Computational Statistics and Data Analysis 191 (2024) 107887

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Nearest neighbors weighted composite likelihood based on pairs for (non-)Gaussian massive spatial data with an application to Tukey-*hh* random fields estimation

Christian Caamaño-Carrillo [a,*], Moreno Bevilacqua [b,c], Cristian López [a], Víctor Morales-Oñate [d,e]

[a] *Departamento de Estadística, Universidad del Bío-Bío, Avda. Collao 1202, Concepción, 4030000, Chile*
[b] *Facultad de Ingeniería y Ciencias, Universidad Adolfo Ibáñez, Avda. Padre Hurtado 750, Viña del Mar, 2520000, Chile*
[c] *Dipartimento di Scienze Ambientali, Informatica e Statistica, Ca' Foscari University of Venice, via Torino 155, Mestre, 30172, Italy*
[d] *Department of Economics, Universidad de las Américas, Vía a Nayón, Quito, 170124, Ecuador*
[e] *School of Economics, Universidad San Francisco de Quito, Diego de Robles s/n, Quito, 170901, Ecuador*

ARTICLE INFO

Keywords:
Covariance estimation
Geostatistics
Large datasets
Vecchia approximation

ABSTRACT

A highly scalable method for (non-)Gaussian random fields estimation is proposed. In particular, a novel (a) symmetric weight function based on nearest neighbors for the method of maximum weighted composite likelihood based on pairs (WCLP) is studied.
The new weight function allows estimating massive (up to millions) spatial datasets and improves the statistical efficiency of the WCLP method using symmetric weights based on distances, as shown in the numerical examples.
As an application of the proposed method, the estimation of a novel non-Gaussian random field named Tukey-*hh* random field that has flexible marginal distributions (possibly skewed and/or heavy-tailed) is considered. In an extensive simulation study the statistical efficiency of the proposed nearest neighbors WCLP method with respect to the WCLP method using weights based on distances is explored when estimating the parameters of the Tukey-*hh* random field. In the Gaussian case the proposed method is compared with the Vecchia approximation from computational and statistical viewpoints. Finally, the effectiveness of the proposed methodology is illustrated by estimating a large dataset of mean temperatures in South-America. The proposed methodology has been implemented in an open-source package for the R statistical environment.

1. Introduction

Many applications of statistics across a wide range of disciplines rely on the estimation of the spatial dependence of a physical process based on irregularly spaced observations and then predict the process at some unknown spatial locations. Gaussian random fields (RFs) are among the most popular tools for analyzing data in spatial statistics (Banerjee et al., 2004; Cressie and Wikle, 2011; Stein, 1999) and several other disciplines, such as machine learning and image analysis, as well as in other branches of applied mathematics including numerical analysis and interpolation theory.

* Corresponding author.
E-mail address: chcaaman@ubiobio.cl (C. Caamaño-Carrillo).

https://doi.org/10.1016/j.csda.2023.107887
Received 25 November 2022; Received in revised form 25 October 2023; Accepted 30 October 2023

Available online 10 November 2023

Unfortunately, practical use of Gaussian RFs has two potential problems. The first problem is from a computational viewpoint. The estimation of Gaussian RFs with the maximum likelihood (ML) method involves $O(n^3)$ operations and $O(n^2)$ memory storage, if n is the number of location sites, which can be computationally impractical when n is only moderately large. This fact motivates the search for estimation methods with a good balance between statistical efficiency and computational complexity. Different estimation methods have been proposed in the recent years to deal with this goal. Among them methods based on low rank structure on the covariance matrix (Banerjee et al., 2008; Cressie and Johannesson, 2008; Stein, 2008), based on tapered covariance matrix (Furrer et al., 2013; Kaufman et al., 2008; Stein, 2013), based on composite likelihood (Bevilacqua and Gaetan, 2015; Eidsvik et al., 2014; Varin et al., 2011), based on approximation using Markov Gaussian RFs (Lindgren et al., 2011) based on multiresolution approximations (Katzfuss, 2017; Katzfuss and Gong, 2020) to name just a few. A general framework that includes several proposals based on Vecchia approximation (Vecchia, 1988) has recently been proposed in Katzfuss and Guinness (2021). For an extensive review see Heaton et al. (2019) and the references therein.

The second problem is from a modeling viewpoint. Indeed, in many geostatistical applications, including climatology, oceanography, the environment and the study of natural resources, the Gaussian framework is unrealistic because the observed data have specific features such as asymmetry and/or heavy tails. One popular approach for modeling this kind of data is the hierarchical model proposed by Diggle et al. (1998) that can be viewed as a generalized linear mixed model (Diggle and Ribeiro, 2007; Diggle and Giorgi, 2019). 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, this kind of construction has some drawbacks. For instance the underlying conditional independence assumption leads to a "forced" nugget effect (Gelfand and Schliep, 2016) that is a discontinuity at the origin of the associated covariance function and this can be troublesome when modeling spatial data displaying some kind of continuity. A scalable method of estimation for this kind of models based on the Vecchia-Laplace approximation has been proposed in Zilber and Katzfuss (2021). Wallin and Bolin (2015) proposed non-Gaussian RFs derived from stochastic partial differential equations to model non-Gaussian spatial data. However, this approach is restricted to the Matérn covariance model with an integer smoothness parameter and its statistical properties are much less understood than those of Gaussian RFs.

A very flexible class of non-Gaussian RFs that solve these potential drawbacks can be obtained through a suitable transformation of one or independent copies of (transformed) Gaussian RFs sharing a common correlation function. Specifically, let $Z = \{Z(\boldsymbol{s}), \boldsymbol{s} \in A\}$, $A \subset \mathbb{R}^d$ a Gaussian RF and let $Y = \{Y(\boldsymbol{s}), \boldsymbol{s} \in A\}$ a RF defined through the transformation

$$Y(\boldsymbol{s}) = f(g_1(Z_1(\boldsymbol{s})), g_2(Z_2(\boldsymbol{s})), ...g_q(Z_q(\boldsymbol{s}))), \quad q \geq 1, \tag{1}$$

where $Z_1, \ldots, Z_q$, are independent copies of Z and $f : \mathbb{R}^q \to \mathbb{R}$ and $g_1, \ldots, g_q$ with $g_i : \mathbb{R} \to \mathbb{R}$ are suitable functions. The class (1) includes several examples of non-Gaussian RFs proposed in the literature such us Bernoulli RFs (Heagerty and Lele, 1998), skew-Gaussian RFs (Zhang and El-Shaarawi, 2010), Tukey $g-h$ RFs Xua and Genton (2017), Student-t RFs (Bevilacqua et al., 2021), Two Piece RFs (Bevilacqua et al., 2022a), Weibull RFs (Bevilacqua et al., 2020) or Poisson RFs (Morales-Navarrete et al., 2022) to mention just a few. In addition, the so-called class of trans-Gaussian RFs (see for instance DeOliveira et al. (1997) and Allcroft and Glasbey (2003)) or the general class of non-Gaussian-RFs based on Gaussian Copula (Kazianka and Pilz, 2010; Masarotto and Varin, 2012; Gräler, 2014) and chi-square Copula (Bárdossy, 2006) belong to the class (1). The general class of non-Gaussian RFs in (1) is a convenient approach since geometrical properties such as mean square continuity and differentiability can be inherited from the underlying Gaussian RF by using flexible correlation models such as the Matérn (Stein, 1999) of the generalized Wendland model (Bevilacqua et al., 2019).

When estimating non-Gaussian RFs such as those belonging to class (1), the computational complexity can be even harder than the Gaussian case, depending on the type of transformation involved. If the non-Gaussian RF is obtained through a monotonic transformation of a Gaussian RF ($q = 1$) and, the inverse transformation has a closed form, then the computation of the multivariate density requires, as in the Gaussian case, $O(n^3)$ operations and $O(n^2)$ memory storage. Some notable examples are Log-Gaussian RFs (Oliveira, 2006) or the class of Gaussian copula RFs (Kazianka and Pilz, 2010; Masarotto and Varin, 2012) or the sinh-arcsinh RFs proposed in Yan et al. (2020) and Blasi et al. (2022). If the transformation is not monotonic and/or involves independent copies of (transformed) Gaussian RFs then the associated multivariate distribution can be computationally prohibitive even for a small n. For instance ML estimation of the skew-Gaussian RF proposed in Zhang and El-Shaarawi (2010) requires computation of order $O(2^{n-1})$. Another notable example is the ML estimation of the Bernoulli RFs proposed in Heagerty and Lele (1998). In this case the transformation is not continuous and the likelihood evaluation requires computation of 2^{n-1}, n-dimensional normal integrals. In other cases the likelihood is completely unknown as, for instance, in the t RFs proposed in Bevilacqua et al. (2021) or the Poisson RF in Morales-Navarrete et al. (2022)

To address the abovementioned computational problem we consider the method of composite likelihood (CL) (Lindsay, 1988; Varin et al., 2011). CL is a general class of objective functions based on the likelihood of marginal or conditional events that has been successfully applied in the recent years when estimating (non-) Gaussian RFs. For instance, in the Gaussian case, Bevilacqua and Gaetan (2015) considered a weighted composite likelihood based on pairs (WCLP hereafter) while Eidsvik et al. (2014) developed a block composite likelihood in a vein similar to Caragea and Smith (2006). Furthermore, the methods proposed, for instance, in Stein et al. (2004) or Guinness (2018), based on Vecchia approximation (Vecchia, 1988), can be viewed as CL methods.

A benefit of using WCLP with respect to other types of CL methods is that in some complex non-Gaussian RFs as those belonging to class (1), the multivariate distribution is unknown and/or difficult to compute but the bivariate density is known and relatively simply to evaluate as for instance in the aforementioned Bernoulli, t, Poisson and skew-Gaussian RFs. In this case, estimation with other types of CL, such as the CL based on independent blocks, is troublesome. As a consequence, WCLP estimation has a broader

applicability than other types of CL and can be performed as long as the bivariate of the non-Gaussian RFs can be evaluated. For this reason hereafter, we focus on WCLP.

The contribution of this paper is twofold. First, a novel asymmetric weight function based on nearest neighbors is proposed for the weights involved in the WCLP estimation method. The main benefits of the proposed nearest neighbors WCLP (NNWCLP hereafter) method are as follows: 1) it improves the statistical efficiency of the WCLP method that uses a symmetric weight function based on distances (DDWCLP hereafter) as proposed for instance in Heagerty and Lele (1998); Varin and Vidoni (2005); Bai et al. (2014); Feng et al. (2014); Bevilacqua and Gaetan (2015) and 2) it allows performing estimation of massive datasets (up to millions) since kd-tree algorithms (Bentley, 1975; Arya et al., 1998) can be exploited to achieve an objective function requiring $O(nm)$ time complexity, where m is the order of nearest neighbors involved, and $O(n)$ memory storage.

The second contribution is a novel non-Gaussian RF that has flexible marginal distributions, possibly skewed and/or heavy-tailed. Our proposal falls into class (1) and is similar to Tukey-gh RFs proposed in Xua and Genton (2017) which is based on a generalization of a RF with Tukey-h marginals (Goerg, 2015). The benefit of our transformation with respect to the Tukey-gh transformation is to possess an explicit inverse, and as a consequence, likelihood-based methods can be readily applied. We provide analytic expressions for the covariance function and for the multivariate probability density function of the proposed Tukey-hh The evaluation of the multivariate pdf, as in the Gaussian case, is computationally expensive for massive datasets. However the bivariate distribution can be easily evaluated and, as a consequence, the NNWCLP method is a suitable estimation tool for this kind of model.

In an extensive simulation study we compare the NNWCLP method versus the DDWCLP method when estimating the parameters of the Tukey-hh RF. It turns out that the NNWCLP method clearly outperforms DDWCLP from a statistical efficiency viewpoint. Then we compare the NNWCLP method with the standard maximum likelihood method. In the purely Gaussian case we also compare the NNWCLP method with some recent improvements (Katzfuss and Guinness, 2021; Guinness, 2021) of Vecchia approximation method originally proposed in Vecchia (1988). It turns out that the proposed method shows a reasonable loss of statistical efficiency with the Vecchia method and, at the same time, a substantial gain in terms of computational time.

Finally we apply the proposed methodology by analyzing a large geo-referenced dataset (approximatively 360,000 data) from the ERA5-Land dataset (Muñoz Sabater et al., 2021) of the mean temperature over the first two months of 2020 in South America.

The methodology considered in this paper has been implemented in the `GeoModels` R package (Bevilacqua et al., 2023) and R code for reproducing the work is available as an online supplement. We want to stress that this paper has been motivated by the first and second "Competition on Spatial Statistics for Large Datasets" (Huang et al., 2021; Abdulah et al., 2022) organized by King Abdullah University of Science and Technology (KAUST). In both competitions the analysis has been performed using the `GeoModels` package achieving very competitive results. For instance, the *GeoModels team* reached the second position in the competition for large spatial datasets (Abdulah et al., 2022).

The remainder of the paper is organized as follows. In Section 2 we review the WCLP estimation method and we introduce the proposed weight function based on nearest neighbors. In Section 3 we introduce the Tukey-hh RFs and provide analytic expressions for the correlation function and the bivariate and multivariate distributions. In Section 4, we present an extensive simulation study to investigate the computational and statistical performance of the NNWCLP method. In Section 5, we present the analysis of the mean temperature data. Finally, in Section 6, we provide some conclusions. All the proofs have been deferred to the Appendix.

2. Nearest neighbors weighted composite likelihood estimation based on pairs

For the rest of the paper, given an RF $Z = \{Z(\boldsymbol{s}), \boldsymbol{s} \in A\}$ defined on $A \subset \mathbb{R}^d$, with $\mathbb{E}(Z(\boldsymbol{s})) = \mu(\boldsymbol{s})$ and $Var(Z(\boldsymbol{s})) = \sigma^2$, we denote by $\rho_Z(\boldsymbol{d}) = Corr(Z(\boldsymbol{s}_i), Z(\boldsymbol{s}_j))$ its correlation function, where $\boldsymbol{d} = \boldsymbol{s}_i - \boldsymbol{s}_j$ is the lag separation vector.

For any set of distinct points $(\boldsymbol{s}_1, \dots, \boldsymbol{s}_n)^T$, $\boldsymbol{s}_i \in A$, $n \in \mathcal{N}$, we denote by $\boldsymbol{Z}_{ij} = (Z(\boldsymbol{s}_i), Z(\boldsymbol{s}_j))^T$, $i \neq j$ and $Z_{i|j} = Z(\boldsymbol{s}_i) \mid Z(\boldsymbol{s}_j) = z_j$ the bivariate random vector and the conditional random variable respectively and we denote by $\boldsymbol{Z} = (Z(\boldsymbol{s}_1), \dots, Z(\boldsymbol{s}_n))^T$ the multivariate random vector. In addition, we denote with $f_{\boldsymbol{Z}_{ij}}$, $f_{Z_{i|j}}$ and $f_{\boldsymbol{Z}}$ the associated probability density functions and we denote f_{Z_k} as the marginal density function of $Z(\boldsymbol{s}_k)$. Finally, we denote Z^* as the standardized RF, *i.e.*, $Z^*(\boldsymbol{s}) := (Z(\boldsymbol{s}) - \mu(\boldsymbol{s}))/\sigma$.

Following Varin et al. (2011) and Lindsay (1988) consider a random vector $\boldsymbol{Z}$ with pdf $f_{\boldsymbol{Z}}(z;\boldsymbol{\theta})$ for some unknown parameter vector $\boldsymbol{\theta}$. Denote by $\{B_1, \dots, B_K\}$ a set of marginal or conditional events with associated log-likelihood $l_k(z;\boldsymbol{\theta}) = log(f_{\boldsymbol{Z}}(z \in B_k, \boldsymbol{\theta}))$, $k = 1, \dots, K$. The log-CL is an objective function defined as a sum of K sub-log-likelihoods

$$CL(\boldsymbol{\theta}) = \sum_{k=1}^{K} l_k(z;\boldsymbol{\theta}) w_k, \tag{2}$$

where w_k are suitable weights that do not depend on $\boldsymbol{\theta}$. The maximum CL estimate is given by $\hat{\boldsymbol{\theta}} = \text{argmax}_{\boldsymbol{\theta}} CL(\boldsymbol{\theta})$.

The WCLP estimation method (Bevilacqua and Gaetan, 2015) is a special case of (2) obtained by setting $B_k = \boldsymbol{Z}_{ij}$ or $B_k = Z_{i|j}$. In the first case we obtain the pairwise marginal log-likelihood $l_{ij}(\boldsymbol{\theta}) = log(f_{\boldsymbol{Z}_{ij}}(z_{ij}, \boldsymbol{\theta}))$ and in the second case we obtain the pairwise conditional log-likelihood $l_{i|j}(\boldsymbol{\theta}) = log(f_{\boldsymbol{Z}_{ij}}(z_{ij}, \boldsymbol{\theta}) / f_{Z_j}(z_j, \boldsymbol{\theta}))$. The corresponding weighted composite log-likelihoods functions are given by:

$$wpl_M(\boldsymbol{\theta}) = \sum_{i=1}^{n} \sum_{j \neq i}^{n} l_{ij}(\boldsymbol{\theta}) w_{ij}, \qquad wpl_C(\boldsymbol{\theta}) = \sum_{i=1}^{n} \sum_{j \neq i}^{n} l_{i|j}(\boldsymbol{\theta}) w_{ij}, \tag{3}$$

and $\hat{\boldsymbol{\theta}}_a = \text{argmax}_{\boldsymbol{\theta}} wpl_a(\boldsymbol{\theta})$, where $a = M, C$ is the associated estimator. Note that, assuming non-zero weights, the computational cost associated with both functions is of order $O(n^2)$. In general, a loss of statistical efficiency is expected for both cases with respect to the

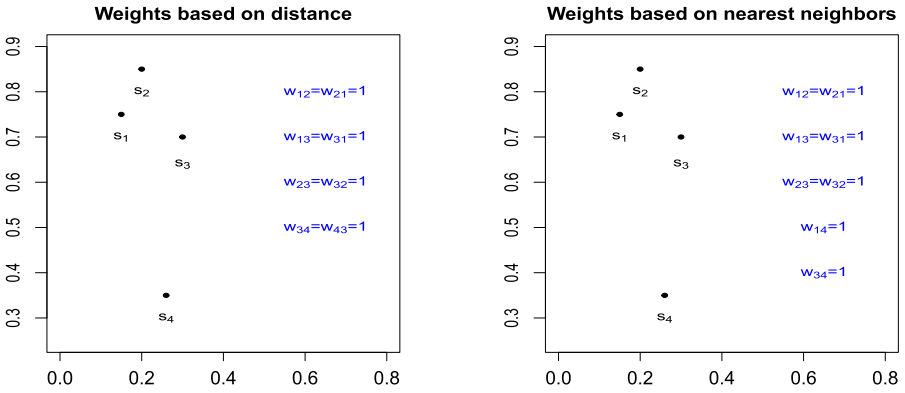

Fig. 1. Left part: location sites of the toy example and the weights selected using the weight function (4) based on distances $w_{ij}(k)$ with $k = 0.36$. Right part: location sites of the example and the weights selected using the weight function (5) based on nearest neighbors $w_{ij}(m)$ with $m = 2$.

ML estimation and the role of the weights w_{ij} is to minimize this loss. Using theory of optimal estimating equations (Heyde, 1997), it can be easily seen (Bevilacqua et al., 2012) that the optimal weights require the computation of the inverse of a $n(n-1) \times n(n-1)$ matrix which is even computationally harder than the requirements for ML estimation. Some approximations of the optimal weights have been proposed in literature as for instance in Li and Sang (2018) and Pace et al. (2019). However the computation of this kind of weights can be computationally demanding for large n.

To avoid this computational problem different authors (Bai et al., 2014; Bevilacqua and Gaetan, 2015; Feng et al., 2014; Heagerty and Lele, 1998; Varin and Vidoni, 2005) have proposed the DDWCLP method that considers the weight function:

$$w_{ij}(k) = \begin{cases} 1 & \|\boldsymbol{s}_i - \boldsymbol{s}_j\| < k \\ 0 & \text{otherwise,} \end{cases} \tag{4}$$

where $k \in \mathbb{R}^+$ is an arbitrary distance greater than the minimum distance of the location points. This kind of weights allows ruling out a certain percentage (depending on k) of the total number of pairs allowing a clear computational gain with respect to the non-zero weighted version. Additionally, it has been shown that this kind of weights improves the statistical efficiency of the method with respect to the use of constant weights (see for instance Joe and Lee (2009), Davis and Yau (2011) and Bevilacqua et al. (2012)).

It should be outlined that the function (4) restricts the weights to be symmetric i.e. $w_{ij}(k) = w_{ji}(k)$ and this is a potential limitation, particularly when considering the conditional likelihood $l_{i|j}(\boldsymbol{\theta})$ which is not symmetric.

Our proposal (NNWCLP) considers weights based on nearest neighbors that can be either symmetric or not symmetric. Specifically, let $N_m(\mathbf{s}_l)$ be the set of neighbors of order $m = 1, 2, \ldots$ of the point $\mathbf{s}_l \in A$. We propose the following weight function:

$$w_{ij}(m) = \begin{cases} 1 & \mathbf{s}_i \in N_m(\mathbf{s}_j) \\ 0 & \text{otherwise,} \end{cases} \tag{5}$$

for $i, j = 1, \ldots, n$ and $i \neq j$. By construction the weights $w_{ij}(m)$ can be either symmetric or not and to illustrate this we consider a simple toy example with four location sites $\mathbf{s}_1 = (0.15, 0.75)^T$, $\mathbf{s}_2 = (0.2, 0.85)^T$, $\mathbf{s}_3 = (0.3, 0.7)^T$ and $\mathbf{s}_4 = (0.26, 0.35)^T$. The left part of Fig. 1 depicts the weights selected using the weight function (4) based on distances $w_{ij}(k)$ and setting $k = 0.36$. On the right part, the weights selected using the weight function (5) based on nearest neighbors $w_{ij}(m)$ with $m = 2$ are depicted (the zero weights are ignored in both cases). It can be appreciated that the number of selected weights is the same in both cases and the two weight functions share most of the weights. However, the NNWCLP method includes the weights $w_{14} = 1$ and $w_{34} = 1$ while the DDWCLP method includes the symmetric weights $w_{34} = 1$ and $w_{43} = 1$.

This simple example shows that the proposed weight function can potentially include weights (and as a consequence pairwise or conditional log-likelihoods) that the method based on distances ignores. Thus, in principle, more information is considered when estimating with NNWCLP than with DDWCLP. An interesting question is whether this implies a gain in statistical efficiency. In Section 4, a simulation study shows that the proposed NNWCLP method actually outperforms the DDWCLP method.

In addition, the proposed weight function is computationally convenient since kd-tree type algorithms (Elseberg et al., 2012; Bentley, 1975; Arya et al., 1998) can be exploited to drastically reduce the computational costs of the WCLP functions in Equation (3). Two preliminary steps are required before the optimization of the WCLP functions: 1) building a kd-tree that typically requires $O(nlog(n))$ time complexity and $O(n)$ associated storage and 2) searching for m nearest neighbors inside the kd-tree that has an $O(mlog(n))$ time complexity. In our implementation in the *R* package `GeoModels` these preliminary steps are performed, using the function `GeoNeighIndex` that exploits the function `knn` of the *R* package *nabor* (Elseberg et al., 2012).

The final step involves the maximization of $wpl_a(\boldsymbol{\theta})$, $a = M, C$ functions in (3) that can be computed in $O(mn)$ time, summing up only the $l_{ij}(\boldsymbol{\theta})$ or $l_{i|j}(\boldsymbol{\theta})$ functions selected through the nearest neighbors weight function.

The maximum XWCLP estimator, where X = NN and DD is given by

$$\widehat{\boldsymbol{\theta}}_a := \text{argmax}_{\boldsymbol{\theta}}\ wpl_a(\boldsymbol{\theta}), \quad a = M, C,$$

and, as in Bevilacqua and Gaetan (2015), under some mixing conditions of the Tukey-*hh* RF and under the assumption that the weight function is compactly supported as in (4) or (5), it can be shown that, under increasing domain asymptotics, $\hat{\theta}_a$ is consistent and asymptotically Gaussian with the asymptotic covariance matrix given by $G_a^{-1}(\boldsymbol{\theta})$ the inverse of the Godambe information $G_a(\boldsymbol{\theta}) := H_a(\boldsymbol{\theta})J_a(\boldsymbol{\theta})^{-1}H_a(\boldsymbol{\theta})$, where $H_a(\boldsymbol{\theta}) := \mathbb{E}[-\nabla^2 wpl_a(\boldsymbol{\theta})]$ and $J_a(\boldsymbol{\theta}) := \mathrm{Var}[\nabla wpl_a(\boldsymbol{\theta})]$. Standard error estimation can be obtained considering the square root diagonal elements of $G_a^{-1}(\hat{\boldsymbol{\theta}})$. Moreover, model selection can be performed by considering the information criterion, defined as

$$\mathrm{PLIC} := -2wpl_a(\hat{\boldsymbol{\theta}}) + 2\mathrm{tr}(H_a(\hat{\boldsymbol{\theta}})G_a^{-1}(\hat{\boldsymbol{\theta}})), \tag{6}$$

which is composite likelihood version of the Akaike information criterion (AIC) (Varin and Vidoni, 2005). Note that, the computation of standard errors and PLIC require evaluation of the matrices $H_a(\hat{\boldsymbol{\theta}})$ and $J_a(\hat{\boldsymbol{\theta}})$. However, the evaluation of $J_a(\hat{\boldsymbol{\theta}})$ is computationally unfeasible for large datasets and in this case subsampling techniques can be used to estimate $J_a(\boldsymbol{\theta})$ as in Bevilacqua et al. (2012) and Heagerty and Lele (1998). A straightforward and more robust alternative that we adopt in Section 4.1 and in Section 5, is parametric bootstrap estimation of $G_a^{-1}(\boldsymbol{\theta})$. Since this technique is simulation based, fast methods of simulation of Gaussian RFs such as circulant embedding or turning bands methods (Emery et al., 2016) are required for large datasets.

3. Tukey-*hh* random fields

Let $G^* = \{G^*(\boldsymbol{s}), \boldsymbol{s} \in A\}$ be a zero mean and unit variance weakly stationary Gaussian RF with correlation function $\rho_{G^*}(\boldsymbol{d})$ and hereafter, with some abuse of notation, we set $\rho(\boldsymbol{d}) := \rho_{G^*}(\boldsymbol{d})$ and $G := G^*$. In addition, hereafter, with $\phi_n(\cdot, \boldsymbol{\mu}, \Sigma)$ we denote the pdf of the Gaussian n-variate distribution with mean $\boldsymbol{\mu}$ and covariance matrix Σ.

Following Xua and Genton (2017), let $T_h^* = \{T_h^*(\boldsymbol{s}), \boldsymbol{s} \in A\}$, with $h \in [0, 1/2)$, be an RF with a standard Tukey-h marginal distribution defined through a monotonic transformation $\tau_h(x) = xe^{\frac{hx^2}{2}}, \quad x \in \mathbb{R}$ of a standard Gaussian RF G as:

$$T_h^*(\boldsymbol{s}) =: \tau_h(G(\boldsymbol{s})). \tag{7}$$

The inverse transformation $\tau_h^{-1}(x)$ can be expressed in terms of the Lambert function *i.e.*, $\tau_h^{-1}(x) = sign(x)\left(\frac{W(hx^2)}{h}\right)^{1/2}$ where $W(\cdot)$ is the Lambert-W function (Goerg, 2015).

This kind of RF has marginal symmetric distributions and the parameter h governs the tail behavior of the RF, with a larger value of h indicating a heavier tail. Specifically the marginal distribution has (asymptotically) a Pareto-heavy tailed distribution with a tail index equal to $1/h$ (Morgenthaler and Tukey, 2000). If $h = 0$ the Gaussian RF G is obtained as the special limit case.

For the Tukey-h RFs $\mathbb{E}(T_h^*(\boldsymbol{s})) = 0$ and $Var(T_h^*(\boldsymbol{s})) = (1-2h)^{-3/2}$ and the correlation function is given by:

$$\rho_{T_h^*}(\boldsymbol{d}) = \frac{\rho(\boldsymbol{d})(1-2h)^{3/2}}{[(1-h)^2 - h^2\rho^2(\boldsymbol{d})]^{3/2}}\,. \tag{8}$$

In addition, the Tukey-h RF has a marginal pdf given by (Goerg, 2015):

$$f_{T_h^*}(t) = \frac{\tau_h^{-1}(t)}{t(1+W(ht^2))}\phi(\tau_h^{-1}(t), 0, 1)\,. \tag{9}$$

Note that $f_{T_h^*}(t)$ is well defined when $t \to 0$ and/or $h \to 0$ using the limits $\lim_{h\to 0}\tau_h^{-1}(t) = t$, $\lim_{h\to 0}W(ht^2) = 0$, $\lim_{t\to 0}\tau_h^{-1}(t) = t$, $\lim_{t\to 0}W(ht^2) = 0$. The multivariate pdf of the vector $\boldsymbol{T}_h^* = (T_h^*(\boldsymbol{s}_1), \ldots, T_h^*(\boldsymbol{s}_N))^T$ is given by:

$$f_{T_h^*}(\boldsymbol{t}) = \frac{\prod_{i=1}^n \tau_h^{-1}(t_i)}{(\prod_{i=1}^n t_i(1+W(ht_i^2))}\phi_n(\tau_h^{-1}(\boldsymbol{t}), \boldsymbol{0}, R_n)\,, \tag{10}$$

where $R_N = [\rho(\boldsymbol{s}_i - \boldsymbol{s}_j)]_{i,j=1}^n$ denotes correlation matrix associated with the underlying correlation function $\rho(\boldsymbol{d})$ and the transformation $\tau_h^{-1}(\boldsymbol{x})$ applies pointwise for a given vector $\boldsymbol{x}$.

To take into account both heavy tails and skewness, a generalization of (7) has been proposed in Xua and Genton (2017) by considering the so-called Tukey-g-h RF $M_{h,g}^* = \{M_{h,g}^*(\boldsymbol{s}), \boldsymbol{s} \in A\}$ defined as:

$$M_{h,g}^*(\boldsymbol{s}) := g^{-1}(e^{gG(\boldsymbol{s})} - 1)e^{\frac{h(G(\boldsymbol{s}))^2}{2}}, \tag{11}$$

where $g \in \mathbb{R}$ is a skewness parameter. The Tukey-h RF is obtained as special case using the limit $lim_{g\to 0}(e^{gx} - 1)/g = x$. Nevertheless, the transformation involved in (11) does not have an explicit inverse and as a consequence (composite) likelihood based methods are not readily applicable. To solve this problem Xua and Genton (2017) proposed to maximize an approximated likelihood function which is basically a multivariate extension of the algorithm proposed in Xu and Genton (2015). This algorithm is based on a linear approximation of the log-likelihood defined on a finite set of equally spaced knots in a specified interval (see Xua and Genton (2017) for the details).

Our proposal considers an alternative generalization of (7) obtained by defining the RF $T_{h_l,h_r}^* = \{T_{h_l,h_r}^*(\boldsymbol{s}), \boldsymbol{s} \in A\}$, with $h_l \in [0, 1/2)$ and $h_r \in [0, 1/2)$, as:

$$T^*_{h_l,h_r}(\boldsymbol{s}) := \begin{cases} \tau_{h_l}(G(\boldsymbol{s})), & G(\boldsymbol{s}) < 0 \\ \tau_{h_r}(G(\boldsymbol{s})), & G(\boldsymbol{s}) \geq 0. \end{cases} \tag{12}$$

The marginal distribution of $T^*_{h_l,h_r}$ is called the Tukey-hh distribution (Morgenthaler and Tukey, 2000) with pdf given by (Goerg, 2015):

$$f_{T^*_{h_l,h_r}}(t) = \frac{\tau^{-1}_{h_l}(t)}{t(1+W(h_l t^2))}\phi(\tau^{-1}_{h_l}(t),0,1)I_{(-\infty,0)}(t) + \frac{\tau^{-1}_{h_r}(t)}{t(1+W(h_r t^2))}\phi(\tau^{-1}_{h_r}(t),0,1)I_{[0,\infty)}(t), \tag{13}$$

where $I_A(x)$ denotes the indicator function of the set A. The properties of the marginal density $f_{T^*_{h_l,h_r}}(t)$ are the same as those of the Tukey-h distribution, except that the left and right tails have to be considered separately. Hereafter we call $T^*_{h_l,h_r}$ a Tukey-hh RF.

The Tukey-h RF is obtained as special case when $h_l = h_r$ and the Gaussian special limit case is obtained when $h_l = h_r \to 0$. For the Tukey-hh RF, the mean and variance, respectively, are given by;

$$\mathbb{E}(T^*_{h_l,h_r}(\boldsymbol{s})) = \frac{h_r - h_l}{\sqrt{2\pi}(1-h_l)(1-h_r)}, \tag{14}$$

$$Var(T^*_{h_l,h_r}(\boldsymbol{s})) = \frac{1}{2}\left[(1-2h_l)^{-3/2} + (1-2h_r)^{-3/2}\right] - \mathbb{E}^2(T^*_{h_l,h_r}(\boldsymbol{s})). \tag{15}$$

From (14) it is apparent that the two-parameters $h_l \in [0, 1/2)$ and $h_r \in [0, 1/2)$ can help correct for skewness through the difference $-1/2 < h_r - h_l < 1/2$, depending on whether $h_l > h_r$ (positive skewness) or $h_l < h_r$ (negative skewness) and for kurtosis (through $max(h_l, h_r)$) (Morgenthaler and Tukey, 2000).

We now study the correlation function of the Tukey-hh RF. Using Lemma (1) in the Appendix, the correlation of the Tukey-hh RF is given by:

$$\rho_{T^*_{h_l,h_r}}(\boldsymbol{d}) = \frac{2\pi(1-h_l)^2(1-h_r)^2\mathbb{E}(T^*_{h_l,h_r}(\boldsymbol{s})T^*_{h_l,h_r}(\boldsymbol{s}+\boldsymbol{d})) - (h_r-h_l)^2}{m(h_l,h_r)}, \tag{16}$$

where a closed form expression of $\mathbb{E}(T^*_{h_l,h_r}(\boldsymbol{s})T^*_{h_l,h_r}(\boldsymbol{s}+\boldsymbol{d}))$ can be found in the Appendix and where $m(h_l,h_r) = \pi(1-h_l)^2(1-h_r)^2((1-2h_l)^{-3/2} + (1-2h_r)^{-3/2}) - (h_r-h_l)^2$. One implication of the correlation function given in Equation (16) is that if the underlying Gaussian RF is weakly stationary then the Tukey-hh RF is also weakly stationary. In addition the RF $T^*_{h_l,h_r}$ is mean-square continuous if the underlying Gaussian RF is mean square continuous since it can be shown that $\mathbb{E}(T^*_{h_l,h_r}(\boldsymbol{s})T^*_{h_l,h_r}(\boldsymbol{s}+\boldsymbol{d})) = \frac{1}{2}[(1-2h_r)^{3/2} + (1-2h_l)^{3/2}]$ and this implies $\rho_{T^*_{h_l,h_r}}(\boldsymbol{0}) = 1$. In addition it can be shown that $T^*_{h_l,h_r}$ inherits the mean square differentiability of underlying Gaussian RF.

Note that a version of the Tukey-hh RFs in Equation (13) that is not mean-square continuous can be obtained by introducing a nugget effect. In our construction a nugget effect can be easily introduced by choosing a discontinuous correlation function of the underlying Gaussian RF that is by replacing $\rho(\boldsymbol{d})$ with $\rho^*(\boldsymbol{d}) = (1-\tau^2)\rho(\boldsymbol{d}) + \tau^2 I(\boldsymbol{d} = \boldsymbol{0})$ where $0 \leq \tau^2 < 1$ represents the underlying nugget effect.

By studying the correlation function of $T^*_{h_l,h_r}$ in (16) some other interesting properties can be shown. For instance a symmetry property with respect to the parameters h_l, h_r exists; that is, $\rho_{T^*_{h_l,h_r}}(\boldsymbol{d}) = \rho_{T^*_{h_r,h_l}}(\boldsymbol{d})$. In addition $\rho_{T^*_{h_l,h_r}}(\boldsymbol{d}) \leq \rho(\boldsymbol{d})$ and $\rho_{T^*_{h_l,h_r}}(\boldsymbol{d}) = 0$ if $\rho(\boldsymbol{d}) = 0$ that is $\rho_{T^*_{h_l,h_r}}(\boldsymbol{d})$ is compactly supported if $\rho(\boldsymbol{d})$ is compactly supported.

To illustrate some examples let us consider a flexible isotropic correlation model for the underlying Gaussian RF that is a reparametrized version of the generalized Wendland correlation function (Gneiting, 2002; Bevilacqua et al., 2019) as proposed in Bevilacqua et al. (2022b). It is defined for $\nu \geq 0$ as:

$$\mathcal{GW}_{\nu,\delta,L(\nu,\delta,\alpha)}(\boldsymbol{d}) = \begin{cases} K\,(U(\boldsymbol{d}))^{\nu+\delta}\,{}_2F_1\left(\frac{\delta}{2}, \frac{\delta+1}{2}; \nu+\delta+1; U(\boldsymbol{d})\right) & 0 \leq \|\boldsymbol{d}\| \leq L(\nu,\delta,\alpha) \\ 0 & \text{otherwise,} \end{cases} \tag{17}$$

where $U(\boldsymbol{d}) := 1 - \left(\frac{\|\boldsymbol{d}\|}{L(\nu,\delta,\alpha)}\right)^2$, $L(\nu,\delta,\alpha) := \alpha\,(\Gamma(\delta+2\nu+1)/\Gamma(\delta))^{\frac{1}{1+2\nu}}$, $K := (2^{-\delta-1}\Gamma^{-1}(2\nu)\Gamma(\nu)\Gamma(2\nu+\delta+1))/\Gamma(\nu+\delta+1)$ and ${}_2F_1\,(a,b,c,x)$ is the Gaussian hypergeometric function (Gradshteyn and Ryzhik, 2007). Here $\alpha > 0$, $\delta \geq (d+1)/2 + \nu$ guarantee the positive definiteness of the model in $\mathbb{R}^d$. Bevilacqua et al. (2022b) show that $\mathcal{GW}_{\nu,\delta,L(\nu,\delta,\alpha)}(\boldsymbol{d}) \to \mathcal{M}_{\nu+0.5,\alpha}(\boldsymbol{d})$ as $\delta \to \infty$ where

$$\mathcal{M}_{\nu,\alpha}(\boldsymbol{d}) = \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{\|\boldsymbol{d}\|}{\alpha}\right)^{\nu}\mathcal{K}_{\nu}\left(\frac{\|\boldsymbol{d}\|}{\alpha}\right) \qquad r \geq 0, \tag{18}$$

is the celebrated Matérn correlation model (Stein, 1999). Thus $\mathcal{GW}_{\nu,\delta,L(\nu,\delta,\alpha)}$ is a flexible correlation model that can switch from compactly supported to globally supported correlation functions. Let us consider the special case $\nu = 0$ that is:

$$\mathcal{GW}_{0,\delta,L(0,\delta,\alpha)}(\boldsymbol{d}) := \begin{cases} \left(1 - \frac{\|\boldsymbol{d}\|}{\delta\alpha}\right)^{\delta}, & 0 \leq \|\boldsymbol{d}\| < \delta\alpha \\ 0 & \text{otherwise.} \end{cases} \tag{19}$$

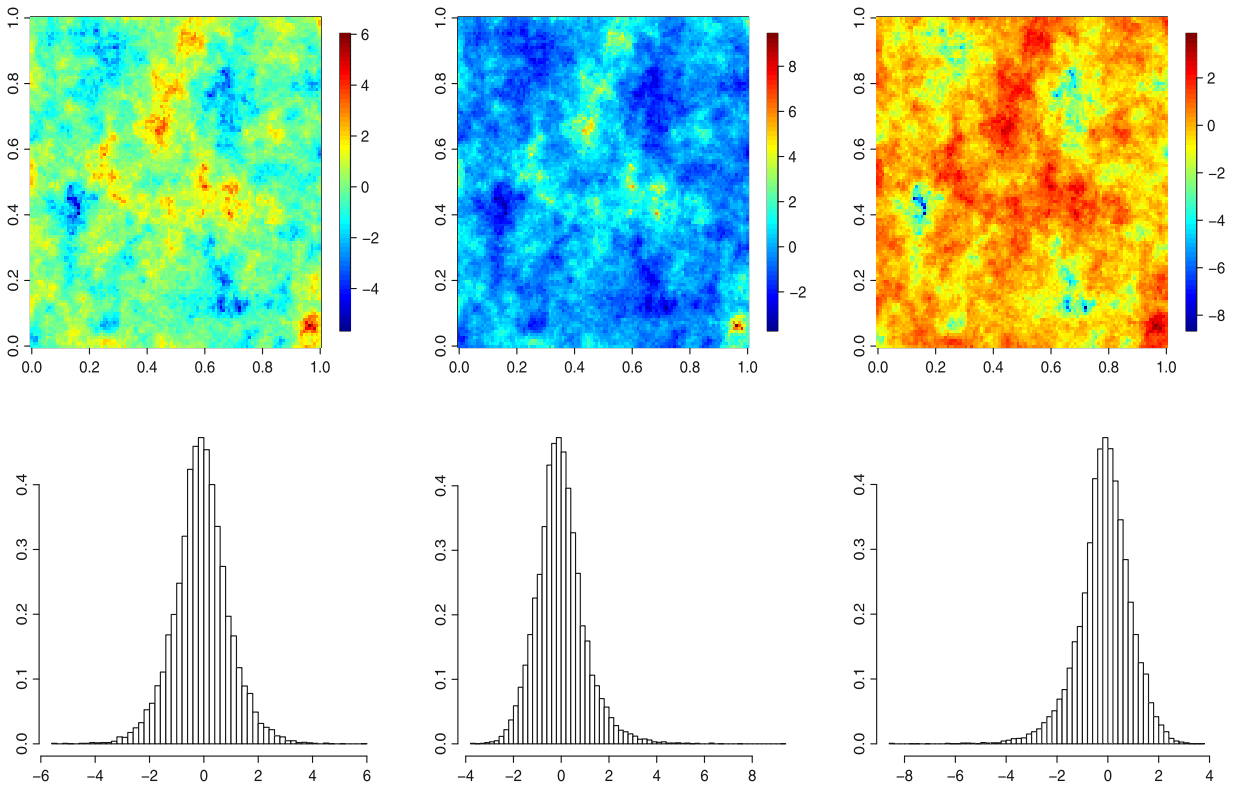

Fig. 2. First column: three realizations of a Tukey-hh RF $T^*_{h_l,h_r}$ with (from left to right) $h_l = h_r = 0.15$ and $h_l = 0.25$ $h_r = 0.05$ and $h_l = 0.05$, $h_r = 0.25$ respectively. Second column: associated histograms.

In the first row of Fig. 2 we depict, from the left to the rights, three realizations on a fine grid of a unit square of a Tukey-hh RF with underlying correlation function $\mathcal{GW}_{0,3.5,L(0,3.5,0.1)}(\boldsymbol{d})$ setting $h_l = h_r = 0.15$ and $h_l = 0.25$, $h_r = 0.05$ and $h_l = 0.05$, $h_r = 0.25$ respectively. The second row depicts the associated histograms showing the flexibility of the Tukey-hh distribution.

We now study the multivariate pdf associated with the Tukey-hh RF. Since the transformation involved in (12) is monotone increasing, the multivariate distribution can be readily obtained. The following theorem gives an explicit closed-form expression for the pdf of the random vector $\boldsymbol{T}^*_{h_l,h_r} = (T^*_{h_l,h_r}(\boldsymbol{s}_1), \ldots, T^*_{h_l,h_r}(\boldsymbol{s}_n))^T$.

Theorem 1. *Let $\boldsymbol{T}^*_{h_l,h_r}$, $h_l, h_r \in [0, 1/2)$ be the $n-$dimensional random vector associated to the Tukey-hh RF with underlying correlation $\rho(\boldsymbol{d})$. Then:*

$$f_{\boldsymbol{T}^*_{h_l,h_r}}(\boldsymbol{t}) = \prod_{i=1}^{n} \frac{\tau^{-1}_{\ell_i}(t_i)}{t_i(1+W(\ell_i t_i^2))} \phi_n(\tau^{-1}_{\boldsymbol{\ell}}(\boldsymbol{t}), \boldsymbol{0}, R_n), \tag{20}$$

where $\tau^{-1}_{\boldsymbol{\ell}}(\boldsymbol{t}) = D(\mathrm{sgn}\{\boldsymbol{t}\}) g_{\boldsymbol{\ell}}(\boldsymbol{t})$ with $g_{\boldsymbol{\ell}}(\boldsymbol{t}) = \left(\sqrt{\frac{W(\ell_1 t_1^2)}{\ell_1}}, \ldots, \sqrt{\frac{W(\ell_n t_n^2)}{\ell_n}} \right)^T$, where $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_n)^T$ with $\ell_k = h_l$ if $t_k < 0$ and $\ell_k = h_r$ if $t_k \geq 0$ for $k = 1, \ldots, n$, $D(\mathrm{sgn}\{\boldsymbol{t}\})$ is the diagonal matrix with elements equal to -1 or 1 depending on the sign of the values of $\boldsymbol{t}$ and $R_n = [\rho(\boldsymbol{s}_i - \boldsymbol{s}_j)]^n_{i,j=1}$.

From Theorem (1) the bivariate pdf can be easily obtained as:

$$f_{\boldsymbol{T}^*_{h_l,h_r;ij}}(t_i, t_j) = \frac{\tau^{-1}_{\ell_i}(t_i)\tau^{-1}_{\ell_j}(t_j)}{t_i t_j (1+W(\ell_i t_i^2))(1+W(\ell_j t_j^2))} \phi_2(\tau^{-1}_{\ell_i,\ell_j}(t_i, t_j), \boldsymbol{0}, R_2), \tag{21}$$

where $\ell_k = h_l$ if $t_k < 0$ and $\ell_k = h_r$ if $t_k \geq 0$ for $k = 1, 2$.

The bivariate pdf in (21) can be easily evaluated since efficient numerical computation of the Lambert-W function can be found in different libraries such as the GNU scientific library (Gough, 2009) and the most important statistical software including R, MATLAB and Python.

Fig. 3 depicts the contour plots of (21) when $\rho(\boldsymbol{d}) = 0.2, 0.5, 0.9$ and for three combinations of the (h_l, h_r) that is $(h_l = 0.2, h_r = 0.2)$ and $(h_l = 0.2, h_r = 0.4)$ and $(h_l = 0.4, h_r = 0.2)$. It turns out that the bivariate Tukey-hh contour lines are not elliptical and when h_l and h_r approach zero the contour plots tend towards an elliptical form, as expected.

Finally, a more flexible model than $T^*_{h_l,h_r}$ can be obtained by defining an RF $T_{h_l,h_r} = \{T_{h_l,h_r}, \boldsymbol{s} \in A\}$ through a location and scale transformation:

$$T_{h_l,h_r}(\boldsymbol{s}) =: \mu(\boldsymbol{s}) + \sigma T^*_{h_l,h_r}(\boldsymbol{s}), \tag{22}$$

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. In addition, non-stationarity can be added into the formulation of (22) by allowing the scale parameter σ to depend on the location $\boldsymbol{s}$. All the properties studied in this section can be easily extended from $T^*_{h_l,h_r}$ to T_{h_l,h_r} including the bivariate density which is given by:

$$f_{T_{h_l,h_r;ij}}(t_i, t_j) = \frac{1}{\sigma^2} f_{T^*_{h_l,h_r;ij}}\left(\frac{t_i - \mu(\boldsymbol{s}_i)}{\sigma}, \frac{t_j - \mu(\boldsymbol{s}_j)}{\sigma}\right). \tag{23}$$

4. Simulation study

In this section, we consider three simulation studies. The first study (Subsection 4.1) compares the proposed NNWCLP method versus the DDWCLP method in terms of statistical efficiency when estimating the parameters of the Tukey-hh RF. In the comparison, we consider both the wpl_a, $a = M, C$ functions. In addition we consider a small simulation study assessing the coverage probabilities of the confidence intervals based on the NNWCLP method.

The second simulation study (Subsection 4.2) focuses on the comparison of the proposed NNWCLP method with the classical maximum likelihood method.

The third simulation study (Subsection 4.3) focuses on Gaussian RF estimation and compares the proposed NNWCLP method with the Vecchia method from statistical and computational viewpoints.

4.1. NNWCLP vs DDWCLP methods when estimating Tukey-hh random fields

As simulation setting we consider a set of n spatial points uniformly distributed on the unit square, $\boldsymbol{s}_i \in A = [0,1]^2$, $i = 1, \ldots, n$. We simulate, using Cholesky decomposition for the underlying Gaussian RF, 1000 realizations of the Tukey-hh RF observed at $n = 500$ spatial location sites. Specifically, we consider the RF in Equation (22) where the mean $\mu(\boldsymbol{s})$ is considered spatially varying through a regression linear model that is:

$$T_{h_l,h_r}(\boldsymbol{s}) = \beta_0 + \beta_1 u(\boldsymbol{s}) + \sigma T^*_{h_l,h_r}(\boldsymbol{s}), \tag{24}$$

where $u(\boldsymbol{s})$ is a standard uniform random variable assumed to be covariate. We set $\beta_0 = 0.5$, $\beta_1 = -0.25$ and $\sigma^2 = 1$, and different combination of the asymmetry/heavy tail parameters are considered that is $h_l = 0.1, 0.2, 0.3$ and $h_r = 0.1, 0.3$. As underlying isotropic parametric correlation model, we consider the model in Equation (19) with $\alpha = 0.06$ and $\delta = 3.5$ where δ is assumed known and fixed. The vector parameters to be estimated are $\boldsymbol{\theta} = (\beta_0, \beta_1, \sigma^2, h_l, h_r, \alpha)^T$.

For wpl_a, $a = M, C$ functions we consider the NNWCLP method by considering the nearest neighbors weight function in Equation (5) with $m = 2, 4, 8, 16$ and the DDWCLP method by considering the weight function based on distances in Equation (4) with $k = 0.03584, 0.05118, 0.07339, 0.10514$, respectively. The values of m and k have been chosen such that the number of pairs involved in the wpl_a, $a = M, C$ estimation is approximatively the same for both weight functions so the comparison between NNWCLP and DDWCLP should be performed for the cases $m = 2$, $k = 0.03584$ then $m = 4$, $k = 0.05118$ and so on.

Table 1 show the bias and root mean squared error (RMSE) of the wpl_M, estimates under the different scenarios when estimating $\boldsymbol{\theta}$ (we highlight the best RMSE in bold for each scenario). It can be appreciated that wpl_M using NNWCLP outperforms DDWCLP in terms of RMSE, for each scenario. In particular, depending on the parameter, it can be appreciated that the best setting is for small values of m ($m = 2$ or $m = 4$ depending on the parameter). Taking into account that the number of pairs involved in the maximization is approximatively the same for both weight functions, this implies that the nearest neighbors weight function select more informative pairs, as suggested in the example in Section 2.

We replicate the same experiment using the wpl_C function. Table 2 show bias and RMSE under the different scenarios when estimating $\boldsymbol{\theta}$. Even in this case, it can be appreciated that maximization of the wpl_C function using NNWCLP overall outperforms wpl_C using DDWCLP in terms of RMSE, for each scenario. In addition, as in the wpl_M case, the distribution of the wpl_C estimates is approximatively symmetric, showing very few outliers for all scenarios. As an example, Fig. 4 show the centered boxplots of the wpl_C estimates using both weight functions, for each parameter, when $h_l = 0.2$ and $h_r = 0.1$.

Comparing Table 1 with Table 2 it can be appreciated that the wpl_C method shows an overall general better performance with respect to the wpl_M method for each parameter. To globally compare wpl_C and wpl_M, we have considered, for each scenario, a measure of global relative efficiency (Davison, 2003), that is:

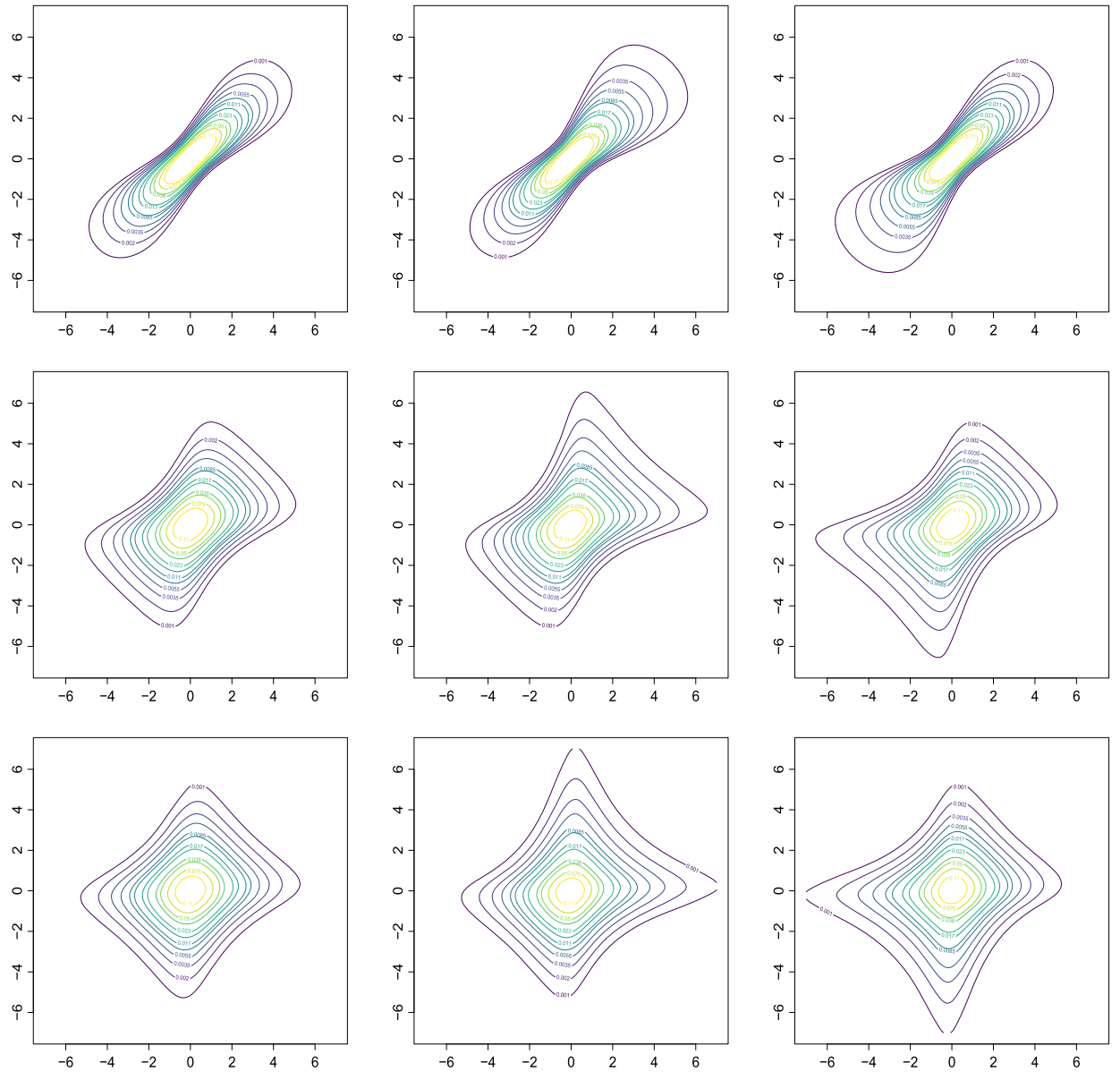

Fig. 3. Contour plots of the bivariate Tukey-*hh* distribution (21) when $(h_l = 0.2, h_r = 0.2)$, $(h_l = 0.2, h_r = 0.4)$ and $(h_l = 0.4, h_r = 0.2)$ (from left to right) and the underlying correlation is $\rho(\boldsymbol{d}) = 0.9$ (first row), $\rho(\boldsymbol{d}) = 0.5$ (second row) and $\rho(\boldsymbol{d}) = 0.2$ (third row), respectively.

$$\text{GRE} = \left(\frac{\det[F^{wpl_C}]^{\frac{1}{2}}}{\det[F^{wpl_M}]^{\frac{1}{2}}} \right)^{1/p}, \tag{25}$$

where $p = 6$ is the number of unknown parameters in $\boldsymbol{\theta}$ and the matrix F^{wpl_a} is the sample mean squared error matrix $F^{wpl_a} = 1000^{-1} \sum_{k=1}^{1000} \left(\hat{\boldsymbol{\theta}}_k^a - \bar{\boldsymbol{\theta}}\right)\left(\hat{\boldsymbol{\theta}}_k^a - \bar{\boldsymbol{\theta}}\right)'$ with $\bar{\boldsymbol{\theta}} = 1000^{-1} \sum_{k=1}^{1000} \hat{\boldsymbol{\theta}}_k^a$, $a = M, C$. Table 3 depicts the GRE results for each scenario and using both weight functions. Since the value of the GRE are overall lower than one, wpl_C outperforms wpl_M using both weight functions and for all the scenarios with a relative efficiency gain between approximatively 10% and 15%.

Summarizing our numerical experiments show that the wpl_a, $a = M, C$ estimators using NNWCLP generally outperform DDWCLP in terms of statistical efficiency and the best efficiencies are achieved with small values of m ($m = 2$ or $m = 4$). In addition the wpl_C estimator shows a better statistical efficiency with respect to the wpl_M estimator irrespective of the weight function.

Another important point related to the NNWCLP method is standard error estimation and the construction of confidence intervals. We perform a small simulation study to assess the coverage probabilities of the confidence intervals based on the NNWCLP method where, as mentioned in section 2, we use parametric bootstrap to estimate $G_a^{-1}(\boldsymbol{\theta})$. In particular, since $G_a^{-1}(\boldsymbol{\theta})$ is the asymptotic variance-covariance matrix of the NNWCLP estimator, the parametric bootstrap estimation of $G_a^{-1}(\boldsymbol{\theta})$ is obtained by simulating K

Table 1
Bias and RMSE when estimating with wpl_M the model in equation (24) with $\beta_0 = 0.5$, $\beta_1 = -0.25$, $\sigma^2 = 1$, for different values of h_l and h_r, with the weight function (5) (NNWCLP) with $m = 2, 4, 8, 16$ and with the weight function (4) (DDWCLP) with $k = 0.03584, 0.05118, 0.07339, 0.10514$. The underlying correlation function is $\mathcal{GW}_{0.3.5,L(0,3.5,\alpha)}(\boldsymbol{d}) = (1 - ||\boldsymbol{d}||/(3.5\alpha))_+^{3.5}$ with $\alpha = 0.06$.

		$h_l = 0.1$				$h_l = 0.2$				$h_l = 0.3$			
		$h_r = 0.1$		$h_r = 0.3$		$h_r = 0.1$		$h_r = 0.3$		$h_r = 0.1$		$h_r = 0.3$	
		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
β_0	$m = 2$	-0.00106	0.12353	-0.00173	**0.12365**	-0.00131	**0.12394**	-0.00192	**0.12317**	-0.00134	**0.12349**	-0.00191	**0.12202**
	$m = 4$	-0.00121	**0.12349**	-0.00163	0.12398	-0.00143	0.12406	-0.00179	0.12365	-0.00143	0.12381	-0.00176	0.12272
	$m = 8$	-0.00086	0.12418	-0.00117	0.12502	-0.00107	0.12494	-0.00134	0.12490	-0.00108	0.12490	-0.00132	0.12418
	$m = 16$	-0.00062	0.12458	-0.00095	0.12578	-0.00092	0.12550	-0.00119	0.12586	-0.00102	0.12570	-0.00126	0.12538
	$k = 0.03584$	-0.00018	0.14471	-0.00121	0.14484	-0.00050	0.14512	-0.00145	0.14426	-0.00049	0.14464	-0.00140	0.14293
	$k = 0.05118$	-0.00044	0.13925	-0.00089	0.13986	-0.00065	0.13986	-0.00105	0.13946	-0.00057	0.13953	-0.00095	0.13835
	$k = 0.07339$	0.00005	0.13587	-0.00035	0.13671	-0.00021	0.13664	-0.00057	0.13657	-0.00020	0.13653	-0.00055	0.13572
	$k = 0.10514$	0.00050	0.13379	0.00010	0.13498	0.00016	0.13468	-0.00020	0.13506	0.00006	0.13487	-0.00028	0.13450
β_1	$m = 2$	0.00086	**0.05206**	0.00083	**0.05477**	0.00082	**0.05367**	0.00078	**0.05657**	0.00075	**0.05468**	0.00070	**0.05779**
	$m = 4$	0.00016	0.05532	0.00005	0.05831	0.00007	0.05701	-0.00008	0.06025	-0.00006	0.05814	-0.00023	0.06156
	$m = 8$	-0.00007	0.06173	-0.00013	0.06473	-0.00020	0.06348	-0.00028	0.06671	-0.00036	0.06465	-0.00046	0.06812
	$m = 16$	0.00015	0.06834	0.00014	0.07120	0.00003	0.07007	-0.00001	0.07321	-0.00016	0.07113	-0.00020	0.07450
	$k = 0.03584$	0.00162	0.05788	0.00145	0.06124	0.00154	0.05992	0.00135	0.06364	0.00145	0.06140	0.00122	0.06535
	$k = 0.05118$	-0.00038	0.05831	-0.00052	0.06181	-0.00041	0.06008	-0.00059	0.06380	-0.00047	0.06132	-0.00067	0.06527
	$k = 0.07339$	-0.00017	0.06403	-0.00028	0.06745	-0.00027	0.06573	-0.00041	0.06950	-0.00042	0.06693	-0.00058	0.07099
	$k = 0.10514$	0.00005	0.07071	0.00001	0.07376	-0.00018	0.07246	-0.00024	0.07589	-0.00043	0.07355	-0.00053	0.07720
α	$m = 2$	-0.00098	0.00775	-0.00098	0.00775	-0.00099	0.00775	-0.00099	0.00775	-0.00100	0.00775	-0.00100	0.00775
	$m = 4$	-0.00118	**0.00707**	-0.00118	**0.00707**	-0.00119	**0.00707**	-0.00118	**0.00707**	-0.00120	**0.00707**	-0.00119	**0.00707**
	$m = 8$	-0.00145	**0.00707**	-0.00144	**0.00707**	-0.00145	**0.00707**	-0.00144	**0.00707**	-0.00146	**0.00707**	-0.00145	**0.00707**
	$m = 16$	-0.00175	0.00775	-0.00174	0.00775	-0.00175	0.00775	-0.00174	0.00775	-0.00176	**0.00707**	-0.00174	**0.00707**
	$k = 0.03584$	-0.00137	0.00949	-0.00140	0.00949	-0.00140	0.00949	-0.00141	0.00949	-0.00143	0.00949	-0.00144	0.00949
	$k = 0.05118$	-0.00157	0.00837	-0.00158	0.00837	-0.00158	0.00837	-0.00159	0.00837	-0.00160	0.00837	-0.00160	0.00837
	$k = 0.07339$	-0.00168	0.00775	-0.00167	0.00775	-0.00168	0.00775	-0.00168	0.00775	-0.00170	0.00775	-0.00169	0.00775
	$k = 0.10514$	-0.00195	0.00775	-0.00195	0.00775	-0.00196	0.00775	-0.00195	0.00775	-0.00197	0.00775	-0.00195	0.00775
σ^2	$m = 2$	0.00198	**0.13259**	0.00250	**0.13943**	0.00347	**0.13649**	0.004002	**0.14307**	0.00477	**0.13982**	0.00533	**0.14612**
	$m = 4$	0.00272	0.13428	0.00368	0.14181	0.00432	0.13860	0.00538	0.14605	0.00566	0.14216	0.00680	0.14933
	$m = 8$	0.00337	0.13557	0.00460	0.14314	0.00497	0.14029	0.00632	0.14772	0.00638	0.14412	0.00781	0.15133
	$m = 16$	0.00405	0.13795	0.00521	0.14564	0.00590	0.14325	0.00719	0.15090	0.00748	0.14744	0.00890	0.15495
	$k = 0.03584$	-0.00153	0.16897	0.00023	0.17816	0.00081	0.17401	0.00298	0.18330	0.00266	0.17819	0.00507	0.18735
	$k = 0.05118$	0.00118	0.15560	0.00326	0.16429	0.00320	0.16016	0.00560	0.16882	0.00484	0.16395	0.00745	0.17242
	$k = 0.07339$	0.00286	0.14896	0.00489	0.15748	0.00476	0.15372	0.00703	0.16208	0.00639	0.15761	0.00888	0.16577
	$k = 0.10514$	0.00228	0.14728	0.00406	0.15576	0.00446	0.15291	0.00650	0.16140	0.00623	0.15719	0.00848	0.16556
h_l	$m = 2$	-0.00788	**0.04313**	-0.00772	**0.04405**	-0.00926	**0.06099**	-0.00902	**0.06221**	-0.01109	**0.07727**	-0.01079	**0.07849**
	$m = 4$	-0.00821	0.04416	-0.00810	0.04506	-0.00964	0.06245	-0.00948	0.06372	-0.01152	0.07893	-0.01133	0.08019
	$m = 8$	-0.00833	0.04593	-0.00827	0.04669	-0.00973	0.06434	-0.00964	0.06550	-0.01173	0.08044	-0.01162	0.08173
	$m = 16$	-0.00873	0.04754	-0.00867	0.04827	-0.01035	0.06626	-0.01025	0.06731	-0.01252	0.08240	-0.01240	0.08185
	$k = 0.03584$	-0.00975	0.05020	-0.00965	0.05128	-0.01214	0.07176	-0.01202	0.07314	-0.01532	0.09017	-0.01517	0.09160
	$k = 0.05118$	-0.00932	0.04848	-0.00930	0.04950	-0.01127	0.06899	-0.01126	0.07043	-0.01393	0.08678	-0.01389	0.08832
	$k = 0.07339$	-0.00917	0.04827	-0.00919	0.04919	-0.01093	0.06841	-0.01093	0.06971	-0.01350	0.08556	-0.01348	0.08689
	$k = 0.10514$	-0.00905	0.04960	-0.00904	0.05040	-0.01103	0.06964	-0.01101	0.07078	-0.01369	0.08660	-0.01366	0.08792
h_r	$m = 2$	-0.00755	**0.04359**	-0.00930	**0.07931**	-0.00768	**0.04427**	-0.00944	**0.08019**	-0.00779	**0.04472**	-0.00955	**0.08081**
	$m = 4$	-0.00781	0.04450	-0.00982	0.08000	-0.00794	0.04517	-0.00997	0.08087	-0.00804	0.04561	-0.01009	0.08149
	$m = 8$	-0.00824	0.04572	-0.01047	0.08130	-0.00836	0.04626	-0.01062	0.08210	-0.00847	0.04680	-0.01075	0.08276
	$m = 16$	-0.00834	0.04754	-0.01043	0.08331	-0.00850	0.04806	-0.01064	0.08408	-0.00862	0.04848	-0.01082	0.08473
	$k = 0.03584$	-0.00894	0.05099	-0.01251	0.09154	-0.00914	0.05167	-0.01284	0.09236	-0.00929	0.05215	-0.01306	0.09295
	$k = 0.05118$	-0.00951	0.04848	-0.01303	0.08741	-0.00968	0.04909	-0.01328	0.08826	-0.00981	0.04960	-0.01345	0.08894
	$k = 0.07339$	-0.00917	0.04868	-0.01216	0.08666	-0.00932	0.04919	-0.01242	0.08741	-0.00945	0.04970	-0.01260	0.08803
	$k = 0.10514$	-0.00881	0.04960	-0.01149	0.08695	-0.00900	0.05010	-0.01180	0.08764	-0.00915	0.05060	-0.01202	0.08826

times a Tukey-hh RF (using the estimated parameters) and then taking the variance covariance matrix of the K estimates. Then, a $100(1-\alpha)\%$ confidence interval is constructed as:

$$(\widehat{\boldsymbol{\theta}}_{a,i} - z_{1-\alpha/2} se(\widehat{\boldsymbol{\theta}}_{a,i}), \widehat{\boldsymbol{\theta}}_{a,i} + z_{1-\alpha/2} se(\widehat{\boldsymbol{\theta}}_{a,i})),$$

where $z_{1-\alpha/2}$ is the quantile of the standard Gaussian distribution, $\widehat{\boldsymbol{\theta}}_{a,i}$ is the $i-$th element of $\widehat{\boldsymbol{\theta}}_a$ and the standard error $se(\widehat{\boldsymbol{\theta}}_{a,i})$ is the square root of the $i-$th diagonal element of the variance covariance matrix bootstrap estimate of $G_a^{-1}(\boldsymbol{\theta})$.

We simulate 100 Tukey-hh RFs using the previous simulation setting considering only the case $h_l = 0.1$ and $h_r = 0.3$ and we estimate with the NNWCLP method using the wpl_C function with $m = 2$. Parametric bootstrap estimation of $G^{-1}(\boldsymbol{\theta})$ has been computed using the `GeoVarestbootstrap` function of the `GeoModels` package setting $K = 200$. The confidence interval coverage probabil-

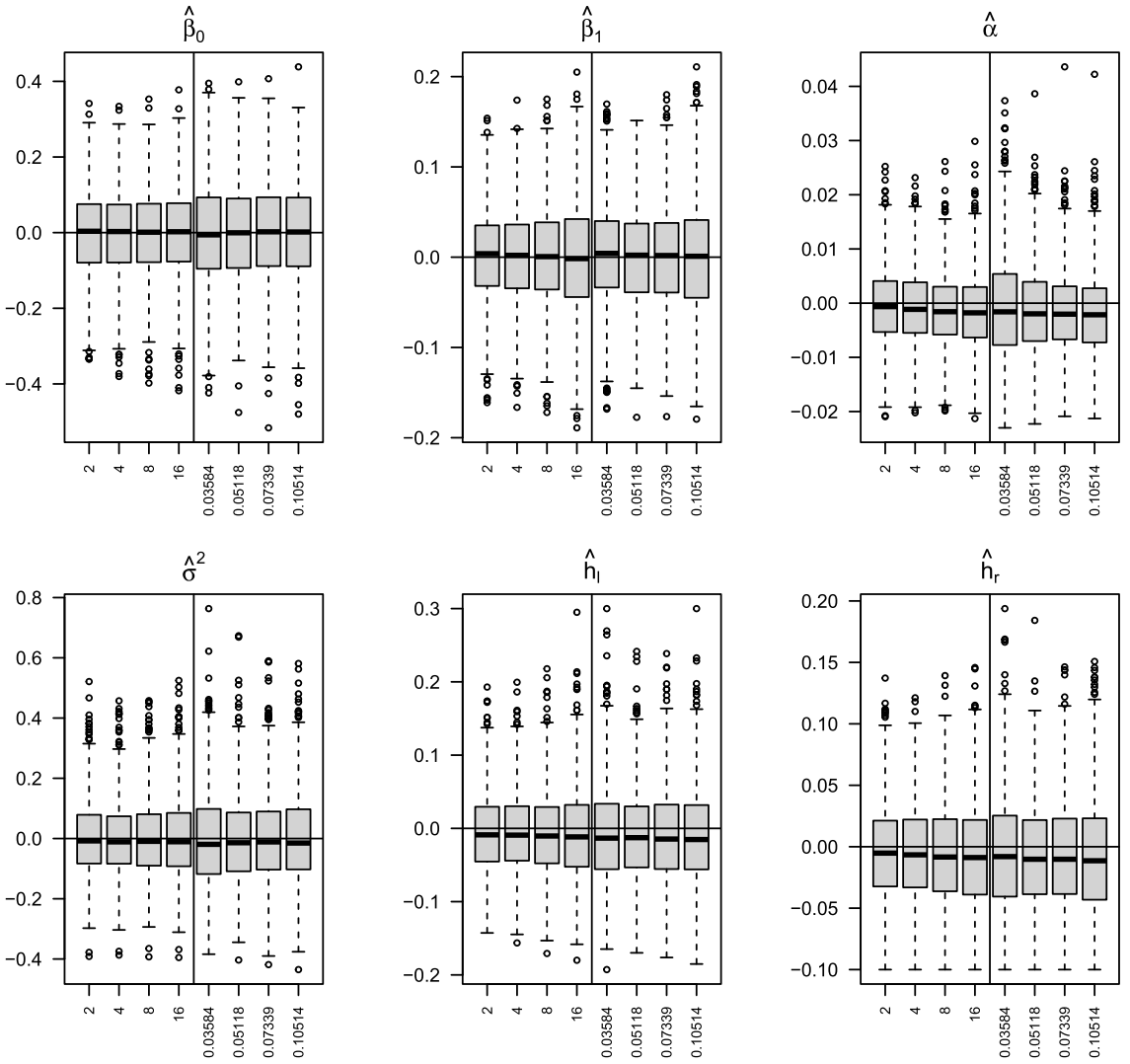

Fig. 4. Centered boxplots of the estimated parameters $\beta_0 = 0.5$, $\beta_1 = -0.25$, $\alpha = 0.05$, $\sigma^2 = 1$, $h_l = 0.2$ and $h_r = 0.1$ (from left to right) when estimating the model (24) with wpl_C using the weight function (5) (NNWCLP) with $m = 2, 4, 8, 16$ and using the weight function (4) (DDWCLP) with $k = 0.03584, 0.05118, 0.07339, 0.10514$ respectively.

ities of $(\beta_0, \beta_1, \sigma^2, h_l, h_r, \alpha)^T$ are reported in Table 4 where CP90 and CP95 represent the coverage probability of the nominal 90% and 95% confidence intervals, respectively. It can be appreciated that the confidence interval based on the parametric bootstrap estimation achieves coverage probability close to the nominal ones.

4.2. Comparing NNWCLP with maximum likelihood method when estimating Tukey-*hh* random fields

We perform a small simulation study comparing NNWCLP method, in particular using the wpl_C function, with standard maximum likelihood using the multivariate density in (20). As simulation setting we consider the same spatial locations of the previous Section, that is we simulate 1000 realizations of the Tukey-*hh* RF observed at $n = 500$ spatial location sites.

We consider the RF in Equation (22) where the mean is considered fixed and equal to zero, $\sigma^2 = 1$ and the asymmetry/heavy tail parameters are fixed as $h_l = 0.2$ and $h_r = 0.1$. Finally, as underlying isotropic parametric correlation model, we consider the same model of the previous Section. For the NNWCLP method we consider the nearest neighbors weight function in Equation (5) with $m = 3$. Finally, the vector parameters to be estimated are $\boldsymbol{\theta} = (\mu, \sigma^2, h_l, h_r, \alpha)^T$. Table 5 depicts RMSE for each parameter, RE and GRE associated to the comparison between NNWCLP and maximum likelihood (ML) methods. Overall, it can be appreciated that NNWCLP performs very well when compared to maximum likelihood with RE and a GRE between 0.9 and 0.1 approximatively.

Table 2
Bias and RMSE when estimating with wpl_C the model in equation (24) with $\beta_0 = 0.5$, $\beta_1 = -0.25$, $\sigma^2 = 1$, for different values of h_l and h_r, with the weight function (5) (NNWCLP) with $m = 2, 4, 8, 16$ and with the weight function (4) (DDWCLP) with $k = 0.03584, 0.05118, 0.07339, 0.10514$, respectively. The underlying correlation function is $\mathcal{GW}_{0.3.5,L(0.3.5,\alpha)}(\boldsymbol{d}) = (1 - ||\boldsymbol{d}||/(3.5\alpha))_+^{3.5}$ with $\alpha = 0.06$.

		$h_l = 0.1$				$h_l = 0.2$				$h_l = 0.3$			
		$h_r = 0.1$		$h_r = 0.3$		$h_r = 0.1$		$h_r = 0.3$		$h_r = 0.1$		$h_r = 0.3$	
		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
β_0	$m = 2$	-0.00171	0.11489	-0.00324	**0.11189**	-0.00174	**0.11345**	-0.00317	**0.10964**	-0.00157	**0.11127**	-0.00289	**0.10696**
	$m = 4$	-0.00090	**0.11485**	-0.00187	0.11269	-0.00086	0.11397	-0.00177	0.11091	-0.00069	0.11229	-0.00155	0.10872
	$m = 8$	-0.00081	0.11580	-0.00156	0.11498	-0.00082	0.11567	-0.00153	0.11397	-0.00068	0.11472	-0.00136	0.11238
	$m = 16$	-0.00098	0.11841	-0.00151	0.11870	-0.00112	0.11887	-0.00159	0.11832	-0.00105	0.11853	-0.00150	0.11730
	$k = 0.03584$	-0.00155	0.14170	-0.00479	0.13740	-0.00135	0.13971	-0.00432	0.13450	-0.00050	0.13624	-0.00329	0.13058
	$k = 0.05118$	-0.00095	0.13635	-0.00238	0.13330	-0.00066	0.13509	-0.00203	0.13088	-0.00007	0.13274	-0.00143	0.12783
	$k = 0.07339$	-0.00012	0.13401	-0.00116	0.13255	-0.00011	0.13364	-0.00111	0.13111	0.00021	0.13221	-0.00078	0.12896
	$k = 0.10514$	0.00052	0.13263	-0.00025	0.13266	0.00032	0.13297	-0.00041	0.13206	0.00038	0.13244	-0.00034	0.13081
β_1	$m = 2$	0.00143	**0.04960**	0.00128	**0.05196**	0.00139	**0.05099**	0.00123	**0.05367**	0.00132	**0.05196**	0.00115	**0.05477**
	$m = 4$	0.00091	0.05040	0.00076	0.05310	0.00081	0.05187	0.00064	0.05486	0.00070	0.05292	0.00051	0.05604
	$m = 8$	0.00044	0.05486	0.00025	0.05771	0.00035	0.05639	0.00014	0.05958	0.00022	0.05753	-0.00001	0.06083
	$m = 16$	0.00005	0.06173	-0.00006	0.06458	-0.00007	0.06340	-0.00020	0.06648	-0.00023	0.06450	-0.00038	0.06775
	$k = 0.03584$	0.00234	0.05486	0.00207	0.05779	0.00228	0.05657	0.00200	0.05992	0.00222	0.05779	0.00192	0.06132
	$k = 0.05118$	0.00018	0.05339	-0.00005	0.05648	0.00015	0.05495	-0.00011	0.05840	0.00011	0.05604	-0.00016	0.05967
	$k = 0.07339$	0.00016	0.05701	-0.00005	0.06025	0.00007	0.05857	-0.00016	0.06221	-0.00004	0.05975	-0.00031	0.06356
	$k = 0.10514$	0.00009	0.06380	-0.00003	0.06693	-0.00012	0.06550	-0.00026	0.06899	-0.00034	0.06663	-0.00051	0.07036
α	$m = 2$	-0.00044	**0.00707**	-0.00037	**0.00707**	-0.00041	**0.00707**	-0.00034	**0.00707**	-0.00038	**0.00707**	-0.00031	**0.00707**
	$m = 4$	-0.00082	**0.00707**	-0.00078	**0.00707**	-0.00080	**0.00707**	-0.00076	**0.00707**	-0.00079	**0.00707**	-0.00074	**0.00707**
	$m = 8$	-0.00120	**0.00707**	-0.00117	**0.00707**	-0.00119	**0.00707**	-0.00116	**0.00707**	-0.00118	**0.00707**	-0.00115	**0.00707**
	$m = 16$	-0.00158	**0.00707**	-0.00156	**0.00707**	-0.00158	**0.00707**	-0.00155	**0.00707**	-0.00157	**0.00707**	-0.00154	**0.00707**
	$k = 0.03584$	-0.00076	0.01000	-0.00065	0.01000	-0.00070	0.01000	-0.00059	0.01000	-0.00068	0.01000	-0.00057	0.01000
	$k = 0.05118$	-0.00124	0.00837	-0.00119	0.00837	-0.00122	0.00837	-0.00116	0.00837	-0.00121	0.00837	-0.00114	0.00837
	$k = 0.07339$	-0.00148	0.00775	-0.00145	0.00775	-0.00147	0.00775	-0.00144	0.00775	-0.00147	0.00775	-0.00143	0.00775
	$k = 0.10514$	-0.00184	0.00775	-0.00183	0.00775	-0.00184	0.00775	-0.00182	0.00775	-0.00184	0.00775	-0.00182	0.00775
σ^2	$m = 2$	-0.00049	0.12198	-0.00009	0.12946	0.00048	0.12490	0.00084	0.13214	0.00142	0.12767	0.00178	0.13472
	$m = 4$	-0.00186	**0.12021**	-0.00152	**0.12716**	-0.00105	**0.12317**	-0.00079	**0.12985**	-0.00023	**0.12594**	-0.00001	**0.13236**
	$m = 8$	-0.00024	0.12462	0.00013	0.13142	0.00075	0.12826	0.00111	0.13491	0.00165	0.13142	0.00201	0.13780
	$m = 16$	0.00138	0.13050	0.00204	0.13780	0.00273	0.13506	0.00344	0.14223	0.00390	0.13871	0.00466	0.14574
	$k = 0.03584$	-0.00502	0.16081	-0.00398	0.16873	-0.00323	0.16453	-0.00194	0.17210	-0.00210	0.16778	-0.00067	0.17510
	$k = 0.05118$	-0.00293	0.14704	-0.00152	0.15463	-0.00155	0.15040	0.00001	0.15758	-0.00047	0.15343	0.00120	0.16028
	$k = 0.07339$	-0.00051	0.14227	0.00088	0.15010	0.00089	0.14598	0.00242	0.15346	0.00204	0.14913	0.00368	0.15636
	$k = 0.10514$	-0.00014	0.14255	0.00121	0.15060	0.00159	0.14731	0.00309	0.15531	0.00299	0.15106	0.00462	0.15884
h_l	$m = 2$	-0.00511	**0.03937**	-0.00468	**0.04062**	-0.00571	**0.05477**	-0.00508	**0.05604**	-0.00653	**0.07007**	-0.00573	**0.07113**
	$m = 4$	-0.00586	0.03950	-0.00556	**0.04062**	-0.00679	0.05523	-0.00634	0.05630	-0.00794	0.07071	-0.00736	0.07176
	$m = 8$	-0.00691	0.04159	-0.00667	0.04243	-0.00806	0.05805	-0.00772	0.05908	-0.00953	0.07362	-0.00908	0.07463
	$m = 16$	-0.00771	0.04472	-0.00756	0.04550	-0.00903	0.06221	-0.00880	0.06332	-0.01083	0.07785	-0.01054	0.07893
	$k = 0.03584$	-0.00670	0.04722	-0.00599	0.04858	-0.00789	0.06768	-0.00686	0.06921	-0.01000	0.08556	-0.00875	0.08660
	$k = 0.05118$	-0.00743	0.04416	-0.00717	0.04528	-0.00885	0.06332	-0.00845	0.06450	-0.01078	0.08136	-0.01019	0.08246
	$k = 0.07339$	-0.00802	0.04472	-0.00785	0.04572	-0.00953	0.06364	-0.00928	0.06481	-0.01163	0.08099	-0.01125	0.08210
	$k = 0.10514$	-0.00845	0.04712	-0.00834	0.04806	-0.01024	0.06648	-0.01006	0.06768	-0.01261	0.08355	-0.01235	0.08479
h_r	$m = 2$	-0.00470	0.04000	-0.00588	0.07225	-0.00469	0.04062	-0.00580	0.07280	-0.00470	0.04111	-0.00573	0.07308
	$m = 4$	-0.00527	**0.03962**	-0.00648	**0.07176**	-0.00526	**0.04025**	-0.00638	**0.07246**	-0.00527	**0.04074**	-0.00630	**0.07287**
	$m = 8$	-0.00657	0.04195	-0.00815	0.07503	-0.00659	0.04254	-0.00811	0.07583	-0.00661	0.04301	-0.00807	0.07635
	$m = 16$	-0.00746	0.04483	-0.00933	0.07925	-0.00754	0.04539	-0.00940	0.08000	-0.00761	0.04583	-0.00945	0.08062
	$k = 0.03584$	-0.00578	0.04817	-0.00857	0.08701	-0.00572	0.04889	-0.00833	0.08769	-0.00564	0.04940	-0.00806	0.08803
	$k = 0.05118$	-0.00754	0.04427	-0.01060	0.08142	-0.00755	0.04494	-0.01051	0.08204	-0.00754	0.04550	-0.01037	0.08240
	$k = 0.07339$	-0.00793	0.04528	-0.01064	0.08191	-0.00800	0.04583	-0.01069	0.08258	-0.00804	0.04637	-0.01068	0.08307
	$k = 0.10514$	-0.00820	0.04722	-0.01067	0.08390	-0.00832	0.04775	-0.01085	0.08462	-0.00841	0.04827	-0.01095	0.08515

4.3. Comparing NNWCLP with the Vecchia method for Gaussian random field estimation

The basic idea of the Vecchia method (Vecchia, 1988; Katzfuss and Guinness, 2021) is to replace the multivariate Gaussian distribution with a product of Gaussian conditional distributions, in which each conditional distribution conditions on only a small subset of previous observations. This kind of approximation depends basically on a specified ordering of the spatial location vector and the size of the conditioning vectors m. Generally, the larger is the m, the more accurate and computationally expensive the approximation is.

Table 3
Global Relative efficiency (GRE) defined in equation (25) between wpl_M and wpl_C methods using the weight function (5) (NNWCLP) with $m = 2, 4, 8, 16$ and using the weight function (4) (DDWCLP) with $k = 0.03584, 0.05118, 0.07339, 0.10514$, respectively, under different scenarios.

	$h_l = 0.1$		$h_l = 0.2$		$h_l = 0.3$	
	$h_r = 0.1$	$h_r = 0.3$	$h_r = 0.1$	$h_r = 0.3$	$h_r = 0.1$	$h_r = 0.3$
$m = 2$	0.93935	0.93595	0.93391	0.93033	0.93288	0.92928
$k = 0.03584$	0.96164	0.95724	0.95639	0.95161	0.95319	0.94782
$m = 4$	0.92106	0.91800	0.91603	0.91263	0.91584	0.91260
$k = 0.05118$	0.94312	0.93977	0.93933	0.93550	0.93916	0.93525
$m = 8$	0.92762	0.92611	0.92487	0.92334	0.92556	0.92411
$k = 0.07339$	0.94712	0.94579	0.94480	0.94340	0.94521	0.94395
$m = 16$	0.94774	0.94846	0.94668	0.94738	0.94667	0.94734
$k = 0.10514$	0.96052	0.96105	0.95968	0.96010	0.95998	0.96047

Table 4
The coverage probability of the nominal 90% and 95% confidence intervals (CP90 and CP95 respectively) associated with the NNWCLP method (wpl_C function with $m = 2$) of $(\beta_0, \beta_1, \sigma^2, h_l, h_r, \alpha)^T$ constructed using parametric bootstrap estimation of the asymptotic variance covariance matrix.

	β_0	β_1	α	σ^2	h_l	h_r
CP90	0.91	0.89	0.90	0.92	0.86	0.89
CP95	0.95	0.94	0.94	0.96	0.92	0.93

Table 5
RMSE for each parameter and relative efficiency (RE) between NNWCLP method (wpl_C) using the weight function (5) with $m = 3$ and maximum likelihood method. The Global Relative efficiency (GRE) is 0.94368.

	μ	σ^2	h_l	h_r	α
NNWCLP	0.10925	0.12044	0.05311	0.03903	0.00630
ML	0.10912	0.12416	0.05229	0.04116	0.00707
RE	1.00122	0.96999	1.01555	0.94829	0.89067

The Vecchia method by construction specifies an approximation of a valid likelihood function that corresponds to a specific data generating process, as opposed to WCLP, and it can be used for simulating and predicting purposes. In general, it can be computed in $O(nm^3)$ time and with $O(nm^2)$ memory burden.

Although the Vecchia method can in principle be used to estimate the proposed Tukey-*hh* RF, we compare the NNWCLP with the Vecchia method when estimating Gaussian RFs.

Specifically, we consider a simulation setting with $n = 5000$ spatial points uniformly distributed on the unit square, $\boldsymbol{s}_i \in A = [0, 1]^2$, $i = 1, \ldots, n$. We simulate, 500 realizations of a zero mean and unit variance Gaussian RF with Matérn correlation model in Equation (18) that is we consider the RF $G(\boldsymbol{s}) = \mu + \sigma G^*(\boldsymbol{s})$ with $\mu = 0$, $\sigma^2 = 1$ and $\rho_{G^*}(\mathbf{h}) = \mathcal{M}_{\nu,\alpha}(\boldsymbol{d})$. The correlation parameters are set $\nu = 0.5, 1.5$ and $\alpha = 0.050, 0.0316$ respectively. This parameter setting guarantees that the practical range of the correlation function is approximatively 0.15. We consider two scenarios: in the first we assume that the smoothness parameter is known and fixed and in the second we assume it unknown.

For each simulated dataset we perform the estimation with the NNWCLP method (using the wpl_C function) and the Vecchia method. To stress the comparison we consider a recent efficient proposal of Guinness (2021) that exploits a Fisher-scoring algorithm to find the maximum of the Vecchia approximation using a specific type of ordering (maximum-minimum ordering, see Guinness (2018) for details). We use the `fit_model` R function of the `GpGp` R package that implements the method of Guinness (2021) and the `GeoFit` R function in the `GeoModels` R package using a specified algorithm of optimization (we use the `nlminb` R function (Gay, 1990)) in our example). The starting values used are the same for both R functions.

Tables 6 and 7 depict the results of the simulation study when estimating with the NNWCLP and Vecchia methods using $m = 2, 4, 8, 16$, assuming the smoothness parameter known and unknown, respectively. The results are reported in terms of RMSE and relative efficiency (RE) for each parameter. In addition, the bottom part of both tables shows the results in terms of global relative efficiency (GRE) as defined in Equation (25). The lowest RMSE values across the different choices of m are reported in bold for each parameter.

As a general pattern, it can be appreciated that the best choice for the Vecchia method is $m = 16$ for each parameter, as expected. For the NNWCLP method the best choice of m generally depends on the type of parameter.

Another general pattern that can be outlined is that the REs of the mean and variance parameter are very close to 1 (including greater than 1 in some cases) indicating that the NNWCLP method performs very well if compared to the Vecchia method, irrespective of the choice of m and the value of ν.

Table 6
RMSE and associated relative efficiency (RE) when estimating the parameters of a Gaussian RF with Matérn correlation and smoothness parameter $\nu = 0.5, 1.5$ (ν is assumed known) using NNWCLP and Vecchia methods with increasing $m = 2, 4, 8, 16$. The bottom line shows information on the global relative efficiency (GRE).

		$\nu = 0.5$				$\nu = 1.5$			
		$m = 2$	$m = 4$	$m = 8$	$m = 16$	$m = 2$	$m = 4$	$m = 8$	$m = 16$
μ	NNWCLP	0.11348	0.11277	0.11226	**0.11138**	0.12983	0.12612	0.12393	**0.12263**
	Vecchia	0.11960	0.11542	0.11475	**0.11149**	0.13298	0.13237	0.12671	**0.12048**
	RE	1.05393	1.02349	1.02218	1.00099	1.02426	1.04956	1.02243	0.98247
α	NNWCLP	**0.00474**	**0.00474**	0.00477	0.00497	0.00187	**0.00180**	0.00181	0.00195
	Vecchia	0.00481	0.00465	0.00465	**0.00445**	0.00141	0.00136	0.00126	**0.00115**
	RE	1.01477	0.98101	0.97484	0.89537	0.75401	0.75556	0.69613	0.58974
σ^2	NNWCLP	0.08669	0.08485	**0.08430**	0.08451	0.12043	0.11390	**0.11233**	0.11274
	Vecchia	0.08713	0.08564	0.08614	**0.08246**	0.10882	0.10983	0.10545	**0.09755**
	RE	1.00508	1.00931	1.02183	0.97574	0.90360	0.96427	0.93875	0.86527
	GRE	1.00747	0.97145	0.95190	0.87194	0.84060	0.77120	0.71119	0.63092

Table 7
RMSE and associated relative efficiency (RE) when estimating the parameters of a Gaussian RF with Matérn correlation and smoothness parameter $\nu = 0.5, 1.5$ (ν is assumed unknown) using NNWCLP and Vecchia method with increasing $m = 2, 4, 8, 16$. The bottom line shows information on global relative efficiency (GRE).

		$\nu = 0.5$				$\nu = 1.5$			
		$m = 2$	$m = 4$	$m = 8$	$m = 16$	$m = 2$	$m = 4$	$m = 8$	$m = 16$
μ	NNWCLP	0.11394	0.11307	0.11236	**0.11176**	0.13767	0.12840	0.12537	**0.12314**
	Vecchia	0.11971	0.11519	0.11479	**0.11151**	0.13271	0.13234	0.12649	**0.12040**
	RE	1.05064	1.01875	1.02163	0.99776	0.96398	1.03069	1.00893	0.97775
α	NNWCLP	0.00668	0.00625	**0.00597**	0.00621	0.00507	0.00414	0.00359	**0.00356**
	Vecchia	0.00645	0.00598	0.00588	**0.00566**	0.00254	0.00220	0.00187	**0.00169**
	RE	0.96557	0.95680	0.98492	0.91143	0.50099	0.53140	0.52089	0.47472
σ^2	NNWCLP	0.08685	0.08543	0.08433	**0.08429**	0.12887	0.11846	0.11454	**0.11338**
	Vecchia	0.08764	0.08676	0.08729	**0.08357**	0.11828	0.11978	0.11278	**0.10303**
	RE	1.00910	1.01557	1.03510	0.99146	0.91782	1.01114	0.98463	0.90871
ν	NNWCLP	0.02180	0.01937	**0.01816**	0.01919	0.33952	0.21636	0.14736	**0.13375**
	Vecchia	0.01720	0.01486	0.01420	**0.01390**	0.04798	0.03285	0.02564	**0.02359**
	RE	0.78900	0.76717	0.78194	0.72434	0.14132	0.15183	0.17400	0.17637
	GRE	0.93404	0.91972	0.92401	0.86902	0.37729	0.40847	0.46559	0.45243

From the computational point of view, the cases $m = 2$ or $m = 4$ and $m = 8$ are the more interesting settings for both methods. The NNWCLP method is very competitive under these settings when $\nu = 0.5$ is assumed known. For instance when $m = 2$ and $\nu = 0.5$ is fixed the NNWCLP has approximatively the same performance of the Vecchia method (GRE = 1.00747). However, a loss of global relative efficiency can be appreciated when $m = 2$ and $\nu = 1.5$ is assumed known (GRE = 0.84060). In addition, the relative efficiencies of the NNWCLP method generally worsen when the smoothness parameter is assumed unknown. If $\nu = 0.5$ this loss of efficiency is still reasonable (for instance GRE = 0.93404 if $m = 2$). Nevertheless, when increasing the smoothness parameter the loss is much more severe (GRE = 0.46559 if $m = 4$). Observing the efficiencies of each parameter it is apparent that this loss of efficiency is mainly due to the poor performance of the NNWCLP method when estimating the smoothness parameter ν compared with Vecchia method. In particular, note that the efficiency of the NNWCLP method when $\nu = 1.5$ clearly improves when increasing m.

Summarizing, as an overall comment the NNWCLP method shows a general reasonable loss of statistical efficiency in comparison to the Vecchia method, particularly when $\nu = 0.5$. This loss of efficiency is much more apparent when estimating and/or increasing the smoothness parameter. However, from a computational viewpoint the NNWCLP method clearly outperforms Vecchia method. To give an idea of the computational gains of the NNWCLP method we compare both methods in terms of R elapsed time using the `system.time` function. For the comparison we have considered the `GPvecchia` R package (Katzfuss et al., 2020). Both methods require a preliminary step. For instance in the Vecchia method a neighborhood structure that depends on m (the size of the conditioning vector) and the type of ordering (we consider a maxmin order as proposed in Guinness (2018)) must be created before the optimization step. For the NNWCLP method the preliminary step involves, as explained in Section 2, building a kd-tree and searching for m nearest neighbors inside the kd-tree. The preliminary step can be performed with the `GeoNeighIndex` function of the `GeoModels` package for the NNWCLP method and the `vecchia_specify` function of the `GPvecchia` package for the Vecchia method. For the second and final step we compare the time needed for the evaluation of the `vecchia_likelihood` R function that implements a likelihood approximation based on Vecchia method versus the evaluation of the NNWCLP function.

Table 8 shows the seconds needed for these two steps when increasing n, the number of location sites, and when $m = 2, 4, 8, 16$. All calculations were carried out on a 2.4 GHz processor with 16 GB of memory.

Table 8
Elapsed time (seconds) needed for the first preliminary step (1) and the objective function evaluation (2) when estimating a spatial dataset with increasing size $n = 10000, 50000, \ldots, 1200000$ using the Vecchia method and the NNWCLP method using $m = 2, 4, 8, 16$.

		$m=2$		$m=4$		$m=8$		$m=16$	
n		NNWCLP	Vecchia	NNWCLP	Vecchia	NNWCLP	Vecchia	NNWCLP	Vecchia
10000	1	0.01	0.61	0.01	0.87	0.02	1.22	0.05	1.87
	2	0.01	0.48	0.02	0.60	0.03	0.90	0.07	1.85
50000	1	0.06	3.42	0.10	3.97	0.14	5.40	0.34	3.42
	2	0.04	2.49	0.07	2.76	0.15	4.34	0.24	2.23
100000	1	0.12	6.80	0.17	7.90	0.27	10.78	0.80	18.43
	2	0.07	5.24	0.16	4.96	0.29	10.46	0.60	20.12
300000	1	0.33	25.65	0.49	27.75	1.03	38.68	2.21	57.95
	2	0.27	16.62	0.52	22.64	1.65	29.43	4.23	49.23
600000	1	1.31	51.37	1.78	64.84	2.63	83.16	4.98	137.93
	2	0.49	35.19	1.23	123.12	1.71	170.08	3.86	201.64
1200000	1	2.52	118.66	3.36	141.51	5.17	170.33	6.12	272.23
	2	1.04	100.74	1.98	111.98	3.99	182.76	9.65	300.32

Focusing on small neighbors, i.e. $m = 2$ or $m = 4$, the gain in terms of computational time (summing up the times of the two steps) with respect to the Vecchia method can be huge when estimating massive spatial dataset. For instance, when $m = 2$ NNWCLP is approximately 60 time faster than Vecchia method and when $m = 4$ NNWCLP is 50 time faster than Vecchia method when estimating on a dataset of size 1200000.

In addition the NNWCLP method, similar to the Vecchia method, can be parallelized to further reduce the computational burden associated with estimation step. A version of the `GeoModels` package (currently available only for macOS at the time) that implements a parallelized NNWCLP method can be found at https://vmoprojs.github.io/GeoModels-page.

As a final comment, NNWCLP exhibits a good balance between statistical efficiency and computational complexity. Compared with the Vecchia method the NNWCLP shows an overall loss of statistical efficiency when estimating a Gaussian RF with Matérn correlation. The loss of efficiency in general tends to increase when estimating the smoothness parameter and/or when considering differentiable RFs. However, the proposed method shows clear computational gains compared to the Vecchia method. As a consequence the proposed estimation method can be an effective solution when analyzing massive datasets.

5. Application

We consider data from the ERA5-Land product Muñoz Sabater et al. (2021) which is a reanalysis dataset providing a consistent view of the evolution of land variables over several decades (1950 to present) at a very fine resolution of the earth that can be downloaded from the section Datasets in https://earthengine.google.com/.

In particular, we focus on hourly surface temperature of January and February 2020 over the region delimited by longitude (in decimal degree) (-85,-34) and latitude (-40,12) which correspond to the greater part of South-America (363827 locations sites). Finally, for each location site, we consider the overall mean of January and February.

Following Li and Zhang (2011), we first detrend the data using splines to remove the cyclic pattern of both variables along the longitude and latitude directions, and we treat the residuals as a realization from an RF. Fig. 5, from left to right, depicts the colored map, the histogram and the empirical semivariogram of the residuals. The graphics suggest that a weakly stationary RF with flexible marginal distribution is potentially an appropriate model for our data. In our analysis we consider three RFs with increasing level of complexity:

1. a Gaussian RF,

$$G(\boldsymbol{s}) = \mu + \sigma G^*(\boldsymbol{s}).$$

2. a Tukey-h RF,

$$T_h(\boldsymbol{s}) = \mu + \sigma T_h^*(\boldsymbol{s}), \quad h \in [0, 0.5),$$

where T_h^* is a Tukey-h RF defined in Equation (7).

3. a Tukey-hh RF

$$T_{h_l,h_r}(\boldsymbol{s}) = \mu + \sigma T_{h_l,h_r}^*(\boldsymbol{s}), \quad h_l, h_r \in [0, 0.5),$$

where T_{h_l,h_r}^* is a Tukey-hh RF defined in Equation (12).

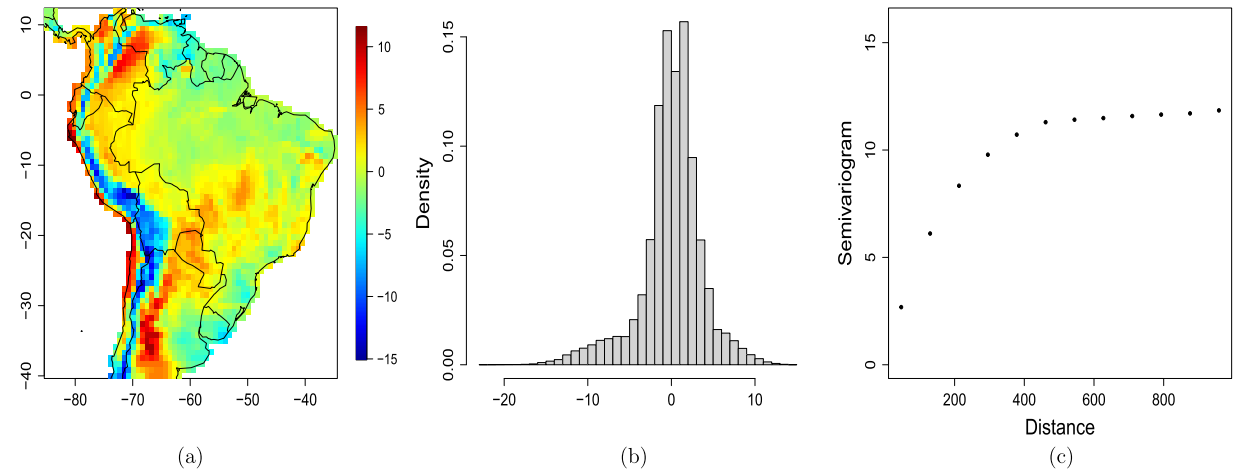

Fig. 5. From left to right: colored map, normalized histogram and empirical semivariogram of the land surface temperature data residuals. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

Table 9
NNWCLP estimation (for $m = 8$ and $m = 16$) with associated standard error for the Gaussian, Tukey-h and Tukey-hh RFs and associated PLIC and RMSE values.

		μ	σ^2	h	h_l	h_r	α	ν	PLIC	RMSE
$m = 8$	Gaussian	0.0234	13.8484				178.96	0.6694	6094235	0.33187
		(0.4461)	(1.269)				(13.705)	(0.0045)		
	Tukey-h	0.9458	7.0187	0.2120			165.70	0.7152	5337899	0.33077
		(0.0269)	(0.6764)	(0.0210)			(13.042)	(0.0054)		
	Tukey-hh	0.4620	6.7776		0.1349	0.2584	158.82	0.7155	5331609	0.33075
		(0.05475)	(0.6309)		(0.0129)	(0.0258)	(11.485)	(0.0062)		
$m = 16$	Gaussian	0.0247	13.812				186.41	0.6565	14303492	0.33220
		(0.4490)	(1.350)				(15.180)	(0.0042)		
	Tukey-h	0.9599	6.8732	0.2196			207.57	0.6581	12751849	0.33211
		(0.0309)	(0.7601)	(0.0249)			(19.042)	(0.0043)		
	Tukey-hh	0.4783	6.6382		0.1382	0.2681	197.39	0.6592	12738045	0.33207
		(0.0755)	(0.6788)		(0.0150)	(0.0291)	(16.655)	(0.0043)		

As underlying correlation model for the RFs G^*, T_h^* and T_{h_l,h_r}^* we consider a Matérn correlation model $\rho_{G^*}(\mathbf{h}) = \mathcal{M}_{\nu,\alpha}(\boldsymbol{d})$. We apply the NNWCLP method for the estimation of the three RFs by considering the weight function (5) with $m = 8$ and $m = 16$. The choice of m follows the empirical evidence from Section 4. Table 9 summarizes the results of the estimates, including their standard error computed using parametric bootstrap. In addition the values of the PLIC defined in (6) are reported. It can be appreciated that the values of the smoothness parameter and spatial scale parameters are quite similar for the three models. However the variance parameter change drastically for the Tukey's RFs compared to the Gaussian RFs. More importantly, the PLIC value for $m = 8, 16$ selects the proposed Tukey-hh RF. Fig. 6 depicts the normalized histogram of the land surface temperature data residuals versus the estimated marginal density function for the Gaussian RF, Tukey-h RF and Tukey-hh RF and, in addition, the empirical semivariogram versus the associated estimated semi-variograms.

We also want to assess the predictive performances of the three models. To do so, we apply a cross validation technique that is we randomly choose 90% of the spatial locations for the parameter estimation and use the remaining 10% for the predictions. We repeat this procedure 100 times, recording the square root of the mean squared error (RMSE) prediction each time.

Since the size of the dataset is very large, the prediction is performed using the best local linear predictor using 100 neighborhoods and using the estimation of the correlation functions $\rho_{G^*}(\mathbf{h})$, $\rho_{T_h^*}(\boldsymbol{d})$ and $\rho_{T_{h_l,h_r}^*}(\boldsymbol{d})$ respectively. Table 9, depicts the empirical mean of the 100 RMSEs obtained. It turns out that the model with the best prediction performance is the proposed Tukey-hh RF. In addition the Tukey-hh RF estimated using NNWCLP with $m = 8$ achieves the best prediction performance.

6. Concluding remarks

In this paper we have focused on weighted composite likelihood based on pairs, a useful estimation method that has a broad applicability when estimating complex (non)-Gaussian RFs. The asymmetric weight function based on nearest neighbors that we have proposed in this paper has been shown to be an effective solution when estimating a Tukey-hh RF, a novel flexible model for spatial data that has flexible marginal distributions, possibly skewed and/or heavy-tailed.

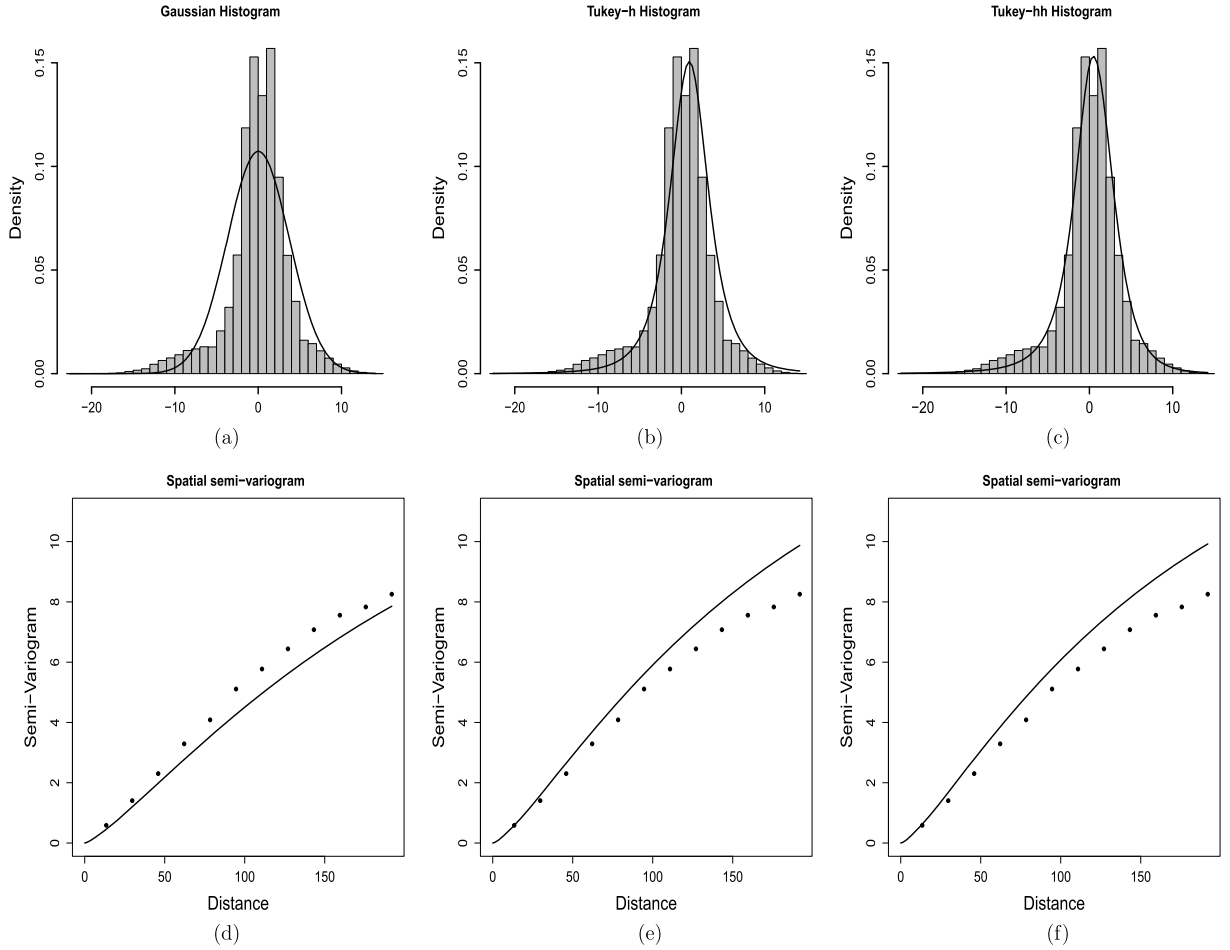

Fig. 6. Top part (from left to right): normalized histogram of the land surface temperature data residuals versus the estimated marginal density function for a) Gaussian RF, b) Tukey-*h* RF and c) Tukey-*hh* RF. Bottom part (from left to right): empirical semivariogram of the land surface temperature data residuals versus the estimated semi-variogram for d) Gaussian RF, e) Tukey-*h* RF and f) Tukey-*hh* RF.

On the one hand we have shown, through numerical examples, that the proposed weight function outperforms the symmetric weight function based on distances from a statistical efficiency viewpoint. On the other hand the computational benefits obtained using the proposed weight function allow estimating massive (up to millions) spatial datasets. This is because kd-tree algorithms can be exploited to achieve an objective function requiring $O(nm)$ time complexity and $O(n)$ memory storage where the best choice of m (the order of the nearest neighbors involved) is typically between 2 and 16, as shown in the numerical examples. Compared to the Vecchia approximation, the proposed method shows a general reasonable loss of statistical efficiency which is more apparent when estimating differentiable Gaussian RFs. However, from a computational point of view, the proposed method clearly outperforms the Vecchia approximation. As a consequence we believe that the proposed method represents an effective solution with a good balance between statistical efficiency and computational complexity when estimating (non)-Gaussian massive datasets. Finally, although we treat the spatial case in this paper, the proposed weight function can be easily extended to the space-time context by considering a spatio-temporal neighborhood.

Acknowledgement

Christian Caamaño-Carrillo was partially supported by grant FONDECYT 11220066 from the Chilean government and DIUBB 2120538 IF/R from the University of Bio-Bio. Moreno Bevilacqua acknowledges financial support from grant FONDECYT 1200068 and ANID/PIA/ANILLOS ACT210096 and ANID project Data Observatory Foundation DO210001 from the Chilean government and project MATH-AMSUD 22-MATH-06 (AMSUD220041). Víctor Morales-Oñate acknowledges support from the Data Science Research Group-CIDED of Escuela Superior Politécnica de Chimborazo and from the Territorial Development, Business and Innovation Research Group-DeTEI of Universidad Técnica de Ambato.

Appendix A. Proofs

A.1. Lemma 1

Lemma 1. *Let $T^*_{h_l,h_r}$, with $h_l, h_r \in [0, 1/2)$ be the Tukey-hh RF defined in (12). Then:*

$$\begin{aligned}\mathbb{E}(T^*_{h_l,h_r}(\boldsymbol{s})T^*_{h_l,h_r}(\boldsymbol{s}+\boldsymbol{d})) = &\frac{g_1(\boldsymbol{d},h_l)u\left(\frac{\rho^2(\boldsymbol{d})}{g_1^2(\boldsymbol{d},h_l)}\right)}{2\pi g_2^{3/2}(\boldsymbol{d},h_l)} + \frac{\rho(\boldsymbol{d})}{2g^{3/2}(\boldsymbol{d},h_l,h_r)} + \frac{\rho(\boldsymbol{d})}{4g_2^{3/2}(\boldsymbol{d},h_l)} \\ &+ \frac{g_1(\boldsymbol{d},h_r)u\left(\frac{\rho^2(\boldsymbol{d})}{g_1^2(\boldsymbol{d},h_r)}\right)}{2\pi g_2^{3/2}(\boldsymbol{d},h_r)} + \frac{\rho(\boldsymbol{d})}{4g_2^{3/2}(\boldsymbol{d},h_r)} \\ &- \frac{[g_1(\boldsymbol{d},h_l)g_1(\boldsymbol{d},h_r)]^{1/2}u\left(\frac{\rho^2(\boldsymbol{d})}{g_1(\boldsymbol{d},h_l)g_1(\boldsymbol{d},h_r)}\right)}{\pi g^{3/2}(\boldsymbol{d},h_l,h_r)},\end{aligned} \tag{A.1}$$

where $g_1(\boldsymbol{d},x) = 1-(1-\rho^2(\boldsymbol{d}))x$, $g_2(\boldsymbol{d},x) = (1-x)^2 - x^2\rho^2(\boldsymbol{d})$, $g(\boldsymbol{d},h_l,h_r) = 1-h_l-h_r+(1-\rho^2(\boldsymbol{d}))h_lh_r$ *and* $u(x) = \sqrt{1-x}+\sqrt{x}arcsin(\sqrt{x})$.

Proof. In the proof we make use of some special functions such as the hypergeometric Gaussian function ${}_2F_1(a,b;c;x)$, the confluent hypergeometric function ${}_1F_1(a;b;x)$ and the parabolic cylinder function $D_v(x)$ (see Gradshteyn and Ryzhik (2007) for their definitions). Setting, $\mathbb{E}(T_i^*, T_j^*) = \mathbb{E}(T^*_{h_l,h_r}(\boldsymbol{s}_i)T^*_{h_l,h_r}(\boldsymbol{s}_j))$ with $\boldsymbol{d} = \boldsymbol{s}_i - \boldsymbol{s}_j$ we have:

$$\begin{aligned}\mathbb{E}(T_i^*, T_j^*) = &\mathbb{E}\left[G(\boldsymbol{s}_i)G(\boldsymbol{s}_j)e^{\frac{h_r(G(\boldsymbol{s}_i))^2}{2}+\frac{h_r(G(\boldsymbol{s}_j))^2}{2}}\right] + \mathbb{E}\left[G(\boldsymbol{s}_i)G(\boldsymbol{s}_j)e^{\frac{h_l(G(\boldsymbol{s}_i))^2}{2}+\frac{h_l(G(\boldsymbol{s}_j))^2}{2}}\right] \\ &+ \mathbb{E}\left[G(\boldsymbol{s}_i)G(\boldsymbol{s}_j)e^{\frac{h_r(G(\boldsymbol{s}_i))^2}{2}+\frac{h_l(G(\boldsymbol{s}_j))^2}{2}}\right] + \mathbb{E}\left[G(\boldsymbol{s}_i)G(\boldsymbol{s}_j)e^{\frac{h_l(G(\boldsymbol{s}_i))^2}{2}+\frac{h_r(G(\boldsymbol{s}_j))^2}{2}}\right] \\ = &\frac{1}{2\pi(1-\rho^2(\boldsymbol{d}))^{1/2}}\int_{\mathbb{R}_+^2} g_ig_je^{-\frac{1}{2(1-\rho^2(\boldsymbol{d}))}\left[g_i^2+g_j^2-2\rho(\boldsymbol{d})g_ig_j\right]}e^{\frac{h_rg_i^2}{2}+\frac{h_rg_j^2}{2}}dg_idg_j \\ &+ \frac{1}{2\pi(1-\rho^2(\boldsymbol{d}))^{1/2}}\int_{\mathbb{R}_-^2} g_ig_je^{-\frac{1}{2(1-\rho^2(\boldsymbol{d}))}\left[g_i^2+g_j^2-2\rho(\boldsymbol{d})g_ig_j\right]}e^{\frac{h_lg_i^2}{2}+\frac{h_lg_j^2}{2}}dg_idg_j \\ &+ \frac{1}{2\pi(1-\rho^2(\boldsymbol{d}))^{1/2}}\int_{\mathbb{R}_+}\int_{\mathbb{R}_-} g_ig_je^{-\frac{1}{2(1-\rho^2(\boldsymbol{d}))}\left[g_i^2+g_j^2-2\rho(\boldsymbol{d})g_ig_j\right]}e^{\frac{h_rg_i^2}{2}+\frac{h_lg_j^2}{2}}dg_idg_j \\ &+ \frac{1}{2\pi(1-\rho^2(\boldsymbol{d}))^{1/2}}\int_{\mathbb{R}_-}\int_{\mathbb{R}_+} g_ig_je^{-\frac{1}{2(1-\rho^2(\boldsymbol{d}))}\left[g_i^2+g_j^2-2\rho(\boldsymbol{d})g_ig_j\right]}e^{\frac{h_lg_i^2}{2}+\frac{h_rg_j^2}{2}}dg_idg_j \\ = &A_1+A_2+A_3+A_4.\end{aligned} \tag{A.2}$$

Taking the first integral A_1 and using (3.462.1) of Gradshteyn and Ryzhik (2007), we obtain,

$$\begin{aligned}A_1 = &\frac{1}{2\pi(1-\rho^2(\boldsymbol{d}))^{1/2}}\int_{\mathbb{R}_+} g_je^{-\frac{[1-(1-\rho^2(\boldsymbol{d}))h_r]g_j^2}{2(1-\rho^2(\boldsymbol{d}))}}\left[\int_{\mathbb{R}_+} g_ie^{\left[-\frac{[1-(1-\rho^2(\boldsymbol{d}))h_r]g_i^2}{2(1-\rho^2(\boldsymbol{d}))}+\frac{\rho(\boldsymbol{d})g_ig_j}{(1-\rho^2(\boldsymbol{d}))}\right]}dg_i\right]dg_j \\ = &\frac{1}{2\pi(1-\rho^2(\boldsymbol{d}))^{1/2}}\left[\frac{(1-\rho^2(\boldsymbol{d}))}{1-(1-\rho^2(\boldsymbol{d}))h_r}\right]\int_{\mathbb{R}_+} g_je^{-\left[\frac{[1-(1-\rho^2(\boldsymbol{d}))h_r]}{2(1-\rho^2(\boldsymbol{d}))}-\frac{\rho^2(\boldsymbol{d})}{4(1-\rho^2(\boldsymbol{d}))[1-(1-\rho^2(\boldsymbol{d}))h_r]}\right]g_j^2} \\ &\times D_{-2}\left(-\frac{\rho(\boldsymbol{d})g_j}{\sqrt{(1-\rho^2(\boldsymbol{d}))[1-(1-\rho^2(\boldsymbol{d}))h_r]}}\right)dg_j.\end{aligned} \tag{A.3}$$

Now, considering (9.240) of Gradshteyn and Ryzhik (2007):

$$\begin{aligned}D_{-2}\left(-\frac{\rho(\boldsymbol{d})g_j}{\sqrt{(1-\rho^2(\boldsymbol{d}))[1-(1-\rho^2(\boldsymbol{d}))h_r]}}\right) = &e^{-\frac{\rho^2(\boldsymbol{d})g_j^2}{4(1-\rho^2(\boldsymbol{d}))[1-(1-\rho^2(\boldsymbol{d}))h_r]}} \\ &\times {}_1F_1\left(1;\frac{1}{2};\frac{\rho^2(\boldsymbol{d})g_j^2}{2(1-\rho^2(\boldsymbol{d}))[1-(1-\rho^2(\boldsymbol{d}))h_r]}\right)\end{aligned}$$

$$+\frac{\sqrt{2\pi}\rho(\boldsymbol{d})g_j e^{-\frac{\rho^2(\boldsymbol{d})g_j^2}{4(1-\rho^2(\boldsymbol{d}))[1-(1-\rho^2(\boldsymbol{d}))h_r]}}}{2\sqrt{(1-\rho^2(\boldsymbol{d}))[1-(1-\rho^2(\boldsymbol{d}))h_r]}}$$
$$\times {}_1F_1\left(\frac{3}{2};\frac{3}{2};\frac{\rho^2(\boldsymbol{d})g_j^2}{2(1-\rho^2(\boldsymbol{d}))[1-(1-\rho^2(\boldsymbol{d}))h_r]}\right), \tag{A.4}$$

combining equations (A.4) and the integral of (A.3) and using (7.621.4) of Gradshteyn and Ryzhik (2007), we obtain

$$\begin{aligned}
A_1 &= \frac{(1-\rho^2(\boldsymbol{d}))^{1/2}}{2\pi[1-(1-\rho^2(\boldsymbol{d}))h_r]}\int_{\mathbb{R}_+} g_j e^{-\frac{[1-(1-\rho^2(\boldsymbol{d}))h_r]g_j^2}{2(1-\rho^2(\boldsymbol{d}))}} {}_1F_1\left(1;\frac{1}{2};\frac{\rho^2(\boldsymbol{d})g_j^2}{2(1-\rho^2(\boldsymbol{d}))[1-(1-\rho^2(\boldsymbol{d}))h_r]}\right)dg_j \\
&+ \frac{\sqrt{2\pi}\rho(\boldsymbol{d})}{4\pi[1-(1-\rho^2(\boldsymbol{d}))h_r]^{3/2}}\int_{\mathbb{R}_+} g_j^2 e^{-\frac{[1-(1-\rho^2(\boldsymbol{d}))h_r]g_j^2}{2(1-\rho^2(\boldsymbol{d}))}} {}_1F_1\left(\frac{3}{2};\frac{3}{2};\frac{\rho^2(\boldsymbol{d})g_j^2}{2(1-\rho^2(\boldsymbol{d}))[1-(1-\rho^2(\boldsymbol{d}))h_r]}\right)dg_j \\
&= \frac{(1-\rho^2(\boldsymbol{d}))^{3/2}}{2\pi[1-(1-\rho^2(\boldsymbol{d}))h_r]^2}{}_2F_1\left(1,1;\frac{1}{2};\frac{\rho^2(\boldsymbol{d})}{[1-(1-\rho^2(\boldsymbol{d}))h_r]^2}\right) \\
&+ \frac{\rho(\boldsymbol{d})(1-\rho^2(\boldsymbol{d}))^{3/2}}{4[1-(1-\rho^2(\boldsymbol{d}))h_r]^3}{}_2F_1\left(\frac{3}{2},\frac{3}{2};\frac{3}{2};\frac{\rho^2(\boldsymbol{d})}{[1-(1-\rho^2(\boldsymbol{d}))h_r]^2}\right).
\end{aligned} \tag{A.5}$$

Using Euler transformation and the identity ${}_2F_1\left(\frac{3}{2},\frac{3}{2};\frac{3}{2};x\right)=(1-x)^{-3/2}$, we obtain

$$A_1 = \frac{[1-(1-\rho^2(\boldsymbol{d}))h_r]}{2\pi[(1-h_r)^2-\rho^2(\boldsymbol{d})h_r^2]^{3/2}}{}_2F_1\left(-\frac{1}{2},-\frac{1}{2};\frac{1}{2};\frac{\rho^2(\boldsymbol{d})}{[1-(1-\rho^2(\boldsymbol{d}))h_r]^2}\right)+\frac{\rho(\boldsymbol{d})}{4[(1-h_r)^2-\rho^2(\boldsymbol{d})h_r^2]^{3/2}}. \tag{A.6}$$

Similarly, A_2, A_3 and A_4 in (A.2), are given by:

$$\begin{aligned}
A_2 &= \frac{[1-(1-\rho^2(\boldsymbol{d}))h_l]}{2\pi[(1-h_l)^2-\rho^2(\boldsymbol{d})h_l^2]^{3/2}}{}_2F_1\left(-\frac{1}{2},-\frac{1}{2};\frac{1}{2};\frac{\rho^2(\boldsymbol{d})}{[1-(1-\rho^2(\boldsymbol{d}))h_l]^2}\right)+\frac{\rho(\boldsymbol{d})}{4[(1-h_l)^2-\rho^2(\boldsymbol{d})h_l^2]^{3/2}}. \\
A_3 &= -\frac{([1-(1-\rho^2(\boldsymbol{d}))h_l][1-(1-\rho^2(\boldsymbol{d}))h_r])^{1/2}}{2\pi[1-h_r-h_l-(1-\rho^2(\boldsymbol{d}))h_rh_l]^{3/2}}{}_2F_1\left(-\frac{1}{2},-\frac{1}{2};\frac{1}{2};\frac{\rho^2(\boldsymbol{d})}{[1-(1-\rho^2(\boldsymbol{d}))h_r][1-(1-\rho^2(\boldsymbol{d}))h_l]}\right) \\
&+ \frac{\rho(\boldsymbol{d})}{4[1-h_r-h_l-(1-\rho^2(\boldsymbol{d}))h_rh_l]^{3/2}}. \\
A_4 &= -\frac{([1-(1-\rho^2(\boldsymbol{d}))h_l][1-(1-\rho^2(\boldsymbol{d}))h_r])^{1/2}}{2\pi[1-h_r-h_l-(1-\rho^2(\boldsymbol{d}))h_rh_l]^{3/2}}{}_2F_1\left(-\frac{1}{2},-\frac{1}{2};\frac{1}{2};\frac{\rho^2(\boldsymbol{d})}{[1-(1-\rho^2(\boldsymbol{d}))h_r][1-(1-\rho^2(\boldsymbol{d}))h_l]}\right) \\
&+ \frac{\rho(\boldsymbol{d})}{4[1-h_r-h_l-(1-\rho^2(\boldsymbol{d}))h_rh_l]^{3/2}}.
\end{aligned} \tag{A.7}$$

Finally, combining equations (A.6) and (A.7), and using the identity ${}_2F_1\left(-\frac{1}{2},-\frac{1}{2};\frac{1}{2};x\right)=\sqrt{1-x}+\sqrt{x}arcsin(\sqrt{x})$ we obtain $\mathbb{E}(T^*_{h_l,h_r}(\boldsymbol{s}_i)T^*_{h_l,h_r}(\boldsymbol{s}_j))$. □

References

Abdulah, S., Alamri, F., Nag, P., Sun, Y., Ltaief, H., Keyes, D.E., Genton, M.G., 2022. The second competition on spatial statistics for large datasets. J. Data Sci. 20, 439–460. https://doi.org/10.6339/22-JDS1076.

Allcroft, D.J., Glasbey, C.A., 2003. A latent Gaussian Markov random-field model for spatiotemporal rainfall disaggregation. J. R. Stat. Soc., Ser. C, Appl. Stat. 52, 487–498.

Arya, S., Mount, D.M., Netanyahu, N.S., Silverman, R., Wu, A.Y., 1998. An optimal algorithm for approximate nearest neighbor searching. J. ACM 45, 891–923.

Bai, Y., Kang, J., Peter, X., Song, K., 2014. Efficient pairwise composite likelihood estimation for spatial-clustered data. Biometrics 70 (3), 661–670.

Banerjee, S., Carlin, B.P., Gelfand, A.E., 2004. Hierarchical Modeling and Analysis for Spatial Data. Chapman & Hall/CRC Press, Boca Raton, FL.

Banerjee, S., Gelfand, A.E., Finley, A.O., Sang, H., 2008. Gaussian predictive process models for large spatial data sets. J. R. Stat. Soc. B 70, 825–848.

Bárdossy, A., 2006. Copula-based geostatistical models for groundwater quality parameters. Water Resour. Res. 42.

Bentley, J.L., 1975. Multidimensional binary search trees used for associative search. Commun. ACM 18, 309–517.

Bevilacqua, M., Gaetan, C., 2015. Comparing composite likelihood methods based on pairs for spatial Gaussian random fields. Stat. Comput. 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. J. Am. Stat. Assoc. 107, 268–280.

Bevilacqua, M., Faouzi, T., Furrer, R., Porcu, E., 2019. Estimation and prediction using generalized Wendland functions under fixed domain asymptotics. Ann. Stat. 47, 828–856.

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

Bevilacqua, M., Caamaño-Carrillo, C., Arellano-Valle, R.B., Morales-Oñate, V., 2021. Non-Gaussian geostatistical modeling using (skew) t processes. Scand. J. Stat. 48, 212–245.

Bevilacqua, M., Caamaño-Carrillo, C., Arellano-Valle, R.B., Gómez, C., 2022a. A class of random fields with two-piece marginal distributions for modeling point-referenced data with spatial outliers. Test 31, 644–674.

Bevilacqua, M., Caamaño-Carrillo, C., Porcu, E., 2022b. Unifying compactly supported and Matérn covariance functions in spatial statistics. J. Multivar. Anal. 189, 104949.

Bevilacqua, M., Morales-Oñate, V., Caamaño-Carrillo, C., 2023. GeoModels: procedures for Gaussian and non Gaussian geostatistical (large) data analysis. https://CRAN.R-project.org/package=GeoModels. r package version 1.1.3.
Blasi, F., Caamaño-Carrillo, C., Bevilacqua, M., Furrer, R., 2022. A selective view of climatological data and likelihood estimation. Spat. Stat., 100596.
Caragea, P., Smith, R., 2006. Approximate likelihoods for spatial processes. Technical Report. Department of Statistics, Iowa State University.
Cressie, N., Johannesson, G., 2008. Fixed rank Kriging for very large spatial data sets. J. R. Stat. Soc. B 70, 209–226.
Cressie, N., Wikle, C., 2011. Statistics for Spatio-Temporal Data. Wiley Series in Probability and Statistics. Wiley, New York.
Davis, R., Yau, C.-Y., 2011. Comments on pairwise likelihood in time series models. Stat. Sin. 21, 255–277.
Davison, A.C., 2003. Statistical Models. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge.
DeOliveira, V., Kedem, B., Short, D.A., 1997. Bayesian prediction of transformed Gaussian random fields. J. Am. Stat. Assoc. 92, 1422–1433.
Diggle, P., Giorgi, E., 2019. Model-Based Geostatistics for Global Public Health: Methods and Applications. Chapman & Hall/CRC Interdisciplinary Statistics Series. Chapman and Hall/CRC Press, New York.
Diggle, P., Tawn, J., Moyeed, R., 1998. Model-based geostatistics. J. R. Stat. Soc., Ser. C, Appl. Stat. 47, 299–350.
Diggle, P.J., Ribeiro, P.J., 2007. Model-Based Geostatistics. Springer, New York.
Eidsvik, J., Shaby, B., Reich, B., Wheeler, M., Niemi, J., 2014. Estimation and prediction in spatial models with block composite likelihoods. J. Comput. Graph. Stat. 29, 295–315.
Elseberg, J., Magnenat, S., Siegwart, R., Nüchter, A., 2012. Comparison of nearest-neighbor-search strategies and implementations for efficient shape registration. J. Softw. Eng. Robot. 3, 2–12.
Emery, X., Arroyo, D., Porcu, E., 2016. An improved spectral turning-bands algorithm for simulating stationary vector Gaussian random fields. Stoch. Environ. Res. Risk Assess. 30, 1863–1873.
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.
Furrer, R., Genton, M.G., Nychka, D., 2013. Covariance tapering for interpolation of large spatial datasets. J. Comput. Graph. Stat. 15, 502–523.
Gay, D.M., 1990. Usage summary for selected optimization routines. Computing Science Technical Report 153. AT&T Bell Laboratories, Murray Hill.
Gelfand, A.E., Schliep, E.M., 2016. Spatial statistics and Gaussian processes: a beautiful marriage. Spat. Stat. 18, 86–104.
Gneiting, T., 2002. Compactly supported correlation functions. J. Multivar. Anal. 83, 493–508.
Goerg, G.M., 2015. The Lambert way to gaussianize heavy-tailed data with the inverse of Tukey's h transformation as a special case. Sci. World J., 1–16.
Gough, B., 2009. GNU Scientific Library Reference Manual. Network Theory Ltd.
Gradshteyn, I., Ryzhik, I., 2007. Table of Integrals, Series, and Products, 7 ed. Academic Press, New York.
Gräler, B., 2014. Modelling skewed spatial random fields through the spatial vine copula. Spat. Stat. 10, 87–102.
Guinness, J., 2018. Permutation and grouping methods for sharpening Gaussian process approximations. Technometrics 60, 415–429.
Guinness, J., 2021. Gaussian process learning via Fisher scoring of Vecchia's approximation. Stat. Comput. 31.
Heagerty, P., Lele, S., 1998. A composite likelihood approach to binary spatial data. J. Am. Stat. Assoc. 93, 1099–1111.
Heaton, M.J., Datta, A., Finley, A.O., Furrer, R., Guinness, J., Guhaniyogi, R., Gerber, F., Gramacy, R.B., Hammerling, D., Katzfuss, M., Lindgren, F., Nychka, D.W., Sun, F., Zammit-Mangion, A., 2019. A case study competition among methods for analyzing large spatial data. J. Agric. Biol. Environ. Stat. 24, 398–425.
Heyde, C., 1997. Quasi-Likelihood and Its Application: A General Approach to Optimal Parameter Estimation. Springer, New York.
Huang, H., Abdulah, S., Sun, Y., Hatem, L., Keyes, D., Genton, M., 2021. Competition on spatial statistics for large datasets. J. Agric. Biol. Environ. Stat. 24, 580–595.
Joe, H., Lee, Y., 2009. On weighting of bivariate margins in pairwise likelihood. J. Multivar. Anal. 100, 670–685.
Katzfuss, M., 2017. A multi-resolution approximation for massive spatial datasets. J. Am. Stat. Assoc. 112, 201–214.
Katzfuss, M., Gong, W., 2020. A class of multi-resolution approximations for large spatial datasets. Stat. Sin. 112, 2203–2226.
Katzfuss, M., Guinness, J., 2021. A general framework for Vecchia approximations of Gaussian processes. Stat. Sci. 36, 124–141.
Katzfuss, M., Jurek, M., Zilber, D., Gong GPvecchia, W., 2020. Scalable Gaussian-process approximations. https://CRAN.R-project.org/package=GPvecchia. r package version 0.1.3.
Kaufman, C.G., Schervish, M.J., Nychka, D.W., 2008. Covariance tapering for likelihood-based estimation in large spatial data sets. J. Am. Stat. Assoc. 103, 1545–1555.
Kazianka, H., Pilz, J., 2010. Copula-based geostatistical modeling of continuous and discrete data including covariates. Stoch. Environ. Res. Risk Assess. 24, 661–673.
Li, B., Zhang, H., 2011. An approach to modeling asymmetric multivariate spatial covariance structures. J. Multivar. Anal. 102, 1445–1453.
Li, F., Sang, H., 2018. On approximating optimal weighted composite likelihood method for spatial models. Stat 7, e194.
Lindgren, F., Rue, H., Lindström, J., 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. J. R. Stat. Soc. B 73, 423–498.
Lindsay, B., 1988. Composite likelihood methods. Contemp. Math. 80, 221–239.
Masarotto, G., Varin, C., 2012. Gaussian copula marginal regression. Electron. J. Stat. 6, 1517–1549.
Morales-Navarrete, D., Bevilacqua, M., Caamaño-Carrillo, C., Castro, L., 2022. Modelling point referenced spatial count data: a Poisson process approach. J. Am. Stat. Assoc., 1–25.
Morgenthaler, S., Tukey, J.W., 2000. Fitting quantiles: doubling, HR, HQ, and HHH distributions. J. Comput. Graph. Stat. 9, 180–195.
Muñoz Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D.G., Piles, M., Rodríguez-Fernández, N.J., Zsoter, E., Buontempo, C., Thépaut, J.-N., 2021. ERA5-land: a state-of-the-art global reanalysis dataset for land applications. Earth Syst. Sci. Data 13, 4349–4383.
Oliveira, V.D., 2006. On optimal point and block prediction in log-Gaussian random fields. Scand. J. Stat. 33, 523–540.
Pace, L., Salvan, A., Sartori, N., 2019. Efficient composite likelihood for a scalar parameter of interest. Stat 8, e222.
Stein, M., 1999. Interpolation of Spatial Data. Some Theory of Kriging. Springer-Verlag, New York.
Stein, M., 2008. A modeling approach for large spatial datasets. J. Korean Stat. Soc. 37, 3–10.
Stein, M., 2013. Statistical properties of covariance tapers. J. Comput. Graph. Stat. 22, 866–885.
Stein, M., Chi, Z., Welty, L., 2004. Approximating likelihoods for large spatial data sets. J. R. Stat. Soc. B 66, 275–296.
Varin, C., Vidoni, P., 2005. A note on composite likelihood inference and model selection. Biometrika 52, 519–528.
Varin, C., Reid, N., Firth, D., 2011. An overview of composite likelihood methods. Stat. Sin. 21, 5–42.
Vecchia, A., 1988. Estimation and model identification for continuous spatial processes. J. R. Stat. Soc. B 50, 297–312.
Wallin, J., Bolin, D., 2015. Geostatistical modelling using non-Gaussian Matérn fields. Scand. J. Stat. 42, 872–890.
Xu, G., Genton, M.G., 2015. Efficient maximum approximated likelihood inference for Tukey's g-and-h distribution. Comput. Stat. Data Anal. 91, 78–91. https://doi.org/10.1016/j.csda.2015.06.002. https://www.sciencedirect.com/science/article/pii/S0167947315001401.
Xua, G., Genton, M.G., 2017. Tukey g-and-h random fields. J. Am. Stat. Assoc. 112, 1236–1249.
Yan, Y., Jeong, J., Genton, M., 2020. Multivariate transformed Gaussian processes. Jpn. J. Stat. Data Sci. 3, 129–152.
Zhang, H., El-Shaarawi, A., 2010. On spatial skew-Gaussian processes and applications. Environmetrics 21 (1), 33–47.
Zilber, D., Katzfuss, M., 2021. Vecchia–Laplace approximations of generalized Gaussian processes for big non-Gaussian spatial data. Comput. Stat. Data Anal. 153, 107081.