**Noname manuscript No.**
(will be inserted by the editor)

# A class of random fields with two-piece marginal distributions for modeling point-referenced data with spatial outliers

**Moreno Bevilacqua · Christian Caamaño-Carrillo · Reinaldo B. Arellano-Valle · Camilo Gómez**

Received: date / Accepted: date

**Abstract** In this paper, we propose a new class of non-Gaussian random fields named two-piece random fields. The proposed class allows to generate random fields that have flexible marginal distributions, possibly skewed and/or heavy-tailed and, as a consequence, has a wide range of applications. We study the second-order properties of this class and provide analytical expressions for the bivariate distribution and the associated correlation functions. We exemplify our general construction by studying two examples: two-piece Gaussian and two-piece Tukey-$h$ random fields.
An interesting feature of the proposed class is that it offers a specific type of dependence that can be useful when modeling data displaying spatial outliers, a property that has been somewhat ignored from modeling viewpoint in the literature for spatial point referenced data.
Since the likelihood function involves analytically intractable integrals, we adopt the weighted pairwise likelihood as a method of estimation. The effectiveness of our methodology is illustrated with simulation experiments as well as with the analysis of a georeferenced dataset of mean temperatures in Middle East.

**Keywords** Asymmetric random fields · Composite likelihood · Spatial outliers · Tukey-$h$ distribution

Facultad de Ingeniería y Ciencias, Universidad Adolfo Ibáñez
Av. Padre Hurtado 750, Viña del Mar, Chile
E-mail: moreno.bevilacqua@uai.cl

Departamento de Estadística, Universidad del Bío-Bío
Av. Collao 1202, Concepción, Chile

Departamento de Estadística, Pontificia Universidad Católica de Chile
Vicuña Mackenna 4860, Santiago, Chile

Departamento de Estadística, Universidad of Valparaíso
Gran Bretaña 1091, Valparaíso, Chile

## 1 Introduction

Today, statistical analyses with one or more variables indexed in space and/or time are useful in different areas, such as environmental sciences and engineering fields (Banerjee et al., 2004; Cressie and Wikle, 2011). Gaussian random fields (RFs) play a central role in providing a building block model for these kinds of data. Although the use of Gaussian RF models simplifies spatial analysis, the normality assumption might be overly restrictive in obtaining an accurate representation of the data structure. For instance, in many geostatistical applications, such as oceanography, the environment and the study of natural resources, the assumption of Gaussianity is unrealistic because the observed data can be highly skewed and/or can display heavy tails. On the other hand, although the marginal Gaussian distribution can be realistic, the type of dependence induced by the multivariate Gaussian distribution can be too restrictive in some circumstances.

In recent years, different approaches have been proposed in order to analyse non-Gaussian spatial data. A general approach is the class of transformed Gaussian RF(s) obtained by a nonlinear transformation of one or independent copies of a Gaussian RF. This is a convenient approach since this kind of construction guarantees that the correlation of the non-Gaussian RF depends on the correlation of the underlying Gaussian RF that can be modeled using flexible correlation models such as the Matern (Stein, 1999) of the Generalized Wendland model (Bevilacqua et al., 2019a). In addition, depending on the type of transformation, the geometric properties of the non-Gaussian RF can be inherited from the underlying Gaussian RF. Some examples of this class can be found in DeOliveira (2006) for log-Gaussian marginal distribution, Zhang and El-Shaarawi (2010), for skew-Gaussian marginal distribution, Xua and Genton (2017) for Tukey $g-h$ marginal distribution, Bevilacqua et al. (2021) for skew-$t$ marginal distribution, just to mention a few. A general method that allows to construct RFs with arbitrary marginal distribution and falls into the class of transformed Gaussian RF(s) is the Gaussian Copula (Kazianka and Pilz, 2010; Masarotto and Varin, 2012; Gräler, 2014).

Another general approach that has been widely used for non-Gaussian spatial data is the spatial generalized linear mixed models as proposed in Diggle et al. (1998). Under this framework, non-Gaussian models for spatial data can be specified using a link function and a latent Gaussian RF through a conditionally independent assumption. However these kind of models have some drawbacks. For instance, the resulting non-Gaussian RFs have a "forced" nugget effect that implies no mean square continuity (Gelfand and Schliep, 2016). This makes spatial generalized linear mixed models unsuitable when analysing spatial data exhibiting continuous realization behaviour.

In this paper, we propose a novel general class of RFs defined on the real line with flexible two-piece marginal distributions (Fernández and Steel, 1998; Arellano-Valle et al., 2005; Rubio and Steel, 2020). Two-piece distribution is a general mechanism that allows to obtain skewed distributions by transforming a symmetric distribution (Jones, 2015). In the univariate case, there is exten-

sive literature on two-piece distributions (Fechner, 1897; Fernández and Steel, 1998; Mudholkar and Hutson, 2000; Wallis, 2014). Our idea relies on generalizing the stochastic representation of the two-piece random variable given in Arellano-Valle et al. (2005) to obtain RFs with marginal distributions of this type. We exemplify our general construction by studying two examples: two-piece Gaussian RFs that allow to obtain (a)symmetric marginal distributions and two-piece Tukey-$h$ RFs (Xua and Genton, 2017) that allow to obtain marginal distributions with both (a)symmetric and heavy tails. The proposed examples are obtained by transforming two independent Gaussian RFs, i.e. they fall into the class of transformed Gaussian RFs.

Besides the flexible marginal distribution behavior of the proposed class, an additional interesting feature is that it offers a specific type of dependence that can be useful when modeling data displaying spatial outliers. A spatial outlier is a spatially referenced variable whose non-spatial attribute values is significantly different from those of other spatially referenced variables in its spatial neighborhood. As outlined in Chen et al. (2008), in contrast to traditional outliers, spatial outliers do not necessarily deviate from the remainder of the whole data set. In this regard they are also called "local outliers" since they are derived from spatial local comparisons. Different algorithms have been proposed for the detection of spatial outliers, in particular in the area of spatial data mining (Haslett et al., 1991; Haining, 1993; Shekhar et al., 2001; Lu et al., 2003; Kou et al., 2007; Chen et al., 2008; Ernst and Haesbroeck, 2017; Singh and Lalitha, 2018). However, from modeling point of view this kind of feature has been somewhat ignored in the literature at least for spatial point referenced data and there are no spatial RF models able to capture spatial outliers to the best of our knowledge.

To better explain the presence of spatial outliers in a spatial point referenced dataset, let us consider a realization of three zero-mean and unit-variance RFs with Gaussian marginal distribution, at $\mathbf{s}_1, \ldots, \mathbf{s}_n$, $n = 3000$ location sites uniformly distributed in the unit square. In particular we consider:

(a) a Gaussian RF with an exponential correlation, *i.e.* $\rho(r) = e^{-3r/0.5}$;
(b) a Gaussian RF with an exponential correlation with a nugget effect, *i.e.* $\rho(r) = e^{-3r/0.5}0.7 + 0.3\mathbb{1}_0(r)$;
(c) a two-piece Gaussian RF (see Section 3) when the parameter of asymmetry is set to zero and the underlying correlation is $\rho(r) = e^{-3r/0.5}$.

Here $r$ is the euclidean distance and $\mathbb{1}_x(y) = 1$ if $x = y$ and 0 otherwise is the indicator function. The boxplots of the three datasets in Figure 1 (first row) shows that the three realizations 'look' Gaussian, as expected.
A useful graphic tool to visualize the presence of spatial outlier is the $h-$scatterplot. An $h$-scatterplot plots all possible pairs of data values $z(\mathbf{s}_i), \ldots, z(\mathbf{s}_n)$ whose locations have distances belonging to a certain interval i.e. considers the plot of the pairs $\{(z(\mathbf{s}_i), z(\mathbf{s}_j)) : a - \epsilon \leq ||\mathbf{s}_i - \mathbf{s}_j|| \leq a + \epsilon\}$ with $a > 0$, $\epsilon \geq 0$. Alternatively, an $h$-scatterplot based on neighborhoods can be considered. In this case the pairs involved in the plot are $\{(z(\mathbf{s}_i), z(\mathbf{s}_j)) : \mathbf{s}_i \in N_m(\mathbf{s}_j)\}$ where

$N_m(\mathbf{s}_l)$ is the $m$ nearest neighbor set of $z(\mathbf{s}_l)$. We use the second type of plot through the paper.

In Figure 1 (second row), the $h$-scatterplot of case c) shows a clear different type of dependence with the respect to the case a) and b). In particular the presence of spatial outliers can be appreciated since for a fixed neighborhood order $m = 5, 15, 30, 60$, the majority of the pairs are concordant i.e. positive (negative) values of one value tend to be associated with positive (negative) values of the other value. However, some pairs display a discordant dependence resulting in a specific $X$-shape of the $h$-scatterplots (this is more apparent for small neighborhood that is $m = 5, 15$). As outlined by a Referee, some copula models as the $t$ copula (Joe, 2014) or the copula proposed in Cote and Genest (2019) exhibit the same type of dependence. However, as shown in Genton and Zhang (2012) these kind of models are not identifiable with just one realization of the RF which is the typical case for spatial data.

To confirm the presence of spatial outliers suggested by the $h$-scatterplots, we apply one of the algorithm proposed in Chen et al. (2008) that allows to detect spatial outliers. In particular, we use the median algorithm which is a variant of the algorithm proposed originally in Shekhar et al. (2001). This algorithm, for each point referenced data $z(\mathbf{s}_i)$, considers the median $\texttt{Med}_k(\mathbf{s}_i)$ of $N_k(\mathbf{s}_i)$. Then, after standardizing the residuals $z(\mathbf{s}_i) - \texttt{Med}_k(\mathbf{s}_i)$, the value $z(\mathbf{s}_i)$ is detected as spatial outlier if the absolute value of the associated standardized residual is bigger than $q_{1-\alpha/2}$ where $q_x$ is the quantile of order $x$ a standard Gaussian distribution. In Figure 1 (third row), we highlight the detected spatial outliers with a small black circle (we set $\alpha = 0.001$ and $k = 15$ in the algorithm). For the cases a) and b) the algorithm detects only 5 and 2 spatial outliers points respectively. However for the case c), *i.e* the proposed two-piece Gaussian RF, the algorithm detects 71 spatial outliers.

This simple illustrative example shows that the proposed two-piece Gaussian RF, which is a special case of a more general class studied in Section 2, is able to generate datasets with spatial outliers with respect to a Gaussian RF irrespective of the nugget effect. As a consequence, it can be a suitable model when analysing point referenced spatial data displaying a certain number of spatial outliers.

Regarding the estimation of these kind of models, it must be said that the likelihood function involves analytically intractable integrals and likelihood-based estimation methods are unfeasible. For this reason we consider the method of composite likelihood (Lindsay, 1988; Varin et al., 2011) that has been successfully applied in the recent years when estimating complex models. Composite likelihood is a a general class of objective functions based on the likelihood of marginal or conditional events. This kind of estimation method can be helpful when it is difficult to evaluate or to specify the full likelihood. In particular the weighted pairwise likelihood (*wpl*) method has been widely used for the estimation of complex non-Gaussian RFs because in many circumstances the multivariate distribution is unknown and/or difficult to compute but the bivariate density is known and relatively simple to evaluate. Some examples are the Bernoulli RF in Heagerty and Lele (1998) or the general

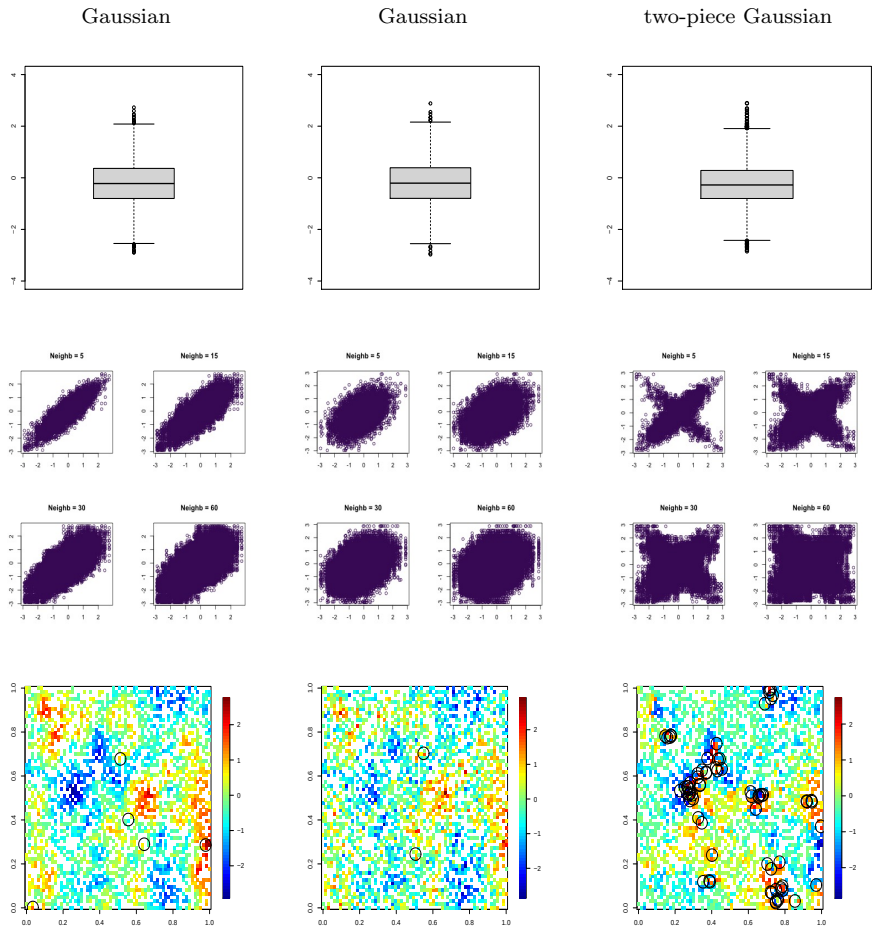



**Fig. 1** First row: boxplots of the simulated datasets. Second row: four $h$-scatterplots of the simulated datasets. Third row: color map of the simulated datasets with detected spatial outliers. The graphics are associated with a Gaussian RF (first column), a Gaussian RF with a nugget effect (second column), a two-piece Gaussian RF (third column).

class of spatial generalized mixed models (Varin and Vidoni, 2005) of the $t$ RF proposed in Bevilacqua et al. (2021) just to name a few. Other examples can be found in Feng et al. (2014) and Alegria et al. (2017).

Exploiting the bivariate distributions of the proposed two-piece Gaussian and two-piece Tukey-$h$ RFs, we explore in a simulation study the use of the *wpl* method for the joint estimation of the skewness, tail and correlation dependence parameters. Additionally, we compare the performance of the optimal linear predictor of the two-piece Gaussian RF with the optimal predictor of the (misspecified) Gaussian RF.

Finally we apply the proposed methodology by analysing a real georeferenced dataset of mean temperatures in Middle East where the proposed two-piece Gaussian and two-piece Tukey-$h$ RF models are compared with the standard Gaussian and Tukey-$h$ RF models. The methods proposed in this paper are implemented in the `R` (R Core Team, 2020) package `GeoModels` (Bevilacqua et al., 2019b) and R code for reproducing the work is available as an online supplement.

The remainder of the paper is organized as follows. In Section 2, we introduce the general class of RFs with two-piece marginal distributions, and we study the associated second-order properties and bivariate distributions. In Section 3, we present the first example of the proposed class, *i.e.* the two-piece Gaussian RF. In Section 4, we study the second example of the proposed class, *i.e.* the two-piece Tukey-$h$ RF. In Section 5, we present two simulation studies that investigates the performance of the *wpl* method when estimating the proposed two-piece RFs and compares the performance of the optimal linear predictor of the two-piece Gaussian RF with the optimal predictor of the (misspecified) Gaussian RF. In Section 6, we analyse a real dataset of mean temperatures in Middle East. Finally, in Section 7, we present some conclusions.

## 2 A class of random fields with two-piece marginal distributions

For the rest of the paper, given a weakly stationary RF $Q = \{Q(\boldsymbol{s}), \boldsymbol{s} \in A\}$ we denote by $\rho_Q(\boldsymbol{h}) = Corr(Q(\boldsymbol{s}_i), Q(\boldsymbol{s}_j))$ its correlation function, where $\boldsymbol{h} = \boldsymbol{s}_i - \boldsymbol{s}_j$ is the lag separation vector.

For any set of distinct points $(\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n)^T$, $n \in \mathcal{N}$, we denote by $\boldsymbol{Q}_{ij} = (Q(\boldsymbol{s}_i), Q(\boldsymbol{s}_j))^T$, $i \neq j$, the bivariate random vector and by $\boldsymbol{Q} = (Q(\boldsymbol{s}_1), \ldots, Q(\boldsymbol{s}_n))^T$ the multivariate random vector. Moreover, we denote with $f_Q$ and $F_Q$ the marginal probability density function (pdf) and cumulative distribution function (cdf) of $Q(\boldsymbol{s})$, respectively, with $f_{\boldsymbol{Q}_{ij}}$, the pdf of $\boldsymbol{Q}_{ij}$. Finally with $f_{|\boldsymbol{Q}_{ij}|}$ we denote the density of $(|Q(\boldsymbol{s}_i)|, |Q(\boldsymbol{s}_j)|)^T$.

Let $G = \{G(\boldsymbol{s}), \boldsymbol{s} \in A\}$, be a zero-mean and unit variance Gaussian RF with correlation $\rho_G(\boldsymbol{h})$ and with some abuse of notation, we set $\rho(\boldsymbol{h}) := \rho_G(\boldsymbol{h})$. Since, in what follows, the proposed non-Gaussian RFs are obtained through the transformation of (independent copies of) $G$, henceforth, we call $G$ and $\rho(\boldsymbol{h})$ the underlying Gaussian RF and underlying correlation model, respectively.

Our proposal considers a general class of RFs with an (a)symmetric marginal distribution of the two-piece type (Arellano-Valle et al., 2005). The basic idea behind these kind of marginal (a)symmetric models is to form a distribution by joining two half symmetric distributions with different scale parameters.

Let $H = \{H(\boldsymbol{s}), \boldsymbol{s} \in A\}$ be a stationary RF, and consider the function $g(\eta) := \frac{a(\eta)}{a(\eta)+b(\eta)}$, where $a(\eta)$ and $b(\eta)$ are known and positive asymmetry functions, and $\eta$ is the asymmetry parameter defined on some subset of $\mathbb{R}$. The first step of our construction considers a discrete RF $K_\eta = \{K_\eta(\boldsymbol{s}), \boldsymbol{s} \in A\}$

| $K_\eta(\boldsymbol{s}_i)$ \ $K_\eta(\boldsymbol{s}_j)$ | | $a(\eta)$ | $-b(\eta)$ | |
|---|---|---|---|---|
| | | $a(\eta)$ | $-b(\eta)$ | |
| $a(\eta)$ | $p(\boldsymbol{h})$ | $g(\eta)-p(\boldsymbol{h})$ | $g(\eta)$ | |
| $-b(\eta)$ | $g(\eta)-p(\boldsymbol{h})$ | $1-2g(\eta)+p(\boldsymbol{h})$ | $1-g(\eta)$ | |
| | $g(\eta)$ | $1-g(\eta)$ | 1 | |

**Table 1** Bivariate distribution of $\boldsymbol{K}_{\eta;ij} = (K_\eta(\boldsymbol{s}_i), K_\eta(\boldsymbol{s}_j))^T$.

defined as:

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

where $q_x$ is the quantile of order $x$ of the marginal distribution of $H$.

Then, by construction, $Pr(K_\eta(\boldsymbol{s}) = a(\eta)) = g(\eta) = 1 - Pr(K_\eta(\boldsymbol{s}) = -b(\eta))$. In principle, $H$ can be any type of stationary RF; however, in what follows, we assume that $H \equiv G$, the underlying Gaussian RF with correlation model $\rho(\boldsymbol{h})$.

Under this setting, $\mathbb{E}(K_\eta(\boldsymbol{s})) = a(\eta) - b(\eta)$, $Var(K_\eta(\boldsymbol{s})) = a(\eta)b(\eta)$ and, from the bivariate distribution of $\boldsymbol{K}_{\eta;ij}$ in Table 1, we have:

$$\mathbb{E}(K_\eta(\boldsymbol{s}_i)K_\eta(\boldsymbol{s}_j)) = (a(\eta)+b(\eta))^2 p(\boldsymbol{h}) - 2a(\eta)b(\eta)g(\eta) + b^2(\eta)(1-2g(\eta)), \tag{2}$$

where

$$p(\boldsymbol{h}) := Pr(K_\eta(\boldsymbol{s}_i) = a(\eta), K_\eta(\boldsymbol{s}_j) = a(\eta)) = \Phi_2\left(q_{g(\eta)}, q_{g(\eta)}; \rho(\boldsymbol{h})\right). \tag{3}$$

Here $\Phi_2(\cdot;\cdot,x)$ is the cdf of a bivariate standard Gaussian distribution with correlation $x$. Note that $0.25 \leq p(\boldsymbol{h}) \leq 0.5$ and it is approximatively equals to 0.25 (0.5) for very low correlations (high correlations).

The second step of our construction considers $X = \{X(\boldsymbol{s}), \boldsymbol{s} \in A\}$, a real-valued stationary zero-mean RF independent of $K_\eta$ with symmetric marginal distributions.

The final step is a generalization of the stochastic representation given in Arellano-Valle et al. (2005), that is our proposal considers a class of RFs $P_\eta = \{P_\eta(\boldsymbol{s}), \boldsymbol{s} \in A\}$ defined using the two RFs defined at the first and second step:

$$P_\eta(\boldsymbol{s}) := |X(\boldsymbol{s})| K_\eta(\boldsymbol{s}). \tag{4}$$

This construction defines a RF with a marginal distribution belonging to a general class of marginal (a)symmetric distributions of the two-piece type that includes the models proposed in Fernández and Steel (1998) and Mudholkar and Hutson (2000) as special cases.

The choice of the marginally symmetric RF $X$ in (4) leads to different RFs with an (a)symmetric marginal distributions and possibly heavy tails. The marginal *pdf* of $P_\eta$ is given by:

$$f_{P_\eta}(p) = \frac{1}{a(\eta)+b(\eta)} \left[ f_{|X(\boldsymbol{s})|}\left(\frac{p}{a(\eta)}\right) I_{[0,\infty)}(p) + f_{|X(\boldsymbol{s})|}\left(-\frac{p}{b(\eta)}\right) I_{(-\infty,0)}(p) \right] \tag{5}$$

or equivalently

$$f_{P_\eta}(p) = \frac{2}{a(\eta)+b(\eta)}\left[f_{X(\boldsymbol{s})}\left(\frac{p}{a(\eta)}\right) I_{[0,\infty)}(p) + f_{X(\boldsymbol{s})}\left(\frac{p}{b(\eta)}\right) I_{(-\infty,0)}(p)\right] \tag{6}$$

where $I_A(x)$ denotes the indicator function of the set $A$. Note that the special symmetric case is obtained when $a(\eta) = b(\eta)$.

The marginal mean and variance of the general two-piece class of RFs are respectively given by:

$$\mathbb{E}(P_\eta(\boldsymbol{s})) = (a(\eta) - b(\eta))\mathbb{E}(|X(\boldsymbol{s})|), \tag{7}$$

$$Var(P_\eta(\boldsymbol{s})) = (a^2(\eta)+b^2(\eta)-a(\eta)b(\eta))\mathbb{E}(|X(\boldsymbol{s})|^2)-(a(\eta)-b(\eta))^2(\mathbb{E}(|X(\boldsymbol{s})|))^2, \tag{8}$$

For $n$ sites, $(\boldsymbol{s}_1,\dots,\boldsymbol{s}_n)^T$, the joint distribution of the random vector $\boldsymbol{P} = (P_\eta(\boldsymbol{s}_1),\dots,P_\eta(\boldsymbol{s}_k))^T$ depends on the joint distribution of the random vector $\boldsymbol{K} = (K_\eta(\boldsymbol{s}_1),\dots,K_\eta(\boldsymbol{s}_k))^T$ that can be derived from the multivariate Gaussian cdf $\Phi_n(\cdot;R_n)$, with $R_n = [\rho(\boldsymbol{s}_i-\boldsymbol{s}_j)]_{i,j=1}^n$. While the probability can be estimated by fast and accurate quadrature for small $n$ (Genz and Bretz, 2009), when $n$ is large it requires Monte Carlo sampling (see Azzimonti and Ginsbourger, 2018, for instance). This fact makes a likelihood approach impractical from computational viewpoint when estimating the parameters of $P_\eta$.

For this reason, we address the estimation problem using *wpl* (see Section 4). This method of estimation requires the knowledge of the bivariate *pdf* that in our case is given by:

$$\begin{aligned}
f_{P_{\eta;ij}}(p_i,p_j) = &\frac{p(\boldsymbol{h})}{a^2(\eta)} f_{|\boldsymbol{X}_{ij}|}\left(\frac{p_i}{a(\eta)},\frac{p_j}{a(\eta)}\right) I_{[0,\infty)\times[0,\infty)}(p_i,p_j)\\
&+ \frac{g(\eta)-p(\boldsymbol{h})}{a(\eta)b(\eta)} f_{|\boldsymbol{X}_{ij}|}\left(\frac{p_i}{a(\eta)},-\frac{p_j}{b(\eta)}\right) I_{[0,\infty)\times(-\infty,0)}(p_i,p_j)\\
&+ \frac{g(\eta)-p(\boldsymbol{h})}{a(\eta)b(\eta)} f_{|\boldsymbol{X}_{ij}|}\left(-\frac{p_i}{b(\eta)},\frac{p_j}{a(\eta)}\right) I_{(-\infty,0)\times[0,\infty)}(p_i,p_j)\\
&+ \frac{(1+p(\boldsymbol{h})-2g(\eta))}{b^2(\eta)} f_{|\boldsymbol{X}_{ij}|}\left(-\frac{p_i}{b(\eta)},-\frac{p_j}{b(\eta)}\right) I_{(-\infty,0)\times(-\infty,0)}(p_i,p_j).
\end{aligned} \tag{9}$$

The computation of $f_{P_{\eta;ij}}(p_i,p_j)$ depends on $p(\boldsymbol{h})$ that requires the computation of bivariate Gaussian probabilities. They can be easily calculated using the most important statistical softwares including R, MATLAB and Python. In particular, the R package `GeoModels` use the R package `pbivnorm` (Genz and Kenkel, 2015) that implements the methods proposed in Gentz (1992).

Finally, the correlation function of $P_\eta$ can be easily obtained as:

$$\rho_{P_\eta}(\boldsymbol{h}) = \frac{\mathbb{E}(|X(\boldsymbol{s}_i)||X(\boldsymbol{s}_j)|)\mathbb{E}(K_\eta(\boldsymbol{s}_i)K_\eta(\boldsymbol{s}_j)) - \mathbb{E}(|X(\boldsymbol{s})|)^2\mathbb{E}(K_\eta(\boldsymbol{s}))^2}{Var(P_\eta(\boldsymbol{s}))}, \tag{10}$$

where $\mathbb{E}(K_\eta(\boldsymbol{s}_i)K_\eta(\boldsymbol{s}_j))$ is given in (2).

From (7), (8) and (10) it is apparent that the knowledge of the mean, variance and correlation function of $P_\eta$ depends on the knowledge of $\mathbb{E}(|X(\boldsymbol{s})|^a)$

for $a = 1, 2$ and $\mathbb{E}(|X(\boldsymbol{s}_i)||X(\boldsymbol{s}_j)|)$. Additionally, the bivariate pdf of $P_\eta$ depends on the distribution of the bivariate random vector $|\boldsymbol{X}_{ij}| = (|X(\boldsymbol{s}_i)|, |X(\boldsymbol{s}_j)|)^T$. In Sections 3 and 4, we study two examples where these quantities can be explicitly calculated.

Different choices of $a(\eta)$ and $b(\eta)$ can be considered. In our examples in Section 3 and 4, we consider $a(\eta) = 1 - \eta$ and $b(\eta) = 1 + \eta$, where $\eta \in (-1, 1)$ is the skewness parameter and $\eta = 0$ correspond to the symmetric case. This is motivated by a more stable behaviour with respect to other choices of the symmetry functions (Arellano-Valle et al., 2005).

Note that a simplified version of the general class in Equation (4) can be obtained assuming $K_\eta(\boldsymbol{s}_i) \perp K_\eta(\boldsymbol{s}_j)$ for $i \neq j$, as suggested in Arellano-Valle et al. (2005). Under this assumption, $P_\eta$ is still a RF with two-piece marginal distributions; however, in this case, it can be shown that the resulting correlation function exhibits a discontinuity at the origin, and as a consequence, the RF is not mean square continuous (Stein, 1999). This leads to a RF with a 'forced' nugget effect that is not suitable for spatial data exhibiting continuous realization behaviour.

In our construction, if $X$ is obtained as a transformation of a Gaussian RF with underlying correlation function $\rho(\mathbf{h})$ as in Section 3 and 4, then a nugget effect can be easily introduced by choosing a correlation function discontinuous at the origin *i.e.* by replacing $\rho(\mathbf{h})$ with $\rho^*(\mathbf{h}) = \rho(\mathbf{h})(1 - \tau^2) + \tau^2 \mathbb{1}_0(||\mathbf{h}||)$, where $0 \leq \tau^2 < 1$. In addition, using the correlation function of $P_\eta$, it can be shown that the RF is mean square continuous but is not mean square differentiable. As a consequence, the RF does not inherit the mean square differentiability of the parent RF. This can be a drawback when using two-piece RFs for modelling spatial data exhibiting smoothed behaviour.

Finally, a more flexible model than $P_\eta$ can be obtained by defining a new RF $Z = \{Z(\boldsymbol{s}), \boldsymbol{s} \in A\}$ through a location and scale transformation:

$$Z(\boldsymbol{s}) = \mu(\boldsymbol{s}) + \sigma P_\eta(\boldsymbol{s}) \tag{11}$$

where $\mu(\boldsymbol{s})$ is the location dependent mean and $\sigma > 0$ is a scale parameter.

A typical parametric specification for the mean is given by $\mu(\boldsymbol{s}) = X(\boldsymbol{s})^T \boldsymbol{\beta}$ where $X(\boldsymbol{s}) \in \mathbb{R}^k$ is a vector of covariates and $\boldsymbol{\beta} \in \mathbb{R}^k$ but other types of parametric or nonparametric functions can be considered. Additional non-stationarity can be added into the formulation of (11) by allowing the scale parameter $\sigma$ to depend on the location $\boldsymbol{s}$. All the properties studied in this Section can be easily extended from $P_\eta$ to $Z$ including the bivariate density which is given by :

$$f_{\boldsymbol{Z}_{ij}}(z_i, z_j) = \frac{1}{\sigma^2} f_{\boldsymbol{P}_{\eta;ij}} \left( \frac{z_i - \mu(\boldsymbol{s}_i)}{\sigma}, \frac{z_j - \mu(\boldsymbol{s}_j)}{\sigma} \right). \tag{12}$$

## 3 Two-piece Gaussian random fields

In this Section we focus in an important example of the class proposed in Equation (4). In particular we study a RF with two-piece Gaussian marginal

distribution (Wallis, 2014) obtained when $X$ in (4) is a standard Gaussian RF $G$ with underlying correlation $\rho(\mathbf{h})$ that is:

$$P_\eta(\boldsymbol{s}) = |G(\boldsymbol{s})|K_\eta(\boldsymbol{s}). \tag{13}$$

We highlight that $G$ and $K_\eta$ share the same underlying correlation $\rho(\mathbf{h})$. In principle, the underlying correlation structure of $G$ need not coincide with that of $K_\eta$. However, if we would use a different correlation function, separate inference on its parameters would be needed complicating the estimation procedure.

Using (6), the marginal pdf is given by:

$$f_{P_\eta(\boldsymbol{s})} = \phi\left(\frac{p}{1-\eta}\right) I_{[0,\infty)}(p) + \phi\left(\frac{p}{1+\eta}\right) I_{(-\infty,0)}(p)$$

where $\phi(\cdot)$ is the pdf of the standard Gaussian distribution. In this case, $\mathbb{E}(|G(\boldsymbol{s})|) = \sqrt{2/\pi}$ and $\mathbb{E}(|G(\boldsymbol{s})|^2) = 1$ and, from Zhang and El-Shaarawi (2010), we have:

$$\mathbb{E}(|G(\boldsymbol{s}_i)||G(\boldsymbol{s}_j)|) = \frac{2}{\pi}\left[(1-\rho^2(\boldsymbol{h}))^{1/2} + \rho(\boldsymbol{h})\arcsin(\rho(\boldsymbol{h}))\right]. \tag{14}$$

Then, using (7) and (8), we have $\mathbb{E}(P_\eta(\boldsymbol{s})) = -2\eta\sqrt{2/\pi}$, $Var(P_\eta(\boldsymbol{s})) = 1 + 3\eta^2 - 8\eta^2/\pi$ and combining (10) and (14), the correlation function of the two-piece Gaussian RF $P_\eta$ is given by:

$$\rho_{P_\eta}(\boldsymbol{h}) = \frac{2\left[(1-\rho^2(\boldsymbol{h}))^{1/2} + \rho(\boldsymbol{h})\arcsin(\rho(\boldsymbol{h}))\right](3\eta^2 + 2\eta + 4p(\boldsymbol{h}) - 1) - 8\eta^2}{3\eta^2\pi - 8\eta^2 + \pi}. \tag{15}$$

Finally, the bivariate pdf of $\boldsymbol{P}_{\eta;ij}$ can be easily obtained using (9) with the pdf of the bivariate random vector $(|G(\boldsymbol{s}_i)|, |G(\boldsymbol{s}_j)|)^T$ given by (Murthy, 2015):

$$f_{|G_{ij}|}(g_i, g_j) = \frac{2}{\pi(1-\rho^2(\boldsymbol{h}))^{1/2}}\left[e^{-\frac{(g_i^2+g_j^2-2\rho(\boldsymbol{h})g_ig_j)}{2(1-\rho^2(\boldsymbol{h}))}} + e^{-\frac{(g_i^2+g_j^2+2\rho(\boldsymbol{h})g_ig_j)}{2(1-\rho^2(\boldsymbol{h}))}}\right] I_{(0,\infty)\times(0,\infty)}(g_i, g_j) \tag{16}$$

It can be easily seen from (15) that $\rho(\boldsymbol{h}) = 0$ if and only if $\rho_{P_\eta}(\boldsymbol{h}) = 0$. Since $\rho(\boldsymbol{h}) = 0$ also implies that the pdf of $\boldsymbol{P}_{\eta;ij}$ can be written as the product of two two-piece Gaussian pdfs, we have that the pairwise zero correlation of $P_\eta$ implies pairwise independence of $P_\eta$ as in the Gaussian case.

As outlined in the Introduction, if $\eta = 0$, we have an interesting example where the marginal distribution is Gaussian, but the bivariate is not and the the case $\eta \neq 0$ allows to obtain an asymmetric marginal distribution. In both cases, a particular type of bivariate dependence is obtained. To visualize this, we consider the associated contour plots and compare them with the contour plots of a skew-Gaussian RF (Zhang and El-Shaarawi, 2010), which is a natural

competitor since both marginal models allow to model skewness. Its stochastic representation is given by:

$$S_\beta(\boldsymbol{s}) = \beta|G_1(\boldsymbol{s})| + \sigma G_2(\boldsymbol{s}) \tag{17}$$

where $G_1$ and $G_2$ are independent copies of $G$ sharing a common underlying correlation function $\rho(\boldsymbol{h})$, $\beta \in \mathbb{R}$ is an asymmetry parameter, $\sigma > 0$ is a scale parameter. Marginally, $S_\beta(\boldsymbol{s})$ follows a skew-Gaussian distribution $SN(0, \omega^2, \alpha)$ (Azzalini and Capitanio, 2014) with $\omega^2 = (\beta^2 + \sigma^2)/\sigma^2$ and $\alpha = \beta/\sigma$ and its pdf is given by:

$$f_{S_\beta}(y) = \frac{2}{\omega}\phi\left(\frac{y}{\omega}\right)\Phi\left(\alpha\frac{y}{\omega}\right).$$

In addition, $\mathbb{E}(S_\beta(\boldsymbol{s})) = (\beta/\sigma)(2/\pi)^{1/2}$, $\text{var}(S_\beta(\boldsymbol{s})) = 1 + (\beta/\sigma)^2(1 - 2/\pi)$ and the correlation function is given by:

$$\rho_{S_\beta}(\boldsymbol{h}, \beta, \sigma) = \frac{2\beta^2}{\pi\sigma^2 + \beta^2(\pi - 2)}\left((1 - \rho^2(\boldsymbol{h}))^{1/2} + \rho(\boldsymbol{h})\arcsin(\rho(\boldsymbol{h})) - 1\right) + \frac{\sigma^2\rho(\boldsymbol{h})}{\sigma^2 + \beta^2(1 - 2/\pi)}. \tag{18}$$

In the comparison, we choose the parameters in such a way that the associated marginal pdfs are approximatively equal. Specifically, for the two-piece Gaussian RF $P_\eta$, we set $\eta = 0, 0.2, 0.6$ and for the skew-Gaussian RF $S_\beta$ we set $\beta = 0, -0.8, -1.2$, $\sigma = 1$.

The upper part of Figure 2 depicts the contour plots of the bivariate density of the two-piece Gaussian RF and the bottom part displays the contour plots of the bivariate density of the skew-Gaussian RF (Alegria et al., 2017) when the underlying correlation is $\rho(\mathbf{h}) = 0.8$, under the specified setting. In the symmetric case ($\eta = 0$ and $\beta = 0$, respectively), it is apparent that the contour plot of the two-piece Gaussian RF displays a clear $X$-shape. On the contrary, in the skew-Gaussian case the contour-plot is elliptical since the Gaussian RF is a special case of the skew-Gaussian RF when $\beta = 0$. Clearly, even in the non-symmetric case ($\eta = 0.2, 0.6$ and $\beta = -0.8, -1.2$ respectively), the dependence is of the same type for the two-piece Gaussian and skew-Gaussian RFs. Figure 3 displays the same comparison but with correlation $\rho(\mathbf{h}) = 0.1$. In this case, the $X$-shape of the contour plots tends to disappear as expected since we are approaching the zero correlation, which implies independence, as previously highlighted.

Finally, Figure 4 (first column) depicts a realization of $P_{0.2}$ and of $S_{-0.8}$ (setting $\sigma = 1$) on a irregular grid in $[0, 1]^2$. As an underlying correlation model, we consider a special case of the generalized Wendland model (Bevilacqua et al., 2019a)

$$\rho_{\alpha,\delta}(\boldsymbol{h}) := (1 - ||\boldsymbol{h}||/\alpha)_+^\delta \tag{19}$$

setting $\alpha = 0.4$ and $\delta = 4$. In both cases we detect the spatial outliers, as in the Introduction, using the algorithm proposed in Chen et al. (2008). They are highlighted with a small black circle. In this example the algorithm detects 57

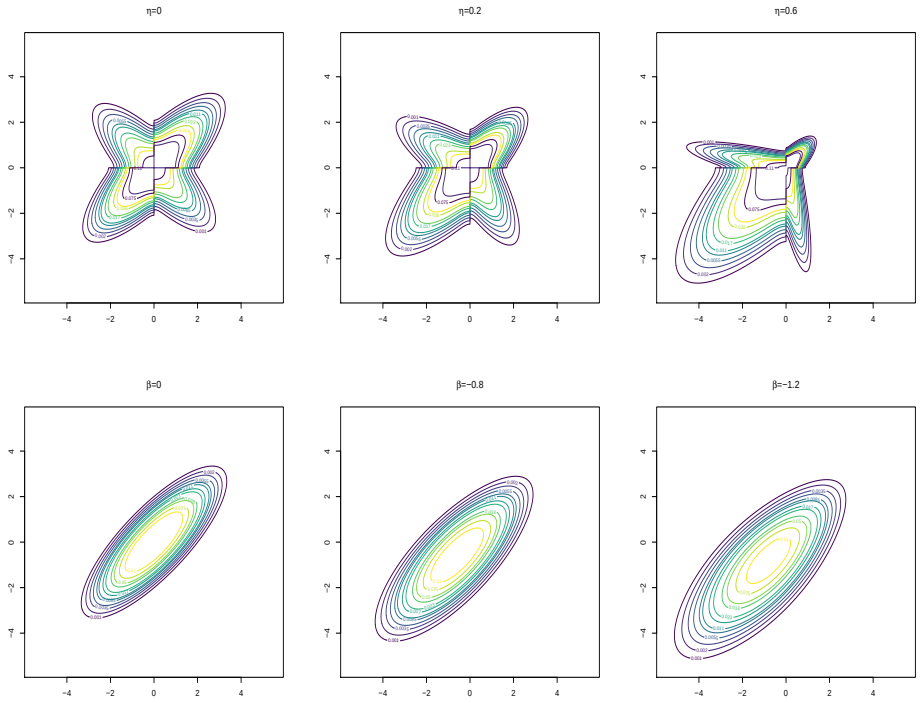



**Fig. 2** Contour plot of the bivariate pdf of the two-piece Gaussian RF $P_\eta$ (upper) when $\eta = 0, 0.2, 0.6$ (from left to right). Contour plot of the bivariate pdf of the skew-Gaussian RF $S_\beta$ (bottom) when $\beta = 0, -0.8, -1.2$ (from left to right). The underlying correlation in both cases is $\rho(\mathbf{h}) = 0.8$.

spatial outliers and 3 spatial outliers for the two-piece Gaussian RF and skew-Gaussian RF respectively. In addition, Figure 4 (second and third column) shows the associated (asymmetric) histograms and the $h-$scatterplots based on neighborhoods that display, as expected, the $X$-shape in particular for $m = 5, 15$.

## 4 Two-piece Tukey-$h$ random fields

More flexible two-piece RFs can be obtained by considering in (4) a RF $X$ with symmetric marginal distribution with heavier tails than those induced by a Gaussian distribution. Two notable examples are the Tukey-$h$ RFs proposed in Xua and Genton (2017) and the $t$ RFs proposed in Bevilacqua et al. (2021). In this Section we focus on the Tukey-$h$ RF.

Let $T_h = \{T_h(\boldsymbol{s}), \boldsymbol{s} \in A\}$ a zero-mean RF with Tukey-$h$ marginal distribution defined through a monotonic transformation of the underlying standard

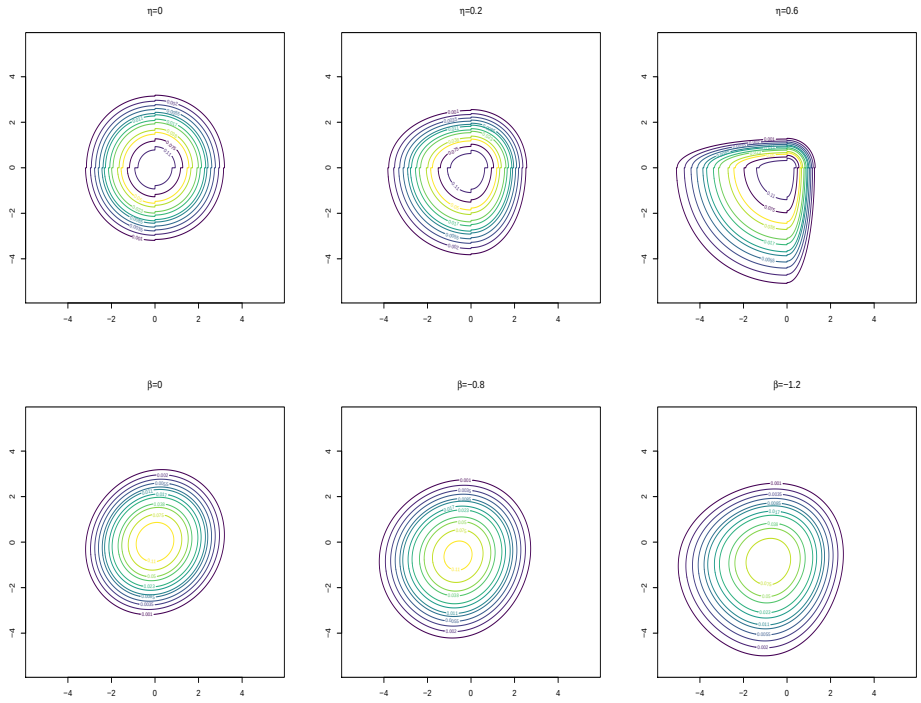



**Fig. 3** The same of Figure 2 with $\rho(\mathbf{h}) = 0.1$.

Gaussian RF $G$ as:

$$T_h(\boldsymbol{s}) =: \tau_h(G(\boldsymbol{s})) = G(\boldsymbol{s})e^{\frac{h(G(\boldsymbol{s}))^2}{2}}. \tag{20}$$

This kind of RF has marginal symmetric distributions and is a special case of the Tukey $g-h$ RF proposed in Xua and Genton (2017). In particular, the parameter $h \in (0, 1/2)$ governs the tail behaviour of the RF, with a larger value of $h$ indicating a heavier tail. The Tukey-$h$ univariate pdf is given by:

$$f_{T_h(\boldsymbol{s})}(t) = \frac{\tau_h^{-1}(t)}{t(1+W(ht^2))}\phi(\tau_h^{-1}(t); 0, 1), \tag{21}$$

where $\tau_h^{-1}(t) = sign(t)\left(\frac{W(ht^2)}{h}\right)^{1/2}$ and $W(\cdot)$ is the Lambert-$W$ function (Goerg, 2015).

Setting $X \equiv T_h$ in (4), we obtain an RF with a two-piece Tukey-$h$ marginal distribution:

$$P_{\eta,h}(\boldsymbol{s}) = |T_h(\boldsymbol{s})|K_\eta(\boldsymbol{s}). \tag{22}$$

Using (6), the marginal pdf of $P_{\eta,h}$ is given by:

$$f_{P_{\eta,h}(\boldsymbol{s})} = f_{T_h(\boldsymbol{s})}(p/(1-\eta))I_{[0,\infty)}(p) + f_{T_h(\boldsymbol{s})}(p/(1+\eta))I_{(-\infty,0)}(p),$$

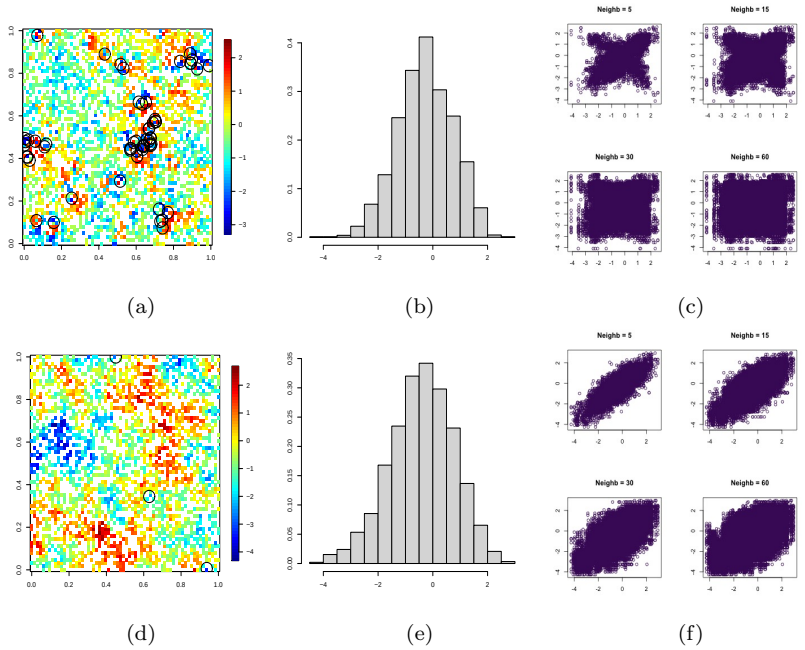



**Fig. 4** First row: (from left to right) a realization of a two-piece Gaussian RF $P_{0.2}$, associated histogram and associated $h$-scatterplots. Second row: (from left to right) a realization of a skew-Gaussian RF $S_{-0.8}$, associated histogram and associated $h$-scatterplots. The underlying correlation model is $\rho(\mathbf{h}) = (1 - ||\boldsymbol{h}||/0.4)_+^4$. Small black circles in a) and d) represent spatial outliers.

and in this case it can be easily shown that $\mathbb{E}(|T_h(\boldsymbol{s})|) = \frac{\sqrt{2}}{\sqrt{\pi}(1-h)}$ and $E(|T_h(\boldsymbol{s})|^2) = (1-2h)^{-3/2}$.

Then using (7) and (8), we have $\mathbb{E}(P_{\eta,h}(\boldsymbol{s})) = -2\eta\sqrt{2/\pi}/(1-h)$ and $Var(P_{\eta,h}(\boldsymbol{s})) = (1+3\eta^2)/(1-2h)^{1.5} - 8\eta^2/(\pi(1-h)^2)$.

The bivariate *pdf* of $P_{\eta,h}$ can be obtained using (9), where the *pdf* of the vector $(|\boldsymbol{T}_h(\boldsymbol{s}_i)|, |\boldsymbol{T}_h(\boldsymbol{s}_j)|)^T$ is given by

$$f_{|\boldsymbol{T}_{h,ij}|}(t_i,t_j) = f_{\boldsymbol{T}_{ij}}(t_i,t_j) + f_{\boldsymbol{T}_{ij}}(-t_i,-t_j) + f_{\boldsymbol{T}_{ij}}(-t_i,t_j) + f_{\boldsymbol{T}_{ij}}(t_i,-t_j) \quad (23)$$

where $f_{\boldsymbol{T}_{ij}}(\cdot,\cdot)$ is the bivariate *pdf* Tukey-$h$ distribution given by Goerg (2015):

$$f_{\boldsymbol{T}_{ij}}(t_i,t_j) = \frac{\tau_h^{-1}(t_i)\tau_h^{-1}(t_j)}{t_i t_j (1+W(ht_i^2))(1+W(ht_j^2))}\phi_2(\tau^{-1}(t_i),\tau^{-1}(t_j),\mathbf{0},R_2). \quad (24)$$

Here, $R_2 = [\rho(\boldsymbol{s}_i - \boldsymbol{s}_j)]_{i,j=1}^2$, the bivariate correlation matrix associated with the underlying correlation function $\rho(\mathbf{h})$.

Figure 5 depicts the contour plot of the bivariate pdf of $P_{\eta,h;ij}$ when $h = 0.1$ and $\eta = 0, 0.2, 0.6$ (from left to right) and $\rho(\mathbf{h}) = 0.8, 0.1$ (the top and bottom, respectively). Note that in the symmetric case $\eta = 0$ the marginal pdf is

Tukey-$h$ as in in Equation (21) but the bivariate pdf is not the bivariate pdf in Equation (24). Also in this example, it is apparent the $X$-shape of the contour plots when the correlation is strong, $\rho(\mathbf{h}) = 0.8$.

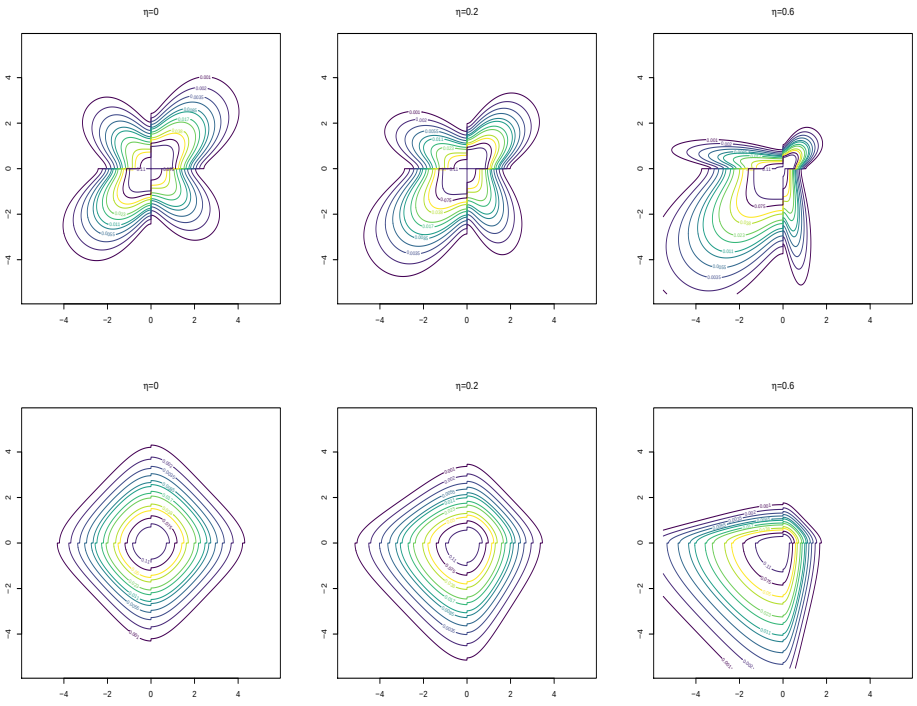



**Fig. 5** Contour plot of the bivariate pdf of $P_{\eta,h;ij}$ that is the bivariate distribution of the two-piece Tukey-$h$ RF when $\eta = 0, 0.2, 0.6$ (from left to right) and the underlying correlation is $\rho(\mathbf{h}) = 0.8$ (first row) and $\rho(\mathbf{h}) = 0.1$ (second row). Overall, the tail parameter is $h = 0.1$.

The following lemma gives the $(1,1)-th$ product moment of the half Tukey-$h$ distribution, which is useful for giving an analytic expression for $\rho_{P_{\nu,\eta}}(\boldsymbol{h})$. It depends on a function $c : [0,1] \to [1,\infty)$ defined as

$$c(x) := \frac{\sqrt{x}arcsin(\sqrt{x}) + \sqrt{1-x}}{(1-x)^{3/2}}, \quad 0 \leq x \leq 1$$

which is a specific instance of the Gauss hypergeometric function (see the proof in the Appendix)

**Lemma 1** *Let $T_h$, $h \in [0, 1/2)$ be a weakly stationary Tukey-h RF with an underlying correlation $\rho(\boldsymbol{h})$. Then:*

$$\mathbb{E}(|T_h(s_i)||T_h(s_j)|) = \frac{2(1-\rho^2(\boldsymbol{h}))^{3/2}}{\pi g^2(\boldsymbol{h},h)} c\left(\frac{\rho^2(\boldsymbol{h})}{g^2(\boldsymbol{h},h)}\right). \tag{25}$$

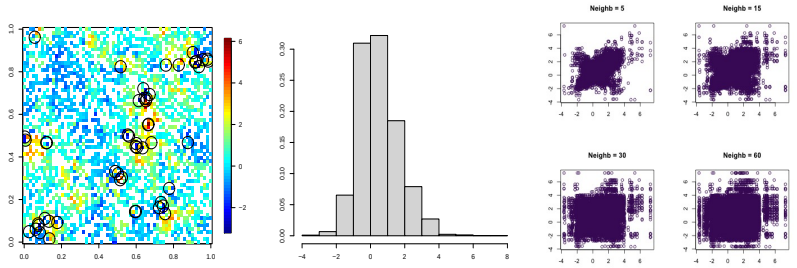
*where* $g(\boldsymbol{h},h)=1-(1-\rho^2(\boldsymbol{h}))h$

We are now ready to give the correlation function of $P_{\eta,h}$.

**Theorem 1** *Let* $P_{\eta,h}$, $h\in[0,1/2)$ *be a stationary RF with two-piece Tukey-h marginals. Then:*

$$\rho_{P_{\eta,h}}(\boldsymbol{h})=\frac{\frac{2(1-\rho^2(\boldsymbol{h}))^{3/2}}{\pi g^2(\boldsymbol{h},h)}c\left(\frac{\rho^2(\boldsymbol{h})}{g^2(\boldsymbol{h},h)}\right)(3\eta^2+2\eta+4\pi_{11}-1)-\frac{8\eta^2}{\pi(1-h)^2}}{\frac{(1+3\eta^2)}{(1-2h)^{3/2}}-\frac{8\eta^2}{\pi(1-h)^2}} \tag{26}$$

*Proof* Combining Equation (10) and Equation (25), we obtain (26).




(a) (b) (c)

**Fig. 6** From left to right: a realization of a two-piece Tukey-$h$ $P_{-0.2,0.1}$, associated histogram and associated $h$-scatterplots. The underlying correlation model is $\rho(\mathbf{h})=(1-||\boldsymbol{h}||/0.4)_+^4$. Small black circles in a) represent spatial outliers.

In Figure 6 we show a realization on a irregular grid in $[0,1]^2$ of a two-piece Tukey-$h$ RF with $h=0.1$ and $\eta=-0.2$, with underlying correlation model given in (19) with $\alpha=0.4$ and $\delta=4$. Also in this case we highlight the detected spatial outliers, using the algorithm proposed in Chen et al. (2008), with a small black circle. In addition, in Figure 6, we show the associated histogram and $h$-scatterplots based on neighborhoods.

Finally, Figure 7 depicts the correlation functions (dotted line) of the two-piece Gaussian $P_\eta$ and two-piece Tukey-$h$ $P_{\eta,h}$ with $\eta=0.2$ and $h=0.1$ RFs in Equations (15) and (26) compared with the underlying Gaussian correlation function $\rho_{0.3,4}(\mathbf{h})$ (continuous line). It can be appreciated that when the underlying correlation is approximatively zero, then $\rho_{P_{0.2}}(\boldsymbol{h})$ and $\rho_{P_{0.2,0.1}}(\boldsymbol{h})$ are approximately zero. Actually, if the underlying correlation model $\rho(\boldsymbol{h})$ is compactly supported, then it can be shown that the resulting two-piece type correlations inherits this feature.

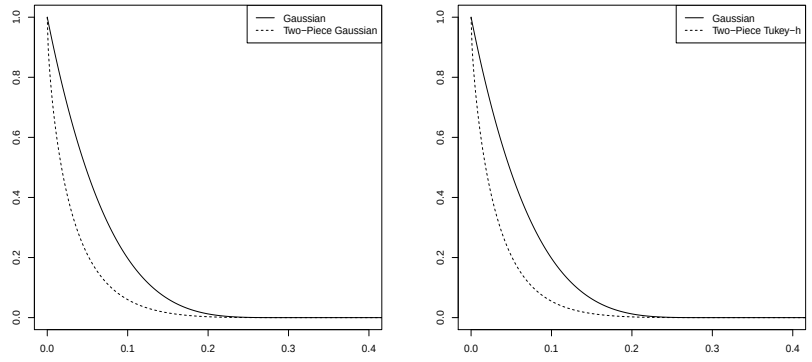



**Fig. 7** From left to right: Correlation functions (dotted line) of the two-piece Gaussian $P_{0.2}$, two-piece Tukey-$h$ $P_{0.2,0.1}$ RFs in Equations (15), (26) compared with the underlying Gaussian correlation function $\rho(\mathbf{h}) = (1 - ||\boldsymbol{h}||/0.4)_+^4$ (continuous line).

## 5 Numerical results

### 5.1 Weighted pairwise likelihood estimation

Let $\boldsymbol{z} = (z_1, \ldots, z_n)^T$ be a realization of a RF $Z = \{Z(\boldsymbol{s}), \boldsymbol{s} \in A \subset \mathbb{R}^2\}$ defined as

$$Z(\boldsymbol{s}) = \mu(\boldsymbol{s}) + \sigma Y(\boldsymbol{s}) \tag{27}$$

observed at distinct spatial locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$, with $\boldsymbol{s}_i \in A$, where $Y = P_\eta, P_{\eta,h}$ can be the two-piece Gaussian or the two-piece Tukey-$h$ RFs defined in Section 3 and 4.

Let $\boldsymbol{\theta} = (\boldsymbol{\mu}, \sigma^2, \eta, \boldsymbol{\alpha}^T)$, be the vector of unknown parameters where $\boldsymbol{\alpha}$ is the vector parameter associated with the underlying parametric correlation model $\rho_{\boldsymbol{\alpha}}(\boldsymbol{h})$, $\sigma^2 > 0$ is the scale parameter, and $\eta \in (-1, 1)$ is the skewness parameter. In addition $\boldsymbol{\mu}$ is the vector parameter associated with the spatial mean $\mu(\boldsymbol{s})$, and here, for simplicity, we assume a spatial constant mean $\mu(\boldsymbol{s}) = \mu$. In the case of two-piece Tukey-$h$ RF we have an additional tail parameter, $h \in (0, 1/2)$.

Composite likelihood is a a general class of objective functions based on the likelihood of marginal or conditional events (Lindsay, 1988; Varin et al., 2011). This kind of estimation method can be helpful when it is difficult to evaluate or to specify the full likelihood. In particular the *wpl* method has been widely used for the estimation of complex non-Gaussian RFs because in some circumstances the multivariate distribution is unknown and/or difficult to compute but the bivariate density is known and relatively simply to evaluate.

Specifically, the method of *wpl* combines the bivariate distributions of all possible distinct pairs of observations. The *wpl* function is given by

$$\mathrm{pl}(\boldsymbol{\theta}) := \sum_{i=1}^{n}\sum_{j=1}^{n} \log(f_{\boldsymbol{Z}_{ij}}(z_i, z_j; \boldsymbol{\theta}))c_{ij},$$

where $f_{\boldsymbol{Z}_{ij}}(z_i, z_j; \boldsymbol{\theta})$ is the bivariate density obtained by coupling the general bivariate density in Equation (9) with Equations (16) and (23) for the two-piece Gaussian or the two-piece Tukey-$h$ RFs respectively and $c_{ij}$ is a nonnegative suitable weight.

In general, a loss of statistical efficiency is expected from *wpl* estimation with respect to the maximum likelihood estimation and the role of the weights $c_{ij}$ is to minimize this loss. The choice of symmetric cut-off weights, namely,

$$c_{ij} = \begin{cases} 1 & \|\boldsymbol{s}_i - \boldsymbol{s}_j\| \leq d_s \\ 0 & \text{otherwise} \end{cases}, \tag{28}$$

for a positive value of $d_s$, can be motivated by its simplicity and by observing that the dependence between observations that are distant is weak. It has been shown that this kind of weights allows to improve both the statistical and computational efficiency of the method (Bevilacqua and Gaetan, 2015; Joe and Lee, 2009). The maximum *wpl* estimator is given by

$$\widehat{\boldsymbol{\theta}} := \mathrm{argmax}_{\boldsymbol{\theta}}\ \mathrm{pl}(\boldsymbol{\theta})$$

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

$$\mathrm{PLIC} := -2\,\mathrm{pl}(\hat{\boldsymbol{\theta}}) + 2\mathrm{tr}(\mathcal{H}_n(\hat{\boldsymbol{\theta}})\mathcal{G}_n^{-1}(\hat{\boldsymbol{\theta}})) \tag{29}$$

which is the composite likelihood version of the Akaike information criterion (Varin and Vidoni, 2005). The computation of standard errors and information criterion requires evaluation of the matrices $\mathcal{H}_n(\hat{\boldsymbol{\theta}})$ and $\mathcal{J}_n(\hat{\boldsymbol{\theta}})$. However, the evaluation of $\mathcal{J}_n(\hat{\boldsymbol{\theta}})$ is computationally unfeasible for large datasets and in this case subsampling techniques can be used to estimate $\mathcal{J}_n(\boldsymbol{\theta})$ as in Bevilacqua et al. (2012) and Heagerty and Lele (1998). A straightforward and more robust alternative is parametric bootstrap estimation of $\mathcal{G}_n^{-1}(\boldsymbol{\theta})$ (Bai et al., 2014; Efron, 1982). We adopt the second strategy in our paper.

To analyse the performance of the *wpl* method, we perform a small simulation study. As a simulation setting, we consider points $\boldsymbol{s}_i \in A = [0,1]^2$, $i = 1, \ldots, N$. Specifically, we simulate $1,000$ realizations of the proposed two

models observed at $N = 500$ spatial location sites uniformly distributed in the unit square. Note that one simulation of the two-piece Gaussian or the two-piece Tukey-$h$ RFs both requires two independent simulations of the underlying Gaussian RF that are performed using Cholesky decomposition.

We consider $\sigma^2 = 1$, a constant mean $\mu = 0$, and we set $h = 0.15$ for Tukey-$h$. Overall, we set the skewness parameter $\eta = 0, 0.5$ that correspond to the symmetric and asymmetric case respectively. As underlying isotropic parametric correlation model, we consider the model in Equation (19) with $\alpha = 0.2$ and $\delta = 4$, where $\delta$ is assumed to be known and fixed. Additionally, in the *wpl* estimation, we consider a cut-off weight function with $d_s = 0.1$.

Table 2 shows the bias and MSE of the *wpl* estimates. As a general comment, the estimates are overall unbiased, and for each model, there are no significant differences between the symmetric case ($\eta = 0$) and the asymmetric case ($\eta = 0.5$) in terms of MSE. Additionally, the MSE of the parameters is not affected by the kind of model except for the $\sigma^2$ parameter. As an illustrative example, Figures 8 and 9 show the boxplots of the *wpl* estimates for the estimation of the two-piece Gaussian (with $\eta = 0.5$) and two-piece Tukey-$h$ (with $\eta = 0$) RF parameters, respectively.

| | two-piece Gaussian | | | | two-piece Tukey-$h$ | | | |
|---|---|---|---|---|---|---|---|---|
| | $\eta = 0$ | | $\eta = 0.5$ | | $\eta = 0$ | | $\eta = 0.5$ | |
| | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE |
| $\hat{\mu}$ | 0.0110 | 0.0025 | 0.0083 | 0.0016 | 0.0082 | 0.0019 | 0.0053 | 0.0016 |
| $\hat{\alpha}$ | −0.0037 | 0.0006 | −0.0035 | 0.0006 | 0.0013 | 0.0007 | 0.0022 | 0.0008 |
| $\hat{\sigma}^2$ | −0.0089 | 0.0099 | −0.0094 | 0.0099 | 0.0065 | 0.0220 | 0.0089 | 0.0225 |
| $\hat{\eta}$ | 0.0078 | 0.0015 | 0.0066 | 0.0011 | 0.0063 | 0.0018 | 0.0060 | 0.0015 |
| $\hat{h}$ | - | - | - | - | −0.0062 | 0.0017 | −0.0064 | 0.0017 |

**Table 2** Bias and MSE when estimating with *wpl* two-piece Gaussian and two-piece Tukey-$h$ RFs, that is, $\mu + \sigma P_\eta(\boldsymbol{s})$ and $\mu + \sigma P_{\eta,h}(\boldsymbol{s})$ respectively, with $\mu = 0$, $\sigma^2 = 1$, $h = 0.15$ for different values of the skewness parameter $\eta = 0, 0.5$, with underlying generalized Wendland correlation function $\rho_{\alpha,4}(\boldsymbol{h}) = (1 - ||\boldsymbol{h}||/\alpha)_+^4$ with $\alpha = 0.2$.

In addition we perform a small simulation study to analyze the performance of the PLIC information criterion (29). We use the previous simulation setting and we generate 100 two-piece Gaussian RFs with $\eta = -0.5, 0, 0.5$ and estimate with *wpl* the Gaussian and two-piece Gaussian RFs. Then, for each RF, we compute the PLIC information criteria using parametric bootstrap, that is we compute PLIC $= -2\,\mathrm{pl}(\hat{\boldsymbol{\theta}}) + 2\mathrm{tr}(\mathcal{H}_n(\hat{\boldsymbol{\theta}})\widehat{\mathcal{G}_n^{-1}(\hat{\boldsymbol{\theta}})})$ where $\widehat{\mathcal{G}_n^{-1}(\hat{\boldsymbol{\theta}})}$ is computed through parametric bootstrap. To be specific, let $\hat{\boldsymbol{\theta}}$ be a parameter estimate for the two-piece Gaussian o Gaussian RF. For $m = 1, \ldots, M$, with $M = 100$, we simulate data $\boldsymbol{z}_m$ with parameter $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, and then we fit the model using $\boldsymbol{z}_m$ to obtain estimate $\hat{\boldsymbol{\theta}}_m$. The parametric bootstrap estimate of the matrix $\mathcal{G}_n^{-1}(\hat{\boldsymbol{\theta}})$, that is $\widehat{\mathcal{G}_n^{-1}(\hat{\boldsymbol{\theta}})}$, is then obtained as the empirical variance-covariance matrix of $\hat{\Theta}_M^T$ where $\hat{\Theta}_M = (\hat{\boldsymbol{\theta}}_1, \ldots, \hat{\boldsymbol{\theta}}_M)$. When $\eta = -0.5, 0.5$ the criteria performs very well and selects 99 times over 100 the correct (two-piece Gaussian) model. When $\eta = 0$ the criteria performs slightly worse and selects 90 times over 100 the correct (two-piece Gaussian) model. This is not surprising since in this

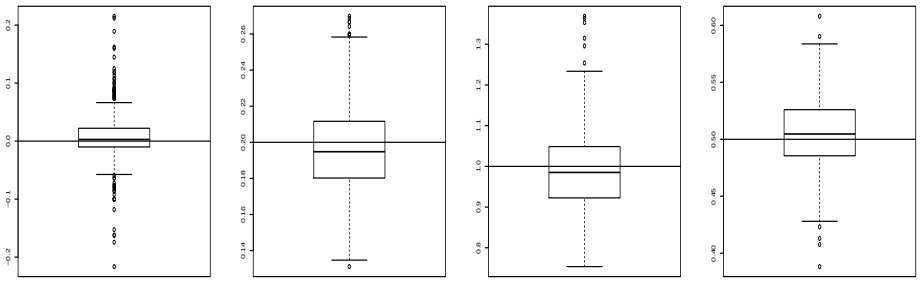


**Fig. 8** Boxplots of *wpl* estimates for $\mu = 0$, $\alpha = 0.2$, $\sigma^2 = 1$ and $\eta = 0.5$ (from left to right) when estimating a two-piece Gaussian RF.

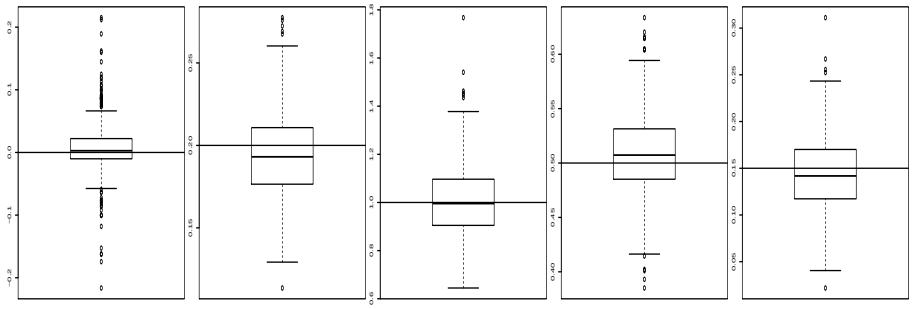


**Fig. 9** Boxplots of *wpl* estimates for $\mu = 0$, $\alpha = 0.2$, $\sigma^2 = 1$, $\eta = 0$ and $h = 0.15$ (from left to right) when estimating a two-piece Tukey-$h$ RF.

case, both RFs share the same marginal Gaussian distribution. Finally, we replicated this numerical experiment by simulating from a Gaussian RF and also in this case the information criteria performs very efficiently, selecting 100 times over 100 the correct (Gaussian) model.

## 5.2 Optimal linear prediction performance

The optimal predictor for the two-piece RF, with respect to the mean squared error criterion, is nonlinear and difficult to evaluate explicitly since it requires the knowledge of the finite dimensional distribution. A practical and less efficient solution can be obtained using the optimal linear predictor. In this Section, we study the performance of the optimal linear predictor of the two-

piece Gaussian RF, $Z(\boldsymbol{s}) = \mu(\boldsymbol{s}) + \sigma P_\eta(\boldsymbol{s})$ and of the optimal predictor of the Gaussian RF, $Z(\boldsymbol{s}) = \mu(\boldsymbol{s}) + \sigma X(\boldsymbol{s})$.

Specifically, we compare the prediction performance of the optimal (linear) predictor at an unknown location $\mathbf{s}_0$ using the vector data $\boldsymbol{z}$ given by:

$$\widehat{Z}(\mathbf{s}_0) = \mu(\mathbf{s}_0) + \boldsymbol{c}^T \Sigma^{-1}(\boldsymbol{z} - \boldsymbol{\mu}), \tag{30}$$

where, for the two-piece Gaussian case, $\mu(\mathbf{s}_0)$, the vector mean $\boldsymbol{\mu}$, the covariance matrix $\Sigma$ and the covariance vector $\boldsymbol{c}$ can be computed using the expression given in Section 3. In particular the covariance matrix, can be computed using the correlation function given in Equation (15).

To compare the prediction performance we adopt a resampling approach and we explore how well the optimal linear predictor works under the two-piece Gaussian RF, compared with the optimal prediction of the (misspecified) Gaussian RF, using root mean squared error (RMSE) and mean absolute error (MAE) as measures of prediction performance. In particular, we consider the following steps:

1. Set $j = 1$. Repeat until $j \leq 100$.
2. Simulate the $j$-th spatial dataset $\boldsymbol{z}_j = (z_j(\boldsymbol{s}_1), \ldots, z_j(\boldsymbol{s}_{500}))^T$ from the proposed two-piece Gaussian RF by considering 500 location sites uniformly distributed on the unit square.
3. Set $k = 1$. Repeat until $k \leq 100$.
4. Randomly split the $j$-th dataset $\boldsymbol{z}_j$ by using 85% of the data (425 observations) for estimation and 15% as the validation dataset (75 observations). We call them $\boldsymbol{z}_j^E$ and $\boldsymbol{z}_j^V$ respectively.
5. Estimate with *wpl* the two-piece Gaussian RF and with maximum likelihood the Gaussian RF using data $\boldsymbol{z}_j^E$.
6. Compute the optimal (linear) predictor for the Gaussian and two-piece Gaussian RFs at the coordinates associated with the validation dataset, given the estimates obtained at the previous step.
7. Compute, for both models, $\text{RMSE}_k = \left[\frac{1}{75}\sum_{i=1}^{75}\left(z_j^V(\boldsymbol{s}_i) - \widehat{Z}_j^V(\boldsymbol{s}_i)\right)^2\right]^{\frac{1}{2}}$ and $\text{MAE}_k = \frac{1}{75}\sum_{i=1}^{75}|z_j^V(\boldsymbol{s}_i) - \widehat{Z}_j^V(\boldsymbol{s}_i)|$ where $\widehat{Z}_j^V(\boldsymbol{s}_i)$ is the optimal (linear) predictor computed at the previous step.
8. $k = k+1$. Compute $\overline{\text{RMSE}}_j = \sum_{k=1}^{100}\text{RMSE}_k/100$ and $\overline{\text{MAE}}_j = \sum_{k=1}^{100}\text{MAE}_k/100$ for the two-piece Gaussian and Gaussian models.
9. $j = j + 1$. Compute $\text{RMSE} = \sum_{j=1}^{100}\overline{\text{RMSE}}_j/100$ and $\text{MAE} = \sum_{j=1}^{100}\overline{\text{MAE}}_j/100$ for the two-piece Gaussian and Gaussian models.

This numerical experiment has been performed by simulating (at step 2) from a two-piece Gaussian RF setting $\mu = 0$, $\sigma^2 = 1$ and $\eta = -0.5, 0, 0.5$ with an underlying correlation model given in (19) with $\alpha = 0.2$ and $\delta = 4$. For the *wpl* estimation of the two-piece Gaussian models, we set $d_s = 0.1$. Table 3, left

part, summarizes the results of our experiment, showing the RMSE and MAE for two-piece Gaussian RF and Gaussian RF when $\eta = 0, 0.5, 0.5$. The optimal linear predictor for the two-Piece Gaussian RF is the best in all scenarios using both RMSE and MAE as measure of prediction performance, including the case $\eta = 0$ (recall that in this case the Gaussian and two-piece Gaussian RFs share the same marginal Gaussian distribution). We finally replicate the numerical experiment simulating from a Gaussian RF at step 2. As expected, in this case, the optimal Gaussian predictor clearly outperforms the two-piece optimal linear predictor in terms of RMSE and MAE (see right part of Table 3). These numerical experiments suggests that if the true model is a Gaussian RF, then predicting with the proposed two-piece Gaussian RF is not recommended.

| | Two-piece Gaussian | | | Gaussian |
|---|---|---|---|---|
| | $\eta = -0.5$ | $\eta = 0$ | $\eta = 0.5$ | |
| $\text{RMSE}_{\text{TP}}$ | 0.9595 | 0.9029 | 0.9553 | 0.763 |
| $\text{RMSE}_{\text{G}}$ | 0.9689 | 0.9189 | 0.9654 | 0.719 |
| $\text{MAE}_{\text{TP}}$ | 0.7413 | 0.6827 | 0.7485 | 0.601 |
| $\text{MAE}_{\text{G}}$ | 0.7543 | 0.6946 | 0.7575 | 0.564 |

**Table 3** Empirical mean of RMSE and MAE associated with the optimal linear predictor of the two-piece Gaussian RF (TP) and with the optimal predictor based on the Gaussian RF (G) when $\eta = -0.5, 0, -0.5$, under the two-piece Gaussian RF (left part) and the Gaussian RF (right part).

## 6 Real data Application

We consider a subset of a global data set of merged mean daily temperature measurements from the Global Surface Summary of Day data (GSOD) with European Climate Assessment &Dataset (ECA&D) data in July 2011. The dataset is described in detail in Kilibarda et al. (2014) and is available in the R package `meteo`. The subset we consider consists of the mean temperature observed on July 4 at the 461 location sites, $z(\boldsymbol{s}_i)$, $i = 1, \ldots, 461$, in the region with longitude $[33, 80]$ and latitude $[23, 40]$ *i.e.* a partial region of Middle East including countries like Saudi Arabia, Iran, Irak and Siria.

Figure 10 (a) shows the observed region and a colored map of the spatial observed points. As in Section 1, using the algorithm proposed in Chen et al. (2008), we highlight the detected spatial outliers with a small black circle. In addition, Figure 10 (b) shows four $h$-scatterplots using $m = 10, 20, 30, 40$ as neighborhood order. Following the same arguments of Section 1, both graphics suggest the presence of some spatial outliers.

Additionally, Figure 10 (c) shows the boxplot of the mean temperature data. It highlights that asymmetry and possibly heavy tails are features that should be considered from a modelling viewpoint. Finally, the empirical semivariogram in Figure 10 (d) suggests that a nugget effect should be also included in the analysis. These preliminary graphical analysis suggest that a RF with flexible marginal distribution and that take into account the presence of spatial outliers is potentially an appropriate model for our data. The proposed two-piece Gaussian and two-piece Tukey-$h$ RFs possess these two specific features. In our analysis we consider seven RFs of the following type:

$$Z(\boldsymbol{s}) = \mu + \sigma R(\boldsymbol{s}), \quad \boldsymbol{s} \in \mathbb{R}^2 \tag{31}$$

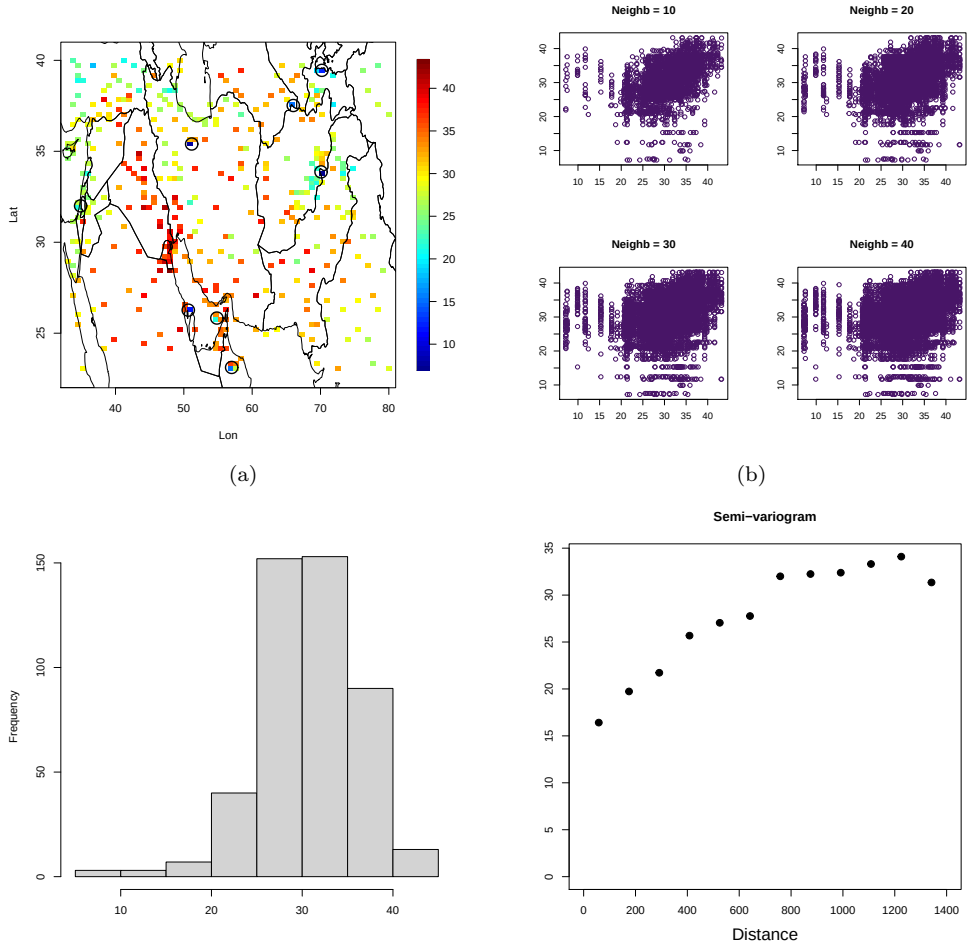
where $R = G, P_0, P_\eta, S_\eta, T_h, P_{0,h}, P_{\eta,h}$ that is we consider a location and scale transformation of respectively:

1) a Gaussian RF.
2) a two-piece Gaussian RF fixing $\eta = 0$ that is with Gaussian marginal distributions.
3) a two-piece Gaussian RF with $\eta \in (-1, 1)$ that is with (a)symmetric marginal distributions.
4) a Skew-Gaussian RF.
5) a Tukey-$h$ RF.
6) a two-piece Tukey-$h$ RF fixing $\eta = 0$ that is with Tukey-$h$ marginal distributions.
7) a two-piece Tukey-$h$ RF with $\eta \in (-1, 1)$ that is with (a)symmetric marginal distributions.

For each of the seven RFs considered, we assume an underlying isotropic correlation function of the exponential type that includes a nugget effect *i.e.* $\rho(\mathbf{h}) = e^{-||\boldsymbol{h}||/\alpha}(1 - \tau^2) + \tau^2 \mathbb{1}_0(||\mathbf{h}||)$, where $0 \leq \tau^2 < 1$.

We estimate the seven RFs with *wpl* using the weight function (28) with $d_s = 280$. The results are summarized in Table 4, where we also report the standard error estimation and the information criterion PLIC computed through

**Fig. 10** a) spatial locations of the data. b) $h$-scatterplot, c) histogram of the data. d) Empirical semivariogram estimation.




(a) (b)

(c) (d)

parametric bootstrap as explained in Section 5. Note that the mean and the spatial scale dependence estimates are quite similar, as expected. Additionally, the estimation of the skewness parameter ($\eta$) and heavy tail parameter ($h$) shows that the deviation from the marginal Gaussian distribution is significant (recall that $\eta \in (-1, 1)$ for the two-piece models and $h \in (0, 0.5)$).

More importantly the PLIC information criterion selects the two-piece type models. In particular, an interesting comparison is between model 1 y 2 and 5 y 6. In the first comparison both models have marginal Gaussian distribution and the same number of parameters. In the second case, both models have marginal Tukey-$h$ distribution and the same number of parameters. However,

the CLIC information criterion selects model 2 and 6 respectively. i.e. the symmetric two-piece type models. A global comparison show that the best selected model is model 6.
Finally, Figure 11 depicts the graphical comparison between the empirical and estimated semivariograms of the residuals for models 1, 2 and 6 i.e. for the Gaussian, two-piece Gaussian ($\eta = 0$) and two-piece Tukey-$h$ ($\eta = 0$) RFs (from left to right). The residuals have been computed as $\hat{R}(\boldsymbol{s}_i) = (Z(\boldsymbol{s}_i) - \hat{\mu})/\hat{\sigma}, \quad i = 1, 2, ..461$ where $\hat{\mu}$ and $\hat{\sigma}$ are *wpl* estimates. Estimated semivariograms for the two-piece models are computed using the correlation functions in Equations (15) and (26)

We want to further evaluate the predictive performances of the seven RFs using RMSE and MAE. As in Section 5.1, we use the following resampling approach: we randomly choose 85% of the spatial locations and we use remaining 15% as data for the predictions. We estimate the seven RFs using the *wpl* method and we use the estimates to compute the optimal linear predictions. Using the predicted values, we then compute the RMSE and MAE. Specifically, for each $j-th$ left-out sample $(z_j^L(\boldsymbol{s}_1), \ldots, \ldots, z_j^L(\boldsymbol{s}_K))$, we compute

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

and

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

where $\widehat{Z}_j^L(\boldsymbol{s}_i)$ is the optimal linear predictor for each of the seven RFs obtained using the $j-th$ *wpl* estimates and $K = 70$. We repeat the approach $j = 1, \ldots, 200$ times and record all RMSEs and MAEs. Finally, we compute the overall mean for each of the seven RFs, which is $\text{RMSE} = \sum_{j=1}^{200}\overline{\text{RMSE}}_j/200$ and $\text{MAE} = \sum_{j=1}^{200}\overline{\text{MAE}}_j/200$. The results are summarized in Table 4. The two-piece models clearly outperforms Gaussian and skew-Gaussian RFs. In particular the model with the best prediction performance is the asymmetric two-piece Tukey-$h$ (model 7).

## 7 Concluding remarks

We introduced a new class of random fields with two-piece marginal distributions for regression and dependence analysis when addressing spatial point referenced data exhibiting skewness and possibly heavy tails. In particular, we focused on two examples: two-piece Gaussian and two-piece Tukey-$h$ random fields. The proposed two-piece random fields have extremely flexible marginal distributions and can, therefore, be applied to a wide range of applications.
The main feature of the proposed class is that it has a specific type of dependence that can be useful when modeling data displaying spatial outliers

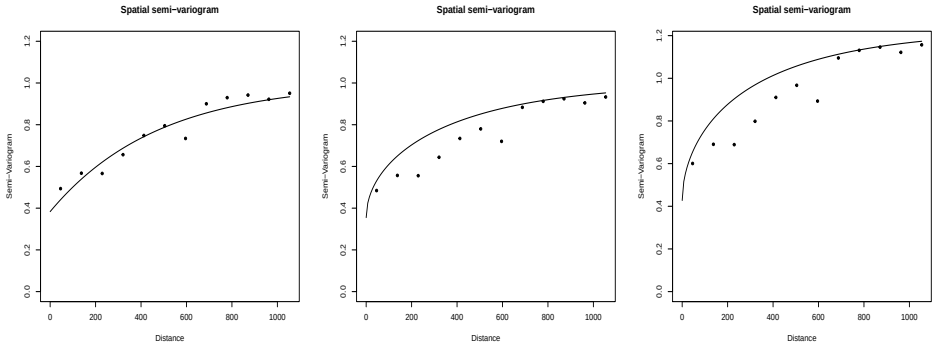



**Fig. 11** Comparison between the empirical and estimated semivariogram of the residuals for the Gaussian, two-piece Gaussian RF and two-piece Tukey-$h$ RFs when $\eta = 0$. (from left to right).

| | $\hat{\eta}$ | $\hat{h}$ | $\hat{\mu}$ | $\hat{\alpha}$ | $\hat{\sigma}^2$ | $\hat{\tau}^2$ | WPL | PLIC | RMSE | MAE |
|---|---|---|---|---|---|---|---|---|---|---|
| Gaussian | - | - | 30.762 (1.721) | 472.074 (323.94) | 34.470 (6.004) | 0.383 (0.104) | −25535.7 | 52385.9 | 4.105 | 2.858 |
| Two-piece Gaussian | 0 | - - | 31.503 (0.142) | 491.139 (118.703) | 35.132 (3.394) | 0.834 (0.114) | −25462.7 | 51069.0 | 4.091 | 2.843 |
| Two-piece Gaussian | 0.228 (0.098) | - | 32.700 (0.099) | 490.581 (63.970) | 33.417 (4.525) | 0.659 (0.151) | −25276.4 | 51203.7 | 4.091 | 2.845 |
| Skew Gaussian | -6.941 (2.025) | - | 36.188 (1.980) | 305.276 (145.88) | 16.662 (6.713) | 0.233 (0.181) | −25403.9 | 52027.9 | 4.177 | 2.871 |
| Tukey-$h$ | - - | 0.087 (0.032) | 31.168 (1.644) | 440.630 (192.314) | 26.335 (6.315) | 0.293 (0.092) | −25308.5 | 52019.5 | 4.097 | 2.849 |
| Two-piece Tukey-$h$ | 0 | 0.063 (0.029) | 31.501 (0.139) | 483.131 (104.215) | 28.333 (4.910) | 0.606 (0.163) | −25319.3 | 50945.9 | 4.093 | 2.846 |
| Two-piece Tukey-$h$ | 0.140 (0.114 ) | 0.045 (0.029 ) | 31.899 (0.137) | 473.237 (39.383) | 29.086 (5.715) | 0.514 (0.157) | −25230.4 | 51297.7 | 4.088 | 2.837 |

**Table 4** Estimates using *wpl* with associated standard error (in parenthesis), when estimating the mean temperatures dataset using the models $1, 2, 3, 4, 5, 6, 7$ (from top to bottom) with underlying exponential correlation model with scale parameter $\alpha$. Last four columns: maximum of the *wpl* function (WPL) infomation criterion(PLIC) and empirical mean of RMSEs and MAEs.

*i.e.* spatial points with values that are significantly different from those of other spatial points in its spatial neighbourhood and that do not necessarily deviate from the remainder of the whole data set (Shekhar et al., 2003). As a consequence, the use of the proposed models is recommended when modelling spatial point referenced data displaying these kind of outliers.

From theory of distribution point of view, if the asymmetry parameter is equal to zero, the proposed models provide new examples of random vectors with marginal distribution of the Gaussian and Tukey-$h$ type whose bivariate or multivariate distribution are not of the same type (Dutta and Genton, 2014).

A limitation for the proposed class is the lack of computationally feasible density outside of the bivariate case that prevents an inference approach based on likelihood methods. However, we showed through some simulation studies that an inferential approach based on the pairwise likelihood is an effective so-

lution for estimating the unknown parameters. Another limitation is the lack of closed form for mean square optimal prediction. However, optimal linear prediction can be considered satisfactory approximation method of prediction (DeOliveira, 2006; Bevilacqua et al., 2020). In Section 5, we have shown through a simulation study that when using optimal linear prediction for the proposed two-piece models, a better prediction performance is obtained with respect to the optimal (misspecified) Gaussian predictor. Finally, the proposed methodology can be easily extended to other continuos spaces such as the space-time framework (Gneiting, 2002; Stein, 2005) or the spherical space (Gneiting, 2013; Porcu et al., 2016) by choosing a suitable underlying correlation function and in can be also extended to discrete spaces by using, for instance, an underling Gaussian Markov random field (Rue and Held, 2005) with a specified precision matrix.

**Acknowledgements** Partial support was provided by FONDECYT grant 1200068, Chile and by ANID – Millennium Science Initiative Program-NCN17_059 and by regional MATH-AmSud program, grant number 20-MATH-03 for Moreno Bevilacqua and by Proyecto Regular Interno DIUBB 2120538 IF/R de la Universidad del Bío-Bío for Christian Caamaño. The authors thank the associate editor, and two referees for their comments and suggestions that led to an improved presentation.

## Appendix

### 7.1 Proof of Lemma 1

*Proof* We make use of some special functions in this proof. In particular the parabolic cylinder function $D_n(x)$, the confluent hypergeometric function ${}_1F_1(a;b;x)$ and the Gaussian hypergeometric function ${}_2F_1(a;b;c;x)$ (see Gradshteyn and Ryzhik (2007) for the definitions of these functions). By definition, we have:

$$\begin{aligned}
\mathbb{E}(|T_h(\boldsymbol{s}_i)||T_h(\boldsymbol{s}_j)|) &= \mathbb{E}(|G(\boldsymbol{s}_i)e^{\frac{h(G(\boldsymbol{s}_i))^2}{2}}||G(\boldsymbol{s}_j)e^{\frac{h(G(\boldsymbol{s}_j))^2}{2}}|) \\
&= \int\limits_{\mathbb{R}^2_+} |g_i e^{\frac{hg_i^2}{2}}||g_j e^{\frac{hg_j^2}{2}}| f_{|\boldsymbol{G}_{ij}|}(g_i,g_j)dg_i dg_j \\
&= \frac{1}{\pi(1-\rho^2(\boldsymbol{h}))^{1/2}} \int\limits_{\mathbb{R}^2_+} g_i g_j e^{-\frac{1}{2(1-\rho^2(\boldsymbol{h}))}\left[g_i^2+g_j^2-2\rho(\boldsymbol{h})g_i g_j\right]} e^{\frac{hg_i^2}{2}+\frac{hg_j^2}{2}} dg_i dg_j \\
&+ \frac{1}{\pi(1-\rho^2(\boldsymbol{h}))^{1/2}} \int\limits_{\mathbb{R}^2_+} g_i g_j e^{-\frac{1}{2(1-\rho^2(\boldsymbol{h}))}\left[g_i^2+g_j^2+2\rho(\boldsymbol{h})g_i g_j\right]} e^{\frac{hg_i^2}{2}+\frac{hg_j^2}{2}} dg_i dg_j \\
&= A_1 + A_2. \qquad (32)
\end{aligned}$$

Taking the first integral of (32) and using (3.462.1) of Gradshteyn and Ryzhik (2007), we obtain

$$A_1 = \frac{1}{\pi(1-\rho^2(\boldsymbol{h}))^{1/2}} \int\limits_{\mathbb{R}_+} g_j e^{-\frac{[1-(1-\rho^2(\boldsymbol{h}))h]g_j^2}{2(1-\rho^2(\boldsymbol{h}))}} \left[ \int\limits_{\mathbb{R}_+} g_i e^{\left[-\frac{[1-(1-\rho^2(\boldsymbol{h}))h]g_i^2}{2(1-\rho^2(\boldsymbol{h}))} + \frac{\rho(\boldsymbol{h})g_i g_j}{(1-\rho^2(\boldsymbol{h}))}\right]} dg_i \right] dg_j$$

$$= \frac{1}{\pi(1-\rho^2(\boldsymbol{h}))^{1/2}} \left[\frac{(1-\rho^2(\boldsymbol{h}))}{1-(1-\rho^2(\boldsymbol{h}))h}\right] \int\limits_{\mathbb{R}_+} g_j e^{-\left[\frac{[1-(1-\rho^2(\boldsymbol{h}))h]}{2(1-\rho^2(\boldsymbol{h}))} - \frac{\rho^2(\boldsymbol{h})}{4(1-\rho^2(\boldsymbol{h}))[1-(1-\rho^2(\boldsymbol{h}))h]}\right] g_j^2}$$

$$\times D_{-2}\left(-\frac{\rho(\boldsymbol{h})g_j}{\sqrt{(1-\rho^2(\boldsymbol{h}))[1-(1-\rho^2(\boldsymbol{h}))h]}}\right) dg_j, \tag{33}$$

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

$$D_{-2}\left(-\frac{\rho(\boldsymbol{h})g_j}{\sqrt{(1-\rho^2(\boldsymbol{h}))[1-(1-\rho^2(\boldsymbol{h}))h]}}\right) = e^{-\frac{\rho^2(\boldsymbol{h})g_j^2}{4(1-\rho^2(\boldsymbol{h}))[1-(1-\rho^2(\boldsymbol{h}))h]}}$$
$$\times {}_1F_1\left(1;\frac{1}{2};\frac{\rho^2(\boldsymbol{h})g_j^2}{2(1-\rho^2(\boldsymbol{h}))[1-(1-\rho^2(\boldsymbol{h}))h]}\right)$$
$$+ \frac{\sqrt{2\pi}\rho(\boldsymbol{h})g_j}{2\sqrt{(1-\rho^2(\boldsymbol{h}))[1-(1-\rho^2(\boldsymbol{h}))h]}} e^{-\frac{\rho^2(\boldsymbol{h})g_j^2}{4(1-\rho^2(\boldsymbol{h}))[1-(1-\rho^2(\boldsymbol{h}))h]}}$$
$$\times {}_1F_1\left(\frac{3}{2};\frac{3}{2};\frac{\rho^2(\boldsymbol{h})g_j^2}{2(1-\rho^2(\boldsymbol{h}))[1-(1-\rho^2(\boldsymbol{h}))h]}\right). \tag{34}$$

by combining Equations (34) and the integral of (33) and using (7.621.4) of Gradshteyn and Ryzhik (2007), we obtain

$$A_1 = \frac{(1-\rho^2(\boldsymbol{h}))^{1/2}}{\pi[1-(1-\rho^2(\boldsymbol{h}))h]} \int\limits_{\mathbb{R}_+} g_j e^{-\frac{[1-(1-\rho^2(\boldsymbol{h}))h]g_j^2}{2(1-\rho^2(\boldsymbol{h}))}} {}_1F_1\left(1;\frac{1}{2};\frac{\rho^2(\boldsymbol{h})g_j^2}{2(1-\rho^2(\boldsymbol{h}))[1-(1-\rho^2(\boldsymbol{h}))h]}\right) dg_j$$
$$+ \frac{\sqrt{2\pi}\rho(\boldsymbol{h})}{2\pi[1-(1-\rho^2(\boldsymbol{h}))h]^{3/2}} \int\limits_{\mathbb{R}_+} g_j^2 e^{-\frac{[1-(1-\rho^2(\boldsymbol{h}))h]g_j^2}{2(1-\rho^2(\boldsymbol{h}))}} {}_1F_1\left(\frac{3}{2};\frac{3}{2};\frac{\rho^2(\boldsymbol{h})g_j^2}{2(1-\rho^2(\boldsymbol{h}))[1-(1-\rho^2(\boldsymbol{h}))h]}\right) dg_j$$
$$= \frac{(1-\rho^2(\boldsymbol{h}))^{3/2}}{\pi[1-(1-\rho^2(\boldsymbol{h}))h]^2} {}_2F_1\left(1,1;\frac{1}{2};\frac{\rho^2(\boldsymbol{h})}{[1-(1-\rho^2(\boldsymbol{h}))h]^2}\right)$$
$$+ \frac{\rho(\boldsymbol{h})(1-\rho^2(\boldsymbol{h}))^{3/2}}{2[1-(1-\rho^2(\boldsymbol{h}))h]^3} {}_2F_1\left(\frac{3}{2},\frac{3}{2};\frac{3}{2};\frac{\rho^2(\boldsymbol{h})}{[1-(1-\rho^2(\boldsymbol{h}))h]^2}\right). \tag{35}$$

Similarly, the second integral of (32) is given by

$$A_2 = \frac{(1-\rho^2(\boldsymbol{h}))^{3/2}}{\pi[1-(1-\rho^2(\boldsymbol{h}))h]^2} {}_2F_1\left(1,1;\frac{1}{2};\frac{\rho^2(\boldsymbol{h})}{[1-(1-\rho^2(\boldsymbol{h}))h]^2}\right)$$
$$- \frac{\rho(\boldsymbol{h})(1-\rho^2(\boldsymbol{h}))^{3/2}}{2[1-(1-\rho^2(\boldsymbol{h}))h]^3} {}_2F_1\left(\frac{3}{2},\frac{3}{2};\frac{3}{2};\frac{\rho^2(\boldsymbol{h})}{[1-(1-\rho^2(\boldsymbol{h}))h]^2}\right). \tag{36}$$

Combining Equations (35), (36) in (32), we obtain

$$\mathbb{E}(|T_h(\boldsymbol{s}_i)||T_h(\boldsymbol{s}_j)|) = \frac{2(1-\rho^2(\boldsymbol{h}))^{3/2}}{\pi[1-(1-\rho^2(\boldsymbol{h}))h]^2} {}_2F_1\left(1,1;\frac{1}{2};\frac{\rho^2(\boldsymbol{h})}{[1-(1-\rho^2(\boldsymbol{h}))h]^2}\right).$$

Finally we use the identity:

$$ {}_2F_1\left(1,1;\frac{1}{2};x\right)=\frac{\sqrt{x}arcsin(\sqrt{x})+\sqrt{1-x}}{(1-x)^{3/2}} $$

## References


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

Arellano-Valle RB, Gómez HW, Quintana FA (2005) Statistical inference for a general class of asymmetric distributions. Journal of Statistical Planning and Inference 128(2):427 – 443

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

Azzimonti D, Ginsbourger D (2018) Estimating orthant probabilities of high dimensional gaussian vectors with an application to set estimation. Journal of Computational and Graphical Statistics 27(2):255–267

Bai Y, Kang J, Song P (2014) Efficient pairwise composite likelihood estimation for spatial-clustered data. Biometrics 7(3):661–670

Banerjee S, Carlin BP, Gelfand AE (2004) Hierarchical Modeling and Analysis for Spatial Data. Chapman & Hall/CRC Press, Boca Raton: FL

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

Bevilacqua M, Gaetan C, Mateu J, Porcu E (2012) Estimating space and space-time covariance functions for large data sets: a weighted composite likelihood approach. Journal of the American Statistical Association 107(497):268–280, DOI 10.1080/01621459.2011.646928

Bevilacqua M, Faouzi T, Furrer R, Porcu E (2019a) Estimation and prediction using generalized wendland covariance functions under fixed domain asymptotics. The Annals of Statistics 47(2):828–856

Bevilacqua M, Morales-Oñate V, Caamaño-Carrillo C (2019b) Geomodels: A package for geostatistical Gaussian and non Gaussian data analysis. https://vmoprojs.github.io/GeoModels-page/, r package version 1.0.3-4

Bevilacqua M, Caamaño-Carrillo C, Gaetan C (2020) On modeling positive continuous data with spatiotemporal dependence. Environmetrics 31(7):e2632

Bevilacqua M, Caamaño-Carrillo C, Arellano-Valle R, Morales-Oñate V (2021) Non-gaussian geostatistical modeling using (skew) t processes. Scandinavian Journal of Statistics 48:212–245

Chen D, Lu C, Kou Y, Chen F (2008) On detecting spatial outliers. Geoinformatica 12:455–475

Cote M, Genest C (2019) Dependence in a background risk model. Journal of Multivariate Analysis 172:28–46

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

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

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

Dutta S, Genton MG (2014) A non-gaussian multivariate distribution with all lower-dimensional gaussians and related families. Journal of Multivariate Analysis 132:82 – 93, DOI https://doi.org/10.1016/j.jmva.2014.07.007, URL http://www.sciencedirect.com/science/article/pii/S0047259X1400164X

Efron B (1982) The jackknife, the bootstrap and other resampling plans. CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM 38

Ernst M, Haesbroeck G (2017) Comparison of local outlier detection techniques in spatial multivariate data. Data Mining and Knowledge Discovery 31:371–399

Fechner GT (1897) Kollektivmasslehre. Engelmann

Feng X, Zhu J, Lin P, Steen-Adams M (2014) Composite likelihood estimation for models of spatial ordinal data and spatial proportional data with zero/one values. Environmetrics 25(8):571–583

Fernández C, Steel M (1998) On bayesian modeling of fat tails and skewness. Journal of the American Statistical Association 93(441):359–371

Gelfand AE, Schliep EM (2016) Spatial statistics and gaussian processes: A beautiful marriage. Spatial Statistics 18:86–104

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

Gentz A (1992) Numerical computation of multivariate normal probabilities. Journal of Computational and Graphical Statistics 1(141–150)

Genz A, Bretz F (2009) Computation of Multivariate Normal and t Probabilities, vol 195. Springer, New York

Genz A, Kenkel B (2015) pbivnorm: Vectorized Bivariate Normal CDF. URL https://cran.r-project.org/package=pbivnorm, r package version 0.6.0

Gneiting T (2002) Nonseparable, stationary covariance functions for space-time data. Journal of the American Statistical Association 97(458):590–600

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

Goerg GM (2015) The lambert way to gaussianize heavy-tailed data with the inverse of tukey's h transformation as a special case. The Scientific World Journal p 1–16

Gradshteyn I, Ryzhik I (2007) Table of Integrals, Series, and Products, 7th edn. Academic Press, New York

Gräler B (2014) Modelling skewed spatial random fields through the spatial vine copula. Spatial Statistics 10:87 –102

Haining R (1993) Spatial Data Analysis in the Social and Environmental Sciences. Cambridge University Press

Haslett J, Brandley R, Craig P, Unwin A, Wills G (1991) Dynamic graphics for exploringspatial data with application to locating global and local

anomalies. The American Statistician 45:234–242
Heagerty P, Lele S (1998) A composite likelihood approach to binary spatial data. Journal of the American Statistical Association 93(443):1099 –1111
Joe H (2014) Dependence Modeling with Copulas. Chapman and Hall/CRC, Boca Raton, FL
Joe H, Lee Y (2009) On weighting of bivariate margins in pairwise likelihood. Journal of Multivariate Analysis 100(4):670–685
Jones MC (2015) On families of distributions with shape parameters. International Statistical Review 83(2):175–192
Kazianka H, Pilz J (2010) Copula-based geostatistical modeling of continuous and discrete data including covariates. Stochastic Environmental Research and Risk Assessment 24:661–673
Kilibarda M, Hengl T, Heuvelink GBM, Gräler B, Pebesma E, Perčec Tadić M, Bajat B (2014) Spatio-temporal interpolation of daily temperatures for global land areas at 1km resolution. Journal of Geophysical Research: Atmospheres 119(5):2294–2313
Kou Y, Lu CT, Dos Santos RF (2007) Spatial outlier detection: A graph-based approach. In: 19th IEEE International Conference on Tools with Artificial Intelligence, vol 1, pp 281–288
Lindsay B (1988) Composite likelihood methods. Contemporary Mathematics 80:221–239
Lu CT, Chen D, Kou Y (2003) Detecting spatial outliers with multiple attributes. In: Proceedings. 15th IEEE International Conference on Tools with Artificial Intelligence, pp 122–128
Masarotto G, Varin C (2012) Gaussian copula marginal regression. Electronic Journal of Statistics 6:1517–1549
Mudholkar GS, Hutson AD (2000) The epsilon skew-normal distribution for analyzing near-normal data. Journal of Statistical Planning and Inference 83(2):291–309
Murthy GSR (2015) A note on multivariate folded normal distribution. Sankhya B 77:108 –113
Porcu E, Bevilacqua M, Genton MG (2016) Spatio-temporal covariance and cross-covariance functions of the great circle distance on a sphere. Journal of the American Statistical Association 111(514):888–898
R Core Team (2020) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL https://www.R-project.org/
Rubio FJ, Steel MFJ (2020) The family of two-piece distributions. Significance 17(1):12–13
Rue H, Held L (2005) Gaussian Markov Random Fields: Theory and Applications. Chapman & Hall, London
Shekhar S, Lu CT, Zhang P (2001) Detecting graph-based spatial outliers: Algorithms and applications (a summary of results). In: Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, New York, NY, USA, KDD '01, pp 371–376

Shekhar S, Lu C, Zhang P (2003) A unified approach to detecting spatial outliers. GeoInformatica 7:139–166

Singh AK, Lalitha S (2018) A novel spatial outlier detection technique. Communications in Statistics - Theory and Methods 47(1):247–257

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

Stein M (2005) Space-time covariance functions. Journal of the American Statistical Association 100(492):310–321

Varin C, Vidoni P (2005) A note on composite likelihood inference and model selection. Biometrika 52(3):519–528

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

Wallis KF (2014) The two-piece normal, binormal, or double gaussian distribution: its origin and rediscoveries. Statistical Science 29(1):106–112

Xua G, Genton MG (2017) Tukey g-and-h random fields. Journal of the American Statistical Association 112(519):1236 –1249

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