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

Christian Caamaño-Carrillo (chcaaman@ubiobio.cl)
University of Bío-Bío

Moreno Bevilacqua
Adolfo Ibáñez University

Cristian López
University of Bío-Bío

Víctor Morales-Oñate
Universidad San Francisco de Quito

Research Article

Keywords: Covariance estimation, Geostatistics, Large datasets, Vecchia approximation

Posted Date: September 20th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2073895/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Nearest neighbours weighted composite likelihood based on pairs for (non-)Gaussian massive spatial data with an application to Tukey-$hh$ random fields estimation

Christian Caamaño-Carrillo$^{1*}$, Moreno Bevilacqua$^{2,3†}$, Cristian López$^{1†}$ and Víctor Morales-Oñate$^{4,5†}$

$^1*$Departamento de Estadística, Universidad del Bío-Bío, Avda. Collao 1202, Concepción, 4030000, Chile.
$^2$Facultad de Ingeniería y Ciencias, Universidad Adolfo Ibáñez, Avda. Padre Hurtado 750, Viña del Mar, 2520000, Chile.
$^3$Dipartimento di Scienze Ambientali, Informatica e Statistica, Ca’ Foscari University of Venice, via Torino 155, Mestre, 30172, Italy.
$^4$Data Analytics Department, Banco Solidario, Amazonas N36-152, Quito, 170135, Ecuador.
$^5$Department of Economics, Universidad San Francisco de Quito, Diego de Robles s/n, Quito, 170901, Ecuador.

*Corresponding author(s). E-mail(s): chcaaman@ubiobio.cl;
Contributing authors: moreno.bevilacqua@uai.cl;
cristiannlpez@hotmail.com; vmorales.ppb@gmail.com;
†These authors contributed equally to this work.

Abstract
In this paper we propose a highly scalable method for (non-)Gaussian random fields estimation. In particular, we propose a novel (a)symmetric weight function based on nearest neighbours for the method of maximum weighted composite likelihood based on pairs (WCLP).
The proposed 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 we consider the estimation of a novel non-Gaussian random field named Tukey-\textit{hh} random field that has flexible marginal distributions, possibly skewed and/or heavy-tailed. In an extensive simulation study we explore the statistical efficiency of the proposed nearest neighbours WCLP method with respect to the WCLP method using weights based on distances when estimating the parameters of the Tukey-\textit{hh} random field. In the Gaussian case we also compare the proposed method 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. Our developments have been implemented in an open-source package for the R statistical environment.

**Keywords:** Covariance estimation, Geostatistics, Large datasets, Vecchia approximation

## 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 [1–3] 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.

Unfortunately, practical use of Gaussian RFs have 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 [4–6], based on tapered covariance matrix [7–9], based on composite likelihood [10–12], based on approximation using Markov Gaussian RFs [13] based on multiresolution approximations [14, 15] to name just a few. A general framework that includes several proposals based on Vecchia approximation [16] has recently been proposed in [17]. For an extensive review see [18] and the references therein.
The second problem is from a modelling 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 modelling these kind of data is the hierarchical model proposed by [19] that can be viewed as a generalized linear mixed model [20, 21]. 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 [22] and this can be troublesome when modelling spatial data displaying some kind of continuity. A scalable method of estimation for these kind of models based on the Vecchia-Laplace approximation has been proposed in [23]. [24] 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 random fields.

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 \( Z = \{Z(s), s \in A\}, A \subset \mathbb{R}^d \) a Gaussian RF and let \( Y = \{Y(s), s \in A\} \) a RF defined through the transformation

\[
Y(s) = f(g_1(Z_1(s)), g_2(Z_2(s)), \ldots, g_q(Z_q(s))), \quad q \geq 1
\]  

where \( Z_1, \ldots, Z_q \), are independent copies of \( Z \) and \( f : \mathbb{R}^q \rightarrow \mathbb{R} \) and \( g_1, \ldots, g_q \) with \( g_i : \mathbb{R} \rightarrow \mathbb{R} \) are suitable functions. The class (1) includes several examples of non-Gaussian RFs proposed in the literature such as Bernoulli RFs [25], skew-Gaussian RFs, [26], Tukey \( g-h \) RFs [27], Student-\( t \) RFs [28], Weibull RFs [29] or Poisson RFs [30] to mention just a few. In addition, the so-called class of trans-Gaussian RFs (see for instance [31] and [32]) or the general class of non-Gaussian-RFs based on Gaussian Copula [33–35] and chi-square Copula [36] 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 [3] of the generalized Wendland model [37].

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 [38] or the class of Gaussian copula RFs [33, 34] or the sinh-arcsinh RFs proposed in [39].
and [40]. 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 [26] requires computation of order $O(2^{n-1})$. Another notable example is the ML estimation of the Bernoulli RFs proposed in [25]. 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 [28] or the Poisson RF in [30].

To address the abovementioned computational problem we consider the method of composite likelihood (CL) [12, 41]. 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, [10] considered a weighted composite likelihood based on pairs (WCLP hereafter) while [11] developed a block composite likelihood in a vein similar to [42]. Furthermore, the methods proposed, for instance, in [43] or [44], based on Vecchia approximation [16], 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 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 [27] which is based on a generalization of an RF with Tukey-$h$ marginals [50]. 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 distribution of the proposed Tukey-$hh$ RFs. It turns out that the evaluation of the multivariate pdf is computationally expensive even for relatively small datasets and, as a consequence, ML estimation is unfeasible. 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 compares 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.

In the purely Gaussian case we also compare the NNWCLP method with some recent improvements [17, 51] of Vecchia approximation method originally proposed in [16]. It turns out that the proposed method show 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 georeferenced dataset (approximatively 360,000 data) from the ERA5-Land dataset [52] 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 [53] 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” [54] organized by King Abdullah University of Science and Technology (KAUST). The analysis in both competitions has been completely performed using the GeoModels package and the methodology proposed in this paper.

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 neighbours. 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 neighbours weighted composite
likelihood estimation based on pairs

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

For any set of distinct points $(s_1, \ldots, s_n)^T$, $s_i \in A$, $n \in \mathbb{N}$, we denote by $Z_{ij} = (Z(s_i), Z(s_j))^T$, $i \neq j$ and $Z_{i|j} = Z(s_i) \mid Z(s_j) = z_j$ the bivariate
random vector and the conditional random variable respectively and we denote by $Z = (Z(s_1), \ldots, Z(s_n))^T$ the multivariate random vector. In addition, we denote with $f_{Z_{ij}}$, $f_{Z_{i|j}}$ and $f_Z$ the associated probability density functions and we denote $f_{Z_k}$ as the marginal density function of $Z(s_k)$. Finally, we denote $Z^*$ as the standardized RF, i.e., $Z^*(s) := (Z(s) - \mu(s))/\sigma$.

Let $B_k$ be a marginal or conditional set of $Z$. The log-CL \[41\] is an objective function defined as a sum of $K$ sub-log-likelihoods

$$CL(\theta) = \sum_{k=1}^{K} l(\theta; B_k)w_k,$$

where $\theta$ is the vector of unknown parameters, $l(\theta; B_k)$ is a log-likelihood calculated by considering only the random variables in $B_k$ and $w_k$ are suitable weights that do not depend on $\theta$. The maximum CL estimate is given by $\hat{\theta} = \text{argmax}_\theta CL(\theta)$.

The WCLP estimation method \[10\] is obtained by setting $B_k = Z_{ij}$ and $B_k = Z_{i|j}$. In the first case we obtain the pairwise marginal log-likelihood $l_{ij} = \log(f_{Z_{ij}})$ and in the second case we obtain the pairwise conditional log-likelihood $l_{i|j} = \log(f_{Z_{i|j}})$. The corresponding weighted composite log-likelihoods functions are given by:

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

and $\hat{\theta}_a = \text{argmax}_\theta wpl_a(\theta)$, 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 ML estimation and the role of the weights $w_{ij}$ is to minimize this loss. Using theory of optimal estimating equations \[55\], it can be easily seen \[56\] that the optimal weights requires 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 \[57\] and \[58\]. However the computation of these kind of weights can be computationally demanding for large $n$.

To avoid this computational problem different authors \[10, 25, 45–47\] have proposed the DDWCLP method that considers the weight function:

$$w_{ij}(k) = \begin{cases} 1 & \|s_i - s_j\| < k \\ 0 & \text{otherwise} \end{cases}. \quad (4)$$

where $k \in \mathbb{R}^+$ is an arbitrary distance greater than the minimum distance of the location points. These 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 these kind of weights improve the statistical efficiency of the method with respect to the use of constant weights (see for instance [59], [60] and [56]).

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 $l_{ij}(\theta)$ which is not symmetric.

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

$$w_{ij}(m) = \begin{cases} 1 & s_i \in N_m(s_j) \\ 0 & \text{otherwise} \end{cases}$$

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 $s_1 = (0.15, 0.75)^T$, $s_2 = (0.2, 0.85)^T$, $s_3 = (0.3, 0.7)^T$ and $s_4 = (0.26, 0.35)^T$. The left part of Figure 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 neighbours $w_{ij}(m)$ 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 [48, 49, 61] 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 $O(n \log(n))$ time complexity and $O(n)$ associated storage and 2) searching for $m$ nearest neighbours inside the kd-tree that has an $O(m \log(n))$ 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 [61].
Figure 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 neighbours \( w_{ij}(m) \) with \( m = 2 \).

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

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

\[
\hat{\theta}_a := \arg\max_{\theta} \ wpl_a(\theta), \quad a = M, C
\]

and, arguing as in [10], under some mixing conditions of the Tukey-hh random field 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 \( G_a^{-1}(\hat{\theta}) \) the inverse of the Godambe information \( G_a(\theta) := H_a(\theta)J_a(\theta)^{-1}H_a(\theta) \), where \( H_a(\theta) := \mathbb{E}[\nabla^2 wpl_a(\theta)] \) and \( J_a(\theta) := \text{Var}[\nabla wpl_a(\theta)] \). Standard error estimation can be obtained considering the square root diagonal elements of \( G_a^{-1}(\hat{\theta}) \). Moreover, model selection can be performed by considering the information criterion, defined as

\[
\text{PLIC} := -2wpl_a(\hat{\theta}) + 2\text{tr}(H_a(\hat{\theta})G_a^{-1}(\hat{\theta})),
\]

which is composite likelihood version of the Akaike information criterion (AIC) [45]. Note that, the computation of standard errors and PLIC require evaluation of the matrices \( H_a(\hat{\theta}) \) and \( J_a(\hat{\theta}) \). However, the evaluation of \( J_a(\hat{\theta}) \) is computationally unfeasible for large datasets and in this case subsampling techniques can be used to estimate \( J_a(\theta) \) as in [56] and [25]. A straightforward and more robust alternative that we adopt in Section 5, is parametric bootstrap estimation of \( G_a^{-1}(\theta) \). Since this technique is simulation based fast
methods of simulation of Gaussian random fields such as circulant embedding or turning bands methods [62] are required for large datasets.

3 Tukey-\(hh\) random fields

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

Following [27], let \(T_h^* = \{T_h^*(s), s \in A\}\), with \(h \in [0, 1/2)\), be an RF with a standard Tukey-\(h\) marginal distribution defined through a monotonic transformation \(\tau_h(x) = xe^{\frac{h}{2}x^2}\), \(x \in \mathbb{R}\) of a standard Gaussian RF \(G\) as:

\[
T_h^*(s) =: \tau_h(G(s)).
\] (6)

The inverse transformation \(\tau_h^{-1}(x)\) can be expressed in terms of the Lambert function \(i.e., \tau_h^{-1}(x) = sign(x) \left(\frac{W(ht^2)}{h}\right)^{1/2}\) where \(W(\cdot)\) is the Lambert-W function [50].

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 \(1/h\) [63]. If \(h = 0\) the Gaussian RF \(G\) is obtained as the special limit case.

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

\[
\rho_{T_h^*}(h) = \frac{\rho(h)(1 - 2h)^{3/2}}{[(1 - h)\rho^2(h) + h^2\rho^2(h)]^{3/2}}.
\] (7)

In addition, the Tukey-\(h\) RF has a marginal pdf given by [50]:

\[
f_{T_h^*}(t) = \frac{\tau_h^{-1}(t)}{t(1 + W(ht^2))} \phi(\tau_h^{-1}(t), 0, 1),
\] (8)

Note that \(f_{T_h^*}(t)\) is well defined when \(t \to 0\) and/ or \(h \to 0\) using the limits \(\lim_{h \to 0} \tau_h^{-1}(t) = t\), \(\lim_{h \to 0} W(ht^2) = 0\), \(\lim_{t \to 0} \tau_h^{-1}(t) = t\) and \(\lim_{t \to 0} W(ht^2) = 0\). The multivariate pdf of the vector \(T_h^* = (T_h^*(s_1), \ldots, T_h^*(s_N))^T\) is given by:

\[
f_{T_h^*}(t) = \frac{\prod_{i=1}^n \tau_h^{-1}(t_i)}{(\prod_{i=1}^n t_i(1 + W(ht_i^2)))} \phi_n(\tau_h^{-1}(t), 0, R_n)
\] (9)

where \(R_N = [\rho(s_i - s_j)]_{i,j=1}^n\) denotes correlation matrix associated with the underlying correlation function \(\rho(h)\) and the transformation \(\tau_h^{-1}(x)\) applies pointwise for a given vector \(x\).
To take into account both heavy tails and skewness, a generalization of (6) has been proposed in [27] by considering the so-called Tukey-\(g\)-\(h\) RF \(M^*_{h,g}(s), s \in A\) defined as:

\[
M^*_{h,g}(s) := g^{-1}(e^{gG(s)} - 1)e^{\frac{h(G(s))^2}{2}}, \tag{10}
\]

where \(g \in \mathbb{R}\) is a skewness parameter. The Tukey-\(h\) RF is obtained as special case using the limit \(\lim_{g \to 0}(e^{gx} - 1)/g = x\). Unfortunately, the transformation involved in (10) does not have an explicit inverse and as a consequence (composite) likelihood based methods are not readily applicable.

Our proposal considers an alternative generalization of (6) obtained by defining the RF \(T^*_{h_l,h_r} = \{T^*_{h_l,h_r}(s), s \in A\}\), with \(h_l \in [0, 1/2)\) and \(h_r \in [0, 1/2)\), as:

\[
T^*_{h_l,h_r}(s) := \begin{cases} 
\tau_{h_l}(G(s)), & G(s) < 0 \\
\tau_{h_r}(G(s)), & G(s) \geq 0 
\end{cases} \tag{11}
\]

The marginal distribution of \(T^*_{h_l,h_r}\) is called the Tukey-\(hh\) distribution [63] with pdf given by [50]:

\[
f_{T^*_{h_l,h_r}}(t) = \frac{\tau_{h_l}^{-1}(t)}{t(1 + W(ht^2))}\phi(\tau_{h_l}^{-1}(t), 0, 1)I_{(0, \infty)}(t) + \frac{\tau_{h_r}^{-1}(t)}{t(1 + W(hr^2))}\phi(\tau_{h_r}^{-1}(t), 0, 1)I_{[0, \infty)}(t), \tag{12}
\]

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:

\[
\mathbb{E}(T^*_{h_l,h_r}(s)) = \frac{h_r - h_l}{\sqrt{2\pi}(1 - h_l)(1 - h_r)} \tag{13}
\]

\[
\text{Var}(T^*_{h_l,h_r}(s)) = \frac{1}{2} \left[ (1 - 2h_l)^{-3/2} + (1 - 2h_r)^{-3/2} \right] - \mathbb{E}^2(T^*_{h_l,h_r}(s)). \tag{14}
\]

From (13) 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 \(\max(h_l, h_r)\)) [63].

We now study the correlation function of the Tukey-\(hh\) RF. The following lemma is useful for giving a closed form expression for the correlation \(\rho_{T^*_{h_l,h_r}}(h)\). The proof can be found in the Appendix.
Lemma 1 Let $T^*_{h_l,h_r}$, with $h_l, h_r \in [0,1/2)$ be the Tukey-hh RF defined in (19). Then:

$$
\mathbb{E}(T^*_{h_l,h_r}(s_i)T^*_{h_l,h_r}(s_j)) = \frac{g_1(h, h_l)u\left(\frac{\rho^2(h)}{g^2_1(h, h_l)}\right)}{2\pi g_2^{3/2}(h, h_l)} + \frac{\rho(h)}{2g_2^{3/2}(h, h_l, h_r)} + \frac{\rho(h)}{4g_2^{3/2}(h, h_l)}
$$

$$
+ \frac{g_1(h, h_r)u\left(\frac{\rho^2(h)}{g^2_1(h, h_r)}\right)}{2\pi g_2^{3/2}(h, h_r)} + \frac{\rho(h)}{4g_2^{3/2}(h, h_r)} - \frac{|g_1(h, h_l)g_1(h, h_r)|^{1/2}u\left(\frac{\rho^2(h)}{g_1(h, h_l)g_1(h, h_r)}\right)}{\pi g_2^{3/2}(h, h_l, h_r)} \tag{15}
$$

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

Using Lemma (1), the correlation of the Tukey-hh RF $\rho_{T^*_{h_l,h_r}}(h)$ is given by:

$$
\rho_{T^*_{h_l,h_r}}(h) = \frac{2\pi(1 - h_l)^2(1 - h_r)^2\mathbb{E}(T^*_{h_l,h_r}(s_i)T^*_{h_l,h_r}(s_j)) - (h_r - h_l)^2}{m(h_l, h_r)} \tag{16}
$$

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 $\mathbb{E}(T^*_{h_l,h_r}(s_i)T^*_{h_l,h_r}(s_i)) = \frac{1}{2}[(1 - 2h_r)^{3/2} + (1 - 2h_l)^{3/2}]$ and this implies $\rho_{T^*_{h_l,h_r}}(0) = 1$.

Note that a version of the Tukey-hh RFs in Equation (12) 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 $\rho(h)$ with $\rho^*(h) = (1 - \tau^2)\rho(h) + \tau^2 I(h = 0)$ 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, $\rho_{T^*_{h_l,h_r}}(h) = \rho_{T^*_{h_r,h_l}}(h)$. In addition $\rho_{T^*_{h_l,h_r}}(h) \leq \rho(h)$ and $\rho_{T^*_{h_l,h_r}}(h) = 0$ if $\rho(h) = 0$ that is $\rho_{T^*_{h_l,h_r}}(h)$ is compactly supported if $\rho(h)$ 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 the generalized Wendland correlation function [37, 64] as proposed in [65]. It is defined for $\nu \geq 0$ as:

$$
\mathcal{W}_{\nu, \delta, L(\nu, \delta, \alpha)}(h) = \begin{cases} 
K(U(h))^{\nu+\delta}F_1\left(\frac{\nu}{2}, \frac{\delta+1}{2}; \nu + \delta + 1; U(h)\right) & 0 \leq \|h\| \leq L(\nu, \delta, \alpha) \\
0 & \text{otherwise},
\end{cases} \tag{17}
$$
where $U(h) := 1 - \left(\frac{\|h\|}{\Gamma(\nu, \delta, \alpha)}\right)^2$, $L(\nu, \delta, \alpha) := \alpha \left(\Gamma(\delta + 2\nu + 1)/\Gamma(\delta)\right)^{\nu+2\nu}$, $K := (2^{-\delta-1}\Gamma^{-1}(2\nu)/\Gamma(2\nu+\delta+1))/\Gamma(\nu+\delta+1)$ and $2F_1(a, b, c, x)$ is the Gaussian hypergeometric function [66]. Here $\alpha > 0$, $\delta \geq (d + 1)/2 + \nu$ guarantee the positive definiteness of the model in $\mathbb{R}^d$. [65] show that $\mathcal{GW}_{\nu, \delta, L(\nu, \delta, \alpha)}(h) \rightarrow \mathcal{M}_{\nu+0.5, \alpha}(h)$ as $\delta \rightarrow \infty$ where

$$
\mathcal{M}_{\nu, \alpha}(h) = \left(\frac{2^{1+\nu}}{\Gamma(\nu)}\right)^{\nu}\mathcal{K}_{\nu} \left(\left(\frac{\|h\|}{\alpha}\right)^\nu\right) r \geq 0.
$$

is the celebrated Matérn correlation model [3]. 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:

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

In the first row of Figure 2 we depict, from the left to the right, three realizations on a fine grid of a unit square of a Tukey-$hh$ RF with underlying correlation function $\mathcal{GW}_{0, 3.5, L(0, 3.5, 0.1)}(h)$ 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. The following theorem gives an explicit closed-form expression for the pdf of the random vector $T_{h_l, h_r}^* = (T_{h_l, h_r}^*(s_1), \ldots, T_{h_l, h_r}^*(s_n))^T$. The proof can be found in the appendix.

**Theorem 1** Let $T_{h_l, h_r}^* = (T_{h_l, h_r}^*(s_1), \ldots, T_{h_l, h_r}^*(s_n))^T$ be the $n$-dimensional random vector associated to the Tukey-$hh$ RF with underlying correlation $\rho(h)$. Then:

$$
f_{T_{h_l, h_r}^*}(t) = \sum_{\ell \in \{-1, +1\}^n} \frac{\Pi_{i=1}^m t_{i}^{-1}(t_i)}{\Pi_{i=1}^m t_i (1 + W(h_t^{-2}t_i^2))} \frac{\Pi_{j=m+1}^n t_{j}^{-1}(t_j)}{\Pi_{j=m+1}^n t_j (1 + W(h_t^{-2}t_j^2))}
$$

$$
\cdot \phi_n(\tau^{-1}(h, t), 0, R_n) \int_{\Pi_{i=1}^m (-\infty, 0)^i \Pi_{j=1}^n (0, \infty)^j} (t_1, \ldots, t_n)
$$

where $I_A(x)$ denotes the indicator function of the set $A$, $R_n = [\rho(s_i - s_j)]_{i,j=1}^n$, $\tau^{-1}(h, t) = D(\ell) \cdot g(h, t)$ with $g(h, t) = \left(W(h_t^{-2}t_i^2)^{-1/2}, \ldots, W(h_t^{-2}t_j^2)^{-1/2}\right)^T$, with $h \in \{h_l, h_r\}$ where $D(\ell)$ are diagonal matrices $D(\ell) = \text{Diag} \{\ell_1, \ldots, \ell_n\}$, with $\ell = (\ell_1, \ldots, \ell_n) \in \{-1, +1\}^n$.

From Theorem 1, it is apparent that likelihood-based methods for the Tukey-$hh$ RFs are impractical from the computational point of view even for
Figure 2  First column: three realization 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.

a relatively small $n$. However when $n = 2$, the bivariate pdf is given by:

$$f_{T_{h_l,h_r}^*}(t_i, t_j) = c(t_i, t_j, h_l, h_r)I_{(-\infty,0) \times (-\infty,0)}(t_i, t_j) + c(t_i, t_j, h_l, h_r)I_{(-\infty,0) \times [0,\infty)}(t_i, t_j) + c(t_i, t_j, h_r, h_l)I_{[0,\infty) \times (-\infty,0)}(t_i, t_j) + c(t_i, t_j, h_r, h_r)I_{[0,\infty) \times [0,\infty)}(t_i, t_j)$$

(21)

where

$$c(t_i, t_j, x, y) = \frac{\tau_{x}^{-1}(t_i)\tau_{y}^{-1}(t_j)}{t_i t_j (1 + W(x t_i^2))(1 + W(y t_j^2))} \phi_2(\tau_{x}^{-1}(t_i), \tau_{y}^{-1}(t_j), 0, R_2)$$

and it 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 [67] and the most important statistical software including R, MATLAB and Python.

Figure 3 depicts the contour plots of (21) when $\rho(h) = 0.2, 0.5, 0.9$ 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 $T_{h_l,h_r} = \{T_{h_l,h_r}, s \in A\}$ through a location and scale transformation:

$$T_{h_l,h_r}(s) =: \mu(s) + \sigma T_{h_l,h_r}^*(s)$$

(22)
where $\mu(s)$ is the location dependent mean and $\sigma > 0$ is a scale parameter. A typical parametric specification for the mean is given by $\mu(s) = X(s)^T \beta$ where $X(s) \in \mathbb{R}^k$ is a vector of covariates and $\beta \in \mathbb{R}^k$ 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 $\sigma$ to depend on the location $s$. 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:

$$f_{T_{h_l, h_r}, ij}(t_i, t_j) = \frac{1}{\sigma^2} f_{T_{h_l, h_r}^*, ij} \left( \frac{t_i - \mu(s_i)}{\sigma}, \frac{t_j - \mu(s_j)}{\sigma} \right).$$  \hspace{1cm} (23)
4 Simulation study

In this section, we consider two 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. The second simulation study (Subsection 4.2) 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, \(s_i \in A = [0, 1]^2, i = 1, \ldots, n\). 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 \(\mu(s)\) is considered spatially varying through a regression linear model that is:

\[
T_{h_l,h_r}(s) = \beta_0 + \beta_1 u(s) + \sigma T^*_l,h_r(s) \tag{24}
\]

where \(u(s)\) 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 \(\delta\) is assumed known and fixed. Then the vector parameters to be estimated are \(\theta = (\beta_0, \beta_1, \sigma^2, h_l, h_r, \alpha)^T\).

For \(wpl_a, a = M, C\) functions we consider the NNWCLP method by considering the nearest neighbours 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 MSE of the \(wpl_M\), estimates under the different scenarios when estimating \(\theta\) (we highlight the best MSE in bold for each scenario). It can be appreciated that \(wpl_M\) using NNWCLP outperforms DDWCLP in terms of MSE, 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 neighbours weight function select more informative pairs, as suggested in the example in Section 2.

We replicate the same experiment using the \( wpl_M \) function. Table 2 show bias and MSE under the different scenarios when estimating \( \theta \). Even in this case, it can be appreciated that maximization of the \( wpl_M \) function using NNWCLP overall outperforms \( wpl_M \) using DDWCLP in terms of MSE, for each scenario. In addition, as in the \( wpl_M \) case, the distribution of the \( wpl_M \) estimates are quite, symmetric, numerically stable contain with very few outliers for all scenarios. As an example, Figure 4 show the centred boxplots of the \( wpl_M \) estimates using both weight functions, for each parameter, for the example in Section 2.

<p>| Table 1 | Bias and MSE 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 = 0.05854, 0.05118, 0.07339, 0.10514 ). The underlying correlation function is ( \mathcal{W}_{0,3,5,0}(h) = (1 - h/(3,5a))^3.5 ) with ( a = 0.06 ). |
| ( h_l = 0.1 ) | ( h_l = 0.3 ) | ( h_l = 0.1 ) | ( h_l = 0.3 ) | ( h_l = 0.1 ) | ( h_l = 0.3 ) |</p>
<table>
<thead>
<tr>
<th>Bias</th>
<th>MSE</th>
<th>Bias</th>
<th>MSE</th>
<th>Bias</th>
<th>MSE</th>
<th>Bias</th>
<th>MSE</th>
<th>Bias</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m = 4 )</td>
<td>0.00016</td>
<td>0.00241</td>
<td>0.00024</td>
<td>0.00395</td>
<td>0.00026</td>
<td>0.00400</td>
<td>0.00028</td>
<td>0.00405</td>
<td>0.00030</td>
</tr>
<tr>
<td>( m = 8 )</td>
<td>0.00016</td>
<td>0.00241</td>
<td>0.00024</td>
<td>0.00395</td>
<td>0.00026</td>
<td>0.00400</td>
<td>0.00028</td>
<td>0.00405</td>
<td>0.00030</td>
</tr>
<tr>
<td>( m = 16 )</td>
<td>0.00016</td>
<td>0.00241</td>
<td>0.00024</td>
<td>0.00395</td>
<td>0.00026</td>
<td>0.00400</td>
<td>0.00028</td>
<td>0.00405</td>
<td>0.00030</td>
</tr>
<tr>
<td>( m = 32 )</td>
<td>0.00016</td>
<td>0.00241</td>
<td>0.00024</td>
<td>0.00395</td>
<td>0.00026</td>
<td>0.00400</td>
<td>0.00028</td>
<td>0.00405</td>
<td>0.00030</td>
</tr>
<tr>
<td>( m = 64 )</td>
<td>0.00016</td>
<td>0.00241</td>
<td>0.00024</td>
<td>0.00395</td>
<td>0.00026</td>
<td>0.00400</td>
<td>0.00028</td>
<td>0.00405</td>
<td>0.00030</td>
</tr>
<tr>
<td>( m = 128 )</td>
<td>0.00016</td>
<td>0.00241</td>
<td>0.00024</td>
<td>0.00395</td>
<td>0.00026</td>
<td>0.00400</td>
<td>0.00028</td>
<td>0.00405</td>
<td>0.00030</td>
</tr>
<tr>
<td>( m = 256 )</td>
<td>0.00016</td>
<td>0.00241</td>
<td>0.00024</td>
<td>0.00395</td>
<td>0.00026</td>
<td>0.00400</td>
<td>0.00028</td>
<td>0.00405</td>
<td>0.00030</td>
</tr>
<tr>
<td>( m = 512 )</td>
<td>0.00016</td>
<td>0.00241</td>
<td>0.00024</td>
<td>0.00395</td>
<td>0.00026</td>
<td>0.00400</td>
<td>0.00028</td>
<td>0.00405</td>
<td>0.00030</td>
</tr>
<tr>
<td>( m = 1024 )</td>
<td>0.00016</td>
<td>0.00241</td>
<td>0.00024</td>
<td>0.00395</td>
<td>0.00026</td>
<td>0.00400</td>
<td>0.00028</td>
<td>0.00405</td>
<td>0.00030</td>
</tr>
<tr>
<td>( m = 2048 )</td>
<td>0.00016</td>
<td>0.00241</td>
<td>0.00024</td>
<td>0.00395</td>
<td>0.00026</td>
<td>0.00400</td>
<td>0.00028</td>
<td>0.00405</td>
<td>0.00030</td>
</tr>
<tr>
<td>( m = 4096 )</td>
<td>0.00016</td>
<td>0.00241</td>
<td>0.00024</td>
<td>0.00395</td>
<td>0.00026</td>
<td>0.00400</td>
<td>0.00028</td>
<td>0.00405</td>
<td>0.00030</td>
</tr>
<tr>
<td>( m = 8192 )</td>
<td>0.00016</td>
<td>0.00241</td>
<td>0.00024</td>
<td>0.00395</td>
<td>0.00026</td>
<td>0.00400</td>
<td>0.00028</td>
<td>0.00405</td>
<td>0.00030</td>
</tr>
<tr>
<td>( m = 16384 )</td>
<td>0.00016</td>
<td>0.00241</td>
<td>0.00024</td>
<td>0.00395</td>
<td>0.00026</td>
<td>0.00400</td>
<td>0.00028</td>
<td>0.00405</td>
<td>0.00030</td>
</tr>
<tr>
<td>( m = 32768 )</td>
<td>0.00016</td>
<td>0.00241</td>
<td>0.00024</td>
<td>0.00395</td>
<td>0.00026</td>
<td>0.00400</td>
<td>0.00028</td>
<td>0.00405</td>
<td>0.00030</td>
</tr>
<tr>
<td>( m = 65536 )</td>
<td>0.00016</td>
<td>0.00241</td>
<td>0.00024</td>
<td>0.00395</td>
<td>0.00026</td>
<td>0.00400</td>
<td>0.00028</td>
<td>0.00405</td>
<td>0.00030</td>
</tr>
<tr>
<td>( m = 131072 )</td>
<td>0.00016</td>
<td>0.00241</td>
<td>0.00024</td>
<td>0.00395</td>
<td>0.00026</td>
<td>0.00400</td>
<td>0.00028</td>
<td>0.00405</td>
<td>0.00030</td>
</tr>
<tr>
<td>( m = 262144 )</td>
<td>0.00016</td>
<td>0.00241</td>
<td>0.00024</td>
<td>0.00395</td>
<td>0.00026</td>
<td>0.00400</td>
<td>0.00028</td>
<td>0.00405</td>
<td>0.00030</td>
</tr>
<tr>
<td>( m = 524288 )</td>
<td>0.00016</td>
<td>0.00241</td>
<td>0.00024</td>
<td>0.00395</td>
<td>0.00026</td>
<td>0.00400</td>
<td>0.00028</td>
<td>0.00405</td>
<td>0.00030</td>
</tr>
</tbody>
</table>

Table 2 shows bias and MSE under the different scenarios when estimating \( \theta \) using the \( wpl_M \) function.
considered, for each scenario, a measure of global relative efficiency [68], that is:

\[ \text{GRE} = \left( \frac{\det[F_{wpl_C}]}{\det[F_{wpl_M}]} \right)^{1/p}, \]  

where \( p = 6 \) is the number of unknown parameters in \( \theta \) and the matrix \( F_{wpl_a} \) is the sample mean squared error matrix \( F_{wpl_a} = 1000^{-1} \sum_{k=1}^{1000} \left( \hat{\theta}^a_k - \bar{\theta} \right) \left( \hat{\theta}^a_k - \bar{\theta} \right)' \) with \( \bar{\theta} = 1000^{-1} \sum_{k=1}^{1000} \hat{\theta}^a_k, a = M, C \). 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\% .

**Figure 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.
Table 2 Bias and MSE 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 \( h \) and \( r \), 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 \( \rho_{0,3,5}(L,(0,3,5)) = (1 - h/(3.5a))_+ \) with \( a = 0.06 \).

<table>
<thead>
<tr>
<th>( h )</th>
<th>( m = 2 )</th>
<th>( m = 4 )</th>
<th>( m = 8 )</th>
<th>( m = 16 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( h_1 = 0.1 )</td>
<td>Bias</td>
<td>MSE</td>
<td>Bias</td>
<td>MSE</td>
</tr>
<tr>
<td>( h_2 = 0.2 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( h_3 = 0.3 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| \( k = 0.03584 \) | | | | | | | | | | |
| \( k = 0.05118 \) | | | | | | | | | | |
| \( k = 0.07339 \) | | | | | | | | | | |
| \( k = 0.10514 \) | | | | | | | | | | |

Table 3 Global Relative efficiency (GRE) defined in equation (25) between \( wpl \)M and the weight function (5) (NNCLP) with \( m = 2, 4, 8, 16 \) and using the weight function (4) (DDCLP) with \( k = 0.03584, 0.05118, 0.07339, 0.10514 \), respectively, under different scenarios.

<table>
<thead>
<tr>
<th>( h_1 = 0.1 )</th>
<th>( h_1 = 0.2 )</th>
<th>( h_1 = 0.3 )</th>
<th>( h_1 = 0.4 )</th>
<th>( h_1 = 0.5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m = 2 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( k = 0.03584 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( k = 0.05118 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( k = 0.07339 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( k = 0.10514 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( m = 4 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( k = 0.03584 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( k = 0.05118 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( k = 0.07339 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( k = 0.10514 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( m = 8 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( k = 0.03584 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( k = 0.05118 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( k = 0.07339 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( k = 0.10514 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( m = 16 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( k = 0.03584 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( k = 0.05118 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( k = 0.07339 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( k = 0.10514 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Resuming our numerical experiments show that the \( wpl_a, a = M, C \) estimators using NNWCLP generally outperforms DDWCLP in terms of statistical efficiency and the best efficiencies are achieved with small values of \( m \). In addition the \( wpl_C \) estimator shows a better statistical efficiency with respect to the \( wpl_M \) estimator irrespective of the weight function.

### 4.2 Comparing NNWCLP with the Vecchia method for Gaussian random field estimation

The basic idea of the Vecchia method \([16, 17]\) 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.

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.

We consider a simulation setting with \( n = 5000 \) spatial points uniformly distributed on the unit square, \( s_i \in A = [0, 1]^2, i = 1, \ldots, n \). 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 \( G(s) = \mu + \sigma G^*(s) \) with \( \mu = 0, \sigma^2 = 1 \) and \( \rho_{G^*}(h) = M_{\nu, \alpha}(h) \). 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 approximately 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 \([51]\) that exploits a Fisher-scoring algorithm to find the maximum of the Vecchia approximation using a specific type of ordering (maximum-minimum ordering, see \([44]\) for details). We use the \texttt{fit_model} R function of the \texttt{GpGp} R package that implements the method of \([51]\) and the \texttt{GeoFit} R function in the \texttt{GeoModels} R package using a specified algorithm of optimization (we use the \texttt{nlminb} R function \([69]\) in our example). The starting values used are the same for both R functions.

Tables 4 and 5 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 MSE 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 MSE values across the different choices of \( m \) are reported in bold for each parameter.
Observing the global index, the Vecchia method is Gaussian RF with Matérn correlation and smoothness parameter 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).

<table>
<thead>
<tr>
<th>$\nu = 0.5$</th>
<th>$\nu = 1.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m = 2$</td>
<td>$m = 4$</td>
</tr>
<tr>
<td>$\mu$</td>
<td>NNWCLP</td>
</tr>
<tr>
<td></td>
<td>Vecchia</td>
</tr>
<tr>
<td></td>
<td>RE</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>NNWCLP</td>
</tr>
<tr>
<td></td>
<td>Vecchia</td>
</tr>
<tr>
<td></td>
<td>RE</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>NNWCLP</td>
</tr>
<tr>
<td></td>
<td>Vecchia</td>
</tr>
<tr>
<td></td>
<td>RE</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\text{det}[F]^{1/2}$</th>
<th>$\nu = 0.5$</th>
<th>$\nu = 1.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>m = 2</td>
<td>m = 4</td>
</tr>
<tr>
<td>$\mu$</td>
<td>NNWCLP</td>
<td>0.00107</td>
</tr>
<tr>
<td></td>
<td>Vecchia</td>
<td>0.00105</td>
</tr>
<tr>
<td></td>
<td>RE</td>
<td>0.97368</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\nu = 0.5$</th>
<th>$\nu = 1.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m = 2$</td>
<td>$m = 4$</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>NNWCLP</td>
</tr>
<tr>
<td></td>
<td>Vecchia</td>
</tr>
<tr>
<td></td>
<td>RE</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>NNWCLP</td>
</tr>
<tr>
<td></td>
<td>Vecchia</td>
</tr>
<tr>
<td></td>
<td>RE</td>
</tr>
<tr>
<td>$\nu$</td>
<td>NNWCLP</td>
</tr>
<tr>
<td></td>
<td>Vecchia</td>
</tr>
<tr>
<td></td>
<td>RE</td>
</tr>
<tr>
<td>$\text{det}[F]^{1/2}$</td>
<td>NNWCLP</td>
</tr>
<tr>
<td></td>
<td>Vecchia</td>
</tr>
<tr>
<td></td>
<td>RE</td>
</tr>
</tbody>
</table>

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. Observing the global index $\text{det}[F]^{1/\nu}$ ($p$ is the number of parameters involved) it can be appreciated that when estimating the smoothness parameter the best choice is $m = 8$ and when the smoothness parameter is fixed the best choice is $m = 2$.

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 \( \nu \).

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=0.973). However, a loss of global relative efficiency can be appreciated when \( m = 2 \) and \( \nu = 1.5 \) is assumed known (GRE=0.810). 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=0.882 if \( m = 4 \)). Nevertheless, when increasing the smoothness parameter the loss is more severe (GRE=0.376 if \( m = 2 \)). 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 compared with Vecchia method.

Resuming, 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 \texttt{system.time} function. For the comparison we have considered the \texttt{GPvecchia} R package \cite{GPvecchia}. Both methods requires a preliminary step. For instance in the Vecchia method a neighbourhood structure that depends on \( m \) (the size of the conditioning vector) and the type of ordering (we consider a maxmin order as proposed in \cite{vecchia}) 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 neighbours inside the kd-tree. The preliminary step can be performed with the \texttt{GeoNeighIndex} function of the \texttt{GeoModels} package for the NNWCLP method and the \texttt{vecchia Specify} function of the \texttt{GPvecchia} package for the Vecchia method. For the second and final step we compare the time needed for the evaluation of the \texttt{vecchia likelihood} R function that implements a likelihood approximation based on Vecchia method versus the evaluation of the NNWCLP function.

Table 6 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.

Focusing on small neighbours, 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 a dataset of size 1200000.
Table 6  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 $m = 2, 4, 8, 16$.

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

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 analysing massive datasets.

5 Application

We consider data from the ERA5-Land product [52] 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 that can be downloaded from https://developers.google.com/earth-engine/datasets/catalog/ECMWF ERA5 LAND HOURLY.

In particular, we focus on the hourly mean temperature of the surface of the Earth over the first two months of 2020 observed at 363827 locations, in the region delimited by longitude (in decimal degree) (-85,-34) and latitude (-40,12) which correspond to the greater part of South-America.
Following [71], 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. Figure (6), from left to right, depicts the coloured 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,
   \[ G(s) = \mu + \sigma G^*(s). \]

2. a Tukey-\(h\) RF,
   \[ T_h(s) = \mu + \sigma T^*_h(s), \quad h \in [0, 0.5). \]

   where \(T^*_h\) is a Tukey-\(h\) RF defined in Equation (6).

3. a Tukey-\(hh\) RF
   \[ T_{h_l,h_r}(s) = \mu + \sigma T^*_{h_l,h_r}(s), \quad h_l, h_r \in [0, 0.5). \]

   where \(T^*_{h_l,h_r}\) is a Tukey-\(hh\) RF defined in Equation (11).

As underlying correlation model for the RFs \(G^*, T^*_h\) and \(T^*_{h_l,h_r}\) we consider a Matérn correlation model \(\rho_{G^*}(h) = M_{\kappa,\alpha}(h)\). 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 7 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-\(hh\) random field.

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 neighbourhoods and using the estimation of the correlation functions \(\rho_{G^*}(h)\), \(\rho_{T^*_h(h)}\) and \(\rho_{T^*_{h_l,h_r}(h)}\) respectively. Clearly, the best (local) linear predictor is optimal (in the mean square sense) for the Gaussian case and linear optimal for Tukey’s RFs. Table 7, depicts the empirical mean of the 100 RMSEs obtained. It turns out that the model with the best prediction performance is the proposed Tukey-\(hh\) random field. In addition the Tukey-\(hh\) random field estimated using NNWCLP with \(m = 8\) achieves the best prediction performance.
Figure 5 From left to right: coloured map, normalized histogram and empirical semivariogram of the land surface temperature data residuals.

Table 7 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.

<table>
<thead>
<tr>
<th></th>
<th>$\mu$</th>
<th>$\sigma^2$</th>
<th>$h$</th>
<th>$h_l$</th>
<th>$h_r$</th>
<th>$\alpha$</th>
<th>$\nu$</th>
<th>PLIC</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m = 8$ Gaussian</td>
<td>0.0234</td>
<td>13.8484</td>
<td>1.269</td>
<td></td>
<td></td>
<td>158.96</td>
<td>0.6694</td>
<td>6094235</td>
<td>0.33187</td>
</tr>
<tr>
<td></td>
<td>(0.4461)</td>
<td></td>
<td>(1.269)</td>
<td></td>
<td></td>
<td>(13.705)</td>
<td>(0.0045)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$m = 8$ Tukey-$h$</td>
<td>0.9458</td>
<td>7.0187</td>
<td>0.2120</td>
<td></td>
<td></td>
<td>165.70</td>
<td>0.7152</td>
<td>5337899</td>
<td>0.33077</td>
</tr>
<tr>
<td></td>
<td>(0.0269)</td>
<td></td>
<td>(0.0210)</td>
<td></td>
<td></td>
<td>(13.042)</td>
<td>(0.0054)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$m = 8$ Tukey-$hh$</td>
<td>0.4620</td>
<td>6.7776</td>
<td>0.1349</td>
<td>0.2584</td>
<td></td>
<td>158.82</td>
<td>0.7155</td>
<td>5331609</td>
<td>0.33075</td>
</tr>
<tr>
<td></td>
<td>(0.0575)</td>
<td></td>
<td>(0.0129)</td>
<td>(0.0258)</td>
<td></td>
<td>(11.485)</td>
<td>(0.0062)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$m = 16$ Gaussian</td>
<td>0.0247</td>
<td>13.812</td>
<td>1.350</td>
<td></td>
<td></td>
<td>186.41</td>
<td>0.6565</td>
<td>14303492</td>
<td>0.33220</td>
</tr>
<tr>
<td></td>
<td>(0.4490)</td>
<td></td>
<td>(1.350)</td>
<td></td>
<td></td>
<td>(15.180)</td>
<td>(0.0042)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$m = 16$ Tukey-$h$</td>
<td>0.9599</td>
<td>6.8732</td>
<td>0.2196</td>
<td></td>
<td></td>
<td>207.57</td>
<td>0.6581</td>
<td>12751849</td>
<td>0.33211</td>
</tr>
<tr>
<td></td>
<td>(0.0309)</td>
<td></td>
<td>(0.0249)</td>
<td></td>
<td></td>
<td>(19.042)</td>
<td>(0.0043)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$m = 16$ Tukey-$hh$</td>
<td>0.4783</td>
<td>6.6382</td>
<td>0.1382</td>
<td>0.2681</td>
<td></td>
<td>197.39</td>
<td>0.6592</td>
<td>12738045</td>
<td>0.33207</td>
</tr>
<tr>
<td></td>
<td>(0.0755)</td>
<td></td>
<td>(0.0150)</td>
<td>(0.0291)</td>
<td></td>
<td>(16.655)</td>
<td>(0.0045)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 neighbours that we have proposed in this paper has been shown to be an effective solution when estimating (non-)Gaussian RFs. 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 when estimating the parameters of the proposed Tukey-$hh$ RF. 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 neighbours involved) is typically between 2 and 8, 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 neighbourhood.
A.1 Proof Lemma 1

Proof In the proof we make use of some special functions such as the hypergeometric Gaussian function $\gamma F_1(a; b; c; x)$, the confluent hypergeometric function $\gamma F_1(a; b; x)$ and the parabolic cylinder function $D_\nu (x)$ (see [66] for their definitions). Setting, $\mathbb{E}(T_i^* T_j^*) = \mathbb{E}(T_i^*; T_j^*)$ with $h = s_i - s_j$ we have:

$$\begin{align*}
\mathbb{E}(T_i^* T_j^*) &= \mathbb{E}\left[G(s_i) G(s_j) e^{\frac{h_r(G(s_i))^2}{2} + \frac{h_r(G(s_j))^2}{2}}\right] + \mathbb{E}\left[G(s_i) G(s_j) e^{\frac{h_l(G(s_i))^2}{2} + \frac{h_l(G(s_j))^2}{2}}\right] \\
&= \frac{1}{2\pi(1 - \rho^2(h))^1/2} \int_{\mathbb{R}^2_+} g_i g_j e^{-\frac{1}{2(1 - \rho^2(h))}\left[g_i^2 + g_j^2 - 2\rho(h)g_i g_j\right]} \left[\frac{h_r g_i^2}{2} + \frac{h_l g_j^2}{2}\right] dg_i dg_j \\
&+ \frac{1}{2\pi(1 - \rho^2(h))^1/2} \int_{\mathbb{R}^2_+} g_i g_j e^{-\frac{1}{2(1 - \rho^2(h))}\left[g_i^2 + g_j^2 - 2\rho(h)g_i g_j\right]} \left[\frac{h_l g_i^2}{2} + \frac{h_r g_j^2}{2}\right] dg_i dg_j \\
&+ \frac{1}{2\pi(1 - \rho^2(h))^1/2} \int_{\mathbb{R}^2_+} g_i g_j e^{-\frac{1}{2(1 - \rho^2(h))}\left[g_i^2 + g_j^2 - 2\rho(h)g_i g_j\right]} \left[\frac{h_r g_i^2}{2} + \frac{h_l g_j^2}{2}\right] dg_i dg_j \\
&= A_1 + A_2 + A_3 + A_4. \quad (A1)
\end{align*}$$

Taking the first integral $A_1$ and using (3.462.1) of [66], we obtain,

$$\begin{align*}
A_1 &= \frac{1}{2\pi(1 - \rho^2(h))^1/2} \int_{\mathbb{R}^2_+} g_j e^{-\frac{1}{2(1 - \rho^2(h))}\left[g_j^2 + \frac{(1 - (1 - \rho^2(h))h_r)g_j^2}{2(1 - \rho^2(h))} + \frac{\rho(h)g_j^2}{1 - (1 - \rho^2(h))h_r}\right]} dg_j \\
&= \frac{1}{2\pi(1 - \rho^2(h))^1/2} \left[\frac{(1 - \rho^2(h))}{1 - (1 - \rho^2(h))h_r}\right] \int_{\mathbb{R}^2_+} g_j e^{-\frac{\rho^2(h)}{2(1 - \rho^2(h))} + \frac{\rho^2(h)}{4(1 - \rho^2(h))[1 - (1 - \rho^2(h))h_r]} g_j^2} dg_j \\
\times D_{-2} \left(-\frac{\rho(h) g_j}{\sqrt{(1 - \rho^2(h))[1 - (1 - \rho^2(h)) h_r]}}\right) dg_j \quad (A2)
\end{align*}$$

Now, considering (9.240) of [66]:

$$D_{-2} \left(-\frac{\rho(h) g_j}{\sqrt{(1 - \rho^2(h))[1 - (1 - \rho^2(h)) h_r]}}\right) = e^{-\frac{\rho^2(h) g_j^2}{4(1 - \rho^2(h))[1 - (1 - \rho^2(h)) h_r]}}$$

Acknowledgements. Moreno Bevilacqua acknowledges financial support from Grant FONDECYT 1200068 and ANID/PIA/ANILLOS ACT210096, Chile from the Chilean government. The work of Christian Caamaño-Carrillo was partially supported by grant FONDECYT 11220066 from the Chilean government and DIUBB 2120538 IF/R from the University of Bío-Bío. Victor Morales-Oñate acknowledges funding 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.
Proof

\[ A_1 = \frac{(1 - \rho^2(h))^{1/2}}{2\pi[1 - (1 - \rho^2(h))h_r]} \int_{\mathbb{R}_+} g_j e^{-\frac{[1 - (1 - \rho^2(h))h_r]g_j^2}{2(1 - \rho^2(h))}} \, dg_j \]
\[ + \frac{\sqrt{2\pi} \rho(h)}{4\pi[1 - (1 - \rho^2(h))h_r]^{3/2}} \int_{\mathbb{R}_+} g_j^2 e^{-\frac{[1 - (1 - \rho^2(h))h_r]g_j^2}{2(1 - \rho^2(h))}} \, dg_j \]
\[ = \frac{(1 - \rho^2(h))^{3/2}}{2\pi[1 - (1 - \rho^2(h))h_r]^{3/2}} 2F_1 \left( \frac{1 - 1/2}{1/2}; \frac{\rho^2(h)}{[1 - (1 - \rho^2(h))h_r]^2} \right) \]
\[ + \frac{\rho(h)(1 - \rho^2(h))^{3/2}}{4[1 - (1 - \rho^2(h))h_r]^{3/2}} 2F_1 \left( \frac{3 - 3/2}{2 - 3/2}; \frac{\rho^2(h)}{[1 - (1 - \rho^2(h))h_r]^2} \right) \]

(A4)

Using Euler transformation and the identity \( 2F_1 \left( \frac{3}{2}, \frac{3}{2}, x \right) = (1 - x)^{-3/2} \), we obtain

\[ A_1 = \frac{[1 - (1 - \rho^2(h))h_r]}{2\pi[(1 - h_r^2 - \rho^2(h))h_r^2]^{3/2}} 2F_1 \left( \frac{-1}{2} - \frac{1}{2}, \frac{1}{2}; \frac{\rho^2(h)}{[1 - (1 - \rho^2(h))h_r]^2} \right) + \frac{\rho(h)}{4[(1 - h_r^2 - \rho^2(h))h_r^2]^{3/2}} \]

(A5)

Similarly, \( A_2, A_3 \) and \( A_4 \) in (A1), are given by:

\[ A_2 = \frac{[1 - (1 - \rho^2(h))h_1]}{2\pi[(1 - h_1^2 - \rho^2(h))h_1^2]^{3/2}} 2F_1 \left( \frac{-1}{2} - \frac{1}{2}, \frac{1}{2}; \frac{\rho^2(h)}{[1 - (1 - \rho^2(h))h_1]^2} \right) + \frac{\rho(h)}{4[(1 - h_1^2 - \rho^2(h))h_1^2]^{3/2}} \]

(A5)

\[ A_3 = -\frac{[(1 - (1 - \rho^2(h))h_1)][1 - (1 - \rho^2(h))h_1)]^{1/2}}{2\pi[h_r - h_1 - (1 - \rho^2(h))h_rh_1]^{3/2}} 2F_1 \left( \frac{-1}{2} - \frac{1}{2}, \frac{1}{2}; \frac{\rho^2(h)}{[1 - (1 - \rho^2(h))h_r][1 - (1 - \rho^2(h))h_1]} \right) \]

\[ + \frac{\rho(h)}{4[1 - h_r - h_1 - (1 - \rho^2(h))h_rh_1]^{3/2}} \]

\[ A_4 = -\frac{[(1 - (1 - \rho^2(h))h_1)][1 - (1 - \rho^2(h))h_1)]^{1/2}}{2\pi[h_r - h_1 - (1 - \rho^2(h))h_rh_1]^{3/2}} 2F_1 \left( \frac{-1}{2} - \frac{1}{2}, \frac{1}{2}; \frac{\rho^2(h)}{[1 - (1 - \rho^2(h))h_r][1 - (1 - \rho^2(h))h_1]} \right) \]

\[ + \frac{\rho(h)}{4[1 - h_r - h_1 - (1 - \rho^2(h))h_rh_1]^{3/2}} \]

Finally, combining equations (A5) and (A6), 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 \( \mathbb{E}(T^*_{h_1,h_r}(s_1)T^*_{h_1,h_r}(s_2)) \).

\[ \square \]

A.2 Proof Theorem 1

Proof Consider \( T^* = (T^*(s_1), \ldots, T^*(s_n))^T \), where \( z = (z(s_1), \ldots, z(s_n))^T \sim N_n(0, R_n) \). In addition, consider the diagonal matrices \( D(\ell) = \text{diag}(\ell_1, \ldots, \ell_n) \), with \( \ell \in \{-1, 1\}^n \), which are such that \( D(\ell)^2 \) is the identity matrix. Since \( \tau^{-1}(h, t) = D(\ell) \circ g(h, t) \) (\( \circ \) is the Hadamard product) with \( g(h, t) = \frac{\rho^2(h)g_j^2}{2(1 - \rho^2(h))[1 - (1 - \rho^2(h))h_r]} \)}
\[
\left( \frac{W(ht_1^2)}{h}, \ldots, \frac{W(ht_n^2)}{h} \right)^{1/2} \]
where \( h \in \{h_1, h_r\} \), we then have
\[
F_{T^*}(t) = P(T^* \leq t) = \sum_{\ell \in \{-1,+1\}^n} \Phi_n(D(\ell) \circ g(h,t); \ 0, R_n) \quad (A7)
\]
Hence, by using the following relation
\[
\frac{\partial^n \Phi_n(D(\ell) \circ g(h,t); \ 0, R_n)}{\partial t_1, \ldots, \partial t_n} = D(\ell) \cdot \phi_n(D(\ell) \circ g(h,t); \ 0, R_n)
\]
\[
\cdot \frac{\partial^n g(h,t)}{\partial t_1, \ldots, \partial t_n} I_m \prod_{i=1}^m (-\infty, 0)^i \prod_{j=1}^{n-m} (0, \infty)^j (t_1, \ldots, t_n) \quad (A8)
\]
where \( I(A) \) denote the indicator function of the set \( A \), with \( m \) negative components and \( n - m \) positive components of vector \( t \).
\[
\frac{\partial^n g(h,t)}{\partial t_1, \ldots, \partial t_n} = g(h,t)\{\text{diag}(t_1, \ldots, t_n)[1 + h \cdot \text{diag}(g(h,t_1), \ldots, g(h,t_n))g(h,t)]\}^{-1}
\]
we find that the joint density of \( T^* \).
\[
f_{T^*_{h_1, h_r}}(t) = D(\ell) \circ g(h,t)\{\text{diag}(t_1, \ldots, t_n)[1 + h \cdot \text{diag}(g(h,t_1), \ldots, g(h,t_n))g(h,t)]\}^{-1}
\]
\[
\cdot \phi_n(D(\ell) \circ \tau(h,t); \ 0, R_n) \quad (A9)
\]
\[
f_{T^*_{h_1, h_r}}(t) = \sum_{\ell \in \{-1,+1\}^n} \frac{\prod_{i=1}^m t_i^{-1}(t_i)}{\prod_{i=1}^m (1 + W(ht_i^2))} \frac{\prod_{j=m+1}^n t_j^{-1}(t_j)}{\prod_{j=m+1}^n (1 + W(ht_j^2))}
\]
\[
\cdot \phi_n(\tau_h^{-1}(t), \ 0, R_n) I_m \prod_{i=1}^m (-\infty, 0)^i \prod_{j=1}^{n-m} (0, \infty)^j (t_1, \ldots, t_n) \quad (A10)
\]
\[
\text{where } I_A(x) \text{ denotes the indicator function of the set } A, \ R_n = [\rho(s_i - s_j)]_{i,j=1}^n, \text{ and the diagonal matrices } D(\ell) = \text{Diag}\{\ell_1, \ldots, \ell_n\}, \text{ with } \ell = (\ell_1, \ldots, \ell_n) \in \{-1,+1\}^n
\]
where \( \tau_h^{-1}(t) = D(\ell) \circ g(h,t) \) with \( h \in \{h_1, h_r\} \).

\[\Box\]

References


[29] Bevilacqua, M., Caamaño-Carrillo, C., Gaetan, C.: On modeling positive
continuous data with spatiotemporal dependence. Environmetrics 31(7), 2632 (2020)


Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- codehh.r