0, \boldsymbol{Z})_{OK} = \hat{\mu} + \boldsymbol{c}^T \mathrm{C}^{-1}(\boldsymbol{Z} - \mathbf{1}\hat{\mu}),$$

and the minimized mean-square prediction error, *i.e*. the ordinary kriging variance is:

$$\sigma^2(\boldsymbol{s}_0)_{OK} = \mathrm{C}(\mathbf{0}) - [\boldsymbol{\lambda}_{OK}^{opt}]^T \boldsymbol{c} + m_{OK}^{opt}.$$

When μ is known ordinary kriging is called simple kriging. Cressie (1993) discusses more complicated versions, such as lognormal and trans-Gaussian kriging, and universal kriging, used in a presence of non-stationary mean field model.

Chapter 2

Space-temporal modelling

2.1 Spatio-Temporal processes

Environmental and geophysical processes such as atmospheric pollutant concentrations, precipitation fields and surface winds are characterized by spatial and temporal variability. In view of the prohibiting costs of spatially and temporally dense monitoring networks, one often aims to develop a statistical model in continuous space and time, based on observations at a limited number of monitoring stations. Examples include environmental monitoring and model assessment for surface ozone levels (Guttorp et al., 1994; Carroll et al., 1997; Meiring et al., 1998; Huang and Hsu, 2004) precipitation forecasts (Amani and Lebel, 1997) and the assessment of wind energy resources (Haslett and Raftery, 1989). Extending geostatistics to the spatio-temporal case implies that the observations

$$Z(\boldsymbol{s},t), \qquad (\boldsymbol{s},t) \in \mathbb{R}^d \times \mathbb{R}$$

indexed in space by $\boldsymbol{s} \in \mathbb{R}^d$ and in time by $t \in \mathbb{R}$, have been captured in the spatio-temporal domain. In keeping with the standards of classical geostatistics, the spatio-temporal distribution of the observations is modeled as a Gaussian distribution. This distribution will be perfectly characterized by defining its first and second-order moments, that is, its expectation and covariance function.

Definition 11. *A spatio-temporal random field $Z(\boldsymbol{s},t)$ is said to be Gaussian if the random vector $\mathbf{Z} = (Z(\boldsymbol{s}_1,t_1),\ldots,Z(\boldsymbol{s}_n,t_n))^T$, for any set of spatio-temporal locations $\{(\boldsymbol{s}_1,t_1),\ldots,(\boldsymbol{s}_n,t_n)\}$, follows a multivariate Gaussian distribution.*

Henceforth, we assume that second moments for the random function exist and are finite. Optimal least-squares prediction, or kriging, then relies on the appropriate specification of the space-time covariance structure. Generally, the covariance between $Z(\boldsymbol{s}_1,t_1)$ and $Z(\boldsymbol{s}_2,t_2)$ depends on the space-time coordinates $(\boldsymbol{s}_1,t_1)$ and $(\boldsymbol{s}_2,t_2)$, and no further structure may exist. In practice, however, estimation and modeling call for simplifying assumptions, such as stationarity, separability, and full symmetry. Specifically

Definition 12. *A spatio-temporal random field $Z(\boldsymbol{s},t)$ is said to have separable covariance if there exist purely spatial and purely temporal covariance functions Cov_S and Cov_T, respectively, such that*

$$Cov(Z(\boldsymbol{s}_1,t_1),Z(\boldsymbol{s}_2,t_2)) = Cov(Z(\boldsymbol{s}_1),Z(\boldsymbol{s}_2))Cov(Z(t_1),Z(t_2)) \tag{2.1}$$

for all space-time coordinates $(\boldsymbol{s}_1,t_1)$ and $(\boldsymbol{s}_2,t_2)$ in $\mathbb{R}^d \times \mathbb{R}$.

The spatio-temporal covariance structure factors into a purely spatial and a purely temporal component, which allows for computationally efficient estimation and inference. Consequently, separable covariance models have been popular even in situations in which they are not physically justifiable. Many statistical tests for separability have been proposed recently and are based on parametric models (Shitan and Brockwell, 1995; Guo and Billard, 1998; Brown et al., 2000), likelihood ratio tests and subsampling (Mitchell et al., 2005) or spectral methods (Scaccia and Martin, 2005; Fuentes, 2005b). A related notion is that of full symmetry (Gneiting, 2002b; Stein, 2005b).

Definition 13. *A spatio-temporal random field $Z(\boldsymbol{s},t)$ has fully symmetric covariance if*

$$Cov(Z(\boldsymbol{s}_1,t_1),Z(\boldsymbol{s}_2,t_2)) = Cov(Z(\boldsymbol{s}_1,t_2),Z(\boldsymbol{s}_2,t_1)) \tag{2.2}$$

for all space-time coordinates $(\boldsymbol{s}_1,t_1)$ and $(\boldsymbol{s}_2,t_2)$ in $\mathbb{R}^d \times \mathbb{R}$.

Atmospheric, environmental and geophysical processes are often under the influence of prevailing air or water flows, resulting in a lack of full symmetry. Transport effects of this type are well-known in the meteorological and hydrological literature and have recently been described by Gneiting (2002b), Stein (2005a) and De Luna and Genton (2005), who considered the Irish wind data of Haslett and Raftery (1989), by Wan et al. (2003) for wind power data, by Huang and Hsu (2004) for surface ozone levels and by Jun and Stein (2004) for atmospheric sulfate concentrations. Separability forms a special case of full symmetry.

Hence, covariance structures that are not fully symmetric are non-separable, and tests for full symmetry (Scaccia and Martin, 2005; Lu and Zimmerman, 2005) can be used to reject separability.

Frequently, trend removal and space deformation techniques (Haslett and Raftery, 1989; Sampson and Guttorp, 1992) allow for a reduction to a stationary covariance structure.

Definition 14. *A spatio-temporal random field $Z(\boldsymbol{s},t)$ has spatially stationary covariance if $Cov(Z(\boldsymbol{s}_1,t_1),Z(\boldsymbol{s}_2,t_2))$ depends on the observation sites $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ only through the spatial separation vector, $\boldsymbol{s}_1-\boldsymbol{s}_2$.*

Definition 15. *A spatio-temporal random field $Z(\boldsymbol{s},t)$ has temporally stationary covariance if $Cov(Z(\boldsymbol{s}_1,t_1),Z(\boldsymbol{s}_2,t_2))$ depends on the observation times t_1 and t_2 only through the temporal lag, t_1-t_2.*

Definition 16. *If a spatio-temporal random field $Z(\boldsymbol{s},t)$ has both spatially and temporally stationary covariance, we say that the process has stationary covariance. Under this assumption, there exists a function C defined on $\mathbb{R}^d\times\mathbb{R}$ such that*

$$Cov(Z(\boldsymbol{s}_1,t_1),Z(\boldsymbol{s}_2,t_2)) = \mathrm{C}(\boldsymbol{s}_1-\boldsymbol{s}_2,t_1-t_2) \tag{2.3}$$

for all $(\boldsymbol{s}_1,t_1)$ and $(\boldsymbol{s}_2,t_2)$ in $\mathbb{R}^d\times\mathbb{R}$.

We call C the space-time covariance function of the process, and its restrictions $\mathrm{C}(\cdot,0)$ and $\mathrm{C}(\mathbf{0},\cdot)$ are purely spatial and purely temporal covariance functions, respectively. For tests of stationarity we point to Fuentes (2005a) and references therein.

2.2 Geostatistical space-time models

2.2.1 Spatio-temporal domains

Geostatistical approaches have been developed to fit random function models in continuous space and time, based on a limited number of spatially and/or temporally dispersed observations. Hence, the natural domain for a geostatistical space-time model is $\mathbb{R}^d\times\mathbb{R}$, where $\mathbb{R}^d$ stands for space and $\mathbb{R}$ for time. Physically, there is clear-cut separation between the spatial and the time dimensions, and a realistic statistical model will take account thereof. This contrasts with a purely mathematical perspective in which $\mathbb{R}^d\times\mathbb{R}=\mathbb{R}^{d+1}$ with no differences between the coordinates. While the latter equality may seem (and indeed is) trivial, it has important implications. In particular, all technical results on spatial covariance functions or on least-square prediction, or kriging, in Euclidean spaces apply directly to space-time problems, simply by separating a vector into its spatial and temporal components.

Henceforth, we focus on covariances structures for spatio-temporal random functions $Z(\boldsymbol{s},t)$ where $(\boldsymbol{s},t) \in \mathbb{R}^d \times \mathbb{R}$. However, other spatio-temporal domains are also relevant in practice. Monitoring data are frequently observed at fixed temporal lags, and it may suffice to model a random function on $\mathbb{R}^d \times \mathbb{Z}$, with time considered discrete. The time autoregressive Gaussian models of Storvik et al. (2002) form a promising tool in this direction. A related result of Stein (2005a) characterizes space-time covariance functions that correspond to a temporal Markov structure. In atmospheric and geophysical applications, the spatial domain of interest is frequently expansive or global and the curvature of the earth needs to be taken into account. In this type of situation, random functions defined on $\mathbb{S}^d \times \mathbb{R}$ or $\mathbb{S}^d \times \mathbb{Z}$ become crucial, where $\mathbb{S}$ denotes a sphere in three-dimensional space. Perhaps the simplest way of defining a random field on $\mathbb{S}^d \times \mathbb{R}$ or $\mathbb{S}^d \times \mathbb{Z}$ is by defining a random function on $\mathbb{R}^3 \times \mathbb{R}$ and restricting it to the desired domain. Gneiting (1999), Stein (2005a,b) and particularly Jun and Stein (2004) discuss these and other ways of defining suitable parametric covariance models on global spatial or spatio-temporal domains.

2.2.2 Stationarity, separability and full symmetry

We consider a generic, typically Gaussian spatio-temporal random function $Z(\boldsymbol{s},t)$ where $(\boldsymbol{s},t) \in \mathbb{R}^d \times \mathbb{R}$. The covariance between $Z(\boldsymbol{s}_1,t_1)$ and $Z(\boldsymbol{s}_2,t_2)$ generally depends on the space-time coordinates $(\boldsymbol{s}_1,t_1)$ and $(\boldsymbol{s}_2,t_2)$, and no further structure may exist. In practice, simplifying assumptions are required, such as the aforementioned notions of stationarity, separability and full symmetry, which were defined in (2.1), (2.2) and (2.3), respectively. In addition, it is sometimes desirable that the space-time process Z has compactly supported covariance, so that $Cov(Z(\boldsymbol{s}_1,t_1),Z(\boldsymbol{s}_2,t_2) = 0$ whenever $||\boldsymbol{s}_1 - \boldsymbol{s}_2||$ and/or $||t_1 - t_2||$ are sufficiently large. Gneiting (2002a) reviews parametric models of compactly supported, stationary and isotropic covariance functions in a spatial setting. Unfortunately, the straightforward idea of thresholding covariances to zero using the product with an indicator function, such as the truncation device of (Haas, 2002, p. 320), yields invalid covariance models. That said, compactly supported covariances are attractive in that they allow for computationally efficient spatio-temporal prediction and simulation. Furrer et al. (2007), for instance, proposed their use for kriging large spatial datasets.

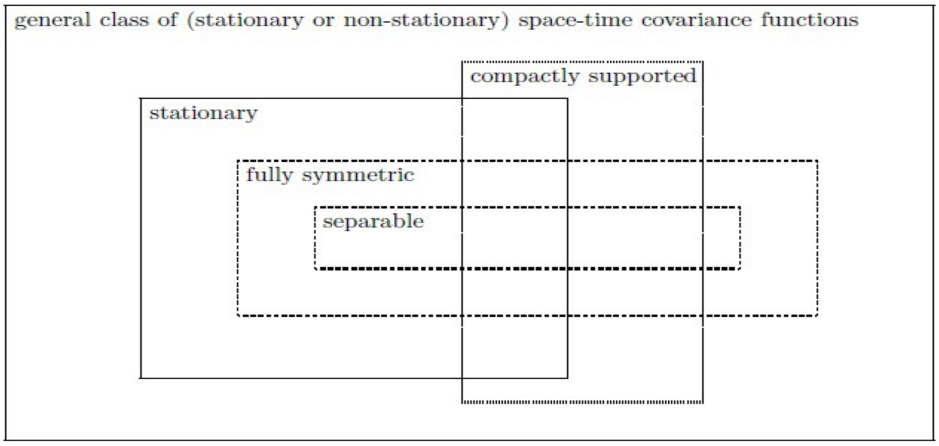


Fig. 2.1 Schematic illustration of the relationships between separable, fully symmetric, stationary and compactly supported covariances within the general class of (stationary or non-stationary) space-time covariance functions. An analogous scheme applies to correlation structures.

Figure 2.1 summarizes the relationships between the various notions in terms of classes of space-time covariance functions, and an analogous scheme applies to correlation structures. The largest class is that of general, stationary or non-stationary covariance functions. A separable covariance can be stationary or non-stationary, and similarly for fully symmetric covariances. However, a separable covariance functions is always fully symmetric, but not vice versa, and this has implications in testing and model fitting. In particular, to reject separability it suffices to reject full symmetry.

Occasionally, the second-oder structure of a spatio-temporal random function is modeled based on variances rather than covariances.

Definition 17. *The spatio-temporal non-stationary variogram is defined as the function*

$$2\gamma((\boldsymbol{s}_1,\boldsymbol{s}_2),(t_1,t_2)) = Var(Z(\boldsymbol{s}_1,t_1) - Z(\boldsymbol{s}_2,t_2)) \tag{2.4}$$

for all $(\boldsymbol{s}_1, t_1)$ *and* $(\boldsymbol{s}_2, t_2)$ *in* $\mathbb{R}^d \times \mathbb{R}$*. While half this quantity is called a semivariogram.*

Separability is not a meaningful assumption for variograms, but we discuss full symmetry and stationarity.

Definition 18. *A spatio-temporal random field* $Z(\boldsymbol{s}, t)$ *has fully symmetric variogram structure if*

$$Var(Z(\boldsymbol{s}_1, t_1) - Z(\boldsymbol{s}_2, t_2)) = Var(Z(\boldsymbol{s}_1, t_2) - Z(\boldsymbol{s}_2, t_1)) \tag{2.5}$$

for all locations $\boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathbb{R}^d$ *and times* $t_1, t_2 \in \mathbb{R}$*.*

Definition 19. *The space-time random function has spatially intrinsically stationary variogram if the non-stationary variogram (2.4) depends on the observation sites* $\boldsymbol{s}_1$ *and* $\boldsymbol{s}_2$ *only through the spatial separation vector,* $\boldsymbol{s}_1 - \boldsymbol{s}_2$*.*

Definition 20. *The space-time random function has temporally intrinsically stationary variogram if (2.4) depends on the observation times* t_1 *and* t_2 *only through the temporal lag,* $t_1 - t_2$*.*

Definition 21. *The process has intrinsically stationary variogram if it has both spatially intrinsically stationary and temporally intrinsically stationary variogram, i.e,*

$$2\gamma(\boldsymbol{s}_1 - \boldsymbol{s}_2, t_1 - t_2) = Var(Z(\boldsymbol{s}_1, t_1) - Z(\boldsymbol{s}_2, t_2))$$

for all $(\boldsymbol{s}_1, t_1)$ *and* $(\boldsymbol{s}_2, t_2)$ *in* $\mathbb{R}^d \times \mathbb{R}$*.*

Variograms exist under slightly weaker assumptions than covariances, and we refer to Gneiting et al. (2001) and Ma (2003b) for the various analogies and correspondences between the two classes of dependence measures. The use of variograms has been particularly popular in purely spatial problems. In discussing geostatistical space-time models, we follow the literature and focus our attention on covariances and correlations.

2.2.3 Positive definiteness

A crucial notion for stationary covariance functions is that of positive definiteness. Specifically, if $Z(\boldsymbol{s}, t)$ is a random function in $\mathbb{R}^d \times \mathbb{R}$, k is a positive integer, and $(\boldsymbol{s}_1, t_1), \ldots, (\boldsymbol{s}_k, t_k)$ are space-time coordinates in $\mathbb{R}^d \times \mathbb{R}$, the covariance matrix of the random vector

$Z(\boldsymbol{s}_1,t_1),\ldots,Z(\boldsymbol{s}_k,t_k))^T$ is nonnegative definite. If the process is stationary with covariance function C on $\mathbb{R}^d \times \mathbb{R}$, this matrix can be written as

$$(\mathrm{C}(\boldsymbol{s}_i - \boldsymbol{s}_j, t_i - t_j))_{i,j=1,\ldots,k} \tag{2.6}$$

A complex-valued function C on $\mathbb{R}^d \times \mathbb{R}$ is called *positive definite* if the matrix in (2.6) is nonnegative definite for all finite collections of space-time coordinates $((\boldsymbol{s}_1,t_1),\ldots,(\boldsymbol{s}_k,t_k)$ in $\mathbb{R}^d \times \mathbb{R}$. Any positive definite function is hermitian, and a positive definite function is real-valued if and only if it is symmetric. It is well known that the class of stationary covariance functions is identical to the class of symmetric positive definite functions, and we use the two terms interchangeably. When we talk of a positive definite function, we explicitly allow for complex-valued functions.

Unfortunately, it can be quite difficult in general to check whether a function is positive definite, and this forms one of the key difficulties in the construction of parametric space-time covariance models. Until a few years ago, geostatistical space-time models were largely based on stationary, separable covariance functions of the form

$$\mathrm{C}(\boldsymbol{h},u) = \mathrm{C}_S(\boldsymbol{h})\mathrm{C}_T(u), \quad (\boldsymbol{h},u) \in \mathbb{R}^d \times \mathbb{R} \tag{2.7}$$

where $\mathrm{C}_S(\boldsymbol{h})$ and $\mathrm{C}_T(u)$ are stationary, purely spatial and purely temporal covariance functions, respectively. A convenient choice is an isotropic model, $\mathrm{C}_S(\boldsymbol{h}) = \mathrm{c}_0(||\boldsymbol{h}||) = \sigma^2\rho(||\boldsymbol{h}||,\boldsymbol{\psi})$, $\boldsymbol{h} \in \mathbb{R}^d$ and $\boldsymbol{\psi} \in \mathbb{R}^p$, where c_0 is one of the standard models in geostatistics, such as the powered exponential class (see, for instance, (Diggle et al., 1998), the Whittle-Matérn class (Whittle, 1954; Matérn, 1986) and the Cauchy class (Gneiting and Schlather, 2004)). These parametric models are listed in section 1.4. Similarly, the temporal covariance function can conveniently be chosen as $\mathrm{C}_T(u) = \mathrm{c}_0(||u||)$, $u \in \mathbb{R}$, where c_0 is another, possibly distinct standard model that guarantees positive definiteness. The product (2.7) yields a positive definite function whenever C_S and C_T are positive definite on $\mathbb{R}^d$ and $\mathbb{R}$, respectively, because products (and also sums, convex combinations and limits) of positive definite functions are positive definite.

2.3 Stationary space-time covariance functions

2.3.1 Bochner's theorem

The celebrated theorem of Bochner (1955) states that a continuous function is positive definite if and only it is the Fourier transform of a finite nonnegative measure. This allows for the following characterization of stationary space-time covariance functions.

Theorem 6. *(Bochner's Theorem). Suppose that* C *is a continuous, bounded, integrable and symmetric function on* $\mathbb{R}^d \times \mathbb{R}$*. Then* C *is a stationary covariance if and only if*

$$\mathrm{C}(\boldsymbol{h},u) = \int\int e^{i(\boldsymbol{h}^T\boldsymbol{\omega}+u\tau)} dF(\boldsymbol{\omega},\tau), \quad (\boldsymbol{h},u) \in \mathbb{R}^d \times \mathbb{R} \tag{2.8}$$

where F is a finite, non-negative and symmetric measure on $\mathbb{R}^d \times \mathbb{R}$*.*

In other words, the class of stationary space-time covariance function on $\mathbb{R}^d \times \mathbb{R}$ is identical to the class of the Fourier transforms of finite, non-negative and symmetric measures on this domain. The measure F in the representation (2.8) is often called the spectral measure. If C is integrable, the spectral measure is absolutely continuous with Lebesgue density

$$f(\boldsymbol{\omega},\tau) = (2\pi)^{-(d+1)} \int\int e^{i(\boldsymbol{h}^T\boldsymbol{\omega}+u\tau)} \mathrm{C}(\boldsymbol{h},u) d\boldsymbol{h} du, \quad (\boldsymbol{\omega},\tau) \in \mathbb{R}^d \times \mathbb{R}$$

and f is called the *spectral density*. If the spectral density exists, the representation (2.8) in Bochner's theorem reduces to

$$\mathrm{C}(\boldsymbol{h},u) = \int\int e^{i(\boldsymbol{h}^T\boldsymbol{\omega}+u\tau)} f(\boldsymbol{\omega},\tau) d\boldsymbol{\omega} d\tau, \quad (\boldsymbol{h},u) \in \mathbb{R}^d \times \mathbb{R}$$

and C and f can be obtained from each other via the Fourier transform.

A stationary space-time covariance function is separable if there exist stationary, purely spatial and purely temporal covariance functions C_S and C_T, respectively, such that (2.7) holds or, equivalently, if we can factor the space-time covariance function as

$$\mathrm{C}(\boldsymbol{h},u) = \frac{\mathrm{C}(\boldsymbol{h},0)\mathrm{C}(\boldsymbol{0},u)}{\mathrm{C}(\boldsymbol{0},0)}$$

for all $(\boldsymbol{h},u) \in \mathbb{R}^d \times \mathbb{R}$ (Mitchell et al., 2005). In spectral terms, a stationary covariance function is separable if and only if the spectral measure factors as a product measure over the spatial and temporal domain, respectively. In particular, if the spectral density exists it factors as a product over the domains. A stationary space-time covariance function is fully symmetric if

$$\mathrm{C}(\boldsymbol{h},u) = \mathrm{C}(\boldsymbol{h},-u) = \mathrm{C}(-\boldsymbol{h},u) = \mathrm{C}(-\boldsymbol{h},-u)$$

for all $(\boldsymbol{h},u) \in \mathbb{R}^d \times \mathbb{R}$. In the purely spatial context, this property is also known as axial symmetry Scaccia and Martin (2005) or reflection symmetry Lu and Zimmerman (2005). For fully symmetric covariances, Bochner's theorem can be specialized as follows

Theorem 7. *Suppose that* C *is a continuous function on* $\mathbb{R}^d \times \mathbb{R}$*. Then* C *is a stationary, fully symmetric covariance if and only if it is of the form*

$$\mathrm{C}(\boldsymbol{h},u) = \int\int \cos(\boldsymbol{h}^T\boldsymbol{\omega})\cos(u\tau)dF(\boldsymbol{\omega},\tau), \quad (\boldsymbol{h},u) \in \mathbb{R}^d \times \mathbb{R} \tag{2.9}$$

where F is a finite, non-negative measure on $\mathbb{R}^d \times \mathbb{R}$.

If C is fully symmetric and the spectral density f exists, then f is fully symmetric, too, that is,

$$f(\boldsymbol{\omega},\tau) = f(\boldsymbol{\omega},-\tau) = f(-\boldsymbol{\omega},\tau) = f(-\boldsymbol{\omega},-\tau)$$

for all $(\boldsymbol{\omega},\tau) \in \mathbb{R}^d \times \mathbb{R}$. A similar characterization applies in terms of general spectral measures. If the space-time covariance function has additional structure, such as spherical symmetry with respect to $\boldsymbol{h} \in \mathbb{R}^d$ for each $u \in \mathbb{R}$, the representation (2.9) can be further specialized. Theorem 2 of Ma (2005) is a result of this type.

The above results apply to continuous functions. In practice, fitted stationary space-time covariance functions often involve a *nugget effect*, that is, a discontinuity at the origin. In the spatio-temporal context, the nugget effect could be purely spatial, purely temporal or spatio-temporal and takes the general form

$$\mathrm{C}(\boldsymbol{h},u) = a\delta_{(\boldsymbol{h},u)=(\mathbf{0},0)} + b\delta_{\boldsymbol{h}=\mathbf{0}} + c\delta_{u=0}, \quad (\boldsymbol{h},u) \in \mathbb{R}^d \times \mathbb{R} \tag{2.10}$$

where a, b and c are nonnegative constants and δ denotes an indicator function. Products and sums of continuous covariances and functions of the form (2.10) are valid covariances, and for all practical purposes functions of this type exhaust the class of valid stationary covariance functions (Gneiting and Sasvári, 1999).

2.3.2 Some non separable fully symmetric models

The following result of Cressie and Huang (1999) characterizes the class of stationary space-time covariance functions under the additional assumption of integrability.

Theorem 8. *(Cressie and Huang Theorem). Suppose that* C *is a continuous, bounded, integrable and symmetric function on* $\mathbb{R}^d \times \mathbb{R}$*. Then* C *is a stationary covariance if and only if*

$$\rho(\boldsymbol{\omega}, u) = \int e^{-i\boldsymbol{h}^T\boldsymbol{\omega}} \mathrm{C}(\boldsymbol{h}, u) d\boldsymbol{h}, \quad u \in \mathbb{R} \tag{2.11}$$

is positive definite for almost all $\boldsymbol{\omega} \in \mathbb{R}^d$*.*

Cressie and Huang (1999) used Theorem 8 to construct stationary space-time covariance functions through closed form Fourier inversion in $\mathbb{R}^d$. Specifically, they considered functions of the form

$$\mathrm{C}(\boldsymbol{h}, u) = \int e^{i\boldsymbol{h}^T\boldsymbol{\omega}} \rho(\boldsymbol{\omega}, u) d\boldsymbol{\omega}, \quad (\boldsymbol{h}, u) \in \mathbb{R}^d \times \mathbb{R}$$

where $\rho(\boldsymbol{\omega}, u)$, $u \in \mathbb{R}$ is a continuous positive definite function for all $\boldsymbol{\omega} \in \mathbb{R}^d$. Gneiting (2002b) gave a criterion that is based on this construction but does not depend on closed form Fourier inversion and does not require integrability. Recall that a continuous function $\varphi(r)$ defined for $r > 0$ or $r \geq 0$ is *completely monotone* if it possesses derivatives $\varphi^{(n)}$ of all orders and $(-1)^n \varphi^{(n)}(r) \geq 0$ for $r > 0$ and $n = 0, 1, 2, \ldots$. Gneiting (2002b) and Ma (2003a) gave various examples of completely monotone functions.

Theorem 9. *(Gneiting). Suppose that* $\varphi(r)$*,* $r \geq 0$*, is a completely monotone function, and that* $\phi(r)$*,* $r \geq 0$*, is a positive function with a completely monotone derivative. Then*

$$\mathrm{C}(\boldsymbol{h}, u) = \frac{1}{\phi(u^2)^{d/2}} \varphi\left(\frac{||\boldsymbol{h}||^2}{\phi(u^2)}\right), \quad (\boldsymbol{h}, u) \in \mathbb{R}^d \times \mathbb{R} \tag{2.12}$$

stationary covariance function on $\mathbb{R}^d \times \mathbb{R}$*.*

The specific choices $\varphi(r) = \sigma^2 \exp(-cr^\gamma)$ and $\phi(r) = (1 + ar^\alpha)^\beta$ recover Eq. (14) of Gneiting (2002b) and yield the parametric family

$$\mathrm{C}(\boldsymbol{h}, u) = \frac{\sigma^2}{(1 + a|u|^{2\alpha})^{\beta d/2}} \exp\left(-\frac{c||\boldsymbol{h}||^{2\gamma}}{(1 + a|u|^{2\alpha})^{\beta\gamma}}\right), \quad (\boldsymbol{h}, u) \in \mathbb{R}^d \times \mathbb{R} \tag{2.13}$$

of stationary space-time covariance functions. Here, a and c are nonnegative scale parameters of time and space, respectively. The smoothness parameters α and γ and the space-time interaction parameter β take values in $(0,1]$, and σ^2 is the variance of the spatio-temporal process. The purely spatial covariance function, $\mathrm{C}(\boldsymbol{h},0)$, is of the powered exponential form, and the purely temporal covariance function, $\mathrm{C}(\mathbf{0},u)$, belongs to the Cauchy class.

Clearly, any stationary covariance of the form (2.12) is fully symmetric. Furthermore, under the assumption of full symmetry the test functions $\rho(\boldsymbol{\omega},u)$ of Theorem 8 are real-valued and symmetric functions of $u \in \mathbb{R}$. If C is not fully symmetric then $\rho(\boldsymbol{\omega},u)$ is generally complex-valued. For instance, the function $\mathrm{C}(\boldsymbol{h},u) = \exp(-h^2 + hu - u^2)$ has Fourier transform proportional to $\exp\left(-\frac{1}{3}\left(\omega^2 + \omega\tau - \tau^2\right)\right)$ and therefore is a stationary covariance in $\mathbb{R} \times \mathbb{R}$. The associated test function $\rho(\omega,u)$, $u \in \mathbb{R}$ is proportional to $\exp\left(-\frac{1}{4}\left(3u^2 + 2i\omega u\right)\right)$ and positive definite yet generally complex-valued.

Non-separable, fully symmetric stationary space-time covariance functions can be constructed as mixtures of separable covariances. The following theorem summarizes relevant results by De Iaco et al. (2002) and Ma (2003a). In view of Theorem 7, the construction is completely general.

Theorem 10. *Let μ be a finite, nonnegative measure on a non-empty set Ψ. Suppose that for each $\boldsymbol{\psi} \in \Psi$, $\mathrm{C}_S^{\boldsymbol{\psi}}$ and $\mathrm{C}_T^{\boldsymbol{\psi}}$ are stationary covariances on $\mathbb{R}^d$ and $\mathbb{R}$, respectively, and suppose that $\mathrm{C}_S^{\boldsymbol{\psi}}(\mathbf{0})\mathrm{C}_T^{\boldsymbol{\psi}}(0)$ has finite integral over Θ. Then*

$$\mathrm{C}(\boldsymbol{h},u) = \int \mathrm{C}_S^{\boldsymbol{\psi}}(\boldsymbol{h})\mathrm{C}_T^{\boldsymbol{\psi}}(u)\,d\mu(\boldsymbol{\psi}), \quad (\boldsymbol{h},u) \in \mathbb{R}^d \times \mathbb{R} \tag{2.14}$$

is a stationary covariance function.

Explicit constructions of the form (2.14) have been reported by various authors. Perhaps the simplest special case is the product-sum model of De Iaco et al. (2001),

$$\mathrm{C}(\boldsymbol{h},u) = a_0\mathrm{C}_S^0(\boldsymbol{h})\mathrm{C}_T^0(u) + a_1\mathrm{C}_S^1(\boldsymbol{h}) + a_2\mathrm{C}_T^2(u), \quad (\boldsymbol{h},u) \in \mathbb{R}^d \times \mathbb{R}$$

where a_0, a_1 and a_2 are nonnegative coefficients and C_S^0, C_S^1 and C_T^0, C_T^2 are stationary, purely spatial and purely temporal covariance functions, respectively. Examples 1 and 2 of De Iaco et al. (2002) consider the particular case of (2.14) in which μ is a gamma distribution on $\Psi = [0,\infty)$ and both $\mathrm{C}_S^{\boldsymbol{\psi}}$ and $\mathrm{C}_T^{\boldsymbol{\psi}}$ are of powered exponential type. This construction yields the parametric family

$$\mathrm{C}(\boldsymbol{h},u) = \sigma^2 \left(1 + \left|\left|\frac{\boldsymbol{h}}{a}\right|\right|^{\alpha} + \left|\frac{u}{b}\right|^{\beta}\right)^{-\gamma}, \quad (\boldsymbol{h},u) \in \mathbb{R}^d \times \mathbb{R} \qquad (2.15)$$

of stationary space-time covariance functions, where $\alpha \in (0,2]$, $\beta \in (0,2]$, $\gamma > 0$, $a > 0$, $b > 0$ and $\sigma > 0$. De Iaco et al. (2002) gave more stringent parameter ranges which are unnecessarily restrictive. Ma (2005) reported various interesting examples of parametric space-time covariance functions that are also based on the mixture representation (2.14).

Chapter 3

Modeling spatial dependent data with Gamma and Weibull marginals

3.1 Random Fields with Gamma and Weibull marginals.

For a given weakly stationary spatial random field $F = \{F(\boldsymbol{s}), \boldsymbol{s} \in D \subset \mathbf{R}^d\}$, let us denote with $\rho_F(\boldsymbol{h}) = Corr(F(\boldsymbol{s}), F(\boldsymbol{s}+\boldsymbol{h}))$ its correlation function and with $\boldsymbol{F}_{ij} = (F(\boldsymbol{s}_i), F(\boldsymbol{s}_j))^T$ its bivariate distribution. In what follows we can equally use a weakly stationary spatial or a spatio-temporal random field $F = \{F(\boldsymbol{s},t), (\boldsymbol{s},t) \in D \subset \mathbf{R}^d \times \mathbf{R}\}$ with space-time correlation function $\rho_F(\boldsymbol{h},u) = Corr(F(\boldsymbol{s},t), F(\boldsymbol{s}+\boldsymbol{h}, t+u))$ and with bivariate distribution $\boldsymbol{F}_{iljk} = (F(\boldsymbol{s}_i, t_l), F(\boldsymbol{s}_j, t_k))^T$.

Nevertheless hereafter, for the sake of notation simplicity, we consider only the spatial case. Nevertheless, since our proposal is completely based on Gaussian random fields, the results presented in this thesis can be easily extended to the spatio-temporal case or spatial data defined on the sphere of radius r, *i.e.* $\mathbb{S}^d_r = \{\boldsymbol{s} \in \mathbb{R}^{d+1} : ||\boldsymbol{s}|| = r\}$. In these cases, space time covariances (Stein, 2005a; Gneiting, 2002c) or covariances defined on the sphere (Gneiting, 2013; Porcu et al., 2016) should be considered.

Let Z_i, $i = 1, \ldots \nu$ an independent finite sequence of zero mean and unit variance second order Gaussian random fields with common correlation function $\rho_Z(\boldsymbol{h})$ that we assume known up to a vector of parameters $\boldsymbol{\theta}$. We define the random process $U_\nu = \{U_\nu(\boldsymbol{s}), \boldsymbol{s} \in D\}$ as

$$U_\nu(\boldsymbol{s}) := \sum_{i=1}^{\nu} Z_i(\boldsymbol{s})^2/\nu. \tag{3.1}$$

Then U_ν is a weakly stationary random field with Gamma marginal distributions denoted by $U_\nu(\boldsymbol{s}) \sim \Gamma(\nu/2, \nu/2)$ for each $\boldsymbol{s}$, where the pairs $\nu/2$, $\nu/2$ are the shape and rate parameter, respectively, *i.e.* $E(U_\nu(\boldsymbol{s})) = 1$, $Var(U_\nu(\boldsymbol{s})) = 2/\nu$ and, using properties of the covariance operator, $\rho_{U_\nu}(\boldsymbol{h}) = \rho_Z^2(\boldsymbol{h})$.

A possible drawback for the $Gamma(\nu/2, \nu/2)$ distribution is that it is a limited model for continuous positive data due to the restrictions to the half integers for the shape parameter. A closer look to the problem reveals that overcoming this limitation is not easy.

The Laplace transform of the random vector $U_\nu(\boldsymbol{s}_1), \dots, U_\nu(\boldsymbol{s}_n)$ is given by (Vere-Jones, 1997,Proposition 4.3)

$$E\left(\exp\left(-\sum_{i=1}^{n}\alpha_i U_\nu(\boldsymbol{s}_i)\right)\right) = \frac{1}{|\boldsymbol{I} + (2/\nu)\boldsymbol{A}\boldsymbol{R}|^{\nu/2}}$$

where $\boldsymbol{I}$ is the $n \times n$ identity matrix, $\boldsymbol{A} = \text{diag}\{\alpha_1, \dots, \alpha_n\}$ and $\boldsymbol{R} = \{\rho_Z(\boldsymbol{s}_i, \boldsymbol{s}_j)\}$ is the correlation matrix of $Z(\boldsymbol{s}_1), \dots, Z(\boldsymbol{s}_n)$. Therefore the finite dimensional distribution of U_ν is a particular instance of the finite dimensional distribution of a β-permanental process Y (Eisenbaum and Kaspi, 2009) for which the Laplace transform satisfies

$$E\left(\exp\left(-\sum_{i=1}^{n}\delta_i Y(\boldsymbol{s}_i)\right)\right) = \frac{1}{|\boldsymbol{I} + \boldsymbol{D}\boldsymbol{B}|^{\beta}}$$

In the last formula β is any positive real, $\boldsymbol{B} = \{b_{ij}\}$ is a real $n \times n$ matrix and $\boldsymbol{D} = \text{diag}\{\delta_1, \dots, \delta_n\}$. The permanental process has Gamma marginal distributions denoted by $Gamma(\beta, 1/b_{ii})$.

So one may be tempted to identify the matrix $\boldsymbol{B}$ with a correlation matrix $\boldsymbol{R}$ that comes from the Gaussian process Z and, possibly, define a process with more flexible marginals. However this is true only for particular instances of Gaussian process Z since the random process $Z^2 = \{Z^2(\boldsymbol{s})\}$ has to be infinitely divisible, (Eisenbaum and Kaspi, 2006).

A necessary and sufficient condition for infinitely divisibility is that the inverse of the covariance matrix is a $\boldsymbol{M}$ matrix, (Marcus, 2014). For instance when $d = 1$ and the correlation model is exponential then the inverse of the covariance is of $\boldsymbol{M}$ type. In this case the shape parameter can be any positive value.

The multivariate distributions of U_ν can be expressed as an infinite series where each element of the series is a double sum of order n involving the computation of Laguerre polynomials and a determinant of a $n \times n$ square matrix. Analytical forms can be derived in some special cases (Royen, 1973). For estimating the unknown parameters, the knowledge of the bivariate distribution of a pair of observations $U_\nu(\boldsymbol{s}_i)$ and $U_\nu(\boldsymbol{s}_j)$ will be important (see Section 3.3). The bivariate distribution of $\boldsymbol{U}_{ij}$ corresponds to Kibble's bivariate Gamma distribution (Kibble, 1998) with density

$$f_{U_{ij}}(\boldsymbol{u}_{ij}) = \frac{2^{-\nu}\nu^{\nu}(u_i u_j)^{\nu/2-1} e^{-\frac{\nu(u_i+u_j)}{2(1-\rho_Z^2(\boldsymbol{h}))}}}{\Gamma\left(\frac{\nu}{2}\right)(1-\rho_Z^2(\boldsymbol{h}))^{\nu/2}} \left(\frac{\nu|\rho_Z(\boldsymbol{h})|\sqrt{u_i u_j}}{2(1-\rho_Z^2(\boldsymbol{h}))}\right)^{1-\nu/2} I_{\nu/2-1}\left(\frac{\nu|\rho_Z(\boldsymbol{h})|\sqrt{u_i u_j}}{(1-\rho_Z^2(\boldsymbol{h}))}\right) \quad (3.2)$$

where $I_\alpha(\cdot)$ denotes the modified Bessel function of the first kind of order α.

A non stationary random field with Gamma marginal can be obtained by considering the transformation

$$W_\nu(\boldsymbol{s}) := \mu(\boldsymbol{s}) U_\nu(\boldsymbol{s}) \quad (3.3)$$

where $\mu(\boldsymbol{s}) > 0$ is a deterministic function, leads to a multiplicative model and define a non stationary random field with Gamma marginal distributions denoted by $W_\nu(\boldsymbol{s}) \sim \Gamma(\nu/2, \nu/(2\mu(\boldsymbol{s})))$ for each $\boldsymbol{s}$, with $E(W_\nu(\boldsymbol{s})) = \mu(\boldsymbol{s})$, $Var(W_\nu(\boldsymbol{s})) = 2\mu(\boldsymbol{s})^2/\nu$ and $\rho_{W_\nu}(\boldsymbol{h}) = \rho_Z^2(\boldsymbol{h})$. Note that, under this parametrization, the mean does not depend on the shape parameter $\nu/2$ and the non-stationarity is due to the heteroscedasticity induced by the relation between the mean and the variance.

In order to consider parametric regression, we assume $log(\mu(\boldsymbol{s})) = X(\boldsymbol{s})^T\boldsymbol{\beta}$, where $X(\boldsymbol{s})$ is a k-dimensional vector of covariates and $\boldsymbol{\beta} \in \mathbb{R}^k$ a vector (unknown) parameter. The canonical link for Gamma distribution in the GLM framework is the inverse link which usually falls out of favour since it can produce negative prediction values. For this reason we adopt a more popular and more attractive alternative that is the the log link. In the following Theorem, we give the pdf of $\boldsymbol{W}_{ij}$ that can be easily obtained from (3.2).

Theorem 11. *Let $U_\nu(\boldsymbol{s}) = \sum_{i=1}^{\nu} Z_i(\boldsymbol{s})^2/\nu$ where Z_i $i = 1,\dots,\nu$ is a independent finite sequence of zero mean unit variance stationary Gaussian random fields with correlation function $\rho_Z(\boldsymbol{h})$ and let $W_\nu(\boldsymbol{s}) = \mu(\boldsymbol{s})U_\nu(\boldsymbol{s})$ with $\mu(\boldsymbol{s}) > 0$ a non random function. Then the pdf of $\boldsymbol{W}_{ij}$ is given by*

$$\begin{aligned} f_{\boldsymbol{W}_{ij}}(\boldsymbol{w}_{ij}) &= \frac{\left(\frac{\nu}{2(\mu(\boldsymbol{s}_i)\mu(\boldsymbol{s}_j))^{\frac{1}{2}}}\right)^{\nu} (w_i w_j)^{\nu/2-1} e^{-\frac{\nu}{2(1-\rho_Z^2(\boldsymbol{h}))}\left(\frac{w_i}{\mu(\boldsymbol{s}_i)}+\frac{w_j}{\mu(\boldsymbol{s}_j)}\right)}}{\Gamma\left(\frac{\nu}{2}\right)(1-\rho_Z^2(\boldsymbol{h}))^{\nu/2}} \\ &\times \left(\frac{\nu|\rho_Z(\boldsymbol{h})|\sqrt{w_i w_j}}{2(\mu(\boldsymbol{s}_i)\mu(\boldsymbol{s}_j))^{\frac{1}{2}}(1-\rho_Z^2(\boldsymbol{h}))}\right)^{1-\nu/2} I_{\nu/2-1}\left(\frac{\nu|\rho_Z(\boldsymbol{h})|\sqrt{w_i w_j}}{(\mu(\boldsymbol{s}_i)\mu(\boldsymbol{s}_j))^{\frac{1}{2}}(1-\rho_Z^2(\boldsymbol{h}))}\right) \quad (3.4) \end{aligned}$$

Proof. Under the transformation $u_i = w_i/\mu(\boldsymbol{s}_i)$ and $u_j = w_j/\mu(\boldsymbol{s}_j)$ in (3.2) with Jacobian $J((u_i,u_j) \to (w_i,w_j)) = (\mu(\boldsymbol{s}_i)\mu(\boldsymbol{s}_j))^{-1}$, the pdf of $\boldsymbol{W}_{ij}$ is given by (3.4). □

If $\rho_Z(\boldsymbol{h}) = 0$, then the bivariate distribution in Theorem 11 can be written as product of two independent random variables $\Gamma(\nu/2, \nu/2\mu(\boldsymbol{s}))$ that is, as in the Gaussian case, no pairwise correlation implies pairwise independence.

Note that when $\nu = 2$ in (3.1), U_2 has marginal distributions which are standard exponential distributions. In such case the bivariate distribution reduces to the Downton's distribution (Downton, 1970) with density

$$f_{\boldsymbol{U}_{ij}}(\boldsymbol{u}_{ij}) = \frac{e^{-\frac{(u_i+u_j)}{(1-\rho_Z^2(\boldsymbol{h}))}}}{(1-\rho_Z^2(\boldsymbol{h}))} I_0\left(\frac{2|\rho_Z(\boldsymbol{h})|\sqrt{u_i u_j}}{(1-\rho_Z^2(\boldsymbol{h}))}\right) \tag{3.5}$$

A random field with Weibull marginal distribution can be obtained by considering the transformation

$$R_\kappa(\boldsymbol{s}) = \frac{\mu(\boldsymbol{s})U_2(\boldsymbol{s})^{1/\kappa}}{\Gamma(1+1/\kappa)} \tag{3.6}$$

where $\mu(\boldsymbol{s}) > 0$ is a deterministic function and $\kappa > 0$ is a shape parameter. As in the previous case, the transformation (3.6) leads to a multiplicative model and define a non stationary random field with Weibull marginal distributions denoted by $R_\kappa(\boldsymbol{s}) \sim Weibull(\mu(\boldsymbol{s})/\Gamma(1+1/\kappa), \kappa)$ for each $\boldsymbol{s}$. Using this specific parametrization $E(R_\kappa(\boldsymbol{s})) = \mu(\boldsymbol{s})$ and $var(R_\kappa(\boldsymbol{s})) = \mu(\boldsymbol{s})^2(\Gamma(1+2/\kappa)\Gamma^{-2}(1+1/\kappa) - 1)$. As in the Gamma case the mean does not depend on the shape parameter and the non-stationarity is induced by the relation between the mean and the variance. We specify, as in the Gamma case, a parametric regression model for the mean assuming $log(\mu(\boldsymbol{s})) = X(\boldsymbol{s})^T\boldsymbol{\beta}$. In the following Theorem, we give the pdf of $\boldsymbol{R}_{ij}$ that can be easily obtained from (3.5).

Theorem 12. *Let $U_\nu(\boldsymbol{s}) = \sum_{i=1}^{\nu} Z_i(\boldsymbol{s})^2/2$ where Z_i $i = 1,\dots,\nu$ is a independent finite sequence of zero mean unit variance weakly stationary Gaussian random fields with common correlation function $\rho_Z(\boldsymbol{h})$ and let $R_\kappa(\boldsymbol{s}) = \lambda(\boldsymbol{s})U_2(\boldsymbol{s})^{1/\kappa}$ with $\lambda(\boldsymbol{s}) = \mu(\boldsymbol{s})/\Gamma(1+1/\kappa) > 0$ a non random function. Then the pdf of $\boldsymbol{R}_{ij}$ is given by*

$$f_{\boldsymbol{R}_{ij}}(\boldsymbol{r}_{ij}) = \frac{\kappa^2\Gamma^{2\kappa}\left(1+\frac{1}{\kappa}\right)(r_i r_j)^{\kappa-1} e^{-\frac{\Gamma^\kappa\left(1+\frac{1}{\kappa}\right)}{(1-\rho_Z^2(\boldsymbol{h}))}\left[\left(\frac{r_i}{\mu(\boldsymbol{s}_i)}\right)^\kappa + \left(\frac{r_j}{\mu(\boldsymbol{s}_j)}\right)^\kappa\right]}}{(\mu(\boldsymbol{s}_i)\mu(\boldsymbol{s}_j))^\kappa(1-\rho_Z^2(\boldsymbol{h}))} I_0\left(\frac{2|\rho_Z(\boldsymbol{h})|(r_i r_j)^{\kappa/2}\Gamma^\kappa\left(1+\frac{1}{\kappa}\right)}{(\mu(\boldsymbol{s}_i)\mu(\boldsymbol{s}_j))^{\kappa/2}(1-\rho_Z^2(\boldsymbol{h}))}\right) \tag{3.7}$$

Proof. Under the transformation $u_i = \Gamma^\kappa\left(1+\frac{1}{\kappa}\right)(r_i/\mu(\boldsymbol{s}_i))^\kappa$ and $u_j = \Gamma^\kappa\left(1+\frac{1}{\kappa}\right)(r_j/\mu(\boldsymbol{s}_j))^\kappa$ in (3.2) with Jacobian $J((u_i,u_j) \to (r_i,r_j)) = \kappa^2\Gamma^{2\kappa}\left(1+\frac{1}{k}\right)(r_i r_j)^{\kappa-1}/(\mu(\boldsymbol{s}_i)\mu(\boldsymbol{s}_j))^\kappa$, the pdf of $\boldsymbol{R}_{ij}$ is given by (3.7). □

If $\rho_Z(\boldsymbol{h}) = 0$, then the bivariate distribution in Theorem 12 can be written as product of two independent random variables $Weibull(\mu(\boldsymbol{s})/\Gamma\left(1+\frac{1}{\kappa}\right), \kappa)$ that is, as in the Gaussian case, no pairwise correlation implies pairwise independence.

3.2 Second-order and geometrical properties

Starting from the equation (3.2), it is easy to show that the correlation function of U_ν, $\rho_{U_\nu}(\boldsymbol{h})$, is equal to $\rho_Z^2(\boldsymbol{h})$ and U_ν and W_ν have the same correlation function. In order to study the correlation function of R_κ we need the following Lemma that gives the $(a,b)-th$ product moments of R_κ. The proof can be found in the Appendix.

Lemma 1. *The $(a,b)-th$ product moment $E(R^a(\boldsymbol{s}_i)R^b(\boldsymbol{s}_j))$ is given by*

$$E(R_\kappa^a(\boldsymbol{s}_i)R_\kappa^b(\boldsymbol{s}_j)) = \frac{\mu(\boldsymbol{s}_i)^a\mu(\boldsymbol{s}_j)^b}{\Gamma^{a+b}\left(1+\frac{1}{\kappa}\right)}\Gamma\left(1+\frac{a}{\kappa}\right)\Gamma\left(1+\frac{b}{\kappa}\right){}_2F_1\left(-\frac{a}{\kappa},-\frac{b}{\kappa};1;\rho_Z^2(\boldsymbol{h})\right) \tag{3.8}$$

where ${}_pF_q(a_1,a_2,\dots,a_p;b_1,b_2,\dots,b_p;x)$ is the generalized hypergeometric function Gradshteyn and Ryzhik (2007), defined by

$${}_pF_q(a_1,a_2,\dots,a_p;b_1,b_2,\dots,b_q;x) = \sum_{k=0}^{\infty}\frac{(a_1)_k,(a_2)_k,\dots,(a_p)_k}{(b_1)_k,(b_2)_k,\dots,(b_q)_k}\frac{x^k}{k!} \quad \text{for} \quad p,q = 0,1,2,\dots$$

with $(a)_k = \Gamma(a+k)/\Gamma(a)$ for $k \in \mathbb{N}\cup\{0\}$ being the Pochhammer symbol.

Using Lemma 1 we can obtain the correlation function of the Weibull random field.

Theorem 13. *Let $U_2(\boldsymbol{s}) = \sum_{i=1}^2 Z_i(\boldsymbol{s})^2/2$ where Z_i $i = 1,2$ are two independent zero mean unit variance weakly stationary Gaussian random fields with correlation function $\rho_Z(\boldsymbol{h})$ and let $R_\kappa(\boldsymbol{s}) = \mu(\boldsymbol{s})U_2(\boldsymbol{s})^{1/\kappa}/\Gamma(1+1/\kappa)$. Then the correlation function of R_κ is given by:*

$$\rho_{R_\kappa}(\boldsymbol{h}) = \frac{\Gamma^2\left(1+\frac{1}{\kappa}\right)}{\left[\Gamma\left(1+\frac{2}{\kappa}\right)-\Gamma^2\left(1+\frac{1}{\kappa}\right)\right]}\left[{}_2F_1\left(-\frac{1}{\kappa},-\frac{1}{\kappa};1;\rho_Z^2(\boldsymbol{h})\right)-1\right] \tag{3.9}$$

Proof. Since $E(R_\kappa(\boldsymbol{s})) = \mu(\boldsymbol{s})$, $Var(R_\kappa(\boldsymbol{s})) = \mu(\boldsymbol{s})^2(\Gamma(1+2/\kappa)\Gamma^{-2}(1+1/\kappa)-1)$ and setting $a = b = 1$ in Lemma 1, $\rho_{R_\kappa}(\boldsymbol{h})$ can be easily obtained. □

Unlike the Gamma case, the correlation of the Weibull random field depends on the shape parameter. From (3.6), we have the relation $\rho_{R_1}(\boldsymbol{h}) = \rho_{U_\nu}(\boldsymbol{h}) = \rho_Z^2(\boldsymbol{h})$. Also, can we show that $0 \le \rho_{W_\nu}(\boldsymbol{h}) \le |\rho_Z(\boldsymbol{h})|$ and $0 \le \rho_{R_\kappa}(\boldsymbol{h}) \le |\rho_Z(\boldsymbol{h})|$.

If $\mu(\boldsymbol{s}) = \mu$ both Gamma and Weibull are weakly stationary. Under this assumption, the following Theorem states that, the geometrical properties of the common latent Gaussian random field are strictly linked to the geometric properties of the Weibull and Gamma random fields (Gaetan and Guyon, 2009, Section 1.4). The proof can be found in the Appendix.

Theorem 14. *Let $U_\nu(\boldsymbol{s}) = \sum_{i=1}^{\nu} Z_i(\boldsymbol{s})^2/\nu$ where Z_i $i = 1,\dots,\nu$ is a independent finite sequence of zero mean unit variance weakly stationary Gaussian random fields with correlation function $\rho_Z(\boldsymbol{h})$ and let $W_\nu(\boldsymbol{s}) = \mu U_\nu(\boldsymbol{s})$ and let $R_\kappa(\boldsymbol{s}) = \mu U_2(\boldsymbol{s})^{1/\kappa}/\Gamma(1+1/\kappa)$. Then we have:*

a) W_ν and R_κ are also weakly stationary;

b) W_ν and R_κ are mean-square continuous if and only if Z is mean-square continuous;

c) W_ν and R_κ are k-times mean-square differentiable if Z is k-times mean-square differentiable

d) The sample paths of W_ν and R_κ are continuous and differentiable if the sample paths of Z are continuous and differentiable.

We exemplify some geometric features of the random field R_κ by means of the Matérn correlation function and Generalized Wendland correlation function defined in section 1.4.

Figures 3.1 and 3.2 show six realizations of the stationary random field R_κ on $D = [0,1]^2$ setting $\mu(\boldsymbol{s}) = 1$ and using the same seed for the pseudo-random generation.

We have firstly considered a Matérn correlation function $\mathscr{M}_{\upsilon,\alpha}(\boldsymbol{h})$ with three different parametrization for the smoothness parameter $\upsilon = 0.5, 1.5, 2.5$, the shape parameter $\kappa = 10, 3, 1$. The values of the range parameters, $\alpha = 0.034, 0.042, 0.067$, have been chosen in order to obtain a practical range equal to 0.2. The corresponding correlation function $\rho_{R_\nu}(\boldsymbol{h})$ are plotted (from left to right) in the center panel of Figure 3.1 and the bottom panel reports the histograms of the observations.

The same simulation experiment have been conducted for the Generalized Wendland correlation function $\rho_Z(\boldsymbol{h}) = \mathscr{GW}_{\gamma,\delta,\alpha}(\boldsymbol{h})$ (see Figure 3.2). Here we have set three different values of the parameters $\gamma = 0, 1, 2$ that correspond to the same degree of the smoothness of the Matérn correlation function and common values for $\alpha = 0.2$ and $\delta = 4.5$. The shape of the marginal distribution does not change with respect the previous experiment, *i.e.*, $\kappa = 10, 3, 1$.

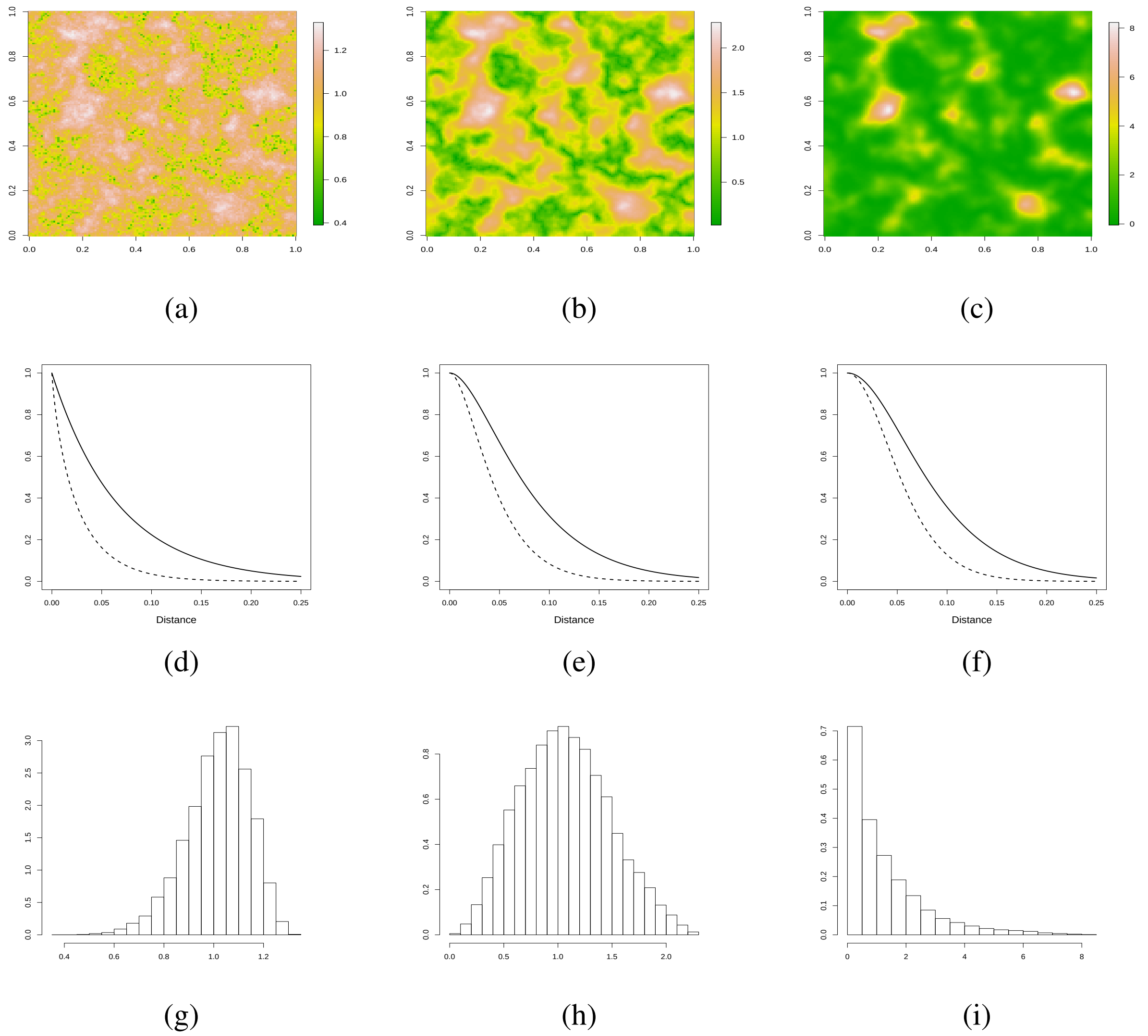


Fig. 3.1 Top: realizations of a stationary random process R_κ with $\mu(\boldsymbol{s}) = 1$ and correlation given under the image. Center: comparison between $\mathscr{M}_{\upsilon,\alpha}(\boldsymbol{h})$ (solid line) and the corresponding $\rho_{R_\kappa}(\boldsymbol{h})$ (dashed line) for different values of κ, υ, α and ($\kappa = 10, 3, 1$, $\upsilon = 0.5, 1.5, 2.5$ and $\alpha = 0.034, 0.042, 0.067$ from left to right). Bottom: histograms of the observations.

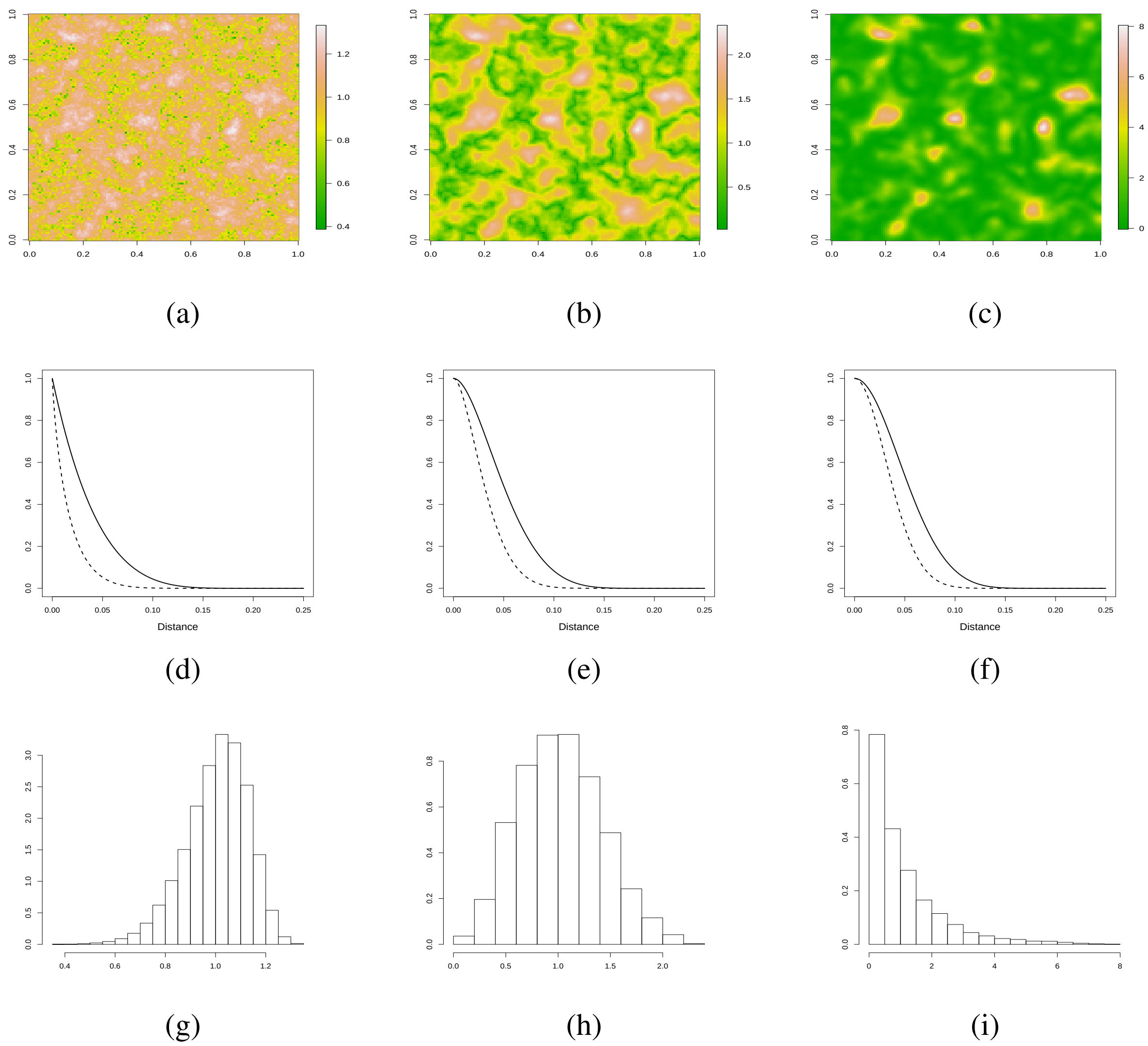


Fig. 3.2 Top: realizations of a stationary random process W with $\mu(s) = 1$ and correlation given under the image. Center: comparison between $\rho_Z(\boldsymbol{h}) = \mathcal{GW}_{\gamma,4.5,0.2}(\boldsymbol{h})$ (solid line) and the corresponding $\rho_{R_\kappa}(\boldsymbol{h})$ (dashed line) for different values of γ and κ ($\gamma = 0, 1, 2$, $\kappa = 10, 3, 1$ from left to right). Bottom: histograms of the observations.

It is apparent that the correlation $\rho_{R_\nu}(\boldsymbol{h})$ inherits the change of the differentiability at the origin from $\rho_Z(\boldsymbol{h})$ when increasing γ. This changes have consequences on the geometrical properties of the associated random processes. In fact the smoothness of the realizations increase with γ. Note also the flexibility of this kind of random process when modeling positive data since both positive and negative skewness can be achieved with different values of κ, see Figures 3.2-(g-i).

3.3 Estimation and prediction

3.3.1 Pairwise Likelihood-based inference

Given $\boldsymbol{w} = \{w(\boldsymbol{s}_1,t_1),\ldots,w(\boldsymbol{s}_N,t_T)\}^\top$ or $\boldsymbol{r} = \{r(\boldsymbol{s}_1,t_1),\ldots,r(\boldsymbol{s}_N,t_T)\}^\top$ a realization from a spatial ($T = 1$) or spatio-temporal ($T > 1$) non stationary spatio-temporal Gamma or Weibull random field W_ν and R_κ defined in Section 3.1, estimation of regression, shape and correlation parameters using maximum likelihood can be very challenging from computational point of view. For instance, in the Gamma case, the finite dimensional distribution can be expressed as an infinite series where each element of the series is a double sum of order NT involving the computation of Laguerre Polynomials and a determinant of a $NT \times NT$ square matrix (Krishnamoorthy and Parthasarathy, 1951; Krishnaiah and Rao, 1961).

In this case composite likelihood, and in particular pairwise likelihood (Lindsay, 1988; Varin et al., 2011), can be a useful computational device in order to estimate the parameters of both W_ν and R_κ combines the bivariate distributions of all possible distinct pairs of observations. For the sake of notation simplicity, we describe the method of pairwise likelihood in the spatial case ($T = 1$). Generalization to the space-time case can be easily obtained as in Bevilacqua et al. (2012). We define

$$\ell_{\boldsymbol{X}^*;ij}(\boldsymbol{\theta}) \equiv log(f_{\boldsymbol{X}^*_{ij}}(\boldsymbol{x}^*_{ij}),\boldsymbol{\theta}), \quad \boldsymbol{X}^* = \boldsymbol{W},\boldsymbol{R}$$

the loglikelihood associated to the bivariate distribution in equation (3.4) or (3.7). Here $\boldsymbol{\theta} = (\nu,\boldsymbol{\beta},\boldsymbol{\psi})^T$ in the Gamma case or $\boldsymbol{\theta} = (\kappa,\boldsymbol{\beta},\boldsymbol{\psi})^T$ in the Weibull case, $\boldsymbol{\beta}$ is the vector of regression parameters and $\boldsymbol{\psi}$ the vector of correlation parameters of the common correlation model of the latents Gaussian random fields in equation (3.1).

The pairwise weighted composite likelihood objective function is then given by:

$$\ell_{\boldsymbol{X}^*}(\boldsymbol{\theta}) = \sum_{i=1}^{N-1}\sum_{j=i+1}^{N} \ell_{\boldsymbol{X}^*;ij}(\boldsymbol{\theta})w^*_{ij}, \quad \boldsymbol{X}^* = \boldsymbol{W},\boldsymbol{R} \tag{3.10}$$

Then the maximum pairwise weighted composite likelihood estimator is given by $\widehat{\boldsymbol{\theta}}_{\boldsymbol{X}^*} = \text{argmax}_{\boldsymbol{\theta}\in A}\,\ell_{\boldsymbol{X}^*}(\boldsymbol{\theta})$ with $A \subseteq \Theta$ the parametric space. A distinctive feature of $\ell_{\boldsymbol{X}^*}(\boldsymbol{\theta})$ is that the associated estimating function, $\nabla\ell_{\boldsymbol{X}^*}(\boldsymbol{\theta})$ where ∇ denotes the vector differential operator, is unbiased and w^*_{ij} is a non-negative weight. Following the lines of Bevilacqua et al. (2012) and Bevilacqua and Gaetan (2015) it can be shown that, under increasing domain asymptotic, $\widehat{\boldsymbol{\theta}}_{\boldsymbol{X}^*}$ is consistent and asymptotically Gaussian with asymptotic covariance matrix given by

$$G^{-1}_{\boldsymbol{X}^*}(\boldsymbol{\theta}) = H^{-1}_{\boldsymbol{X}^*}(\boldsymbol{\theta})J_{\boldsymbol{X}^*}(\boldsymbol{\theta})H^{-1}_{\boldsymbol{X}^*}(\boldsymbol{\theta})^T, \quad \boldsymbol{X}^* = \boldsymbol{W},\boldsymbol{R} \tag{3.11}$$

where $G(\boldsymbol{\theta})$ is the *Godambe* information matrix and

$$H_{\boldsymbol{X}^*}(\boldsymbol{\theta}) = -E[\nabla^2\ell_{\boldsymbol{X}^*}(\boldsymbol{\theta})], \qquad J_{\boldsymbol{X}^*}(\boldsymbol{\theta}) = E[\nabla\ell_{\boldsymbol{X}^*}(\boldsymbol{\theta})\nabla\ell_{\boldsymbol{X}^*}(\boldsymbol{\theta})^T], \quad \boldsymbol{X}^* = \boldsymbol{W},\boldsymbol{R}$$

In addition, the Akaike information criterion, used for model selection, in the composite likelihood context assumes the following form:

$$CLIC_{\boldsymbol{X}^*} = -2\ell_{\boldsymbol{X}^*}(\boldsymbol{\theta}) + 2\text{tr}\{H_{\boldsymbol{X}^*}(\boldsymbol{\theta})G^{-1}_{\boldsymbol{X}^*}(\boldsymbol{\theta})\}, \quad \boldsymbol{X}^* = \boldsymbol{W},\boldsymbol{R} \tag{3.12}$$

where $\text{tr}(\cdot)$ is the trace of the matrix Varin and Vidoni (2005).

The evaluation of the standard error and information criteria requires consistent estimation of $G^{-1}_{\boldsymbol{X}^*}(\boldsymbol{\theta})$ and it can be obtained through the plug-in estimates $H_{\boldsymbol{X}^*}(\hat{\boldsymbol{\theta}}_{\boldsymbol{X}^*})$ and $J_{\boldsymbol{X}^*}(\hat{\boldsymbol{\theta}}_{\boldsymbol{X}^*})$. Nevertheless the latter becomes computationally unfeasible for large data sets since it is of order $O(N^4)$. In order to estimate $J_{\boldsymbol{X}^*}(\hat{\boldsymbol{\theta}}_{\boldsymbol{X}^*})$ we use a spatial subsampling method proposed in Heagerty and Lumley (2000).

Provided that $(W^*)^{-1}J_{\boldsymbol{X}^*}(\hat{\boldsymbol{\theta}})$ converges to a matrix $J^*_{\boldsymbol{X}^*}$ as $N \to \infty$, where $W^* = \sum_{(i,j)\in D} w^*_{ij}$ we use subsampling method in order to obtain an estimate $\hat{J}^*_{\boldsymbol{X}^*}$ of $J^*_{\boldsymbol{X}^*}$ and then estimate $J_{\boldsymbol{X}^*}(\hat{\boldsymbol{\theta}}_{\boldsymbol{X}^*})$ by $W^*\hat{J}^*_{\boldsymbol{X}^*}$. Given $S_1,\dots,S_f$ spatial subsets of the spatial observational region the estimator is given by:

$$\widehat{J}^*_{\boldsymbol{X}^*} = \frac{1}{f}\sum_{k=1}^{f}\frac{1}{(W^*)^{(k)}}\sum_{\substack{(i,j)\in S_k\\(i',j')\in S_k}}[\nabla\ell_{\boldsymbol{X}^*}(\hat{\boldsymbol{\theta}}_{\boldsymbol{X}^*})]_{ij}[\nabla\ell_{\boldsymbol{X}^*}(\hat{\boldsymbol{\theta}}_{\boldsymbol{X}^*})^T]_{i'j}w^*_{ij}w^*_{i'j'}, \quad \boldsymbol{X}^* = \boldsymbol{W},\boldsymbol{R} \tag{3.13}$$

where $(W^*)^{(k)} = \sum_{(i,j)\in S_k} w^*_{ij}$. The subsets are derived gathering the points that fall in a collection of (possibly) overlapping spatial subregions of the same shape of the region of observations but of smaller volume Lee and Lahiri (2002). Finally, the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}_{\boldsymbol{X}^*}$ can then be estimated using the subsampling approximation:

$$G_{\boldsymbol{X}^*}^{-1}(\hat{\boldsymbol{\theta}}_{\boldsymbol{X}^*}) = W^* H_{\boldsymbol{X}^*}(\hat{\boldsymbol{\theta}}_{\boldsymbol{X}^*})^{-1} \hat{J}^*_{\boldsymbol{X}^*} H_{\boldsymbol{X}^*}(\hat{\boldsymbol{\theta}}_{\boldsymbol{X}^*})^{-1}, \quad \boldsymbol{X}^* = \boldsymbol{W}, \boldsymbol{R}$$

and standard error estimation of each parameter is computed taking the square root of the diagonal elements of $G_{\boldsymbol{X}^*}^{-1}(\hat{\boldsymbol{\theta}}_{\boldsymbol{X}^*})$. If N is relatively small, parametric bootstrap can be a valid alternative for the estimation of the standard error (Varin et al., 2011).

The role of the weights in $\ell_{\boldsymbol{X}^*}(\boldsymbol{\theta})$ is to improve the statistical efficiency. A compactly supported weight function, *i.e.*

$$w^*_{ij} = \begin{cases} 1, & ||\boldsymbol{s}_j - \boldsymbol{s}_j|| \le d_s; \\ 0, & \text{otherwise} \end{cases} \tag{3.14}$$

has evident computational advantages. It has been shown in Joe and Lee (2009), Davis and Yau (2011) and Bevilacqua et al. (2012) that compactly supported simple or smooth weight functions allow to improve considerably the statistical efficiency of the pairwise likelihood estimation.

3.3.2 Linear prediction

We consider the problem of optimal linear prediction at some spatial $\boldsymbol{s}_0$ or spatio-temporal $(\boldsymbol{s}_0, t_0)$ location site. The computational issues highlighted in the previous subsection forestall the use of the conditional distributions for the prediction. Therefore we choose a suboptimal solution based on linear predictors. More precisely, we borrow the ideas from Monestiez et al. (2006), Oliveira (2014), and Bellier et al. (2010) for defining two linear predictors at some unobserved location $\boldsymbol{s}_0$ based on the observations at locations $\boldsymbol{s}_0, \ldots, \boldsymbol{s}_n$. As outlined in Oliveira (2014) the method uses a form of optimal linear prediction that mixes aspects of simple and ordinary kriging.

Let us consider a realization from the non-stationary Gamma or Weibull random field defined in Section 3.1 that is $\boldsymbol{x} = \{x(\boldsymbol{s}_1, t_1), \ldots, x(\boldsymbol{s}_N, t_T)\}^\top$, $\boldsymbol{x} = \boldsymbol{w}, \boldsymbol{r}$. Again we consider the spatial case ($T = 1$) for the sake of notation simplicity. If the mean $\mu(\boldsymbol{s}_i)$, $i = 1, \ldots, n$ is known then residuals in the form of ratios, rather than the usual differences, can be obtained as:

$$O_\nu(\boldsymbol{s}_i) = \frac{W_\nu(\boldsymbol{s}_i)}{\mu(\boldsymbol{s}_i)}, \quad O_\kappa(\boldsymbol{s}_i) = \frac{R_\kappa(\boldsymbol{s}_i)}{\mu(\boldsymbol{s}_i)}$$

where $\boldsymbol{O}_z = [O_z(\boldsymbol{s}_i)]_{i=1}^n$ with $z = \nu, \kappa$ is a realization from a stationary Gamma or Weibull random field with mean equal to 1. Note that the residuals have marginal distributions $Gamma(\nu/2, \nu/2)$ or $Weibull(\kappa, \Gamma^{-1}(1+1/\kappa))$, respectively.

Let us considered the family of predictors of the form:

$$\mathbf{F}(\boldsymbol{a}_z) = \{\mu(\boldsymbol{s}_0)\boldsymbol{a}_z^T\boldsymbol{O}_z : \quad \boldsymbol{a}_z \in \mathbf{R}^n, \quad \boldsymbol{a}^T\mathbf{1} = 1\} \tag{3.15}$$

where $\mathbf{1} = (1, \dots 1)^T$ and the goal is to find the best predictor (in mean square error sense) within this family under the contrasts $\boldsymbol{a}_z^T\mathbf{1} = 1$ with respect to $\boldsymbol{a}_z$. The solution is given by $\boldsymbol{a}_z^* = \Sigma_z^{-1}\left(\boldsymbol{c}_z + \frac{(1-\mathbf{1}\Sigma^{-1}\boldsymbol{c}_z)}{\mathbf{1}^T\Sigma_z^{-1}\mathbf{1}}\mathbf{1}\right)$ and the optimal predictor is $\mu(\boldsymbol{s}_0)\boldsymbol{a}_z^{*T}\boldsymbol{O}_x$ with associated $MSE(\boldsymbol{a}_z^*) = \mu^2(\boldsymbol{s}_0)(\sigma^2_{z;0,0} + [\boldsymbol{a}_z^*]^T\boldsymbol{c}_z + m^*)$, where $m^* = \frac{(1-\mathbf{1}^T\Sigma_z^{-1}\boldsymbol{c}_z)}{\mathbf{1}^T\Sigma_z^{-1}\mathbf{1}}$. Here $\Sigma_z = [\sigma_{z;ij}]_{i,j=1}^{n,n}$ and $\boldsymbol{c}_z = [\sigma_{z,0i}]_{i=1}^n$ where $\sigma_{z:ij}$ is the generic element of covariance matrix and $\sigma_{z,0i}$ is the generic element expressing the covariance between the data and the point to predict. Another unbiased predictor is given by considering the following simpler class:

$$\mathbf{S}(\boldsymbol{b}_z) = \{\mu(\boldsymbol{s}_0)(\boldsymbol{b}_z^T\boldsymbol{O}_z + k) : \quad \boldsymbol{b}_z \in \mathbf{R}^n,\ k \in \mathbf{R}\} \tag{3.16}$$

In this case the optimal linear unbiased predictor is given by $\mu(\boldsymbol{s}_0)(1 + \boldsymbol{c}_z^T\Sigma_z^{-1}(O_z - \mathbf{1}))$ with associated mean square error $\mu^2(\boldsymbol{s}_0)(\sigma^2_{z;00} - \boldsymbol{c}_z^T\Sigma_z^{-1}\boldsymbol{c}_z)$ where $\sigma^2_{z;00}$, $z = \nu, \kappa$ is the variance associated to the Gamma or Weibull random field.

In practice $\mu(\boldsymbol{s}_0)$, Σ_z and $\boldsymbol{c}_z$ are unknown and need to be estimated in order to compute the two predictors and the associated mean square errors. We propose to use $\widehat{\mu(\boldsymbol{s}_0)} = e^{X_0(\boldsymbol{s})^T\hat{\boldsymbol{\beta}}}$ and $\widehat{\sigma_{\nu;ij}} = 2\widehat{\mu(\boldsymbol{s}_i)^2}\rho_Z^2(d_{ij}, \hat{\psi})/\hat{\nu}$ in the Gamma case and

$$\widehat{\sigma_{\kappa;ij}} = \widehat{\mu(\boldsymbol{s}_i)^2}\left[{}_2F_1\left(-\frac{1}{\hat{\kappa}}, -\frac{1}{\hat{\kappa}}; 1; \rho_Z^2(d_{ij}; \hat{\boldsymbol{\psi}})\right) - 1\right]$$

in the Weibull case where $d_{ij} = ||\boldsymbol{s}_i - \boldsymbol{s}_j||$, and $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\psi}}$, $\hat{\nu}$, $\hat{\kappa}$ are pairwise likelihood estimates.

Similarly we can obtain an estimation $\widehat{\boldsymbol{c}}_z$ for $\boldsymbol{c}_z$. A simulation study with Poisson data in Oliveira (2014) show that the two predictors are practically equivalent in terms of bias and mean square prediction error.

3.4 Numerical results

In this section, we start by describing the performance of the weighted pairwise composite likelihood method of estimation when estimating non stationary gamma and Weibull random fields both in the spatial and spatio-temporal setting. Then we apply our methodology to the analysis of Ireland wind speed data Haslett and Raftery (1989).

3.4.1 Simulation study

In this subsection we study the finite sample properties of the weighted pairwise composite likelihood method described in Section 3.3, when estimating the non stationary Gamma and Weibull random fields described in Section 3.1. We deal with both spatial and spatio-temporal setting. Specifically:

1. We simulate, using Cholesky decomposition, 500 realization of a Gamma and Weibull random fields on $N = 1000$ location sites uniformly distributed in the unit square and $k = 2$ that is we set $\boldsymbol{X\beta}$ with $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$, $\beta_0 = 0.5$, $\beta_1 = 0.25$ where $\boldsymbol{X}$ is $N \times 2$ matrix with first column the unit vector and the second column an iid realization from a $U(0,1)$. Then we fix $\nu = 3,6,9$ for the Gamma case and $\kappa = 0.5,3,10$ for the Weibull case as shape parameters. As isotropic parametric correlation model we consider special cases of Matérn and Generalized Wendland class in equations (1.7) and (1.8)

$$\rho(\boldsymbol{h}) = \mathscr{M}_{0.5,\alpha}(\boldsymbol{h}) = e^{-||\mathbf{h}||/\alpha} \tag{3.17}$$

$$\rho(\boldsymbol{h}) = \mathscr{GW}_{0,4,\alpha}(\boldsymbol{h}) = (1 - ||\mathbf{h}||/\alpha)_+^4 \tag{3.18}$$

with $\alpha = 0.15/3$ in the first model and $\alpha = 0.15$ in the second model. In the pairwise likelihood estimation we consider a cut off weight function as in (3.14) setting $d_s = 0.1$. Table 3.1 and 3.2 show bias and mean square error (MSE) associated to regression, shape and correlation parameters. In the gamma case, the shape parameters are assumed to be known. In appendix B we show a code example the simulation and estimating spatial Weibull random field with Generalized Wendland correlation function and $\tau = 0$

2. We simulate, using Cholesky decomposition, 500 realization of a Gamma and Weibull random fields on $N = 100$ location sites uniformly distributed in the unit

square, $t_1 = 1, t_2 = 2, \dots, t_{20} = 20$ *i.e.* $T = 20$ temporal instants and $k = 2$ that is we set $\boldsymbol{X\beta}$ with $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$, $\beta_0 = -1$, $\beta_1 = 2$ where X is $NT \times 2$ matrix with first column the unit vector and the second column an iid realization from a $U(0,1)$. Then we fix $\nu = 3, 6, 9$ for the Gamma case and $\kappa = 0.5, 3, 10$ for the Weibull case as in Scenario 1. As isotropic in space and symmetric in time non-separable correlation models we consider two non separable models,

$$\rho(\boldsymbol{h},u) = \frac{1}{(1+|u|/\alpha_t)^{2.5}} \mathscr{M}_{0.5,\alpha_s(1+|u|/\alpha_t)^{\frac{\tau_1}{2}}}(\boldsymbol{h}) = \frac{1}{(1+|u|/\alpha_t)^{2.5}} exp\left\{-\frac{||\mathbf{h}||}{\alpha_s(1+|u|/\alpha_t)^{\frac{\tau_1}{2}}}\right\} \tag{3.19}$$

$$\rho(\boldsymbol{h},u) = \frac{1}{(1+|u|/\alpha_t)^{2.5}} \mathscr{GW}_{0,4,\alpha_s(1+|u|/\alpha_t)^{\tau_2}}(\boldsymbol{h}) = \frac{1}{(1+|u|/\alpha_t)^{2.5}} \left(1-\frac{||\mathbf{h}||}{\alpha_s(1+|u|/\alpha_t)^{\tau_2}}\right)_+^4 \tag{3.20}$$

with $\boldsymbol{h} = \boldsymbol{s}_i - \boldsymbol{s}_j$, $u = t_i - t_j$, $\alpha_s > 0$, $\alpha_t > 0$ and where $\tau_i \in [0,1]$ with $i = 1,2$ are non-separability parameters. We set $\alpha_s = 0.15/3$ in the first model and $\alpha_s = 0.15$ in the second model $\alpha_t = 1$, $\tau_1 = 1$ and $\tau_2 = 0.5$

This first is a special case of a more general class proposed in Gneiting (2002c) that we call Gneting-exponetial model and the second is a special case of a more general model proposed in Gneiting (2013) that we call Gneting-Askey model. Both can be viewed as a generalization to the space time context of the spatial covariance model in equations (3.17) and (3.18). In the pairwise likelihood estimation we consider a a space time version of the weight function as in (3.14) setting $d_s = 0.1$, $d_t = 2$, which bypasses an explosion of the number of pairwise terms and shifts focus to relatively short-range distances where dependence matters most. Tables 3.3 and 3.4 show bias and MSE associated to regression, shape and correlation parameters. Also in this case, the shape parameters are assumed to be known in the gamma case.

As general comment the estimates are overall approximatively unbiased for both spatial and the spatio-temporal Scenarios (see tables 3.1 to 3.4). MSE is affected by the magnitude of the shape parameter for both the Gamma and Weibull case. Specifically MSE seems to decreass when the shape parameter increase. Therefore, the precision of estimates of the regression parameters benefit of the increment of the shape parameter. Finally not surprising is the fact that the temporal parameter α_t is worst estimated since we have less observations along the temporal dimension.

With respect to the distribution of the weighted pairwise likelihood estimates, the figure 3.3, as an example, depicts the boxplots of the estimates for a Weibull random field with $\beta_0 = -1$, $\beta_1 = 2$, $\kappa = 3$ using a Gneiting-Askey correlation model with $\alpha_s = 0.15$ and $\alpha_t = 1$. It is apparent that, the distribution of the estimates are quite symmetric numerically stable with very few outliers.

ν	3				6				9			
	Exp		Askey		Exp		Askey		Exp		Askey	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\hat{\beta}_0$	−0.00056	0.00524	6.21e-05	0.00369	−0.00033	0.00251	0.00030	0.00178	0.00140	0.00179	0.00131	0.00120
$\hat{\beta}_1$	−0.00147	0.00686	−0.00113	0.00739	−0.00127	0.00364	−0.00122	0.00373	−0.00174	0.00272	−0.00224	0.00244
$\hat{\alpha}$	−0.00093	3.93e-05	−0.00176	0.00023	−0.00100	3.37e-05	−0.00162	0.00022	−0.00082	3.45e-05	−0.00074	0.00021

Table 3.1 Gamma random field estimation with pairwise likelihood: bias and mean square error for regression and spatial scale parameters under Scenario 1 using Exponential and Askey correlation model for $\nu = 3, 6, 9$.

κ	0.5				3				10			
	Exp		Askey		Exp		Askey		Exp		Askey	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\hat{\beta}_0$	−0.00063	0.03053	0.00216	0.02254	0.00076	0.00087	0.00080	0.00063	0.00029	8.32e-05	0.00027	6.02e-05
$\hat{\beta}_1$	−0.00774	0.04057	−0.00838	0.04342	−0.00128	0.00113	−0.00140	0.00121	−0.00040	0.00010	−0.00042	0.00011
$\hat{\alpha}$	−0.00078	5.12e-05	−0.00108	0.00030	−0.00078	5.11e-05	−0.00108	0.00030	−0.00081	5.05e-05	−0.00108	0.00030
$\hat{\kappa}$	0.00189	0.00028	0.00114	0.00022	0.01133	0.00995	0.00682	0.00784	0.03899	0.10979	0.02273	0.08709

Table 3.2 Weibull random field estimation with pairwise likelihood: bias and mean square error for regression, shape and spatial scale parameters under Scenario 1 using Exponential and Askey correlation model for $\kappa = 0.5, 3, 10$.

ν	3				6				9			
	Exp		Askey		Exp		Askey		Exp		Askey	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\hat{\beta}_0$	0.00046	0.00172	0.00026	0.00161	0.00054	0.00082	−0.00081	0.00078	−0.00040	0.00057	9.99e-05	0.00049
$\hat{\beta}_1$	−0.00225	0.00427	0.00040	0.00453	−0.00117	0.00208	0.00149	0.00220	0.00027	0.00155	0.00033	0.00144
$\hat{\alpha}_s$	−0.00026	2.38e-05	−0.00052	0.00027	-3.10e-05	2.28e-05	−0.00099	0.00025	5.72e-05	2.33e-05	−0.00069	0.00023
$\hat{\alpha}_t$	−0.11850	0.23182	−0.10210	0.18942	−0.11669	0.23601	−0.09286	0.18515	−0.11802	0.23227	−0.08549	0.17330

Table 3.3 Gamma random field estimation with pairwise likelihood: bias and mean square error for regression and spatial scale parameters under Scenario 2 using Gneiting-Exponential and Gneiting-Askey correlation model for $\nu = 3, 6, 9$.

κ	0.5				3				10			
	Exp		Askey		Exp		Askey		Exp		Askey	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\hat{\beta}_0$	0.00080	0.01037	−0.00612	0.00997	0.00048	0.00030	−0.00069	0.00029	0.00019	2.66e-05	−0.00020	2.67e-05
$\hat{\beta}_1$	−0.00483	0.02434	0.00506	0.02670	−0.00083	0.00068	0.00085	0.00074	−0.00028	5.70e-05	0.00028	6.58e-05
$\hat{\alpha}_s$	-6.6e-06	2.46e-05	−0.00061	0.00027	-3.31e-06	2.46e-05	−0.00061	0.00027	4.80e-05	2.49e-05	−0.00062	0.00027
$\hat{\alpha}_t$	−0.11145	0.23587	−0.10852	0.20146	−0.11047	0.23524	−0.11199	0.20656	−0.16984	0.30154	−0.15047	0.24689
$\hat{\kappa}$	0.00070	9.09e-05	0.00069	9.19e-05	0.00418	0.00327	0.00410	0.00331	0.01370	0.03651	0.01369	0.03670

Table 3.4 Weibull random field estimation with pairwise likelihood: bias and mean square error for regression, shape and spatial scale parameters under Scenario 2 using Gneiting-Exponential and Gneiting-Askey correlation model for $\kappa = 0.5, 3, 10$.

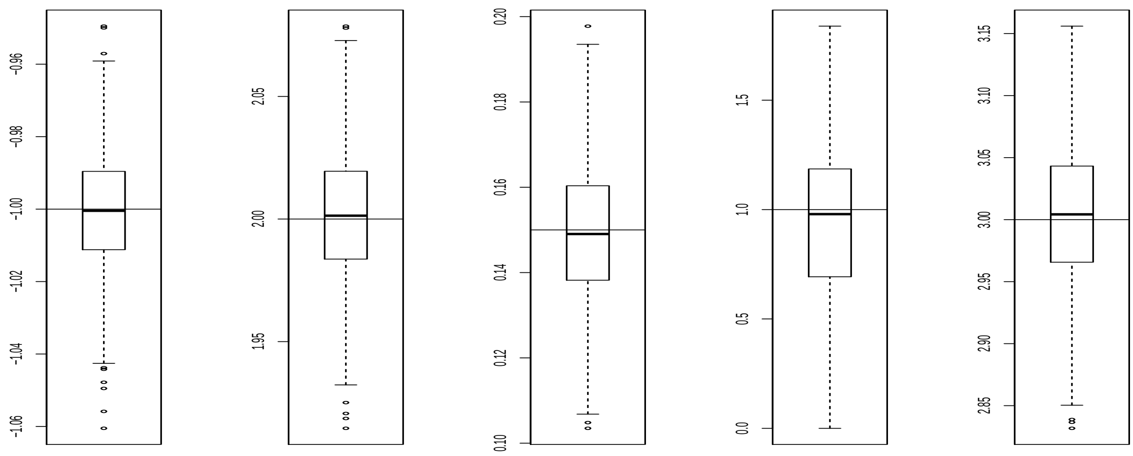


Fig. 3.3 Boxplots of weighted pairwise likelihood estimations for parameters $\beta_0 = -1, \beta_1 = 2, \alpha_s = 0.15, \alpha_t = 1, \kappa = 3$ (from left to right) under Scenario 2 when estimating a Weibull random field using an Gneiting-Askey correlation model.

3.4.2 Irish wind speed data

In this subsection we consider the dataset of the daily average wind speeds $y(\boldsymbol{s}_i,t)$, $i=1,\dots,12$, $t=1,\dots,6574$, observed for a period $1961-1978$, at twelve synoptic meteorological stations in the Republic of Ireland. The dataset can be obtained from the R package `gstat`. A in-depth analysis has been carried out for the first time by Haslett and Raftery (1989) and later the data were considered in several papers (see for instance Gneiting (2002b), Stein et al. (2004), Stein (2005a) and De Luna and Genton (2005)). The original time series are positive, except for few zeroes, the marginal distributions are noticeably asymmetric and the standard deviations are correlated with means that very over seasons.

For these reason the previous statistical analysis have suggested to preliminary transform the data and then extract a common periodic component. These transformations yield to approximately stationary time-series, with bell-shaped marginal distributions, that can be modelled by a Gaussian random fields. Note that in the aforementioned papers Rosslare's station was removed for stationarity reasons, as recommended by Haslett and Raftery (1989). Nevertheless we consider it in our analysis since our approach takes into account possible non stationarity.

The statistical analysis of this kind of data is challenging for two reasons. First, data are clearly not Gaussian and second reason is the size of the dataset ($78,888$ observations) can be computationally demanding. Weighted pairwise likelihood is generally an appealing estimation method when dealing with large data set. In order to further reduce the computational burden associate to the estimation task, we consider an implementation based on OpenCL in a GPGPU (General Purpose Graphical processing unit) framework, as a part of `GeoModels` an upcoming R package.

In order to asses the predictive performances of the different models we split the data set into training period consisting of year $1961-1970$, with $n_{tr}=3289$ days, and a test period for the remaining years, $n_{te}=3285$ days.

Differently from the previous attempts, we adapt the W_ν and R_κ random fields using the original data and replacing the sixteen zeros with the mean speed of the previous and next day. The seasonal variations in the level and in the variability are modelled by the common trend

$$\log\{\mu(\boldsymbol{s},t;\boldsymbol{\beta})\}=\beta_0+\sum_{l=1}^{3}\beta_{1,l}\cos\left(\frac{2\pi lt}{365}\right)+\beta_{2,l}\sin\left(\frac{2\pi lt}{365}\right)$$

Following Gneiting (2002b) we restrict our attention to the short distance and short term dependence using the correlation of the Gneiting-Exponential type

$$\rho(\boldsymbol{h},u;\boldsymbol{\psi}) = \frac{1}{(1+|u|/\alpha_t)^{1.5}}\exp\left\{-\frac{||\boldsymbol{h}||}{\alpha_s(1+|u|/\alpha_t)^{\frac{\tau_1}{2}}}\right\} \tag{3.21}$$

where $\boldsymbol{\psi} = (\alpha_s, \alpha_t, \tau_1)$, with different values of $\tau_1 = 0, 0.5, 1$. Note that the space-time separability, *i.e.* $\rho(\boldsymbol{h},u) = \rho_S(\boldsymbol{h})\rho_T(u)$ for two correlation functions ρ_S and ρ_T, is inherited only by the W_ν random field.

A natural competitor of our models is the Log-Gaussian random field $L = \{L(\boldsymbol{s},t)\}$ namely $L(\boldsymbol{s},t) = \mu(\boldsymbol{s},t)\exp\{\sigma Z(\boldsymbol{s},t)\}$, $\sigma > 0$, where $Z(\boldsymbol{s},t)$ is a zero mean, unit variance Gaussian random field with correlation $\rho(\boldsymbol{h},u)$ (see, for instance, Oliveira, 2006). Note that $E(L(\boldsymbol{s},t)) = \mu^*(\boldsymbol{s},t;\boldsymbol{\beta})$ and

$$\log\{\mu^*(\boldsymbol{s},t;\boldsymbol{\beta})\} = \beta_0^* + \sum_{l=1}^{3}\beta_{1,l}\cos\left(\frac{2\pi lt}{365}\right) + \beta_{2,l}\sin\left(\frac{2\pi lt}{365}\right)$$

with $\beta_0^* = \beta_0 + \sigma^2/2$.

The sample size prevents the use of a full likelihood approach even for the Log-Gaussian model. Therefore the estimation of the marginal and dependence parameters is performed by using the weighted pairwise likelihood, with cut-off weight one if $d_s \leq 350$ and $d_t \leq 1$ and 0 otherwise.

Our implementation, in the `GeoModels` package, based on C and OpenCL allows to speed-up five times the computation of the weigthed pairwise likelihood estimator with respect to a standard C implementation. In particular one evaluation of the function needs approximatively 1.1 and 4.9 seconds with or without OpenCL implementation where the time of evaluation (in seconds) has been measured in terms of elapsed time on a laptop with 2.4 GHz processor and 16 GB of memory.

In Table 3.5 we reports the estimates with the standard errors computed using the sub-sampling technique in Bevilacqua et al. (2012) and, for each model, the associate maximum pairwise likelihood. Actually we have estimated several specification of W_ν with $\nu = 4, 5, \ldots, 9$ but we report our best finding ($\nu = 6$).

The estimated trends are not dissimilar for the three models. Instead it is evident that the estimates of α_s and α_t entail a stronger spatial and temporal range for Gamma and Weibull models with respect to the Log-Gaussian ones.

	$\tau_1 = 0$			$\tau_1 = 0.5$			$\tau_1 = 1$		
	Gamma	Weibull	LogGaussian	Gamma	Weibull	LogGaussian	Gamma	Weibull	LogGaussian
$\hat{\beta}_0$	2.34306 (0.00060)	2.33832 (0.00085)	2.17973 (0.00042)	2.34297 (0.00064)	2.33820 (0.00087)	2.17959 (0.00037)	2.34298 (0.00054)	2.33832 (0.00086)	2.17949 (0.00043)
$\hat{\beta}_{11}$	0.05309 (0.02912)	0.04377 (0.02707)	0.07396 (0.00662)	0.05328 (0.02900)	0.04384 (0.02709)	0.07378 (0.00455)	0.05310 (0.02462)	0.04400 (0.02709)	0.07364 (0.00445)
$\hat{\beta}_{21}$	0.13675 (0.01058)	0.14575 (0.00754)	0.10606 (0.00526)	0.13724 (0.01063)	0.14576 (0.00756)	0.10589 (0.00376)	0.13725 (0.00904)	0.14574 (0.00756)	0.10580 (0.00381)
$\hat{\beta}_{12}$	−0.00749 (0.19440)	−0.01020 (0.11137)	−0.00979 (0.05039)	−0.00783 (0.18658)	−0.01022 (0.11152)	−0.00992 (0.03406)	−0.00787 (0.15741)	−0.01030 (0.11087)	−0.00992 (0.03329)
$\hat{\beta}_{22}$	−0.02883 (0.05149)	−0.02947 (0.03992)	−0.03445 (0.01372)	−0.02906 (0.05139)	−0.02945 (0.04006)	−0.03444 (0.00925)	−0.02913 (0.04330)	−0.02960 (0.04000)	−0.03451 (0.00889)
$\hat{\beta}_{13}$	−0.01531 (0.09667)	−0.01287 (0.09048)	−0.02577 (0.01910)	−0.01552 (0.09579)	−0.01290 (0.09056)	−0.02587 (0.01302)	−0.01556 (0.08073)	−0.01259 (0.09308)	−0.02578 (0.01279)
$\hat{\beta}_{23}$	−0.00281 (0.51692)	−0.00700 (0.16094)	0.00036 (1.38056)	−0.00283 (0.51496)	−0.00701 (0.16109)	0.00026 (1.30178)	−0.00276 (0.44714)	−0.00693 (0.16340)	0.00021 (1.59120)
$\hat{\kappa}$		1.96259 (0.00165)			1.96293 (0.00172)			1.96255 (0.00167)	
$\hat{\sigma}^2$			0.40478 (0.00038)			0.40459 (0.00027)			0.40443 (0.00025)
$\hat{\alpha}_s$	640.707 (5.72426)	582.796 (7.67767)	244.388 (0.23188)	641.975 (5.32311)	579.545 (8.26608)	243.371 (0.35126)	642.491 (5.66679)	578.782 (8.18816)	242.233 (0.51468)
$\hat{\alpha}_t$	9.59909 (0.13280)	8.24511 (0.23854)	3.66026 (0.02768)	9.42928 (0.13604)	8.12055 (0.23568)	3.52091 (0.02675)	9.27699 (0.10322)	7.97968 (0.23589)	3.38582 (0.02869)
CLIC	8886705	**8151609**	12805228	8870918	**8152061**	12283246	8640458	**8156138**	14300738

Table 3.5 Pairwise likelihood estimates for Gamma with $\nu = 6$, Weibull and LogGaussian random field and associated standard errors for regression, shape and correlation parameters.

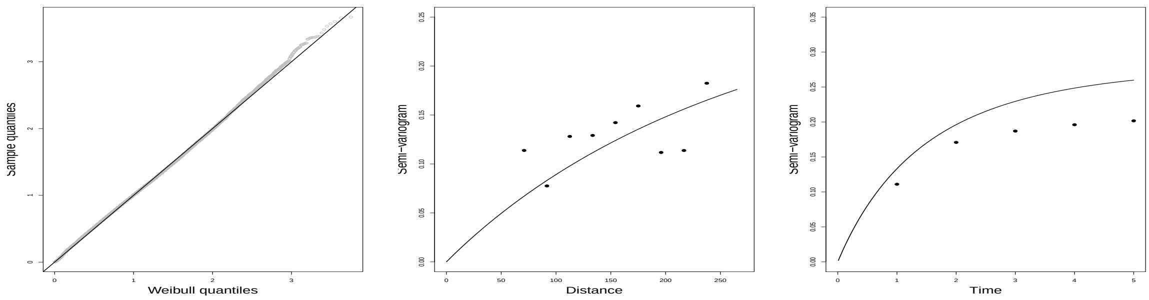
The Weibull random field achieve the highest pairwise likelihoods and the Log-Gaussian random field is the worst model from this point of view. A comparison of the three models in terms of CLIC leads to a clear discrimination among them, in favour of the first ones.

A crucial step in any statistical investigation is the assessment of the adequacy of the models proposed and fitted to the data under analysis and this step is usually based on residual models.

We remind that $M(\boldsymbol{s},t) = Y(\boldsymbol{s},t)/\mu(\boldsymbol{s},t;\boldsymbol{\beta})$ is a random variable with distribution Gamma, Weibull or Log-Gaussian. Thus the residuals

$$\hat{m}(\boldsymbol{s}_i,t) = \frac{y(\boldsymbol{s}_i,t)}{\mu(\boldsymbol{s}_i,t;\hat{\boldsymbol{\beta}})}$$

can be used for the checking the goodness-of-fit of models. The marginal fitting of the Weibull model is extremely good, according the Figure 3.4-(a) in which we compare the quantiles of the residuals with the quantiles of the distribution $Weibull(\hat{\kappa},\Gamma^{-1}(1+1/\hat{\kappa}))$. Moreover, the choice (3.21) as a model for $\rho(\boldsymbol{h},u;\boldsymbol{\psi})$ is supported by the comparison of the empirical semi-variograms of the residuals with the estimated semi-variograms Figure 3.4-(b-c).



(a) (b) (c)

Fig. 3.4 (a) Q-Q plot of the residuals against the estimated quantiles of a Weibull distribution. Empirical semi-variogram (dotted points) of the residuals and estimated theoretical semi-variogram (solid line) along the spatial distance (b) and the time-lag (c)

We gain a better insight if we compare the prediction performance of the three models. In doing this, we considerer one-day ahead prediction $\hat{Y}(\boldsymbol{s}_i,t)$ of the wind speed in the test period using the observations for the last three days $\hat{y}(\boldsymbol{s}_i,t-k)$, $i=1,\dots,12$ and $k=1,\dots,3$. The root-mean square error (RMSEA) and the mean absolute error (MAE) for each station, namely

$$RMSE(\boldsymbol{s}_i)=\left(\frac{1}{n_{te}}\sum_{t=n_{tr}+1}^{n_{te}+n_{tr}}(y(\boldsymbol{s}_i,t)-\hat{Y}(\boldsymbol{s}_i,t))^2\right)^{\frac{1}{2}}, \quad MAE(\boldsymbol{s}_i)=\frac{1}{n_{te}}\left(\sum_{t=n_{te}+1}^{n_{te}+n_{tr}}|y(\boldsymbol{s}_i,t)-\hat{Y}(\boldsymbol{s}_i,t)|\right)$$

are customary measures of the prediction performance. For the Gamma and the Weibull model we have reported the results obtained by using the simple kriging predictor (3.15) and we have not found differences with respect to use the alternative ordinary kriging predictor (3.16). Instead the Log-Gaussian model we have chosen the optimal predictor in Oliveira (2006). As benchmark, we have also considered the naïve predictor $\widehat{Y}(s_i,t)=y(s_i,t-1)$, that uses the observation recorded the day before at the station.

Even though the simple kriging predictor is an suboptimal solution, Gamma and Weibull models outperform again the Log-Gaussian model as we can infer from the last column in Tables 3.6-3.7. Overall, the three models performs better than the naïve predictor.

	RPT	VAL	ROS	KIL	SHA	BIR	DUB	CLA	MUL	CLO	BEL	MAL	Total
Naïve	5.551	4.989	5.051	3.404	4.539	3.663	4.289	4.211	3.859	4.246	5.499	6.230	4.711
$\tau_1 = 0$													
Gamma	4.859	4.375	4.366	**3.437**	4.003	**3.489**	3.815	**3.791**	**3.450**	**3.813**	4.900	5.827	**4.234**
Weibull	4.839	**4.364**	**4.319**	3.611	**4.000**	3.599	**3.799**	3.812	3.464	3.824	4.924	5.949	4.265
Log-Gaussian	**4.773**	4.398	4.322	3.553	4.064	3.638	3.885	3.914	3.593	3.939	**4.847**	**5.757**	4.268
$\tau_1 = 0.5$													
Gamma	4.854	4.376	4.353	**3.489**	4.006	**3.526**	3.809	**3.802**	**3.458**	**3.821**	4.905	5.842	**4.243**
Weibull	4.837	**4.367**	**4.313**	3.670	**4.004**	3.639	**3.796**	3.823	3.475	3.833	4.931	5.971	4.278
Log-Gaussian	**4.762**	4.405	4.296	3.675	4.081	3.730	3.879	3.948	3.627	3.970	**4.853**	**5.782**	4.293
$\tau_1 = 1$													
Gamma	4.849	4.379	4.343	**3.544**	4.011	**3.565**	3.805	**3.815**	**3.469**	**3.830**	4.910	5.857	**4.253**
Weibull	4.834	**4.370**	4.307	3.736	**4.009**	3.685	**3.794**	3.835	3.486	3.844	4.938	5.991	4.292
Log-Gaussian	**4.753**	4.415	**4.283**	3.815	4.105	3.836	3.882	3.989	3.671	4.010	**4.866**	**5.818**	4.326

Table 3.6 Root-mean-square-error (RMSE) for one-day ahead predictors of the wind speed at the meteorological stations during the test period (1971-1978). In bold we report the predictor with the lowest RMSE.

	RPT	VAL	ROS	KIL	SHA	BIR	DUB	CLA	MUL	CLO	BEL	MAL	Total
Naïve	4.301	3.834	3.868	2.543	3.465	2.833	3.311	3.254	3.015	3.266	4.269	4.866	3.647
$\tau_1 = 0$													
Gamma	3.771	**3.448**	3.381	**2.816**	**3.148**	**2.816**	**3.029**	**3.039**	**2.759**	**3.043**	3.841	4.551	**3.304**
Weibull	3.752	3.458	**3.339**	3.009	3.163	2.923	3.030	3.078	2.782	3.084	3.866	4.632	3.343
Log-Gaussian	**3.734**	3.478	3.390	2.876	3.222	2.910	3.097	3.126	2.882	3.154	**3.824**	**4.492**	3.349
$\tau_1 = 0.5$													
Gamma	3.766	**3.450**	3.371	**2.867**	**3.153**	**2.847**	**3.025**	**3.049**	**2.767**	**3.052**	3.845	4.561	**3.313**
Weibull	3.749	3.460	**3.332**	3.064	3.168	2.957	3.029	3.088	2.791	3.094	3.872	4.648	3.354
Log-Gaussian	**3.720**	3.483	3.364	2.990	3.240	2.988	3.096	3.156	2.910	3.182	**3.826**	**4.506**	3.372
$\tau_1 = 1$													
Gamma	3.761	**3.452**	3.362	**2.920**	**3.159**	**2.879**	**3.023**	**3.060**	**2.776**	**3.062**	3.849	4.571	**3.323**
Weibull	3.746	3.462	**3.327**	3.115	3.174	2.989	3.027	3.096	2.801	3.104	3.877	4.663	3.368
Log-Gaussian	**3.709**	3.490	3.349	3.120	3.263	3.078	3.103	3.192	2.944	3.218	**3.834**	**4.532**	3.403

Table 3.7 Mean absolute error (MAE) for one-day ahead predictors of the wind speed at the meteorological stations during the test period (1971-1978). In bold we report the predictor with the lowest MAE.

Chapter 4

Modeling spatial dependent data with symmetric and asymmetric *t* marginal distribution

4.1 Random Fields with *t* marginals.

Let G_i, $i = 1, \dots \nu$, $\nu \geq 1$, be an independent sequence of zero mean and unit variance second order Gaussian random fields with common correlation function $\rho_G(\boldsymbol{h})$ and define $W_\nu(\boldsymbol{s}) := \sum_{i=1}^{\nu} G_i(\boldsymbol{s})^2/\nu$. Then W is a weakly stationary Gamma random field with marginal distribution denoted by $W_\nu(\boldsymbol{s}) \sim \Gamma(\nu/2, \nu/2)$ for each $\boldsymbol{s}$, with $E(W_\nu(\boldsymbol{s})) = 1$, $Var(W_\nu(\boldsymbol{s})) = 2/\nu$ and $\rho_{W_\nu}(\boldsymbol{h}) = \rho_G^2(\boldsymbol{h})$. The distribution of $\boldsymbol{W}_{ij}$ is defined in the equation (3.2).

Let Z be a zero mean, unit variance weakly stationary Gaussian random field independent of W with correlation function $\rho_Z(\boldsymbol{h})$. We define the following random field with marginal distribution of the t type as:

$$Y_\nu(\boldsymbol{s}) := \mu(\boldsymbol{s}) + \sigma \tilde{R}_\nu(\boldsymbol{s}) Z(\boldsymbol{s}) \tag{4.1}$$

where $\mu(\boldsymbol{s})$ is spatially varying mean, $\sigma > 0$ is a scale parameter and $\tilde{R}_\nu(\boldsymbol{s}) := W_\nu(\boldsymbol{s})^{-\frac{1}{2}}$ is a weakly stationary non-negative random field with marginal distribution given by the square root of the inverse gamma distribution, namely:

$$f_{\tilde{R}_\nu(\boldsymbol{s})}(\tilde{r}) = 2\left(\frac{\nu}{2}\right)^{\nu/2} \tilde{r}^{-\nu-1} e^{-\frac{\nu}{2\tilde{r}^2}} / \Gamma\left(\frac{\nu}{2}\right). \tag{4.2}$$

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, $\boldsymbol{\beta} \in \mathbb{R}^k$ but other types of parametric or non parametric functions can be considered.

In this section we study the second order properties and bivariate distribution of the standardized random field

$$Y_\nu^*(\boldsymbol{s}) := [Y_\nu(\boldsymbol{s}) - \mu(\boldsymbol{s})]/\sigma = \tilde{R}_\nu(\boldsymbol{s})Z(\boldsymbol{s}). \tag{4.3}$$

The second order properties of Y_ν are easily obtained by means of transformation of location and scale. By construction, the marginal distribution of Y_ν^* is of t type:

$$f_{Y_\nu^*(\boldsymbol{s})}(y;\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{y^2}{\nu}\right)^{-\frac{(\nu+1)}{2}} \tag{4.4}$$

with $E(Y_\nu^*(\boldsymbol{s})) = 0$ and $Var(Y_\nu^*(\boldsymbol{s})) = \nu/(\nu-2)$, $\nu > 2$, and the associated correlation function is given by

$$\rho_{Y_\nu^*}(\boldsymbol{h}) = \left(\frac{\nu-2}{\nu}\right)\left[E(\tilde{R}_\nu(\boldsymbol{s}_i)\tilde{R}_\nu(\boldsymbol{s}_j))\rho_Z(\boldsymbol{h})\right] \tag{4.5}$$

In order to find a closed form expression for the correlation function in equation (4.5), we need the bivariate distribution of $\tilde{\boldsymbol{R}}_{ij}$ that can be easily obtained from (3.2) and it is given by:

$$f_{\tilde{\boldsymbol{R}}_{ij}}(\tilde{\boldsymbol{r}}_{ij}) = \frac{2^{-\nu+2}\nu^{\nu}(\tilde{r}_i\tilde{r}_j)^{-\nu-1}e^{-\frac{\nu}{2(1-\rho_G^2(\boldsymbol{h}))}\left(\frac{1}{\tilde{r}_i^2}+\frac{1}{\tilde{r}_j^2}\right)}}{\Gamma\left(\frac{\nu}{2}\right)(1-\rho_G^2(\boldsymbol{h}))^{\nu/2}}\left(\frac{\nu|\rho_G(\boldsymbol{h})|}{2(1-\rho_G^2(\boldsymbol{h}))\tilde{r}_i\tilde{r}_j}\right)^{1-\frac{\nu}{2}} I_{\frac{\nu}{2}-1}\left(\frac{\nu|\rho_G(\boldsymbol{h})|}{(1-\rho_G^2(\boldsymbol{h}))\tilde{r}_i\tilde{r}_j}\right) \tag{4.6}$$

Note that when $\rho_G(\boldsymbol{h}) = 0$, then (4.6) can be written as the product of two independent square root inverse gamma random variables with parameters $\nu/2$, $\nu/2$. The bivariate distribution in equation (4.6) is useful for proving the following Lemma. The proof can be found in the Appendix.

Lemma 2. *Let $W_\nu(\boldsymbol{s}) = \sum_{i=1}^{\nu} G_i(\boldsymbol{s})^2/\nu$ where G_i $i = 1,\dots,\nu$ is a independent finite sequence of zero mean unit variance weakly stationary Gaussian random fields with correlation function $\rho_G(\boldsymbol{h})$ and let $\tilde{R}_\nu(\boldsymbol{s}) = W_\nu(\boldsymbol{s})^{-1/2}$. Then the $(a,b)-th$ product moment*

$E(\tilde{R}_\nu^a(\boldsymbol{s}_i)\tilde{R}_\nu^b(\boldsymbol{s}_j))$ is given by

$$E(\tilde{R}_\nu^a\tilde{R}(\boldsymbol{s}_i)_\nu^b(\boldsymbol{s}_j)) = \frac{2^{-(a+b)/2}\nu^{(a+b)/2}}{\Gamma^2\left(\frac{\nu}{2}\right)}\Gamma\left(\frac{\nu-a}{2}\right)\Gamma\left(\frac{\nu-b}{2}\right){}_2F_1\left(\frac{a}{2},\frac{b}{2};\frac{\nu}{2};\rho_G^2(\boldsymbol{h})\right) \tag{4.7}$$

for $\nu > a$ and $\nu > b$.

Using Lemma 2, we are now ready to give a closed form expression for the correlation function of the t random field defined in (4.3).

Theorem 15. *Let G_i $i = 1,\dots,\nu$, $\nu > 2$ and Z independent copies of a zero mean unit variance weakly stationary Gaussian random fields with correlation $\rho_G(\boldsymbol{h})$ and $\rho_Z(\boldsymbol{h})$ respectively. Let $W_\nu(\boldsymbol{s}) = \sum_{i=1}^{\nu} G_i(\boldsymbol{s})^2/\nu$. Then the correlation function of $Y_\nu^*(\boldsymbol{s}) = W_\nu(\boldsymbol{s})^{-1/2}Z(\boldsymbol{s})$ is given by:*

$$\rho_{Y_\nu^*}(\boldsymbol{h}) = \frac{(\nu-2)\Gamma^2\left(\frac{\nu-1}{2}\right)}{2\Gamma^2\left(\frac{\nu}{2}\right)}\left[{}_2F_1\left(\frac{1}{2},\frac{1}{2};\frac{\nu}{2};\rho_G^2(\boldsymbol{h})\right)\rho_Z(\boldsymbol{h})\right]; \quad \nu > 2. \tag{4.8}$$

Proof. Setting $a = b = 1$ in (4.7) and using it in (4.5) we obtain (4.8). □

In principle $\rho_G(\boldsymbol{h})$ and $\rho_Z(\boldsymbol{h})$ are two different correlation functions. Nevertheless, in order to obtain a more parsimonious model, we assume a common correlation function $\rho(\boldsymbol{h}) := \rho_G(\boldsymbol{h}) = \rho_Z(\boldsymbol{h})$. We will call $\rho(\boldsymbol{h})$ the underlying correlation function of the t random field Y_ν^*. The correlation of the t random field only depends on ν and $\rho(\boldsymbol{h})$.

Under this assumption, the following Theorem states that, it can be easily shown that geometrical properties such as stationarity, mean-square continuity and degrees of mean-square differentiability can be inherited by the t random field from the common properties of the Gaussian random fields G_i, $i = 1,\dots,\nu$ and Z. The proof can be found in the Appendix and is similar to the theorem 14.

Theorem 16. *Let G_i $i = 1,\dots,\nu$, $\nu > 2$ and Z independent copies of a zero mean unit variance weakly stationary Gaussian random fields with correlation $\rho(\boldsymbol{h})$. Let $W_\nu(\boldsymbol{s}) = \sum_{i=1}^{\nu} G_i(\boldsymbol{s})^2/\nu$ and let $Y_\nu^*(\boldsymbol{s}) = W_\nu(\boldsymbol{s})^{-1/2}Z(\boldsymbol{s})$. Then:*

a) Y_ν^* is also weakly stationary.

b) Y_ν^* is mean-square continuous if and only if G_i and Z are mean-square continuous;

c) Y_ν^* is k-times mean-square differentiable if G_i and Z are k-times mean-square differentiable.

d) *The sample paths of Y_ν^* is continuous and differentiable if the sample paths of G are continuous and differentiable.*

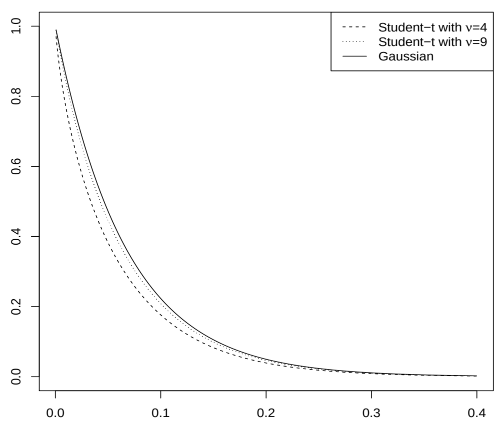
We exemplify some geometric features of the random field Y_ν^* by means of the Matérn correlation function and Generalized Wendland correlation function defined in section 1.4.

Left part of Figure 4.1a compares the correlation function of the t random field given by equation (4.8) with underlying Matérn correlation function $\rho(\boldsymbol{h}) = \mathscr{M}_{0.5,0.2/3}(\boldsymbol{h})$ when $\nu = 4, 9$. It is apparent that when increasing the degrees of freedom $\rho_{Y_\nu^*}(\boldsymbol{h})$ approach $\rho(\boldsymbol{h})$. Indeed, using series expansion of the hypergeometric function, it can be easily shown that

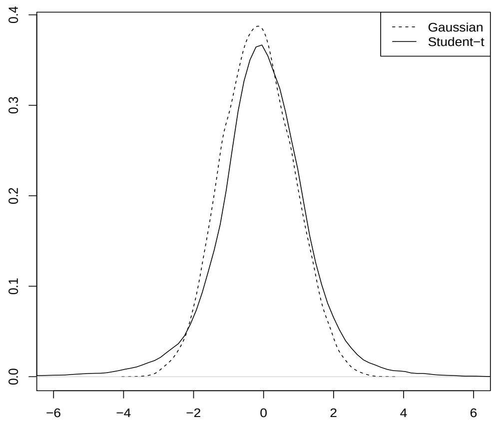
$$\lim_{\nu \to \infty} \rho_{Y_\nu^*}(\boldsymbol{h}) = \rho(\boldsymbol{h})$$

and also, can we show that $0 \leq \rho_{Y_\nu^*}(\boldsymbol{h}) \leq |\rho(\boldsymbol{h})|$ for $\nu > 2$.

On the right part of Figure 4.1-(b) we compare a kernel non parametric density estimation of a realization of a zero mean, unit variance Gaussian random field and a realization of a zero mean t random field when $\nu = 4$ using $\rho(\boldsymbol{h}) = \mathscr{M}_{0.5,0.2/3}(\boldsymbol{h})$.



(a)



(b)

Fig. 4.1 (a) a comparison of the exponential correlation function of a Gaussian RF with the correlation function of a t RF in equation (4.8) with $\nu = 4, 9$. (b) a comparison of a non parametric kernel density estimation of one realization from a zero mean Gaussian RF and an associated zero mean t RF with $\nu = 4$

.

Figure 4.2 compare two realizations of a zero mean unit variance Gaussian random field with Matérn correlation function $\rho(\boldsymbol{h}) = \mathscr{M}_{0.5,0.2}(\boldsymbol{h})$ and Generalized Wendland correlation function $\rho(\boldsymbol{h}) = \mathscr{GW}_{1,4,0.2}(\boldsymbol{h})$ respectively and the associated realizations of a zero mean t random field with $\nu = 4$. In these cases the sample paths of the Gaussian field are zero and one times differentiable. (Bevilacqua et al., 2018). It can be appreciated from Figure 4.2 that this feature is inherited by the t random field.

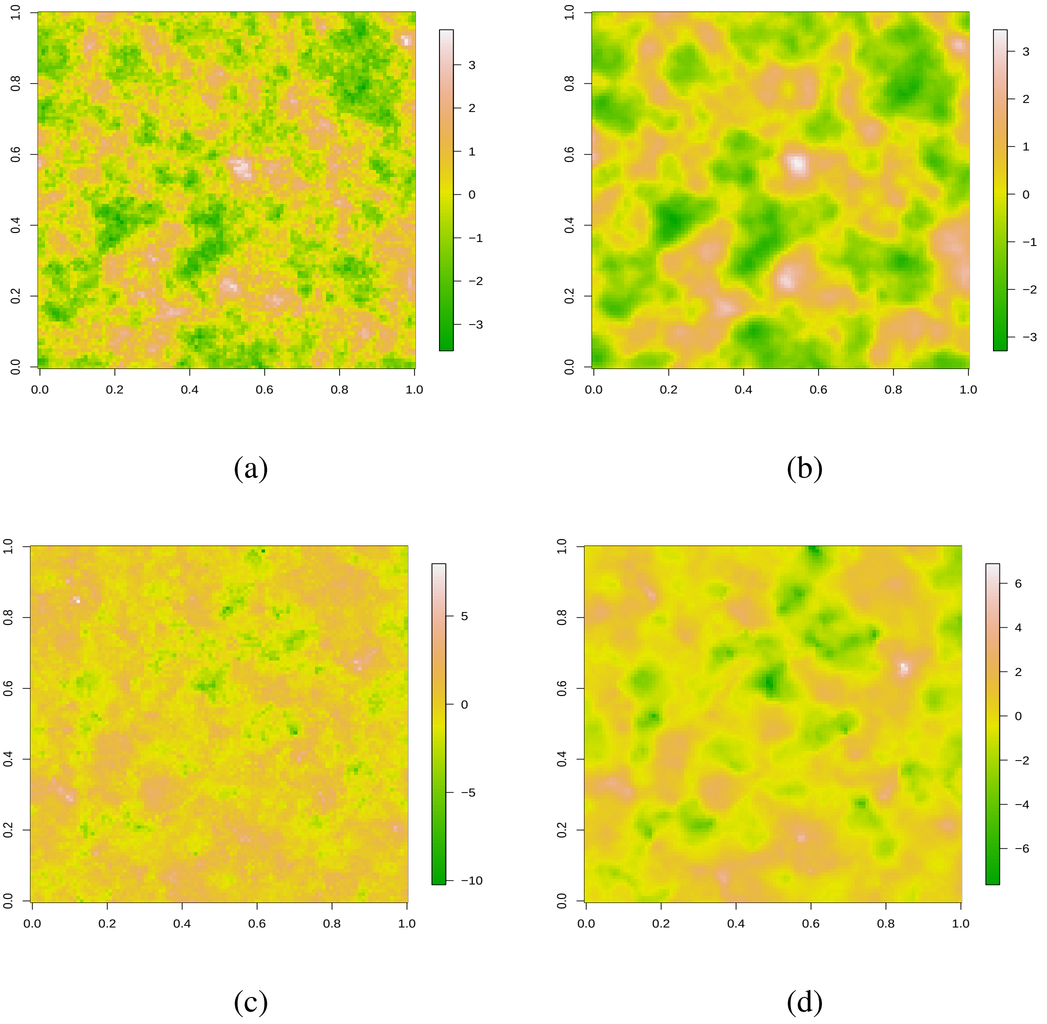


Fig. 4.2 Upper part: two realizations of a zero mean gaussian RF on $[0,1]^2$ with $\rho(\boldsymbol{h}) = \mathscr{GW}_{0,0.2,4}(\boldsymbol{h})$ and $\rho(\boldsymbol{h}) = \mathscr{GW}_{1,0.2,4}(\boldsymbol{h})$ (from left to right). Bottom part: two associated realizations of a t RF with underlying correlation $\rho(\boldsymbol{h})$ when $\nu = 4$.

We now consider the bivariate random vector of the random field Y_ν^* defined by

$$\boldsymbol{Y}_{ij}^* = \boldsymbol{W}_{ij}^{-1/2} \circ \boldsymbol{Z}_{ij}$$

where $\circ$ denotes the Schur product vector. The following Theorem gives the pdf of the bivariate random vector $\boldsymbol{Y}_{ij}^*$ in terms of Appell Hypergeometric function of the fourth type (Gradshteyn and Ryzhik, 2007) defined by

$$F_4(a,b;c,c';w,z) = \sum_{k=0}^{\infty}\sum_{m=0}^{\infty} \frac{(a)_{k+m}(b)_{k+m}w^k z^m}{k!m!(c)_k(c')_m}, \quad |\sqrt{w}|+|\sqrt{z}|<1 \tag{4.9}$$

The proof can be found in the Appendix.

Theorem 17. *Let Y_ν^*, $\nu > 2$ a weakly stationary t random field as defined in equation (4.3) with underlying correlation $\rho(\boldsymbol{h})$. Then:*

$$\begin{aligned} f_{\boldsymbol{Y}_{ij}^*}(\boldsymbol{y}_{ij}) \; &= \; \frac{\nu^\nu[(y_i^2+\nu)(y_j^2+\nu)]^{-(\nu+1)/2}\Gamma^2\left(\frac{\nu+1}{2}\right)}{\pi\Gamma^2\left(\frac{\nu}{2}\right)(1-\rho^2(\boldsymbol{h}))^{-(\nu+1)/2}} F_4\left(\frac{\nu+1}{2},\frac{\nu+1}{2},\frac{1}{2},\frac{\nu}{2};\frac{\rho^2(\boldsymbol{h})y_i^2y_j^2}{(y_i^2+\nu)(y_j^2+\nu)},\frac{\nu^2\rho^2(\boldsymbol{h})}{(y_i^2+\nu)(y_j^2+\nu)}\right) \\ &+ \; \frac{\rho(\boldsymbol{h})y_iy_j\nu^{\nu+2}[(y_i^2+\nu)(y_j^2+\nu)]^{-\nu/2-1}}{2\pi(1-\rho^2(\boldsymbol{h}))^{-(\nu+1)/2}} F_4\left(\frac{\nu}{2}+1,\frac{\nu}{2}+1,\frac{3}{2},\frac{\nu}{2};\frac{\rho^2(\boldsymbol{h})y_i^2y_j^2}{(y_i^2+\nu)(y_j^2+\nu)},\frac{\nu^2\rho^2(\boldsymbol{h})}{(y_i^2+\nu)(y_j^2+\nu)}\right) \end{aligned} \tag{4.10}$$

Note that when $\rho(\boldsymbol{h}) = 0$, according to (4.9) we obtain $F_4(a,b;c,c';0,0) = 1$ and, as a consequence, the pdf in Theorem 17 can be written as product of two independent t random variables with ν degrees of freedom. The existence of such pdf is guaranteed for $\nu = 3,4,\ldots$ and, under infinity divisibility of the multivariate Gamma density, for $\nu > 2$.

4.2 Random fields with asymmetric t marginals

In this section we first review the skew Gaussian random field proposed in Zhang and El-Shaarawi (2010). For this random field we give the finite dimensional distribution generalizing previous results in Alegría et al. (2017). Starting from definition and properties of the t random field Y_ν^* with underlying correlation $\rho(\boldsymbol{h})$, we propose a random field with skew-t marginals obtained mixing a skew Gaussian random field with an inverse square root Gamma random field as defined in section 4.1. Then we propose a general construction for random fields with asymmetric marginal distribution of the two piece type, obtained mixing a random field with half-t marginal and with a specific discrete random field.

4.2.1 Skew-Gaussian Random fields

In our first approach, the basic idea is to replace the Gaussian random field in (4.1) with a skew-Gaussian random field. According to Zhang and El-Shaarawi (2010) we define a skew-Gaussian random field as:

$$\tilde{U}_\eta(\boldsymbol{s}) := g(\boldsymbol{s}) + \eta|X_1(\boldsymbol{s})| + \omega X_2(\boldsymbol{s}), \quad \boldsymbol{s} \in A \subset \mathbb{R}^d \tag{4.11}$$

where $\eta \in \mathbb{R}$, X_i $i = 1,2$ are two independent weakly stationary zero mean unit variance Gaussian random fields with a common correlation functions $\rho(\boldsymbol{h})$. Here $g(\boldsymbol{s})$ is spatially varying mean that can be specified as $g(\boldsymbol{s}) = X(\boldsymbol{s})^T\boldsymbol{\beta}$ where $X(\boldsymbol{s}) \in \mathbb{R}^k$ is a vector of covariates, $\boldsymbol{\beta} \in \mathbb{R}^k$. The parameters η and ω allow to model the asymmetry and variance of the random field simultaneously.

The random field $\tilde{U}_\eta$ has marginal distribution of the skew Gaussian type given by

$$f_{\tilde{U}_\eta(\boldsymbol{s})}(\tilde{u}) = \frac{2}{(\eta^2+\omega^2)^{1/2}}\phi\left(\frac{(\tilde{u}-g(\boldsymbol{s}))}{(\eta^2+\omega^2)^{1/2}}\right)\Phi\left(\frac{\eta(\tilde{u}-g(\boldsymbol{s}))}{\omega(\eta^2+\omega^2)^{1/2}}\right) \tag{4.12}$$

with $E(\tilde{U}_\eta(\boldsymbol{s})) = g(\boldsymbol{s}) + \eta(2/\pi)^{1/2}$, $Var(\tilde{U}_\eta(\boldsymbol{s})) = \omega^2 + \eta^2(1-2/\pi)$ and with correlation function given by Zhang and El-Shaarawi (2010):

$$\rho_{\tilde{U}_\eta}(\boldsymbol{h}) = \frac{2\eta^2}{\pi\omega^2+\eta^2(\pi-2)}\Big((1-\rho^2(\boldsymbol{h}))^{1/2} + \rho(\boldsymbol{h})\arcsin(\rho(\boldsymbol{h})) - 1\Big) + \frac{\omega^2\rho(\boldsymbol{h})}{\omega^2+\eta^2(1-2/\pi)}. \tag{4.13}$$

The following Theorem generalize the results in Alegría et al. (2017) and give an explicit closed form expression for the finite-dimensional distribution of $\tilde{U}_\eta$. Here we suppose $\eta \neq 0$ (the case with zero skewness is reduced to the Gaussian scenario). The proof can be found in the Appendix.

Theorem 18. *Let* $\tilde{U}_\eta(\boldsymbol{s}) = g(\boldsymbol{s}) + \eta|X_1(\boldsymbol{s})| + \omega X_2(\boldsymbol{s})$ *where* $X_i(\boldsymbol{s})$ $i = 1,2$ *are two independent weakly-stationary zero mean unit variance random fields with correlation function* $\rho(\boldsymbol{h})$ *and correlation matrix* Ω. *Then*

$$f_{\tilde{\boldsymbol{U}}}(\tilde{\boldsymbol{u}}) = 2\sum_{l=1}^{2^{n-1}} \phi_n(\tilde{\boldsymbol{u}}-\boldsymbol{\alpha};\boldsymbol{A}_l)\Phi_n(\boldsymbol{c}_l;\boldsymbol{0},\boldsymbol{B}_l) \tag{4.14}$$

where

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

4.2.2 Skew-t Random fields

We now consider $g(\boldsymbol{s})=0$ and $\omega=1$ in (4.11) and define the following random field

$$G_{\nu,\eta}(\boldsymbol{s}) := \mu(\boldsymbol{s})+\sigma\tilde{R}_\nu(\boldsymbol{s})\tilde{U}_\eta(\boldsymbol{s}) \tag{4.15}$$

This type of construction is similar to (4.1) but now we replace a zero mean unit variance Gaussian random field with a standard Skew-Gaussian random field. As in the previous Section, we consider the standardized random field

$$G^*_{\nu,\eta}(\boldsymbol{s}) := [G_{\nu,\eta}(\boldsymbol{s})-\mu(\boldsymbol{s})]/\sigma=\tilde{R}_\nu(\boldsymbol{s})\tilde{U}_\eta(\boldsymbol{s}). \tag{4.16}$$

and we focus on the second order properties of $G^*_{\nu,\eta}$. The associated properties of $G_{\nu,\eta}$ are easily obtained by means of transformation of location and scale. The marginal distribution of $G^*_{\nu,\eta}$ is given by (Azzalini and Capitanio, 2014):

$$f_{G^*_{\nu,\eta}}(\boldsymbol{s})(g)=2f_{Y^*_\nu(\boldsymbol{s})}(g;\nu)F_{Y^*_\nu(\boldsymbol{s})}\left(\eta g\sqrt{\frac{\nu+1}{\nu+g^2}};\nu+1\right) \tag{4.17}$$

where $F_X(\cdot;\cdot)$ denotes the cdf of (4.3) with $E(G^*_{\nu,\eta}(\boldsymbol{s}))=\frac{\sqrt{\nu}\Gamma\left(\frac{\nu-1}{2}\right)\eta}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}$, if $\nu>1$ and $Var(G^*_{\nu,\eta}(\boldsymbol{s}))=\left[\frac{\nu}{\nu-2}(1+\eta^2)-\frac{\nu\Gamma^2\left(\frac{\nu-1}{2}\right)\eta^2}{\pi\Gamma^2\left(\frac{\nu}{2}\right)}\right]$, if $\nu>2$. If $\eta=0$, (4.17) reduces to the usual t density given in (4.4) and if $\nu\to\infty$, (4.17) converges to the standard skew Gaussian distribution. Using (4.7) and (4.13) we easily obtain the correlation function of skew-t random field.

Theorem 19. *Let G_i $i=1,\dots,\nu$, $\nu>2$, X_1, X_2 independent copies of a zero mean unit variance weakly stationary Gaussian random field with correlation $\rho(\boldsymbol{h})$. Let $\tilde{U}_\eta(\boldsymbol{s})=\eta|X_1(\boldsymbol{s})|+X_2(\boldsymbol{s})$ and $W_\nu(\boldsymbol{s})=\sum_{i=1}^\nu G_i(\boldsymbol{s})^2/\nu$. Then the correlation function of $G^*_{\nu,\eta}(\boldsymbol{s})=$*

$\tilde{R}_\nu(\boldsymbol{s})\tilde{U}_\eta(\boldsymbol{s})$ *is given by*

$$\rho_{G^*_{\nu,\eta}}(\boldsymbol{h}) = \frac{\pi(\nu-2)\Gamma^2\left(\frac{\nu-1}{2}\right)}{2\left[\pi\Gamma^2\left(\frac{\nu}{2}\right)(1+\eta^2)-\eta^2(\nu-2)\Gamma^2\left(\frac{\nu-1}{2}\right)\right]}\left[{}_2F_1\left(\frac{1}{2},\frac{1}{2};\frac{\nu}{2};\rho^2(\boldsymbol{h})\right)\left\{(1+\eta^2(1-\frac{2}{\pi}))\rho_{\tilde{U}_\eta}(\boldsymbol{h})+\frac{2\eta^2}{\pi}\right\}-\frac{2\eta^2}{\pi}\right] \quad (4.18)$$

Note that, since $\tilde{U}_\eta$ and Y^*_ν share the same underlying correlation model, the correlation function $\rho_{G^*_{\nu,\eta}}(\boldsymbol{h})$ only depends on ν, η and $\rho(\boldsymbol{h})$. Moreover $\rho_{G^*_{\nu,\eta}}(\boldsymbol{h}) = \rho_{G^*_{\nu,-\eta}}(\boldsymbol{h})$ that is the correlation is invariant with respect to positive or negative asymmetry. As in the Theorem 16, it can be easily shown that geometrical properties such as stationarity, mean-square continuity and degrees of mean-square differentiability can be inherited by the skew-t random field from the common Gaussian random field. Since the bivariate distribution of a skew-t random field is difficult to calculate, we propose another approach to construct a asymmetric t random field.

4.2.3 Two piece t random fields

In the second approach we introduce the stochastic representation of a general class of asymmetric univariate distributions given in Arellano-Valle et al. (2005). Let H is a zero mean unit variance Gaussian random field with correlation $\rho(\boldsymbol{h})$. We define the discrete random field

$$K_\eta(\boldsymbol{s}) := \begin{cases} 1-\eta, & \text{if} \quad H(\boldsymbol{s}) < q_{\frac{1-\eta}{2}} \\ -1-\eta, & \text{if} \quad H(\boldsymbol{s}) \geq q_{\frac{1-\eta}{2}} \end{cases}$$

with $|\eta| < 1$ the asymmetry parameter and q_x is the quantile of order x of the standard Gaussian distribution. By construction, $Pr(K_\eta(\boldsymbol{s}) = 1-\eta) = \frac{1-\eta}{2}$ and $Pr(K_\eta(\boldsymbol{s}) = -1-\eta) = \frac{1+\eta}{2}$. Moreover, it can be easily shown that the distribution of $\boldsymbol{K}_{ij}$ is completely characterized by

$$p_{11} := \Phi_2\left(q_{\frac{1-\eta}{2}}, q_{\frac{1-\eta}{2}}, \rho(\boldsymbol{h})\right)$$

and η where $\Phi_2(\cdot,\cdot,x)$ is the cdf of a bivariate standard Gaussian distribution with correlation x. Now we define the following random field

$$P_{\nu,\eta}(\boldsymbol{s}) := \mu(\boldsymbol{s}) + \sigma|Y^*_\nu(\boldsymbol{s})|K_\eta(\boldsymbol{s}) \quad (4.19)$$

where $|Y^*_\nu(\boldsymbol{s})|$ define a half-t random field with marginal distribution given by

$$f_{|Y^*_\nu(\boldsymbol{s})|}(|y|;\nu) = 2f_{Y^*_\nu(\boldsymbol{s})}(y;\nu), \quad y > 0$$

with $E(|Y^*_\nu(\boldsymbol{s})|) = \frac{\sqrt{\nu}\Gamma\left(\frac{\nu-1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}$ and $Var(|Y^*_\nu(\boldsymbol{s})|) = \frac{\nu}{\nu-2} - \frac{\nu\Gamma^2\left(\frac{\nu-1}{2}\right)}{\pi\Gamma^2\left(\frac{\nu}{2}\right)}$, $\nu > 2$. The marginal distribution of $P_{\nu,\eta}$ is given by

$$f_{P_{\nu,\eta}(\boldsymbol{s})}(y) = \frac{2}{\sigma}\left[f_{|Y^*_\nu(\boldsymbol{s})|}\left(\frac{y-\mu(\boldsymbol{s})}{\sigma(1-\eta)};\nu\right)I(y \geq \mu(\boldsymbol{s})) + f_{|Y^*_\nu(\boldsymbol{s})|}\left(\frac{y-\mu(\boldsymbol{s})}{\sigma(1+\eta)};\nu\right)I(y < \mu(\boldsymbol{s}))\right] \tag{4.20}$$

the so called two piece t distribution (Arellano-Valle et al., 2005). As in the previous section we focus on the standardized version:

$$P^*_{\nu,\eta}(\boldsymbol{s}) := \frac{P_{\nu,\eta}(\boldsymbol{s}) - \mu(\boldsymbol{s})}{\sigma} = |Y^*_\nu(\boldsymbol{s})|K_\eta(\boldsymbol{s}) \tag{4.21}$$

with $E(P^*_{\nu,\eta}(\boldsymbol{s})) = -\frac{2\eta\sqrt{\nu}\Gamma\left(\frac{\nu-1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}$ and $Var(P^*_{\nu,\eta}(\boldsymbol{s})) = \frac{\nu(3\eta^2+1)}{\nu-2} - \frac{4\eta^2\nu\Gamma^2\left(\frac{\nu-1}{2}\right)}{\pi\Gamma^2\left(\frac{\nu}{2}\right)}$.

The following Theorem gives the pdf of the bivariate random vector $|\boldsymbol{Y}^*_{ij}|$. The proof can be found in the Appendix.

Theorem 20. *Let Y^*_ν, $\nu > 2$ a weakly stationary t random field as defined in equation (4.3) with underlying correlation $\rho(\boldsymbol{h})$. Then the pdf of $|\boldsymbol{Y}^*_{ij}|$ is given by*

$$f_{|\boldsymbol{Y}^*_{ij}|}(|\boldsymbol{y}_{ij}|) = \frac{\nu^\nu[(y_i^2+\nu)(y_j^2+\nu)]^{-(\nu+1)/2}\Gamma^2\left(\frac{\nu+1}{2}\right)}{\pi\Gamma^2\left(\frac{\nu}{2}\right)(1-\rho^2(\boldsymbol{h}))^{-(\nu+1)/2}} F_4\left(\frac{\nu+1}{2},\frac{\nu+1}{2},\frac{1}{2},\frac{\nu}{2};\frac{\rho^2(\boldsymbol{h})y_i^2y_j^2}{(y_i^2+\nu)(y_j^2+\nu)},\frac{\nu^2\rho^2(\boldsymbol{h})}{(y_i^2+\nu)(y_j^2+\nu)}\right) \tag{4.22}$$

If $\rho(\boldsymbol{h}) = 0$, then the bivariate distribution in Theorem 20 can be written as product of two independent random variables half-t that is, as in the Gaussian case, no pairwise correlation implies pairwise independence.

The following Lemma gives the $(a,b)-th$ product moment of the half-t distribution and it is useful for giving a closed form expression for $\rho_{P^*_{\nu,\eta}}(\boldsymbol{h})$. The proof can be found in the Appendix.

Lemma 3. *The $(a,b)-th$ product moment of the half-t distribution is given by*

$$E(|Y^*_\nu(\boldsymbol{s}_i)|^a|Y^*_\nu(\boldsymbol{s}_j)|^b) = \frac{\nu^{\frac{a+b}{2}}\Gamma\left(\frac{a+1}{2}\right)\Gamma\left(\frac{\nu-a}{2}\right)\Gamma\left(\frac{b+1}{2}\right)\Gamma\left(\frac{\nu-b}{2}\right)}{\pi\Gamma^2\left(\frac{\nu}{2}\right)} {}_2F_1\left(\frac{a}{2},\frac{b}{2};\frac{\nu}{2};\rho^2(\boldsymbol{h})\right) {}_2F_1\left(-\frac{a}{2},-\frac{b}{2};\frac{1}{2};\rho^2(\boldsymbol{h})\right) \tag{4.23}$$

We are now ready to give the correlation function of the skew-t random field $P^*_{\nu,\eta}$. As in the first approach, since K_η and Y^*_ν share the same underlying correlation model, the correlation function of $P^*_{\nu,\eta}$ only depends on ν, η and $\rho(\boldsymbol{h})$.

Theorem 21. *Let $P^*_{\nu,\eta}$, $\nu > 2$, $\eta \in (-1,1)$ a weakly stationary two piece t random field as defined in equation (4.21). Then:*

$$\rho_{P^*_{\nu,\eta}}(\boldsymbol{h}) = A(\nu,\eta)\left[{}_2F_1\left(\frac{1}{2},\frac{1}{2};\frac{\nu}{2};\rho^2(\boldsymbol{h})\right)\left(\rho(\boldsymbol{h})arcsin(\rho(\boldsymbol{h})) + (1-\rho^2(\boldsymbol{h}))^{\frac{1}{2}}\right)(3\eta^2+2\eta+4p_{11}-1) - 4\eta^2\right], \tag{4.24}$$

where $A(\nu,\eta) = \frac{\nu(\nu-2)\Gamma^2\left(\frac{\nu-1}{2}\right)}{\nu\pi\Gamma^2(\frac{\nu}{2})(3\eta^2+1)-4\eta^2\nu(\nu-2)\Gamma^2(\frac{\nu-1}{2})}$.

Proof. Since $\rho_{P^*_{\nu,\eta}}(\boldsymbol{h}) = \frac{E(|Y^*_\nu(\boldsymbol{s}_i)||Y^*_\nu(\boldsymbol{s}_j)|)E(K_\eta(\boldsymbol{s}_i)K_\eta(\boldsymbol{s}_j)) - E^2(|Y^*_{\nu,\eta}(\boldsymbol{s})|)E^2(K_\eta(\boldsymbol{s}))}{Var(P^*_{\nu,\eta}(\boldsymbol{s}))}$, setting $a=b=1$ and $a=1$, $b=0$ in (4.23) and considering $E(K_\eta(\boldsymbol{s}_i)K_\eta(\boldsymbol{s}_j)) = 3\eta^2+2\eta+4p_{11}-1$, $E(K_\eta(\boldsymbol{s})) = -2\eta$ and using the identity ${}_2F_1\left(-\frac{1}{2},-\frac{1}{2};\frac{1}{2};\rho^2(\boldsymbol{h})\right) = \rho(\boldsymbol{h})arcsin(\rho(\boldsymbol{h})) + (1-\rho^2(\boldsymbol{h}))^{\frac{1}{2}}$, we obtain (4.24). □

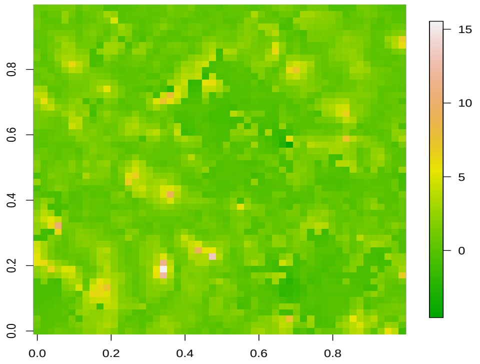
Note that $\rho_{P^*_{\nu,\eta}}(\boldsymbol{h}) = \rho_{P^*_{\nu,-\eta}}(\boldsymbol{h})$, for $\eta \in (-1,1)$. Also in this case it can be easily shown that geometrical properties such as stationarity, mean-square continuity and degrees of mean-square differentiability can be inherited by the two piece t random field from the common Gaussian random field.

Figure 4.3 (upper part) compare two realizations of a two piece t random field $P^*_{4,\eta}$ with $\eta = -0.5, 0.5$ (from left to right) with underlying Generalized Wendland correlation $\rho(\boldsymbol{h}) = \mathcal{GW}_{1,4,0.5}(\boldsymbol{h})$. In the same Figure 4.3 (central part) compares correlation function $\rho_{P^*_{4,\eta}}(\boldsymbol{h})$ with $\eta = -0.5, 0.5$ of a two piece t random field (from left to right) with correlation function $\rho_{Y^*_4}(\boldsymbol{h})$ of a t random field Y^*_4. It is interesting to note that $\rho_{P^*_{4,\eta}}(\boldsymbol{h})$ can be greater and lower of $\rho_{Y^*_4}(\boldsymbol{h})$. Finally, in the Figure 4.3 (bottom part) we compare a kernel non parametric density estimation of a realization of a two piece t random field $P^*_{4,\eta}$ with $\eta = -0.5, 0.5$ (from left to right) and a t random field Y^*_4.

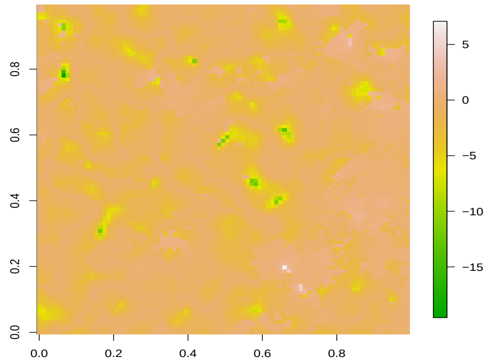
We now consider the bivariate random vector of the random field $P^*_{\nu,\eta}$ defined by

$$\boldsymbol{P}^*_{ij} := |\boldsymbol{Y}^*_{ij}| \circ \boldsymbol{K}_{ij}$$

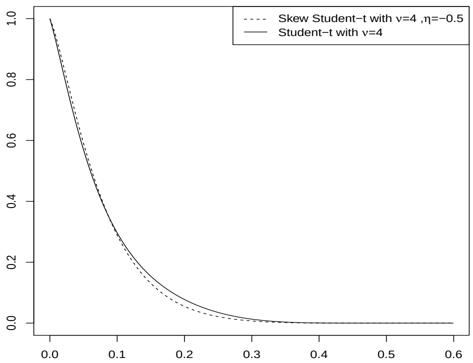
where $\circ$ denotes the Schur product vector. The following Theorem gives the pdf of the bivariate random vector of $\boldsymbol{P}^*_{ij}$. The proof is omitted since it is straightforward.



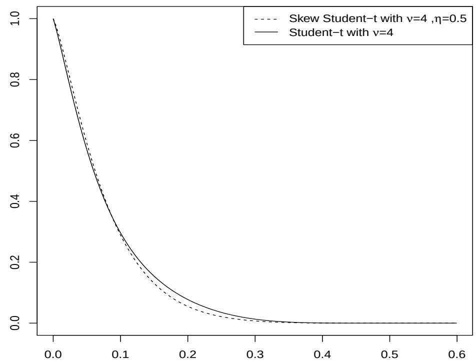
(a)



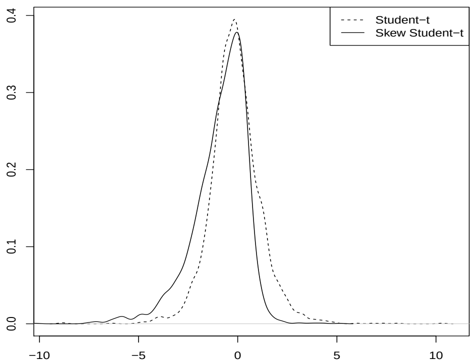
(b)



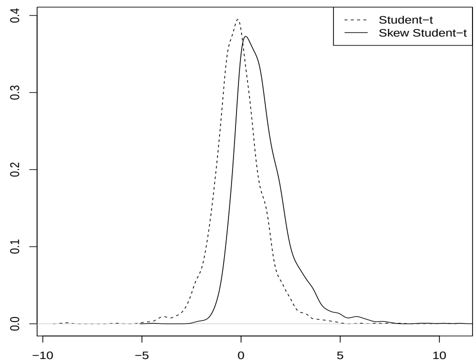
(c)



(d)



(e)



(f)

Fig. 4.3 Upper part: two realizations on $[0,1]^2$ of $P^*_{4,\eta}$ with $\eta = -0.5, 0.5$ (from left to right) with underlying correlation $\rho(\boldsymbol{h}) = \mathscr{GW}_{1,4,0.5}(\boldsymbol{h})$. Central part: comparison between $\rho_{P^*_{4,-0.5}}(\boldsymbol{h})$ and $\rho_{Y^*_4}(\boldsymbol{h})$ and $\rho_{P^*_{4,0.5}}(\boldsymbol{h})$ and $\rho_{Y^*_4}(\boldsymbol{h})$. Bottom part:a comparison between a non parametric kernel density estimation of the two realizations in the upper part and of two realizations of a t RF with the same underlying correlation function.

Theorem 22. *Let $P^*_{\nu,\eta}$, $\nu > 2$, $\eta \in (-1,1)$ a weakly stationary two piece t random field as defined in equation (4.21). Then:*

$$\begin{aligned} f_{\boldsymbol{P}^*_{ij}}(\boldsymbol{p}_{ij}) &= 4\Bigg[\frac{p_{11}}{(1-\eta)^2} f_{|\boldsymbol{Y}^*_{ij}|}\left(\frac{p_i}{1-\eta}, \frac{p_j}{1-\eta}\right) I(p_i \geq 0, p_j \geq 0) \\ &+ \frac{1-\eta-2p_{11}}{2(1-\eta^2)} f_{|\boldsymbol{Y}^*_{ij}|}\left(\frac{p_i}{1-\eta}, \frac{p_j}{1+\eta}\right) I(p_i \geq 0, p_j < 0) \\ &+ \frac{1-\eta-2p_{11}}{2(1-\eta^2)} f_{|\boldsymbol{Y}^*_{ij}|}\left(\frac{p_i}{1+\eta}, \frac{p_j}{1-\eta}\right) I(p_i < 0, p_j \geq 0) \\ &+ \frac{p_{11}+\eta}{(1+\eta)^2} f_{|\boldsymbol{Y}^*_{ij}|}\left(\frac{p_i}{1+\eta}, \frac{p_j}{1+\eta}\right) I(p_i < 0, p_j < 0)\Bigg] \end{aligned} \tag{4.25}$$

If $\rho(\boldsymbol{h}) = 0$, then the bivariate distribution in Theorem 22 can be written as product of two independent random variables two piece t that is, as in the Gaussian case, no pairwise correlation implies pairwise independence.

4.3 Simulation study

In this section, we analyze the performance of the weighted pairwise composite likelihood method of estimation when estimating t and two piece t random fields both in the spatial setting.

For the simulation study, we consider two scenarios. In the first scenario, we focus on $k = 2$ and we estimate all the parameters except for the parameter λ that we consider known and fixed. For the t and two peace t, following (Ciccio and Monti, 2011), we use a reparametrization by using the inverse degrees of freedom, $\lambda = 1/\nu$. In the second scenario, we focus on $k = 1$ and an exponential correlation function. Following the remarks in section 4.1, in this case $\lambda \in (0, 0.5)$ and we consider estimation of the inverse of degree of freedom, regression, asymmetry, variance and correlation parameters for the t and two piece t random fields. Specifically:

1. We simulate, using Cholesky decomposition, 500 realization of a t and two piece t random fields on $N = 1000$ location sites uniformly distributed in the unit square and $k = 2$ that is we set $\boldsymbol{X\beta}$ with $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$, $\beta_0 = 0.5$, $\beta_1 = -0.25$ where $\boldsymbol{X}$ is $N \times 2$ matrix with first column the unit vector and the second column an iid realization from a $U(0,1)$. Then we fix $\lambda = 1/\nu$, $\nu = 3, 6, 9$, for the t and two piece t cases, $\eta = 0.5$ for the two piece t case and $\sigma^2 = 1$ for the variance parameter. As isotropic parametric correlation model we consider special cases of Matérn and Generalized

Wendland class in equations (1.7) and (1.8)

$$\rho(\boldsymbol{h}) = \mathscr{M}_{0.5,\alpha}(\boldsymbol{h}) = e^{-||\mathbf{h}||/\alpha} \tag{4.26}$$

$$\rho(\boldsymbol{h}) = \mathscr{GW}_{0,4,\alpha}(\boldsymbol{h}) = (1 - ||\mathbf{h}||/\alpha)_{+}^{4} \tag{4.27}$$

with $\alpha = 0.15/3$ in the first model and $\alpha = 0.15$ in the second model. In the pairwise likelihood estimation we consider a cut off weight function as in (3.14) setting $d_s = 0.02$. Table 4.1 and 4.2 show bias and mean square error (MSE) associated to regression, asymmetry, variance and correlation parameters using both correlation models. In the t and two piece t cases, the degrees of freedom are assumed to be known.

2. We simulate, using Cholesky decomposition, 500 realization realization of a t and two piece t random fields in $N = 200$ points $0.005, 0.01, \ldots, 1$, $k = 2$ that is we set $\boldsymbol{X\beta}$ with $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$, $\beta_0 = 0.5$, $\beta_1 = -0.25$ where $\boldsymbol{X}$ is $N \times 2$ matrix with first column the unit vector and the second column an iid realization from a $U(0,1)$. Then we fix $\lambda = 1/\nu$, $\nu = 3, 6, 9$, for the t and two piece t cases, $\eta = 0.5$ for the two piece t case and $\sigma^2 = 1$ for the variance parameter.

   As isotropic parametric correlation model we consider the case Matérn defined in equation (4.26) with $\alpha = 0.1/3$. In the pairwise likelihood estimation we consider a cut off weight function as in (3.14) setting $d_s = 0.005$. Table 4.3 show bias and mean square error (MSE) associated to regression, degrees of freedom, asymmetry, variance and correlation parameters using correlation model Exponential. In this scenario the degree freedom can be estimated.

As general comment the estimates are overall approximatively unbiased for both Scenarios (see tables 4.1 to 4.3). MSE is affected by the magnitude of the λ parameter for both the t and two piece t case in Scenario 2 (see table 4.3). Specifically MSE seems to decreass when the λ parameter increase. Therefore, the precision of estimates of the regression parameters benefit of the increment of the shape parameter. Under scenario 1, we observe the same situation in the case of two piece t, since in case t the MSE remain stable (see tables 4.1 to 4.2).

With respect to the distribution of the weighted pairwise likelihood estimates, the figure 4.4, as an example, depicts the boxplots of the estimates for a two piece t random field with $\beta_0 = 0.5$, $\beta_1 = -0.25$, $\eta = 0.5$, using a Askey correlation model with $\alpha = 0.15$ and

$\sigma^2 = 1$. It is apparent that, the distribution of the estimates are quite symmetric numerically stable with very few outliers.

λ	1/3				1/6				1/9			
	Exp		Askey		Exp		Askey		Exp		Askey	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\hat{\beta}_0$	0.00735	0.01872	0.00193	0.01176	−0.00216	0.02053	0.00355	0.01112	0.01406	0.01982	0.00115	0.01012
$\hat{\beta}_1$	0.00469	0.00949	0.00541	0.00982	−0.00060	0.00799	−0.00501	0.00970	0.00399	0.00711	0.00322	0.00885
$\hat{\alpha}$	4.71e-06	0.00010	−0.00169	0.00074	−0.00083	9.89e-05	0.00083	0.00077	9.98e-05	0.00011	0.00309	0.00077
$\hat{\sigma}^2$	−0.00401	0.03020	−0.00671	0.02216	−0.01587	0.02368	−0.00597	0.01676	−0.02155	0.02243	−0.00082	0.01587

Table 4.1 t random field estimation with weighted pairwise likelihood: bias and mean square error for regression, spatial scale and variance parameters under Scenario 1 using Exponential and Askey correlation model when $\lambda = 1/\nu$ with $\nu = 3, 6, 9$.

λ	1/3				1/6				1/9			
	Exp		Askey		Exp		Askey		Exp		Askey	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\hat{\beta}_0$	0.00631	0.00110	0.01297	0.00165	0.00620	0.00115	0.00974	0.00160	0.00856	0.00114	0.01176	0.00146
$\hat{\beta}_1$	0.00202	0.00230	0.00690	0.00306	0.00435	0.00226	0.00440	0.00287	0.00055	0.00203	0.00448	0.00246
$\hat{\eta}$	0.01365	0.00255	0.01074	0.00184	0.00897	0.00184	0.00889	0.00143	0.00844	0.00207	0.00697	0.00157
$\hat{\alpha}$	0.00188	0.00012	0.00583	0.00095	0.00219	0.00012	0.00568	0.00086	0.00225	0.00011	0.00675	0.00101
$\hat{\sigma}^2$	0.00811	0.02331	0.00668	0.01851	0.00716	0.02025	0.00037	0.01355	0.00214	0.01668	0.00504	0.01416

Table 4.2 two piece t random field estimation with weighted pairwise likelihood: bias and mean square error for regression, asymmetry, spatial scale and variance parameters under Scenario 1 using Exponential and Askey correlation model when $\lambda = 1/\nu$ with $\nu = 3, 6, 9$.

λ	1/3				1/6				1/9			
	t		two piece t		t		two piece t		t		two piece t	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\hat{\beta}_0$	−0.00825	0.07300	0.00139	0.00093	0.01683	0.06900	0.00018	0.00106	−0.01866	0.06129	−0.00123	0.00082
$\hat{\beta}_1$	−0.00282	0.00446	0.00158	0.00163	−0.00023	0.00343	0.00200	0.00148	0.00014	0.00343	0.00215	0.00125
$\hat{\lambda}$	−0.00881	0.00590	−0.01054	0.00633	−0.00844	0.00387	−0.01073	0.00440	−0.01418	0.00296	−0.00815	0.00316
$\hat{\eta}$			0.01270	0.00604			0.00545	0.00425			0.01213	0.00355
$\hat{\alpha}$	−0.00192	7.54e-05	−0.00070	5.50e-05	−0.00258	8.44e-05	-8.11e-05	6.33e-05	−0.00245	7.97e-05	0.00047	6.44e-05
$\hat{\sigma}^2$	−0.02203	0.09475	0.01829	0.07356	−0.05728	0.08158	0.01283	0.06501	−0.03209	0.07853	0.01971	0.06284

Table 4.3 t and two piece t random field estimation with weighted pairwise likelihood: bias and mean square error for regression, inverse degree of freedom, asymmetry, spatial scale and variance parameters under Scenario 2 using Exponential correlation model when $\lambda = 1/\nu$ with $\nu = 3, 6, 9$.

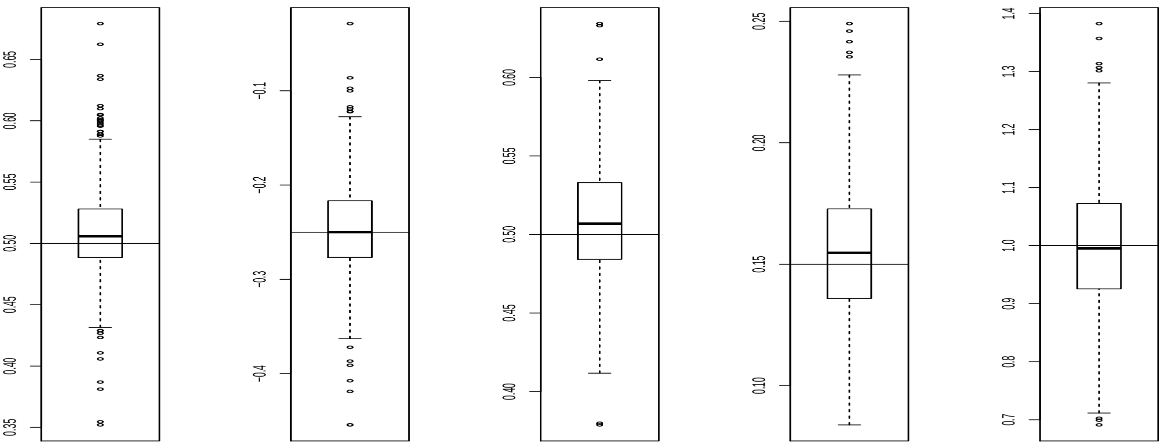


Fig. 4.4 Boxplots of weighted pairwise likelihood estimations for parameters $\beta_0 = 0.5$, $\beta_1 = -0.25$, $\eta = 0.5$, $\alpha = 0.15$, $\sigma^2 = 1$ (from left to right) under Scenario 1 when estimating a two piece t random field with $\nu = 6$ and using Askey correlation model.

Concluding remarks

In this thesis we proposed some models for regression and dependence analysis when dealing with non Gaussian continuous data.

Indeed, 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 negative or positive asymmetry and/or heavy tails.

Specifically we proposed two possibly non stationary random processes with Gamma and Weibull marginals. Both models are obtained transforming a rescaled sum of independent copies of the squared of a 'parent' Gaussian random process.

We then propose a weakly stationary random field with t marginals obtained mixing a Gaussian random field with an inverse square root Gamma random field.

Two possible generalizations allowing for possible asymmetry are considered. In the first approach we obtain a skew-t random field mixing a skew Gaussian with an inverse square root Gamma random field. In the second approach we obtain a two piece t random field mixing a specific discrete random field with a random field with half-t marginals.

For this kind of random fields we studied the associated second order properties and in the stationary case, the geometrical properties. Moreover we study the bivariate distribution. A nice feature of our models is that that nice properties such as stationarity, mean-square continuity, and degrees of mean-square differentiability are inherited from the parent Gaussian random field.

In the thesis the so-called nugget effect is not considered but can be easily induced by choosing a discontinuous correlation function in the parent Gaussian random field.

It is worth noting that a hierarchical spatial model like the generalized linear model (Diggle et al., 1998) always induces discontinuity in the path.

Finally, we remark that even though we have limited to ourselves to a continuous Euclidean space, our models can be extended to the case of spherical domains using covariance models introduced in (Gneiting, 2013; Porcu et al., 2016).

Some limitations are evident in our proposal. For instance, a limitation is the lack of an amenable expression of the density outside of the bivariate case. Nevertheless, we have showed that an inferential approach based on the pairwise likelihood could be a solution for estimating the unknown parameter. Moreover, multivariate probabilities could be evaluated by Monte Carlo method since the random processes can be quickly simulated. However our solution to the conditional prediction, based on a linear predictor, is limited and deserves further consideration.

Finally, the shape parameter of the Gamma and the degrees of freedom of the t random field are not necessarily allow to vary in $\mathbb{R}^+$. The solution of this problem is interesting topic for future research.

References

Abromowitz, M. and I. Stegun (1967). *Handbook of Mathematical Functions*. U.S. Goverment Printing Office.

Alder, R. J. (1981). *The Geometric of Random Fields*. New York: Jonh Wiley & Song.

Alegría, A., S. Caro, M. Bevilacqua, E. Porcu, and J. Clarke (2017). Estimating covariance functions of multivariate skew-gaussian random fields on the sphere. *Spatial Statistics*.

Amani, A. and T. Lebel (1997). Lagrangian kriging for the estimation of sahelian rainfall at small time steps. *Journal of Hydrology 192*, 125–157.

Arellano-Valle, R. B., H. W. Gómez, and F. A. Quintana (2005). Statistical inference for a general class of asymmetric distributions. *Journal of Statistical Planning and Inference 128*(2), 427 – 443.

Azzalini, A. and A. Capitanio (2014). *The Skew-Normal and Related Families*. United States of America by Cambridge University Press, New York.

Banerjee, S., B. P. Carlin, and A. E. Gelfand (2014). Hierarchical modeling and analysis for spatial data (second ed.). *Boca Raton: FL: CRC Press.*.

Bellier, E., P. Monestiez, and C. Guinet (2010). Geostatistical modelling of wildlife populations: a non-stationary hierarchical model for count data. *In: Atkinson, P.M., Lloyd, C.D. (Eds.), geoENV VII-Geostatistics for Environmental Applications. Springer*, 1–12.

Bevilacqua, M., T. Faouzi, R. Furrer, and E. Porcu (2018). Estimation and prediction using generalized wendland functions under fixed domain asymptotics. *The Annals of Statistics*. Forthcoming.

Bevilacqua, M. and C. Gaetan (2015). Comparing composite likelihood methods based on pairs for spatial gaussian random fields. *Statistics and Computing 25*, 877–892.

Bevilacqua, M., C. Gaetan, J. Mateu, and E. Porcu (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.

Bochner, S. (1955). *Harmonic Analysis and the Theory of Probability*. Berkeley and Los Angeles: University of California Press.

Brown, P. E., K. F. Kåresen, G. O. Roberts, and S. Tonellato (2000). Blur-generated non-separable space-time models. *Journal of the Royal Statistical Society, Series B 62*, 847–860.

Brychkov, Y. A. and N. Saad (2017). On some formulas for the appell function $f_2(a,b,b';c,c';w;z)$. *Journal Integral Transforms and Special Functions 25*, 1465 –1483.

Butler, A. and C. Glasbey (2008). A latent gaussian model for compositional data with zeros. *Journal of the Royal Statistical Society, Series C 57*, 505–520.

Carroll, R., R. Chen, E. George, T. Li, H. Newton, H. Schmiediche, and N. Wang (1997). Ozone exposure and population density in harris county, texas (with discussion). *Journal of the American Statistical Association 92*, 392–415.

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

Cressie, N. (1993). *Statistics for Spatial Data* (revised ed.). New York: Wiley.

Cressie, N. and H. Huang (1999). Classes of nonseparable, spatiotemporal stationary covariance functions. *Econometric Theory 2*, 305–330.

Cressie, N. and C. Wikle (2011). *Statistics for Spatio-Temporal Data.* Wiley Series in Probability and Statistics. Wiley.

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

De Iaco, S., D. Myers, and T. Posa (2001). Space-time analysis using a general product-sum model. *Statistics and Probability Letters 52*, 21–28.

De Iaco, S., D. Myers, and T. Posa (2002). Nonseparable space-time covariance models: some parametric families. *Mathematical Geology 34*, 23–42.

De Luna, X. and M. Genton (2005). Predictive spatio-temporal models for spatially sparse environmental data. *Statistica Sinica 15*, 547–568.

DeBastiani, F., A. Cysneiros, M. Uribe-Opazo, and M. Galea (2015). Influence diagnostics in elliptical spatial linear models. *TEST 24*, 322–340.

Diggle, P., J. Tawn, and R. Moyeed (1998). Model-based geostatistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 47*, 299–350.

Downton, F. (1970). Bivariate exponential distributions in reliability theory. *Journal of the Royal Statistical Society, Series B 32*, 408–417.

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

Eisenbaum, N. and H. Kaspi (2009). On permanental processes. *Stochastic Processes and their Applications 119*, 1401–1415.

Fuentes, M. (2005a). A formal test for nonstationarity of spatial stochastic processes. *Journal of Multivariate Analysis 96*, 30–54.

Fuentes, M. (2005b). Testing for separability of spatio-temporal covariance functions. *Journal of Statistical Planning and Inference 136*, 447–466.

Furrer, R., M. Genton, and D. Nychka (2007). Covariance tapering for interpolation of large spatial datasets. *J. Comput. Graph. Statist. 15*, 502–523.

Gaetan, C. and X. Guyon (2009). *Spatial Statistics and Modeling*. Springer.

Genton, M. G. and H. Zhang (2012). Identifiability problems in some non-gaussian spatial random fields. *Chilean Journal of Statistics 3*(171–179).

Gneiting, T. (1999). Correlation functions for atmospheric data analysis. *Quarterly Journal of the Royal Meteorological Society 125*, 2449–2464.

Gneiting, T. (2002a). Compactly supported correlation functions. *Journal of Multivariate Analysis 83*, 493–508.

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

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

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

Gneiting, T. and Z. Sasvári (1999). The characterization problem for isotropic covariance functions. *Mathematical Geology 31*, 105–111.

Gneiting, T., Z. Sasvári, and M. Schlather (2001). Analogies and correspondences between variograms and covariance functions. *Advances in Applied Probability 33*, 617–630.

Gneiting, T. and M. Schlather (2004). Stochastic models that separate fractal dimension and the hurst effect. *SIAM Review 46*, 69–282.

Gradshteyn, I. and I. Ryzhik (2007). *Table of Integrals, Series, and Products* (7 ed.). Academic Press.

Guo, J. H. and L. Billard (1998). Some inference results for causal autoregressive processes on a plane. *Journal of Time Series Analysis 19*, 681–691.

Guttorp, P., W. Meiring, and P. Sampson (1994). A space-time analysis of ground-level ozone. *Environmetrics 5*, 241–254.

Haas, T. (2002). New systems for modeling, estimating, and predicting a multivariate spatiotemporal process. *Environmetrics 13*, 311–332.

Haslett, J. and A. Raftery (1989). Space-time modelling with long-memory dependence: assessing ireland's wind-power resource (with discussion). *Applied Statistics 38*, 1–50.

Heagerty, P. and T. Lumley (2000). Window subsampling of estimating functions with application to regression models. *Journal of the American Statistical Association 95*, 197–211.

Heaton, M. J., A. Datta, A. Finley, R. Furrer, R. Guhaniyogi, F. Gerber, R. B. Gramacy, F. L. D. N. F. S. D. Hammerling, M. Katzfuss, and A. Zammit-Mangion (2017). Methods for analyzing large spatial data: A review and comparison. *ArXiv e-prints*.

Huang, H. and N. Hsu (2004). Modeling transport effects on ground-level ozone using a non-stationary space-time model. *Environmetrics 15*, 251–268.

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

Jun, M. and M. Stein (2004). Statistical comparison of observed and cmaq modeled daily sulfate levels. *Atmospheric Environment 38*, 4427–4436.

Kazianka, H. and J. Pilz (2010). Copula-based geostatistical modeling of continuous and discrete data including covariates. *Stochastic Environmental Research and Risk Assessment 24*, 661–673.

Kibble, W. F. (1998). A two-variate gamma type distribution. *Sankhyā: The Indian Journal of Statistics (1933-1960 5*, 137–150.

Kim, H. and B. Mallick (2004). A bayesian prediction using the skew-gaussian processes. *Journal of Statistical Planning and Inference 120*, 85–101.

Krishnaiah, P. and M. Rao (1961). Remarks on a multivariate gamma distribution. *The American Mathematical Monthly 68*, 342–346.

Krishnamoorthy, A. S. and M. Parthasarathy (1951). A multivariate gamma-type distribution. *Ann. Math. Statist. 22*(4), 549–557.

Lee, Y. and S. Lahiri (2002). Least squares variogram fitting by spatial subsampling. *Journal of the Royal Statistical Society B 64*, 837–854.

Lindgren, F., H. Rue, and J. Lindstrüm (2011). An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B 73*, 423–498.

Lindsay, B. (1988). Composite likelihood methods. *Contemporary Mathematics 80*, 221–239.

Lu, N. and D. Zimmerman (2005). Testing for directional symmetry in spatial dependence using the periodogram. *Journal of Statistical Planning and Inference 129*, 369–385.

Ma, C. (2003a). Families of spatio-temporal stationary covariance models. *Journal of Statistical Planning and Inference 116*, 489–501.

Ma, C. (2003b). Nonstationary covariance functions that model space-time interactions. *Statistics and Probability Letters 61*, 411–419.

Ma, C. (2005). Spatio-temporal variograms and covariance models. *Advances in Applied Probability 37*, 706–725.

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.

Marcus, M. B. (2014). Multivariate gamma distributions. *Electron. Commun. Probab. 19*, 1–10.

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

Matérn, B. (1986). *Spatial Variation: Stochastic Models and their Applications to Some Problems in Forest Surveys and Other Sampling Investigations (2nd ed.).* Heidelberg:Springer.

Meiring, W., P. Guttorp, and P. Sampson (1998). Space-time estimation of grid-cell hourly ozone levels for assessment of a deterministic model. *Environmental and Ecological Statistics 5*, 197–222.

Mitchell, M., M. Genton, and M. Gumpertz (2005). Testing for separability of space-time covariances. *Environmetrics 16*, 819–831.

Monestiez, P., L. Dubroca, E. Bonnin, J. Durbec, and C. Guinet (2006). Geostatistical modelling of spatial distribution of balaenoptera physalus in the northwestern mediterranean sea from sparse count data and heterogeneous observation efforts. *Ecological Modelling 193*, 615–628.

Oliveira, V. D. (2006). On optimal point and block prediction in log-gaussian random fields. *Scandinavian Journal of Statistics 33*, 523–540.

Oliveira, V. D. (2014). Poisson kriging: A closer investigation. *Spatial Statistics 7*, 1–20.

Oliveira, V. D., B. Kedem, and D. A. Short (1997). Bayesian prediction of transformed gaussian random fields. *Journal of the American Statistical Association 92*, 1422–1433.

Porcu, E., M. Bevilacqua, and M. Genton (2016). Spatio-temporal covariance and cross covariance functions of the great circle distance on a sphere. *Journal of the American Statistical Association 111*, 888–898.

Rouhani, S. and T. Hall (1989). Space-time kriging of groundwater data. *Geostatistics, Armstrong, M. (Ed.), Kluwer Academic Publishers 2*, 639–651.

Royen, T. (1973). Multivariate gamma distributions ii. *In Encyclopedia of Statistical Sciences*, 419–425.

Sampson, P. and P. Guttorp (1992). Nonparametric estimation of nonstationary spatial covariance structures. *Journal of the American Statistical Association 87*, 108–119.

Scaccia, L. and R. Martin (2005). Testing axial symmetry and separability of lattice processes. *Journal of Statistical Planning and Inference 131*, 19–39.

Shapiro, A. and J. Botha (1991). Variogram fitting with a conditional class of conditionally nonnegative definite functions. *Computational Statistics and Data Analysis 11*, 87–96.

Shitan, M. and P. Brockwell (1995). An asymptotic test for separability of a spatial autoregressive model. *Communications in Statistics, Theory and Methods 24*, 2027–2040.

Stein, M. (1999). *Interpolation of Spatial Data. Some Theory of Kriging*. New York: Springer-Verlag.

Stein, M. (2005a). Space-time covariance functions. *Journal of the American Statistical Association 100*, 310–321.

Stein, M. (2005b). Statistical methods for regular monitoring data. *Journal of the Royal Statistical Society B 67*, 667–687.

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

Storvik, G., A. Frigessi, and D. Hirst (2002). Stationary space-time Gaussian fields and their time autoregressive representation. *Stochastic Modelling 2*, 139–161.

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

Varin, C. and P. Vidoni (2005). A note on composite likelihood inference and model selection. *Biometrika 52*, 519–528.

Vere-Jones, D. (1997). Alpha-permanents and their applications to multivariate gamma, negative binomialand ordinary binomial distributions. *New Zealand Journal of Mathematics 26*, 125–149.

Wan, Y., M. Milligan, and B. Parsons (2003). Output power correlation between adjacent wind power plants. *Journal of Solar Energy Engineering 125*, 551–555.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika 49*, 305–314.

Yaglom, A. (1987). *Correlation Theory of Stationary and Related Random Functions*. New York: Springer.

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

Appendix A

Proof Lemma 1. Note that equation (3.7) can be rewritten in terms of the hypergeometric function using the identity ${}_0F_1(;b;x) = \Gamma(b)x^{(1-b)/2}I_{b-1}(2\sqrt{x})$. Then by definition of product moments and using series expansion of hypergeometric function ${}_0F_1$, we have:

$$
\begin{aligned}
E(R_\kappa^a(\boldsymbol{s}_i)R_\kappa^b(\boldsymbol{s}_j)) &= \frac{k^2\Gamma^{2\kappa}\left(1+\frac{1}{\kappa}\right)}{(\mu(\boldsymbol{s}_i)\mu(\boldsymbol{s}_j))^\kappa(1-\rho_Z^2(\boldsymbol{h}))} \int\limits_{\mathbb{R}_+^2} r_i^{\kappa+a-1} r_j^{\kappa+b-1} e^{-\frac{\Gamma^\kappa\left(1+\frac{1}{\kappa}\right)}{(1-\rho_Z^2(\boldsymbol{h}))}\left(\frac{r_i}{\mu(\boldsymbol{s}_i)}\right)^\kappa} e^{-\frac{\Gamma^\kappa\left(1+\frac{1}{\kappa}\right)}{(1-\rho_Z^2(\boldsymbol{h}))}\left(\frac{r_j}{\mu(\boldsymbol{s}_j)}\right)^\kappa} \\
&\times \; {}_0F_1\left(1; \frac{\rho_Z^2(\boldsymbol{h})(r_ir_j)^\kappa\Gamma^{2\kappa}\left(1+\frac{1}{\kappa}\right)}{(\mu(\boldsymbol{s}_i)\mu(\boldsymbol{s}_j))^\kappa(1-\rho_Z^2(\boldsymbol{h}))^2}\right) dr_i dr_j \\
&= \frac{\kappa^2\Gamma^{2\kappa}\left(1+\frac{1}{\kappa}\right)}{(\mu(\boldsymbol{s}_i)\mu(\boldsymbol{s}_j))^\kappa(1-\rho_Z^2(\boldsymbol{h}))} \sum_{m=0}^{\infty} \int\limits_{\mathbb{R}_+^2} r_i^{\kappa+a+m\kappa-1} r_j^{\kappa+b+m\kappa-1} e^{-\frac{\Gamma^\kappa\left(1+\frac{1}{\kappa}\right)}{(1-\rho_Z^2(\boldsymbol{h}))}\left(\frac{r_i}{\mu(\boldsymbol{s}_i)}\right)^\kappa} \\
&\times \; e^{-\frac{\Gamma^\kappa\left(1+\frac{1}{\kappa}\right)}{(1-\rho_Z^2(\boldsymbol{h}))}\left(\frac{r_j}{\mu(\boldsymbol{s}_j)}\right)^\kappa} \frac{1}{m!\,(1)_m} \left(\frac{\rho_Z^2(\boldsymbol{h})\Gamma^{2\kappa}\left(1+\frac{1}{\kappa}\right)}{(\mu(\boldsymbol{s}_i)\mu(\boldsymbol{s}_j))^\kappa(1-\rho_Z^2(\boldsymbol{h}))^2}\right)^m dr_i dr_j \\
&= \frac{k^2\Gamma^{2\kappa}\left(1+\frac{1}{\kappa}\right)}{(\mu(\boldsymbol{s}_i)\mu(\boldsymbol{s}_j))^\kappa(1-\rho_Z^2(\boldsymbol{h}))} \sum_{m=0}^{\infty} \frac{I(m)}{m!\,(1)_m} \left(\frac{\rho_Z^2(\boldsymbol{h})\Gamma^{2\kappa}\left(1+\frac{1}{\kappa}\right)}{(\mu(\boldsymbol{s}_i)\mu(\boldsymbol{s}_j))^\kappa(1-\rho_Z^2(\boldsymbol{h}))^2}\right)^m \qquad \text{(A.1)}
\end{aligned}
$$

Using Fubini's Theorem and (3.381.4) of Gradshteyn and Ryzhik (2007), we obtain

$$
\begin{aligned}
I(\kappa) &= \int\limits_{\mathbb{R}_+} r_i^{\kappa+a+m\kappa-1} e^{-\frac{\Gamma^\kappa\left(1+\frac{1}{\kappa}\right)}{(1-\rho_Z^2(\boldsymbol{h}))}\left(\frac{r_i}{\mu(\boldsymbol{s}_i)}\right)^\kappa} dr_i \int\limits_{\mathbb{R}_+} r_j^{\kappa+b+m\kappa-1} e^{-\frac{\Gamma^\kappa\left(1+\frac{1}{\kappa}\right)}{(1-\rho_Z^2(\boldsymbol{h}))}\left(\frac{r_j}{\mu(\boldsymbol{s}_j)}\right)^\kappa} dr_j \\
&= \kappa^{-2}\Gamma\left(1+\frac{a}{\kappa}+m\right)\Gamma\left(1+\frac{b}{\kappa}+m\right)\left(\frac{\mu(\boldsymbol{s}_i)^\kappa(1-\rho_Z^2(\boldsymbol{h}))}{\Gamma^\kappa\left(1+\frac{1}{\kappa}\right)}\right)^{1+\frac{a}{\kappa}+m}\left(\frac{\mu(\boldsymbol{s}_j)^\kappa(1-\rho_Z^2(\boldsymbol{h}))}{\Gamma^\kappa\left(1+\frac{1}{\kappa}\right)}\right)^{1+\frac{b}{\kappa}+m} \qquad \text{(A.2)}
\end{aligned}
$$

and combining equations (A.2) and (A.1), we obtain

$$
E(R_\kappa^a(\boldsymbol{s}_i)R_\kappa^b(\boldsymbol{s}_j)) = \frac{\mu(\boldsymbol{s}_i)^a\mu(\boldsymbol{s}_j)^b(1-\rho_Z^2(\boldsymbol{h}))^{1+(a+b)/\kappa}\Gamma\left(1+\frac{a}{\kappa}\right)\Gamma\left(1+\frac{b}{\kappa}\right)}{\Gamma^{a+b}\left(1+\frac{1}{\kappa}\right)} {}_2F_1\left(1+\frac{a}{\kappa}, 1+\frac{b}{\kappa}; 1; \rho_Z^2(\boldsymbol{h})\right)
$$

Finally, using Euler transformation we have (3.8). □

Proof Theorem 14. We prove the results only for R_κ with correlation $\rho_{R_\kappa}(\boldsymbol{h})$ i.e the Weibull case. The same arguments can be easily used for W_ν with correlation $\rho_{W_\nu}(\boldsymbol{h})$. If Z is second-order stationary with correlation function $\rho_Z(\boldsymbol{h})$ then it is straightforward to see that $\rho_{R_\kappa}(\boldsymbol{h})$ only depends on $\boldsymbol{h}$ and thus R_κ is also second-order stationary.

By (Stein, 1999, Chapter 2.4), the mean-square continuity and the k-times mean-square differentiability of the random field R_κ are equivalent to the continuity and $2k$-times differentiability of $\rho_{R_\kappa}(\boldsymbol{h})$ at $\boldsymbol{h} = \boldsymbol{0}$, respectively. Define the function

$$g(x) = \frac{\Gamma^2\left(1+\frac{1}{\kappa}\right)}{\left[\Gamma\left(1+\frac{2}{\kappa}\right) - \Gamma^2\left(1+\frac{1}{\kappa}\right)\right]}\left[{}_2F_1\left(-\frac{1}{\kappa}, -\frac{1}{\kappa}; 1; x^2\right) - 1\right]$$

It is easy to see that $\rho_{R_\kappa}(\boldsymbol{h}) = g\{\rho_Z(\boldsymbol{h})\}$ and for any $-1 \leq x \leq 1$, $g(x)$ is a continuous and infinitely differentiable function. Hence, $\rho_{R_\kappa}(\boldsymbol{h})$ is continuous at $\boldsymbol{h} = \boldsymbol{0}$ if and only if $\rho_Z(\boldsymbol{h})$ is continuous at $\boldsymbol{h} = \boldsymbol{0}$, which implies that R_κ is mean-square continuous if and only if Z is mean-square continuous. Furthermore, the $2k$-th derivative of $\rho_{R_\kappa}(\boldsymbol{h})$ at $\boldsymbol{h} = \boldsymbol{0}$, (*i.e.*, $\rho_{R_\kappa}^{2k}(\boldsymbol{0})$), exist if $\rho_Z^{2k}(\boldsymbol{0})$ exist, which implies that R_κ is k-times mean-square differentiable if Z is k-times mean-square differentiable. Finally the sample paths continuity and differentiability of R_κ are the some of the parent Gaussian random field. □

Proof Lemma 2. 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 ${}_0F_1$ we have

$$\begin{aligned}
E(\tilde{R}^a(\boldsymbol{s}_i)\tilde{R}^b(\boldsymbol{s}_j)) &= \frac{2^{-\nu+2}\nu^\nu}{\Gamma^2\left(\frac{\nu}{2}\right)(1-\rho_G^2(\boldsymbol{h}))^{\nu/2}} \int_{\mathbb{R}_+^2} \tilde{r}_i^{-\nu+a-1}\tilde{r}_j^{-\nu+b-1} e^{-\frac{\nu}{2(1-\rho_G^2(\boldsymbol{h}))\tilde{r}_i^2}} e^{-\frac{\nu}{2(1-\rho_G^2(\boldsymbol{h}))\tilde{r}_j^2}} \\
&\times\ {}_0F_1\left(\frac{\nu}{2}; \frac{\nu^2\rho_G^2(\boldsymbol{h})}{4(1-\rho_G^2(\boldsymbol{h}))^2\tilde{r}_i^2\tilde{r}_j^2}\right) d\tilde{\boldsymbol{r}}_{ij} \\
&= \frac{2^{-\nu+2}\nu^\nu}{\Gamma^2\left(\frac{\nu}{2}\right)(1-\rho_G^2(\boldsymbol{h}))^{\nu/2}} \sum_{k=0}^{\infty} \int_{\mathbb{R}_+^2} \tilde{r}_i^{-\nu+a-2k-1}\tilde{r}_j^{-\nu+b-2k-1} e^{-\frac{\nu}{2(1-\rho_G^2(\boldsymbol{h}))\tilde{r}_i^2}} e^{-\frac{\nu}{2(1-\rho_G^2(\boldsymbol{h}))\tilde{r}_j^2}} \\
&\times\ \frac{1}{k!\left(\frac{\nu}{2}\right)_k}\left(\frac{\rho_G^2(\boldsymbol{h})\nu^2}{4(1-\rho_G^2(\boldsymbol{h}))}\right)^k d\tilde{\boldsymbol{r}}_{ij} \\
&= \frac{2^{-\nu+2}\nu^\nu}{\Gamma^2\left(\frac{\nu}{2}\right)(1-\rho_G^2(\boldsymbol{h}))^{\nu/2}} \sum_{k=0}^{\infty} \frac{I(k)}{k!\left(\frac{\nu}{2}\right)_k}\left(\frac{\rho_G^2(\boldsymbol{h})\nu^2}{4(1-\rho_G^2(\boldsymbol{h}))}\right)^k
\end{aligned} \tag{A.3}$$

where, using Fubini's Theorem

$$I(k) = \int_{\mathbb{R}_+} \tilde{r}_i^{-\nu+a-2k-1} e^{-\frac{\nu}{2(1-\rho_G^2(\boldsymbol{h}))\tilde{r}_i^2}} d\tilde{r}_i \int_{\mathbb{R}_+} \tilde{r}_j^{\nu+b-2k-1} e^{-\frac{\nu}{2(1-\rho_G^2(\boldsymbol{h}))\tilde{r}_j^2}} d\tilde{r}_j$$

Then using the univariate density (4.2), we obtain

$$I(k) = \Gamma\left(\frac{\nu-a}{2}+k\right)\Gamma\left(\frac{\nu-b}{2}+k\right) 2^{\frac{\nu+a}{2}+k-1} 2^{\frac{\nu+b}{2}+k-1} \left(\frac{(1-\rho_G^2(\boldsymbol{h}))}{\nu}\right)^{\frac{\nu-a}{2}+k} \left(\frac{(1-\rho_Z^2(\boldsymbol{h}))}{\nu}\right)^{\frac{\nu-b}{2}+k} \tag{A.4}$$

and combining equations (A.4) and (A.3), we obtain

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

Finally, using Euler transformation we obtain (4.7). □

Proof Theorem 16. If the weakly stationary Gaussian random fields G_i, $i = 1, \ldots, \nu$ and Z have a common correlation $\rho(\boldsymbol{h})$ then from (4.8) it is straightforward to see that Y_ν^* is also weakly stationary. By (Stein, 1999, Chapter 2.4), the mean-square continuity and the k-times mean-square differentiability of Y_ν^* are equivalent to the continuity and $2k$-times differentiability of $\rho_{Y_\nu^*}(\boldsymbol{h})$ at $\boldsymbol{h} = \boldsymbol{0}$, respectively. Define the function

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

Then $\rho_{Y_\nu^*}(\boldsymbol{h}) = g\{\rho(\boldsymbol{h})\}$ and for any $-1 \le x \le 1$, $g(x)$ is a continuous and infinitely differentiable function. Hence, $\rho_{Y_\nu^*}(\boldsymbol{h})$ is continuous at $\boldsymbol{h} = \boldsymbol{0}$ if and only if $\rho(\boldsymbol{h})$ is continuous at $\boldsymbol{h} = \boldsymbol{0}$, which implies that Y_ν^* is mean-square continuous if and only if G_i and Z are mean-square continuous. Furthermore, the $2k$-th derivative of $\rho_{Y_\nu^*}(\boldsymbol{h})$ at $\boldsymbol{h} = \boldsymbol{0}$, (i.e., $\rho_{Y_\nu^*}^{2k}(\boldsymbol{0})$), exist if $\rho^{2k}(\boldsymbol{0})$ exist, which implies that Y_ν^* is k-times mean-square differentiable if G_i and Z are k-times mean-square differentiable. Finally the sample paths continuity and differentiability of Y_ν^* are the some of the parent Gaussian random field. □

Proof Theorem 17. 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 ${}_0F_1$. Under the transformation $z_i = y_i\sqrt{w_i}$ and

$z_j = y_j\sqrt{w_j}$ with Jacobian $J((z_i,z_j)\to(y_i,y_j)) = (w_iw_j)^{1/2}$, we have:

$$
\begin{aligned}
f_{\boldsymbol{Y}^*_{ij}}(\boldsymbol{y}_{ij}) &= \int_{\mathbb{R}^2_+} f_{\boldsymbol{Z}_{ij}|\boldsymbol{W}_{ij}}(\boldsymbol{z}_{ij}|\boldsymbol{w}_{ij}) f_{\boldsymbol{W}_{ij}}(\boldsymbol{w}_{ij}) J d\boldsymbol{w}_{ij} \\
&= \frac{2^{-\nu}\nu^{\nu}}{2\pi\Gamma^2\left(\frac{\nu}{2}\right)(1-\rho^2(\boldsymbol{h}))^{(\nu+1)/2}} \int_{\mathbb{R}^2_+} (w_iw_j)^{(\nu+1)/2-1} e^{-\frac{1}{2(1-\rho^2(\boldsymbol{h}))}\left[w_iy_i^2+w_jy_j^2-2\rho(\boldsymbol{h})\sqrt{w_iw_j}y_iy_j\right]} e^{-\frac{\nu(w_i+w_j)}{2(1-\rho^2(\boldsymbol{h}))}} \\
&\times\ {}_0F_1\left(\frac{\nu}{2};\frac{\nu^2\rho^2(\boldsymbol{h})w_iw_j}{4(1-\rho^2(\boldsymbol{h}))^2}\right) d\boldsymbol{w}_{ij} \\
&= \frac{2^{-\nu}\nu^{\nu}}{2\pi\Gamma^2\left(\frac{\nu}{2}\right)(1-\rho^2(\boldsymbol{h}))^{(\nu+1)/2}} \int_{\mathbb{R}^2_+} (w_iw_j)^{(\nu+1)/2-1} e^{-\frac{1}{2(1-\rho^2(\boldsymbol{h}))}\left[y_i^2-2\rho(\boldsymbol{h})\sqrt{\frac{w_j}{w_i}}y_iy_j+\nu\right]w_i} e^{-\frac{(y_j^2+\nu)w_j}{2(1-\rho^2(\boldsymbol{h}))}} \\
&\times \sum_{k=0}^{\infty}\frac{1}{k!\left(\frac{\nu}{2}\right)_k}\left(\frac{\nu^2\rho^2(\boldsymbol{h})w_iw_j}{4(1-\rho^2(\boldsymbol{h}))^2}\right)^k d\boldsymbol{w}_{ij} \\
&= \frac{2^{-\nu}\nu^{\nu}}{2\pi\Gamma^2\left(\frac{\nu}{2}\right)(1-\rho^2(\boldsymbol{h}))^{(\nu+1)/2}} \sum_{k=0}^{\infty}\frac{I(k)}{k!\left(\frac{\nu}{2}\right)_k}\left(\frac{\nu^2\rho^2(\boldsymbol{h})}{4(1-\rho^2(\boldsymbol{h}))^2}\right)^k \qquad \text{(A.5)}
\end{aligned}
$$

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

$$
\begin{aligned}
I(k) &= \int_{\mathbb{R}_+} w_j^{(\nu+1)/2+k-1} e^{-\frac{(y_j^2+\nu)w_j}{2(1-\rho^2(\boldsymbol{h}))}} \left[\int_{\mathbb{R}_+} w_i^{(\nu+1)/2+k-1} e^{\left[-\frac{(y_i^2+\nu)}{2(1-\rho^2(\boldsymbol{h}))}w_i - \frac{\rho(\boldsymbol{h})\sqrt{w_j}y_iy_j}{(\rho^2(\boldsymbol{h})-1)}\sqrt{w_i}\right]} dw_i\right] dw_j \\
&= 2\left(\frac{y_i^2+\nu}{(1-\rho^2(\boldsymbol{h}))}\right)^{-\left(\frac{\nu+1}{2}+k\right)} \Gamma(\nu+1+2k)\int_{\mathbb{R}_+} w_j^{(\nu+1)/2+k-1} e^{\left[\frac{\rho^2(\boldsymbol{h})y_i^2y_j^2}{4(1-\rho^2(\boldsymbol{h}))(y_i^2+\nu)} - \frac{(y_j^2+\nu)}{2(1-\rho^2(\boldsymbol{h}))}\right]w_j} \\
&\times\ D_{-(\nu+1+2k)}\left(-\frac{\rho(\boldsymbol{h})y_iy_j\sqrt{w_j}}{\sqrt{(1-\rho^2(\boldsymbol{h}))(y_i^2+\nu)}}\right) dw_j \\
&= 2\left(\frac{y_i^2+\nu}{(1-\rho^2(\boldsymbol{h}))}\right)^{-\left(\frac{\nu+1}{2}+k\right)} \Gamma(\nu+1+2k)A(k) \qquad \text{(A.6)}
\end{aligned}
$$

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

$$
\begin{aligned}
D_{-(\nu+1+2k)}\left(-\frac{\rho(\boldsymbol{h})y_iy_j\sqrt{w_j}}{\sqrt{(1-\rho^2(\boldsymbol{h}))(y_i^2+\nu)}}\right) &= \frac{2^{-(\nu+1)/2+k}\sqrt{\pi}}{\Gamma\left(\frac{\nu}{2}+k+1\right)}e^{-\frac{\rho^2(\boldsymbol{h})y_i^2y_j^2w_j}{4(1-\rho^2(\boldsymbol{h}))(y_i^2+\nu)}} \\
&\times\ {}_1F_1\left(\frac{\nu+1}{2}+k;\frac{1}{2};\frac{\rho^2(\boldsymbol{h})y_i^2y_j^2w_j}{2(1-\rho^2(\boldsymbol{h}))(y_i^2+\nu)}\right) \\
&+\ \frac{2^{-\nu/2-k}\sqrt{\pi}\rho(\boldsymbol{h})y_iy_j\sqrt{w_j}}{\Gamma\left(\frac{\nu+1}{2}+k\right)\sqrt{(1-\rho^2(\boldsymbol{h}))(y_i^2+\nu)}}e^{-\frac{\rho^2(\boldsymbol{h})y_i^2y_j^2w_j}{4(1-\rho^2(\boldsymbol{h}))(y_i^2+\nu)}} \\
&\times\ {}_1F_1\left(\frac{\nu}{2}+k+1;\frac{3}{2};\frac{\rho^2(\boldsymbol{h})y_i^2y_j^2w_j}{2(1-\rho^2(\boldsymbol{h}))(y_i^2+\nu)}\right) \qquad \text{(A.7)}
\end{aligned}
$$

combining equations (A.7) and the integral of (A.6) and using (7.621.4) of Gradshteyn and Ryzhik (2007), we obtain

$$
\begin{aligned}
A(k) &= \int_{\mathbb{R}_+} w_j^{(\nu+1)/2+k-1}e^{-\frac{(y_j^2+\nu)}{2(1-\rho^2(\boldsymbol{h}))}w_j}{}_1F_1\left(\frac{\nu+1}{2}+k;\frac{1}{2};\frac{\rho^2(\boldsymbol{h})y_i^2y_j^2w_j}{2(1-\rho^2(\boldsymbol{h}))(y_i^2+\nu)}\right)dw_j \\
&+ \int_{\mathbb{R}_+} w_j^{\nu/2+k+1-1}e^{-\frac{(y_j^2+\nu)}{2(1-\rho^2(\boldsymbol{h}))}w_j}{}_1F_1\left(\frac{\nu}{2}+k+1;\frac{3}{2};\frac{\rho^2(\boldsymbol{h})y_i^2y_j^2w_j}{2(1-\rho^2(\boldsymbol{h}))(y_i^2+\nu)}\right)dw_j \\
&= \Gamma\left(\frac{\nu+1}{2}+k\right)\left(\frac{y_j^2+\nu}{2(1-\rho^2(\boldsymbol{h}))}\right)^{-\frac{(\nu+1)}{2}-k}{}_2F_1\left(\frac{\nu+1}{2}+k,\frac{\nu+1}{2}+k;\frac{1}{2};\frac{\rho^2(\boldsymbol{h})y_i^2y_j^2}{(y_i^2+\nu)(y_j^2+\nu)}\right) \\
&+ \Gamma\left(\frac{\nu}{2}+k+1\right)\left(\frac{y_j^2+\nu}{2(1-\rho^2(\boldsymbol{h}))}\right)^{-\frac{\nu}{2}-k-1}{}_2F_1\left(\frac{\nu}{2}+k+1,\frac{\nu}{2}+k+1;\frac{3}{2};\frac{\rho^2(\boldsymbol{h})y_i^2y_j^2}{(y_i^2+\nu)(y_j^2+\nu)}\right) \qquad \text{(A.8)}
\end{aligned}
$$

finally, combining equations (A.8), (A.6) and (A.5), we obtain

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

and using the following representation of the F_4 Appell function (Brychkov and Saad, 2017)

$$
F_4(a,b;c,c';w,z)=\sum_{k=0}^{\infty}\frac{(a)_k(b)_kz^k}{k!(c')_k}{}_2F_1(a+k,b+k;c;w),\quad |\sqrt{w}|+|\sqrt{z}|<1 \qquad \text{(A.9)}
$$

we obtain (4.10). □

Proof Theorem 18. Consider $\tilde{\boldsymbol{U}} = (\tilde{U}(\boldsymbol{s}_1),\ldots,\tilde{U}(\boldsymbol{s}_n))^T$, $\boldsymbol{V} = (|X_1(\boldsymbol{s}_1)|,\ldots,|X_1(\boldsymbol{s}_n)|)^T$, $\boldsymbol{Q} = (X_2(\boldsymbol{s}_1),\ldots,X_2(\boldsymbol{s}_n))^T$ where $\boldsymbol{X}_k = (X_k(\boldsymbol{s}_1),\ldots,X_k(\boldsymbol{s}_n))^T \sim N_n(\boldsymbol{0},\Omega)$, for $k = 1,2$, which are assumed to be independent. By definition of the skew-Gaussian random field in (4.11) we have:

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

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

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

To solve this integral we need $f_{\boldsymbol{V}}(\boldsymbol{v})$, *i.e.*, the joint density of $\boldsymbol{V} = (|X_1(\boldsymbol{s}_1)|,\ldots,|X_1(\boldsymbol{s}_n)|)^T$. Let $\boldsymbol{X}_k = (X_1,\ldots,X_n)^T = (X_1(\boldsymbol{s}_1),\ldots,X_1(\boldsymbol{s}_n))^T$ and $\boldsymbol{V} = (|X_1|,\ldots,|X_n|)^T$. Also, consider the diagonal matrices $\boldsymbol{D}(\boldsymbol{l}) = \text{diag}\{l_1,\ldots,l_n\}$, with $\boldsymbol{l} = (l_1,\ldots,l_n) \in \{-1,+1\}^n$, which are such that $\boldsymbol{D}(\boldsymbol{l})^2 = \boldsymbol{D}(\boldsymbol{l})$ for all $\boldsymbol{l} \in \{-1,+1\}^n$. 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

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

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

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

where the last identity comes from the fact that $\Omega_{-l}=\boldsymbol{D}(-\boldsymbol{l})\Omega\boldsymbol{D}(-\boldsymbol{l})=\boldsymbol{D}(\boldsymbol{l})\Omega\boldsymbol{D}(\boldsymbol{l})=\Omega_l$ for all $\boldsymbol{l}\in\{-1,+1\}^n$; e.g. for $n=3$ the sum must be done 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 $\tilde{\boldsymbol{U}}$ is thus given by

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

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

□

Proof Theorem 20. Let $|\boldsymbol{Y}^*_{ij}|$ where the pdf of bivariate random vector $\boldsymbol{Y}^*_{ij}$ is given by the equation (4.10), then we have:

$$\begin{aligned} f_{|\boldsymbol{Y}^*_{ij}|}(|\boldsymbol{y}_{ij}|) &= f_{\boldsymbol{Y}_{ij}}(y_i,y_j)+f_{\boldsymbol{Y}_{ij}}(-y_i,y_j)+f_{\boldsymbol{Y}_{ij}}(y_i,-y_j)+f_{\boldsymbol{Y}_{ij}}(-y_i,-y_j) \\ &= \frac{\nu^\nu[(y_i^2+\nu)(y_j^2+\nu)]^{-(\nu+1)/2}\Gamma^2\left(\frac{\nu+1}{2}\right)}{\pi\Gamma^2\left(\frac{\nu}{2}\right)(1-\rho^2(\boldsymbol{h}))^{-(\nu+1)/2}} F_4\left(\frac{\nu+1}{2},\frac{\nu+1}{2},\frac{1}{2},\frac{\nu}{2};\frac{\rho^2(\boldsymbol{h})y_i^2y_j^2}{(y_i^2+\nu)(y_j^2+\nu)},\frac{\nu^2\rho^2(\boldsymbol{h})}{(y_i^2+\nu)(y_j^2+\nu)}\right) \end{aligned}$$

□

Proof Lemma 3. Let $X_i=|Y^*_\nu(\boldsymbol{s}_i)|$ and $X_j=|Y^*_\nu(\boldsymbol{s}_j)|$. Then using (4.22) we have

$$\begin{aligned} E(X_i^aX_j^b) &= \frac{4\nu^\nu\Gamma^2\left(\frac{\nu+1}{2}\right)}{\pi\Gamma^2\left(\frac{\nu}{2}\right)(1-\rho^2(\boldsymbol{h}))^{-\frac{(\nu+1)}{2}}}\int_{\mathbb{R}^2_+} x_i^ax_j^b[(x_i^2+\nu)(x_j^2+\nu)]^{-\frac{(\nu+1)}{2}} \\ &\times F_4\left(\frac{\nu+1}{2},\frac{\nu+1}{2},\frac{1}{2},\frac{\nu}{2};\frac{\rho^2(\boldsymbol{h})x_i^2x_j^2}{(x_i^2+\nu)(x_j^2+\nu)},\frac{\nu^2\rho^2(\boldsymbol{h})}{(x_i^2+\nu)(x_j^2+\nu)}\right)d\boldsymbol{x}_{ij} \\ &= \frac{4\nu^\nu\Gamma^2\left(\frac{\nu+1}{2}\right)}{\pi\Gamma^2\left(\frac{\nu}{2}\right)(1-\rho^2(\boldsymbol{h}))^{-\frac{\nu+1}{2}}}\sum_{k=0}^{\infty}\frac{\left(\frac{\nu+1}{2}\right)_k^2}{k!\left(\frac{\nu}{2}\right)_k}(\nu^2\rho^2(\boldsymbol{h}))^kI(k) \end{aligned} \tag{A.10}$$

where, using Fubini's Theorem and (3.251.11) of Gradshteyn and Ryzhik (2007), we obtain

$$\begin{aligned} I(k) &= \int_{\mathbb{R}^2_+} x_i^ax_j^b[(x_i^2+\nu)(x_j^2+\nu)]^{-\frac{(\nu+1)}{2}-k}{}_2F_1\left(\frac{\nu+1}{2}+k,\frac{\nu+1}{2}+k;\frac{1}{2};\frac{\rho^2(\boldsymbol{h})x_i^2x_j^2}{(x_i^2+\nu)(x_j^2+\nu)}\right)d\boldsymbol{x}_{ij} \\ &= \sum_{l=0}^{\infty}\frac{\left(\frac{\nu+1}{2}+k\right)_l^2\rho^{2l}(\boldsymbol{h})}{l!\left(\frac{1}{2}\right)_l}\left[\int_{\mathbb{R}_+} x_i^{a+2l}(x_i^2+\nu)^{-\frac{(\nu+1)}{2}-k-l}dx_i\int_{\mathbb{R}_+} x_j^{b+2l}(x_j^2+\nu)^{-\frac{(\nu+1)}{2}-k-l}dx_j\right] \\ &= \sum_{l=0}^{\infty}\frac{\left(\frac{\nu+1}{2}+k\right)_l^2\rho^{2l}(\boldsymbol{h})\nu^{\frac{(a+b)}{2}-\nu-2k}\Gamma\left(\frac{a+1}{2}+l\right)\Gamma\left(\frac{\nu-a}{2}+k\right)\Gamma\left(\frac{b+1}{2}+l\right)\Gamma\left(\frac{\nu-b}{2}+k\right)}{4l!\left(\frac{1}{2}\right)_l\Gamma^2\left(\frac{\nu+1}{2}+k+l\right)} \\ &= \frac{\nu^{\frac{(a+b)}{2}-\nu-2k}\Gamma\left(\frac{a+1}{2}\right)\Gamma\left(\frac{\nu-a}{2}\right)\left(\frac{\nu-a}{2}\right)_k\Gamma\left(\frac{b+1}{2}\right)\Gamma\left(\frac{\nu-b}{2}\right)\left(\frac{\nu-b}{2}\right)_k}{4\Gamma^2\left(\frac{\nu+1}{2}+k\right)}\sum_{l=0}^{\infty}\frac{\left(\frac{a+1}{2}\right)_l\left(\frac{b+1}{2}\right)_l\rho^{2l}(\boldsymbol{h})}{l!\left(\frac{1}{2}\right)_l} \end{aligned} \tag{A.11}$$

and combining equations (A.11) and (A.10), we obtain

$$E(X_i^aX_j^b)=\frac{\nu^{\frac{(a+b)}{2}}\Gamma\left(\frac{a+1}{2}\right)\Gamma\left(\frac{\nu-a}{2}\right)\Gamma\left(\frac{b+1}{2}\right)\Gamma\left(\frac{\nu-b}{2}\right)}{\pi\Gamma^2\left(\frac{\nu}{2}\right)(1-\rho^2(\boldsymbol{h}))^{-\frac{\nu+1}{2}}}{}_2F_1\left(\frac{\nu-a}{2},\frac{\nu-b}{2};\frac{\nu}{2};\rho^2(\boldsymbol{h})\right){}_2F_1\left(\frac{\nu+1}{2},\frac{\nu+1}{2};\frac{1}{2};\rho^2(\boldsymbol{h})\right)$$

Finally, using Euler transformation we obtain (4.23). □

Appendix B

All the simulated and estimates have been carried out using un upcoming version of the R package `GeoModels`. Here we present the code for simulating and estimating with maximum likelihood spatial Weibull random field with Generalized Wendland model with $\tau = 0$.

B.1 Code example

B.1.1 Simulation and estimated spatial Weibull random field with Generalized Wendland model with $\tau = 0$

```
library(GeoModels)
library(spam)
model="Weibull"
# Define the spatial-coordinates of the points:
set.seed(368)
NN=1000
x <- runif(NN, 0, 1)
y <- runif(NN, 0, 1)
coords <- cbind(x,y)

# Set the covariance models parameters:
corrmodel <- "Wend0";sparse=FALSE
mean <- 0.5
mean1=0.25
```

```
sill <- 1

### must be fixed
nugget <- 0
scale <- 0.15
power2=4
shape=10
X=cbind(rep(1,nrow(coords)),runif(nrow(coords)))
param<-list(mean=mean,mean1=mean1,sill=sill,nugget=nugget,
       scale=scale,shape=shape,power2=power2)

# Simulation of the spatial Weibull random field:
set.seed(32)
k=1
nsim=1000
res=res2=NULL

par(mfrow=c(1,2))
while(k<nsim){
data <-GeoSim(coordx=coords,corrmodel=corrmodel,param=param,
       model=model,X=X)$data

fixed<-list(nugget=nugget, sill=sill,power2=power2) ### note that
sill must be fixed

# Starting value for the estimated parameters
start<-list(scale=scale,shape=shape,mean=mean,mean1=mean1)

# Maximum pairwise composite-likelihood fitting of the random
field:

fit2 <- GeoFit(data=data,coordx=coords,corrmodel=corrmodel,
        maxdist=0.1,likelihood="Marginal",type="Pairwise",model=model,
        start=start,fixed=fixed,X=X)
```

```
if(fit2$convergence=="Successful"){ res=rbind(res,fit2$param)
k=k+1}
print(k) }
```