UNIVERSITÀ
DI SIENA
1240

QUADERNI DEL DIPARTIMENTO

DI ECONOMIA POLITICA E STATISTICA

Víctor Morales-Oñate
Federico Crudu
Moreno Bevilacqua

Blockwise Euclidean likelihood for spatio-temporal covariance models

n. 822 – Marzo 2020

Blockwise Euclidean likelihood for spatio-temporal covariance models

Víctor Morales-Oñate*
Banco Solidario

Federico Crudu†
Università di Siena
CRENoS

Moreno Bevilacqua ‡
Universidad de Valparaíso
Millennium Nucleus Center for the Discovery of Structures in Complex Data

Abstract

In this paper we propose a spatio-temporal blockwise Euclidean likelihood method for the estimation of covariance models when dealing with large spatio-temporal Gaussian data. The method uses moment conditions coming from the score of the pairwise composite likelihood. The blockwise approach guarantees considerable computational improvements over the standard pairwise composite likelihood method. In order to further speed up computation we consider a general purpose graphics processing unit implementation using OpenCL. We derive the asymptotic properties of the proposed estimator and we illustrate the finite sample properties of our methodology by means of a simulation study highlighting the computational gains of the OpenCL graphics processing unit implementation. Finally, we apply our estimation method to a wind component data set.

Keywords: Composite likelihood; Euclidean likelihood; Gaussian random fields; Parallel computing; OpenCL

*Risk Division, Data Analytics, Quito, Ecuador, victor.morales@uv.cl

†Department of Economics and Statistics, University of Siena, Piazza San Francesco, 7/8 53100 Siena, Italy, federico.crudu@unisi.it

‡Department of Statistics, Avenida Gran Bretaa 1111 Playa Ancha, Valparaso, Chile, moreno.bevilacqua@uv.cl

1 Introduction

With the advent and expansion of Geographical Information Systems (GIS) along with related software, statisticians today routinely encounter large spatial or spatio-temporal data sets containing one or multiple variables observed across a large number of location sites. This has generated considerable interest in statistical modeling for large geo-referenced spatial and spatio-temporal data; see, for instance, Sherman (2011) and Cressie & Wikle (2015).

Gaussian random fields (RFs) are the cornerstone for this kind of analysis and have been largely used in the past years thanks to a well developed and rich theory. Moreover, they represent the building block for more sophisticated models or non-Gaussian RFs (see, for instance, De Oliveira et al. (1997), Xu & Genton (2017) and Bevilacqua et al. (2020)). The covariance function is a crucial object in Gaussian RF analysis. It is well known, in fact, that, together with the mean, the covariance function completely characterizes the finite dimensional distribution of the RF. Furthermore, it is also well known that the spatio-temporal kriging predictor depends on the knowledge of such covariance function.

Since a covariance function must be positive definite, practical estimation generally requires the selection of some parametric classes of covariances and the corresponding estimation of these parameters. The maximum likelihood method is generally considered the best option for estimating the covariance model parameters. Nevertheless, the evaluation of the objective function under the Gaussian assumption requires the solution of a system of linear equations. For a Gaussian RF observed in n spatio-temporal locations the computational burden is $O(n^3)$, making this method computationally impractical for large data sets. This fact motivates the search for estimation methods with a good balance between computational complexity and statistical efficiency.

Some solutions have been proposed involving approximations of the covariance matrix (Kaufman et al., 2008; Cressie & Johannesson, 2008; Litvinenko et al., 2017), stochastic approximations of the score function (Stein et al., 2013) or approximations based on Markov

random fields (RFs) (Rue & Tjelmeland, 2002; Rue & Held, 2005; Lindgren et al., 2011) or Gaussian predictive process (Banerjee et al., 2008) or on composite likelihood idea (Bevilacqua et al., 2012; Bevilacqua & Gaetan, 2015; Eidsvik et al., 2014; Stein et al., 2004; Bai et al., 2012) among others. For an extensive review see Heaton et al. (2019) and the references therein.

The concept of composite likelihood (CL) refers to a general class of objective functions based on the likelihood of marginal or conditional events (see Lindsay, 1988; Varin et al., 2011, for a recent review). This kind of estimation method has two important features: first, it is generally an appealing estimation method when dealing with large data sets; second, it can be helpful when the specification of the likelihood is difficult. As outlined in Bevilacqua & Gaetan (2015) the class of CL functions is very large and, to the best of our knowledge, there are no clear guidelines on how to chose a specific member of this class for a given estimation problem. In the Gaussian case, if the choice of the CL is driven by computational concerns, the CL based on pairs has clear computational advantages with respect to other types of CL functions.

In a purely spatial context, Bevilacqua et al. (2015) propose a blockwise Euclidean likelihood (EU) method (Antoine et al., 2007; Owen, 2001) for the estimation of a latent Gaussian RF when considering binary data. The moment conditions used in the EU estimator derive from the score function of the CL based on marginal pairs. A feature of this approach is that it is possible to obtain computational benefits over the standard pairwise likelihood depending on the choice of the spatial blocks.

The main advantage of EU estimators is due to their computational simplicity. While similar estimators, such as the empirical likelihood estimator and the exponential tilting estimator (see, e.g.: Nordman & Caragea, 2008; Newey & Smith, 2004; Kitamura, 1997; Qin & Lawless, 1994), are computed via the solution of complicated optimization problems in the parameter of interest and an auxiliary parameter vector, EU estimators are characterized by a closed form solution for the auxiliary parameter and a simple optimization problem based on a quadratic form. This structure makes the EU estimator particularly

appealing for the problem we want to tackle.

The goal of the paper is to modify and extend the approach in Bevilacqua et al. (2015) to the spatio-temporal context and Gaussian data. This generalization implies the construction of (possibly overlapping) spatio-temporal blocks. Different types of blocks should be considered depending on the type of data. For instance, for a few location sites observed in a large number of temporal instants, the use of temporal blocks is the natural choice. The asymptotic properties of the proposed estimator are established under increasing domain asymptotics.

Since the proposed method is highly amenable to parallelization, we reduce the computational complexity by considering an implementation based on the OpenCL language (Stone et al., 2010) in a general purpose graphical processing unit (GPGPU) framework (A. Lee et al., 2010; Suchard et al., 2010). This allows to considerably reduce the computational costs associated to the blockwise EU estimation of the spatio-temporal covariance model.

The remainder of the paper is organized as follows. In Section 2, we introduce the concept of spatio-temporal RF and the pairwise likelihood estimation method. In Section 3, we introduce the blockwise spatio-temporal EU method and we establish the associated asymptotic properties. In Section 4, we investigate the performance of the spatio-temporal blockwise EU estimator in terms of statistical and computational efficiency highlighting the gains induced by the graphics processing unit (GPU) parallelization. In Section 5, we apply our methodology to a data set on Mediterranean wind speed. Finally, in Section 6 we give some conclusions.

2 Spatio-temporal pairwise likelihood

Let $\boldsymbol{l} = (\boldsymbol{s}^\top, t)^\top$ denote a generic spatio-temporal index with $\boldsymbol{l} \in \mathcal{L} = \mathcal{S} \times \mathcal{T}$ with $\mathcal{S} \subset \mathbb{R}^d$ and $\mathcal{T} \subset \mathbb{R}^+$ being our sampling region, and let $Z = \{Z_{\boldsymbol{l}}, \quad \boldsymbol{l} \in \mathcal{L}\}$ be a real-valued spatio-temporal RF (STRF) defined on $\mathcal{L}$. When $\mathcal{T} = \{t_0\}$ then $\mathcal{L} \equiv \mathcal{S}$ and $Z_{\boldsymbol{s}} \equiv Z_{(\boldsymbol{s}^\top, t_0)^\top}$

is a purely spatial RF. When $\mathcal{S}=\{\boldsymbol{s}_0\}$ then $\mathcal{L}\equiv\mathcal{T}$ and $Z_t\equiv Z_{(\boldsymbol{s}_0^\top,t)^\top}$ is a purely temporal RF. The high order of complexity of spatio-temporal interactions calls for simplifying assumptions, such as those of intrinsic or weak stationarity, that have implications on the existence of the moments of the RF.

A STRF Z is second-order (weakly stationary) if $\mathrm{E}[Z_{\boldsymbol{l}}]=\mu$ and $\mathrm{Var}[Z_{\boldsymbol{l}}]=\sigma^2$ are finite constants for all $\boldsymbol{l}\in\mathcal{L}$ and the covariance $\mathrm{Cov}[Z_{\boldsymbol{l}},Z_{\boldsymbol{l}'}]=C(\boldsymbol{h},u)=\sigma^2\rho(\boldsymbol{h},u)$ with $\rho(\cdot,\cdot)$ a positive definite function such that $\rho(\mathbf{0},0)=1$ that only depends on $\boldsymbol{h}=\boldsymbol{s}'-\boldsymbol{s}$ and $u=t'-t$. Isotropy is another very common assumption and also the building block for more sophisticated models. Isotropic spatial RFs have the feature that, for a candidate correlation function $\phi:[0,\infty)\to\mathbb{R}$ and given $\boldsymbol{s}'$, $\boldsymbol{s}$, two arbitrary location sites in $\mathcal{S}$, the correlation function solely depends on the Euclidean distance (denoted $\|\cdot\|$ throughout) that is $\rho(\boldsymbol{h})=\phi(\|\boldsymbol{h}\|)$. Spatio-temporal modeling inherits the assumption of spatial isotropy coupling, through a continuous function, spatial isotropy with temporal symmetry. This is, $\phi:[0,\infty)\times[0,\infty)\to\mathbb{R}$, with $\phi(0,0)=1$, such that $\rho(\boldsymbol{h},u)=\phi\left(\|\boldsymbol{h}\|,|u|\right)$.[1]

In the past years, many parametric models have been proposed in order to model the covariance function of a Gaussian STRF. A possible simple construction is obtained as the product of any valid isotropic spatial and temporal symmetric covariance as for instance:

$$C(\boldsymbol{h},u,\boldsymbol{\theta})=\sigma^2\exp\left(-\frac{||\boldsymbol{h}||}{\alpha_s}-\frac{|u|}{\alpha_t}\right), \tag{1}$$

where $\boldsymbol{\theta}=(\sigma^2,\alpha_s,\alpha_t)^\top$. Here α_s and α_t are positive spatial and temporal scale parameters respectively. This kind of covariance model, called *separable* model, has been criticized for its lack of flexibility. For such a reason, different classes of *non separable* covariance models have been proposed, in order to capture possible spatio-temporal interactions. A special

[1]We will use the notation $|\cdot|$ to indicate both the cardinality of a set and the absolute value of a scalar. Hence, for a generic set $\mathcal{A}$, $|\mathcal{A}|$ is its cardinality, while for a generic scalar a, $|a|$ is its absolute value. The different notation for sets and scalars avoids any potential confusion.

case of the celebrated Gneiting class (Gneiting, 2002) is given by:

$$C(\boldsymbol{h}, u, \boldsymbol{\theta}) = \frac{\sigma^2}{(1 + |u|/\alpha_t)} e^{-\frac{\|\boldsymbol{h}\|}{\alpha_s (1+|u|/\alpha_t)^{\beta/2}}}, \tag{2}$$

where $\boldsymbol{\theta} = (\sigma^2, \alpha_s, \alpha_t, \beta)^\top$. In this case, the parameter $\beta \in [0, 1]$ is a (non) separability parameter. When $\beta = 0$ the covariance model is separable.

Let us assume that $\boldsymbol{z} = \{z_{\boldsymbol{l}_1}, \ldots, z_{\boldsymbol{l}_n}\}^\top$ is a realization of Z and define $\ell_{ij}(\boldsymbol{\theta}) \equiv \log(f_{\boldsymbol{Z}_{ij}}(\boldsymbol{z}_{ij}), \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^{d_{\boldsymbol{\theta}}}$, the loglikelihood associated to the Gaussian bivariate distribution random vector $\boldsymbol{Z}_{ij} = (Z_{\boldsymbol{l}_i}, Z_{\boldsymbol{l}_j})^\top$. The pairwise weighted composite likelihood objective function is then given by

$$pl(\boldsymbol{\theta}) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \ell_{ij}(\boldsymbol{\theta}) w_{ij}, \tag{3}$$

where w_{ij} are suitable positive weights not depending on $\boldsymbol{\theta}$. Then the maximum pairwise weighted composite likelihood estimator is given by $\widehat{\boldsymbol{\theta}}_{PL} = \operatorname{argmax}_{\boldsymbol{\theta} \in \Theta} pl(\boldsymbol{\theta})$.

A distinctive feature of $pl(\boldsymbol{\theta})$ is that the associated estimating function,

$$\nabla pl(\boldsymbol{\theta}) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \nabla \ell_{ij}(\boldsymbol{\theta}) w_{ij},$$

where ∇ denotes the vector differential operator with respect to $\boldsymbol{\theta}$, is unbiased. Let us then define $\boldsymbol{g}_{ij}(\boldsymbol{\theta}) := \nabla \ell_{ij}(\boldsymbol{\theta}) w_{ij}$. Hence,

$$\mathrm{E}[\boldsymbol{g}_{ij}(\boldsymbol{\theta}_0)] = \boldsymbol{0} \tag{4}$$

where $\boldsymbol{\theta}_0$ is unique. Moreover, $\widehat{\boldsymbol{\theta}}_{PL}$ is consistent and its asymptotic distribution, under increasing domain asymptotics, is Gaussian with asymptotic covariance matrix given by $\boldsymbol{G}(\boldsymbol{\theta})^{-1} = \boldsymbol{H}(\boldsymbol{\theta})^{-1} \boldsymbol{J}(\boldsymbol{\theta}) \boldsymbol{H}(\boldsymbol{\theta})^{-1^\top}$ where $\boldsymbol{G}(\boldsymbol{\theta})$ is the Godambe information matrix and $\boldsymbol{H}(\boldsymbol{\theta}) = -\mathrm{E}[\nabla^2 pl(\boldsymbol{\theta})]$, $\boldsymbol{J}(\boldsymbol{\theta}) = \mathrm{E}[\nabla pl(\boldsymbol{\theta}) \nabla pl(\boldsymbol{\theta})^\top]$ (Bevilacqua et al., 2012).

The role of the weights w_{ij} in Equations (3) and (4) is to reduce computational time and

to improve the statistical efficiency of the estimator. As shown in Joe & Lee (2009), Davis & Yau (2011) and Bevilacqua & Gaetan (2015), compactly supported weight functions depending on fixed spatial or spatio-temporal distance, i.e.

$$w_{ij} = \begin{cases} 1 & \|\boldsymbol{s}_i - \boldsymbol{s}_j\| \leq d_s, |t_i - t_j| < d_t, \\ 0 & \text{otherwise} \end{cases}, \tag{5}$$

can significantly improve both the statistical efficiency and the computational complexity of the estimation method.

3 Spatio-temporal blockwise Euclidean likelihood

In what follows we introduce the spatio-temporal blockwise EU (STBEU) under a general spatio-temporal framework for both evenly and unevenly spaced lattice. A similar framework has been considered in Bai et al. (2012) and Nordman & Caragea (2008). The approach is not exactly the same as that of Bevilacqua et al. (2012) and exploits the limiting results of Jenish & Prucha (2009) for RFs.

Let us construct the blockwise version of the moment conditions described in Equation (4). Let $\mathcal{L} \subset \mathbb{R}^d \times \mathbb{R}^+$ be our sampling region, where the generic element $\boldsymbol{l} = (\boldsymbol{s}^\top, t)^\top$ includes both the spatial index and the time index and consider a block length b_n where $b_n^{-1} + \frac{b_n^{2(1+d)}}{n} \to 0$ as $n \to \infty$ and a set $\mathcal{U} = \left(-\frac{1}{2}, \frac{1}{2}\right]^d \times (0, 1]$. Then, a $(1+d)$-dimensional block is defined as

$$\mathcal{B}_{b_n}(\boldsymbol{\kappa}) = \boldsymbol{\kappa} + b_n \mathcal{U}$$

while the associated index set is defined as

$$\mathcal{K}_{b_n} = \{\boldsymbol{\kappa} : \mathcal{B}_{b_n}(\boldsymbol{\kappa}) \subset \mathcal{L}\}$$

with $\boldsymbol{\kappa} \in \mathbb{R}^d \times \mathbb{R}^+$ and $N = |\mathcal{K}_{b_n}|$, the number of blocks. The blockwise version of Equation (4) is

$$\mathrm{E}[\boldsymbol{m}_{\boldsymbol{\kappa}}(\boldsymbol{\theta}_0)] = \mathbf{0} \tag{6}$$

where, for $\mathcal{D}_{b_n}(i, j, \boldsymbol{\kappa}) = \left\{(i, j) : (\boldsymbol{l}_i, \boldsymbol{l}_j) \in \mathcal{B}_{b_n}(\boldsymbol{\kappa}) \cap \mathbb{R}^d \times \mathbb{R}^+\right\}$ and $b_n^{1+d} = |\mathcal{D}_{b_n}|$,

$$\boldsymbol{m}_{\boldsymbol{\kappa}}(\boldsymbol{\theta}) = \frac{1}{b_n^{1+d}} \sum_{\{i,j\} \in \mathcal{D}_{b_n}(i,j,\boldsymbol{\kappa})} \boldsymbol{g}_{ij}(\boldsymbol{\theta})$$

and

$$\widehat{\boldsymbol{m}}(\boldsymbol{\theta}) = \frac{1}{N} \sum_{\boldsymbol{\kappa} \in \mathcal{K}_{b_n}} \boldsymbol{m}_{\boldsymbol{\kappa}}(\boldsymbol{\theta}).$$

The STBEU objective function is defined as

$$R_n(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \frac{1}{2} \sum_{\boldsymbol{\kappa} \in \mathcal{K}_{b_n}} \left(1 + \boldsymbol{\lambda}^\top \boldsymbol{m}_{\boldsymbol{\kappa}}(\boldsymbol{\theta})\right)^2 \tag{7}$$

(see Antoine et al., 2007). From the first order conditions of Equation (7) we can compute an estimator of the auxiliary parameter $\boldsymbol{\lambda}$

$$\frac{\widehat{\boldsymbol{\lambda}}(\boldsymbol{\theta})}{b_n^{1+d}} = -\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\theta})^{-1} \widehat{\boldsymbol{m}}(\boldsymbol{\theta}) \tag{8}$$

with

$$\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\theta}) = \frac{b_n^{1+d}}{N} \sum_{\boldsymbol{\kappa} \in \mathcal{K}_{b_n}} \boldsymbol{m}_{\boldsymbol{\kappa}}(\boldsymbol{\theta}) \boldsymbol{m}_{\boldsymbol{\kappa}}(\boldsymbol{\theta})^\top. \tag{9}$$

By plugging in Equation (8) into Equation (7) we find

$$R_n(\boldsymbol{\theta}, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\theta})) = \frac{N}{2} \left(1 - b_n^{1+d} \widehat{\boldsymbol{m}}(\boldsymbol{\theta})^\top \widehat{\boldsymbol{\Sigma}}(\boldsymbol{\theta})^{-1} \widehat{\boldsymbol{m}}(\boldsymbol{\theta})\right) = \frac{N}{2} \left(1 - b_n^{1+d} Q_n(\boldsymbol{\theta})\right)$$

where $Q_n(\boldsymbol{\theta})$ is implicitly defined. Hence,

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\Theta} Q_n(\boldsymbol{\theta}) \tag{10}$$

is the STBEU estimator for the parameter vector $\boldsymbol{\theta}$.

3.1 Asymptotic results

The asymptotic results are derived by adapting to our problem some results in Jenish & Prucha (2009) (see also Bai et al., 2012).

A1 Let $\mathcal{L} \subset \mathbb{R}^d \times \mathbb{R}^+$ be a possibly unevenly spaced lattice. For any two points $\boldsymbol{l}$ and $\boldsymbol{k}$ in $\mathcal{L}$ their distance is at least d_0. This is, given a distance metric $\rho(\cdot,\cdot)$, we have $\rho(\boldsymbol{l},\boldsymbol{k}) \geq d_0$.

A2 Let $\mathcal{L}_n$ be a sequence of arbitrary subsets of $\mathcal{L}$ such that $|\mathcal{L}_n| \to \infty$ as $n \to \infty$.

A3 The parameter set $\Theta \subset \mathbb{R}^{d_{\boldsymbol{\theta}}}$ is compact and $\boldsymbol{\theta}_0$ is an interior point of Θ.

A4 For some $\delta > 0$ and $e > 0$ and for all $\boldsymbol{\kappa} \in \mathcal{L}_n$,

$$\lim_{e\to\infty} \mathrm{E}\left[\sup_{\boldsymbol{\theta}\in\Theta} \|\boldsymbol{m}_{\boldsymbol{\kappa}}(\boldsymbol{\theta})\|^{2+\delta} \mathbf{1}\{\sup_{\boldsymbol{\theta}\in\Theta} \|\boldsymbol{m}_{\boldsymbol{\kappa}}(\boldsymbol{\theta})\| > e\}\right] = 0$$

where $\mathbf{1}\{\cdot\}$ is the indicator function.

A5 Define $\nabla^{\ell}_{\boldsymbol{\theta}}$ the ℓ-th derivative operator with respect to $\boldsymbol{\theta}$ and $\ell = 0, 1, 2$. Then, (i) $\mathrm{E}\left[\|\nabla_{\boldsymbol{\theta}} \boldsymbol{m}_{\boldsymbol{\kappa}}(\boldsymbol{\theta})\|^{1+\eta}\right] < \infty$ for all $\boldsymbol{l} \in \mathcal{L}_n$, with $\eta > 0$; (ii) $\mathrm{E}\left[\sup_{\boldsymbol{\theta}\in\Theta} \|\nabla^{\ell}_{\boldsymbol{\theta}} \boldsymbol{m}_{\boldsymbol{l}}(\boldsymbol{\theta})\|\right] < \infty$; (iii) let $\nabla^{\ell}_{\boldsymbol{\theta}} \boldsymbol{m}(\boldsymbol{\theta}) = \mathrm{E}\left[\nabla^{\ell}_{\boldsymbol{\theta}} \boldsymbol{m}_{\boldsymbol{l}}(\boldsymbol{\theta})\right]$, then $\nabla_{\boldsymbol{\theta}} \boldsymbol{m}(\boldsymbol{\theta}_0)$ is full column rank; (iv) $\lim \widehat{\boldsymbol{\Sigma}}(\boldsymbol{\theta}) \to \boldsymbol{\Sigma}(\boldsymbol{\theta})$ as $n \to \infty$, a positive definite matrix.

A6 Consider $\mathcal{V} \subseteq \mathcal{L}_n$ and $\mathcal{W} \subseteq \mathcal{L}_n$, let $\sigma(\mathcal{V}) = \sigma(\boldsymbol{z}_{\boldsymbol{l}}, \boldsymbol{l} \in \mathcal{V})$ and $\sigma(\mathcal{W}) = \sigma(\boldsymbol{z}_{\boldsymbol{l}}, \boldsymbol{l} \in \mathcal{W})$ and $\alpha(\mathcal{V},\mathcal{W}) = \alpha(\sigma(\mathcal{V}), \sigma(\mathcal{W}))$. Consider also the set $\mathbb{R}^d \times \mathbb{R}^+$ endowed with the metric $\rho(\boldsymbol{l},\boldsymbol{k}) = \max_{1\leq i\leq 1+d} |l_i - k_i|$. In addition to that define the set distance as

$\rho(\mathcal{V}, \mathcal{W}) = \inf\{\rho(\boldsymbol{l}, \boldsymbol{k}) : \boldsymbol{l} \in \mathcal{V}, \boldsymbol{k} \in \mathcal{W}\}$ for any subset $\mathcal{V}, \mathcal{W} \subset \mathbb{R}^d \times \mathbb{R}^+$. Then, the α-mixing coefficient for the random field is given by

$$\alpha_{p,q}(r) = \sup\left(\alpha(\mathcal{V}, \mathcal{W}), |\mathcal{V}| \leq p, |\mathcal{W}| \leq q, \rho(\mathcal{V}, \mathcal{W}) \geq r\right).$$

We assume that the following conditions hold:

(a) $\sum_{h=1}^{\infty} h^{(1+d)-1} \alpha_{1,1}(h)^{\frac{\delta}{2+\delta}} < \infty$,

(b) $\sum_{h=1}^{\infty} h^{(1+d)-1} \alpha_{p,q}(h) < \infty$ for $p + q \leq 4$,

(c) $\alpha_{1,\infty}(h) = O(h^{-(1+d)-\varepsilon})$ for some $\varepsilon > 0$.

Theorem 1. *Assume A1 to A6 hold. Then,*

1. $\widehat{\boldsymbol{\theta}} \to_p \boldsymbol{\theta}_0$,

2. $\sqrt{n}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \to_d N\left(\mathbf{0}, \boldsymbol{\Omega}(\boldsymbol{\theta}_0)\right)$

where $\boldsymbol{\Omega}(\theta_0) = (\nabla_{\boldsymbol{\theta}} \boldsymbol{m}(\boldsymbol{\theta}_0)^\top \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)^{-1} \nabla_{\boldsymbol{\theta}} \boldsymbol{m}(\boldsymbol{\theta}_0))^{-1}$.

In what follows we discuss some important features of the assumptions used to derive Theorem 1. Assumption A1 defines the structure of the lattice. Even though we allow the lattice to be unevenly spaced, we do not want the points to be too close to each other. Under Assumption A2 the number of points in any subset of $\mathcal{L}$ grows as n grows. Assumption A3 is a standard condition on the parameter space. A4 is an assumption on the tail behavior of the moment condition and it is called uniform $L_{\delta+2}$ integrability. Together with A1, A2 and the α-mixing condition A6 allow us to use a central limit theorem for RFs. A5 is a set of regularity conditions. In particular, A5(i) and A5(ii) allow us to use a uniform law of large numbers, A5(iii) is necessary to guarantee invertibility of the variance covariance matrix of the estimator, while A5(iv) is a condition on the finiteness of the limiting variance covariance matrix of the moment conditions and it is used in the consistency proof.

4 Numerical experiments

4.1 Statistical efficiency

This section compares the relative efficiency of the STBEU with respect to the pairwise likelihood (PL). To this end, we configure two sampling schemes, a regular sampling scheme and an irregular sampling scheme. In the first case, we set a regular grid with unit spacing $[-a, a]^2$ and with $n_s = (2a + 1)^2$ locations in space and n_t in time. In the second case, the setting involves an irregular grid with $n_s = \frac{(2a+1)^2}{2} \times 2$ locations in space uniformly distributed on $[-a, a]^2$ and n_t in time. In both cases we have $N = n_t \times n_s$ spatio-temporal locations and $a \in \mathbb{R}$. In what follows we consider three specific simulation settings:

1. spatial blocks: more space than time locations, $[-8, 8]^2$ and $n_t = 19$, that is $n_s = 289$ and $n_{st} = 5202$;

2. temporal blocks: more time than space locations, $[-2, 2]^2$ and $n_t = 210$, that is $n_s = 25$ and $n_{st} = 5250$;

3. spatio-temporal blocks: balanced spatio-temporal locations, $[-5, 5]^2$ and $n_t = 50$, that is $n_s = 121$ and $n_{st} = 6050$.

Note that *more* means roughly 10 times (or higher) locations more than the other and *balanced* means less than 2 times. Under these settings, we perform 500 simulations of a Gaussian random field with Double Exponential and Gneiting covariance functions as defined in Equations (1) and (2). In both cases we estimate the spatial and temporal scale parameters and the variance parameters that is α_s, α_t and σ^2 respectively. For each simulation setting and covariance model we consider two combinations of parameters, so that we can evaluate the effect of an increasing spatial and temporal dependence through α_s, α_t (specific parameter values are found in Tables 2, 3 and 4).

We also consider the effect of the block length on the efficiency of the STBEU estimator. Following Y. D. Lee & Lahiri (2002) and Bevilacqua et al. (2015), spatial blocks are formed

by the set $[C\sqrt{\gamma}, C\sqrt{\gamma}]^2$ in overlapping and non overlapping cases with C being a positive constant and we chose γ to be the range of the spatial coordinates. Temporal blocks are formed by a sequence of the temporal length spaced by b_t. For example, if the spatial block has length $b_s = 2$, the temporal block length $b_t = 10$, $\gamma = 16$ and $n_t = 50$, then $C = 1/2$ and the prototype spatio-temporal block $\mathcal{U}$ is equal to $(-1/8, 1/8]^2 \times 5$.

We chose $b_s = \{2, 4\}$ for space, $b_t = \{2, 3\}$ for time and $b_{st} = \{4, 9\}$ for spatio-temporal blocking. In the overlapping version, constants o_s and o_t are needed to tune the degree of overlapping. A possible choice for these constants is $o_s = b_s p_s$ and $o_t = b_t p_t$ with $0 < p_s \leq 1$ and $0 < p_t \leq 1$. We set $p = p_s = p_t = 0.5$ for the overlapping case while $p = p_s = p_t = 1$ corresponds to the non overlapping case. Table 1 shows the number of spatio-temporal blocks associated with three settings under the overlapping and non-overlapping version.

Finally, the distances d_s and d_t in the weight function (5) are chosen to be 25% of its corresponding block length.

Blocking		p	
		1	0.5
Spatial	$b_s = 2$	64	225
	$b_s = 4$	16	49
Temporal	$b_t = 2$	105	209
	$b_t = 3$	70	139
Spatio-temporal	$b_{st} = 4$	625	3969
	$b_{st} = 9$	144	800

Table 1: Number of spatial, temporal and spatio-temporal blocks resulting from fixing the block length b and the overlapping parameter $p = p_s = p_t$ used in the simulation study

Figure 1 shows the intuition behind the spatio-temporal blocking procedure. Think of spatio-temporal locations as being a dense block as showed in the upper-left panel of Figure 1 with time represented by depth. Spatial blocking is the upper-right panel: space is divided by the blocking procedure mentioned above such that every block considers all time locations. The lower-left panel represents temporal blocking: time is divided uniformly and all space locations are considered in each block. Finally, the lower-right panel is the spatio-temporal blocking which is a combination of both spatial and temporal

blocking. Note that, regardless of the procedure, every block considers spatio-temporal locations. Say we have more space locations than time locations, then better performance is expected by choosing spatial blocking. The same reasoning applies for temporal blocking or spatio-temporal blocking.

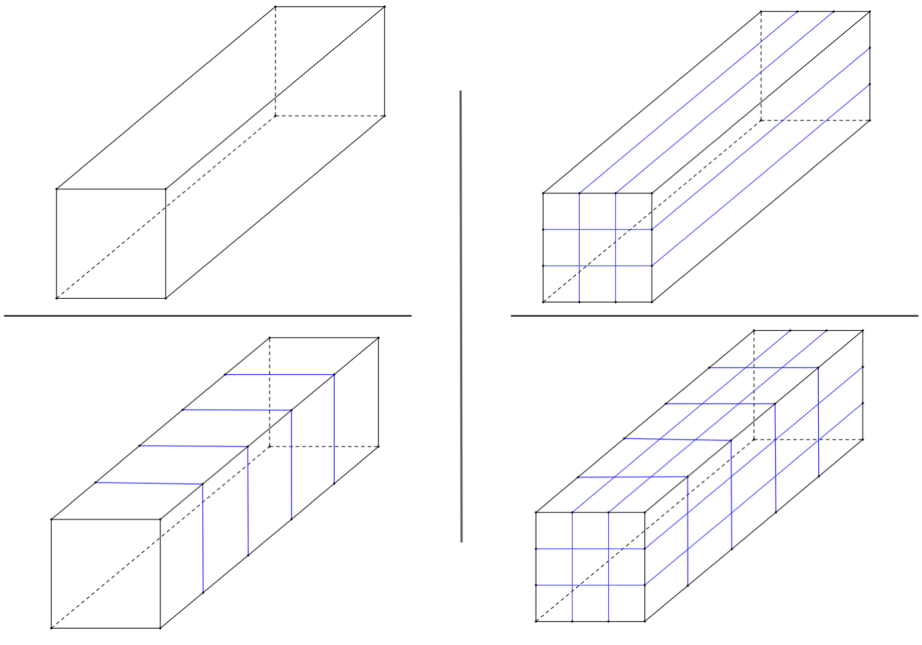

Figure 1: Intuition behind the spatio-temporal blocking procedure

Tables 2, 3 and 4 report the simulation results for the spatial, temporal and spatio-temporal blocking respectively. We measure efficiency in two ways. The first one corresponds to the simulated relative efficiency defined as $SRE = \frac{mse_{PL}}{mse_{STBEU}}$. SRE is reported for every parameter and scenario. The second approach is the simulated total relative efficiency (STRE) as a measure of overall efficiency for the multi-parameter case (Bevilacqua & Gaetan, 2015). The STRE is defined as $STRE = \left(\frac{D_{PL}}{D_{STBEU}}\right)^{1/p}$ where $p = 3$ is the number of parameters of the model, D_{PL} and D_{STBEU} are the determinants of the variance covariance matrices of the PL and STBEU estimators respectively.

The simulation results allow us to make some interesting comments on the performance of the estimators under scrutiny. First of all, we notice that it is difficult to have a clear ranking between STBEU and PL in absolute terms. However, we notice that for certain

	Double exponential				Gneiting			
	Regular		Irregular		Regular		Irregular	
	$b=2$	$b=4$	$b=2$	$b=4$	$b=2$	$b=4$	$b=2$	$b=4$
	$\alpha_s = 1.2/3$ $\alpha_t = 1.2/3$				$\alpha_s = 1.8/3$ $\alpha_t = 1.8/3$			
α_s	0.982	0.894	0.545	0.636	1.213	1.055	0.645	0.749
	(1.035)	(0.979)	(0.354)	(0.472)	(1.218)	(1.23)	(0.426)	(0.659)
α_t	0.896	0.834	0.339	0.563	1.171	0.998	0.579	0.743
	(0.828)	(0.838)	(0.216)	(0.439)	(1.115)	(1.109)	(0.383)	(0.612)
σ^2	0.934	0.898	0.405	0.662	0.917	0.852	0.402	0.662
	(0.918)	(0.957)	(0.288)	(0.444)	(0.898)	(0.945)	(0.284)	(0.544)
$STRE$	0.952	0.901	0.529	0.701	1.054	0.963	0.625	0.778
	(0.939)	(0.937)	(0.391)	(0.536)	(1.039)	(1.054)	(0.47)	(0.68)
	$\alpha_s = 1.2/3$ $\alpha_t = 1.2/19$				$\alpha_s = 1.8/3$ $\alpha_t = 1.8/19$			
α_s	1.142	0.928	0.552	0.619	1.791	1.4	0.788	0.881
	(1.192)	(1.085)	(0.338)	(0.532)	(1.856)	(1.634)	(0.502)	(0.799)
α_t	1.057	0.886	0.483	0.638	1.58	1.245	0.716	0.848
	(1.027)	(0.98)	(0.316)	(0.538)	(1.638)	(1.417)	(0.475)	(0.737)
σ^2	0.918	0.845	0.386	0.624	0.914	0.851	0.381	0.624
	(0.91)	(0.962)	(0.27)	(0.527)	(0.907)	(0.954)	(0.267)	(0.53)
$STRE$	1.038	0.921	0.59	0.723	1.233	1.076	0.69	0.832
	(1.04)	(1.01)	(0.433)	(0.633)	(1.253)	(1.186)	(0.513)	(0.742)

Table 2: Simulated relative efficiency (with respect to the PL) of STBEU estimator under spatial blocking. Relative efficiency is presented for different values of the block length, overlapping-non overlapping (in parentheses) and regular-irregular cases. Rows with $STRE$ caption shows the overall performance.

specifications STBEU clearly outperforms PL. For example, this happens in Table 2 for the $STRE$ when using the Double Exponential correlation function with $b = 2$ in the regular case and for the Gneiting correlation function for almost all the results ($STRE$ and SRE) in the regular case. Similar results are found in Tables 3 and 4. It is worth mentioning that STBEU outperforms PL in some irregular cases as well. Particularly, for α_t in the temporal blocking case using the Gneiting correlation function.

In addition to that, since the computation of STBEU is comparatively time saving, a researcher concerned with speed may be willing to trade off some statistical efficiency in favor of higher computational efficiency. Further details on computational efficiency are presented in Section 4.2. Moreover, consistently with the results in Bevilacqua et al. (2015),

	Double exponential				Gneiting			
	Regular		Irregular		Regular		Irregular	
	$b=2$	$b=3$	$b=2$	$b=3$	$b=2$	$b=3$	$b=2$	$b=3$
	$\alpha_s = 3.1/3$ $\alpha_t = 3.1/3$				$\alpha_s = 4/3$ $\alpha_t = 4/3$			
α_s	1.195	0.704	1.121	0.675	1.125	0.694	1.031	0.622
	(0.572)	(0.393)	(0.516)	(0.361)	(0.528)	(0.377)	(0.462)	(0.332)
α_t	1.427	0.965	1.359	0.852	2.841	2.022	2.103	1.423
	(0.818)	(0.506)	(0.665)	(0.462)	(1.613)	(1.12)	(1.048)	(0.78)
σ^2	1.02	0.655	1.000	0.624	1.018	0.643	1.007	0.613
	(0.462)	(0.32)	(0.44)	(0.309)	(0.454)	(0.322)	(0.436)	(0.304)
$STRE$	1.189	0.841	1.169	0.82	1.325	0.986	1.217	0.884
	(0.706)	(0.515)	(0.668)	(0.486)	(0.8)	(0.608)	(0.701)	(0.533)
	$\alpha_s = 3.1/3$ $\alpha_t = 3.1/19$				$\alpha_s = 4/3$ $\alpha_t = 4/19$			
α_s	1.216	0.709	1.147	0.689	1.069	0.648	1.01	0.617
	(0.576)	(0.39)	(0.535)	(0.374)	(0.492)	(0.352)	(0.462)	(0.329)
α_t	1.763	1.166	1.544	0.967	3.379	2.365	2.542	1.695
	(1.013)	(0.616)	(0.764)	(0.523)	(1.935)	(1.318)	(1.284)	(0.926)
σ^2	1.012	0.647	1.008	0.63	1.02	0.63	1.015	0.617
	(0.463)	(0.323)	(0.452)	(0.32)	(0.45)	(0.321)	(0.441)	(0.304)
$STRE$	1.315	0.916	1.264	0.886	1.401	1.034	1.288	0.936
	(0.776)	(0.558)	(0.726)	(0.52)	(0.844)	(0.636)	(0.746)	(0.559)

Table 3: Simulated relative efficiency (with respect to the PL) of STBEU estimator under temporal blocking. Relative efficiency is presented for different values of the block length, overlapping-non overlapping (in parentheses) and regular-irregular cases. Rows with $STRE$ caption shows the overall performance.

the STBEU tends to perform better when the spatial data are on a regular grid. Finally, we notice that the effect of the block length has a considerable impact on the results. In general, we notice that smaller block lengths tend to provide better results. This suggest that, given an adequate procedure for the selection of the block length in conjunction with our computationally efficient approach, we may obtain further improvements. This problem is relevant and it is the object of future research.

4.2 Computational efficiency

The STBEU estimator is implemented in C and OpenCL (OCL) standard, both interfacing with R. We used a MacBook Pro laptop that has three devices, an Intel Core CPU and

	Double exponential				Gneiting			
	Regular		Irregular		Regular		Irregular	
	$b_{st}=4$	$b_{st}=9$	$b_{st}=4$	$b_{st}=9$	$b_{st}=4$	$b_{st}=9$	$b_{st}=4$	$b_{st}=9$
	$\alpha_s = 0.4/3$ $\alpha_t = 3/3$				$\alpha_s = 0.4/3$ $\alpha_t = 3/19$			
α_s	1.491	0.634	1.028	0.711	1.622	0.895	1.189	0.918
	(0.826)	(0.4)	(0.332)	(0.406)	(1.029)	(0.553)	(0.428)	(0.564)
α_t	1.968	0.896	1.143	0.956	3.257	1.492	1.747	1.422
	(1.08)	(0.502)	(0.385)	(0.534)	(1.78)	(0.844)	(0.647)	(0.844)
σ^2	0.902	0.551	0.557	0.513	0.91	0.543	0.565	0.512
	(0.579)	(0.362)	(0.205)	(0.316)	(0.576)	(0.356)	(0.207)	(0.315)
$STRE$	1.398	0.794	1.035	0.846	1.542	0.904	1.095	0.92
	(0.923)	(0.535)	(0.431)	(0.524)	(1.02)	(0.6)	(0.46)	(0.582)
	$\alpha_s = 1.2/3$ $\alpha_t = 6/3$				$\alpha_s = 1.2/3$ $\alpha_t = 6/19$			
α_s	1.576	0.666	1.086	0.718	0.776	0.52	1.125	0.799
	(0.854)	(0.417)	(0.375)	(0.417)	(0.471)	(0.258)	(0.381)	(0.483)
α_t	2.581	1.165	1.567	1.226	1.675	1.124	2.123	1.575
	(1.435)	(0.644)	(0.554)	(0.688)	(0.938)	(0.555)	(0.811)	(0.928)
σ^2	0.897	0.539	0.568	0.502	0.534	0.463	0.582	0.506
	(0.56)	(0.351)	(0.222)	(0.317)	(0.348)	(0.23)	(0.227)	(0.32)
$STRE$	1.624	0.909	1.205	0.939	0.962	0.755	1.177	0.953
	(1.076)	(0.61)	(0.516)	(0.586)	(0.633)	(0.413)	(0.506)	(0.605)

Table 4: Simulated relative efficiency (with respect to the PL) of STBEU estimator under spatio-temporal blocking. Relative efficiency is presented for different values of the block length, overlapping-non overlapping (in parentheses) and regular-irregular cases. Rows with $STRE$ caption shows the overall performance.

two GPU devices: Intel Iris Pro and AMD Radeon R9 M370X Compute Engine, but we worked in CPU and AMD since they support double precision. Computational efficiency performance is evaluated comparing C vs OpenCL (through R) in two ways: evaluation of $\boldsymbol{g}_{ij}$ from Equation (4) in one block, and the full blockwise approach.

Our AMD device supports OpenCL version 1.2. There are 10 Compute Units (CUs), where each CU contains 16 stream cores, and each stream core houses four processing elements. Thus, each compute unit in the Radeon R9 M370X has 64 (16×4) processing elements (i.e. 640 PE in total)[2]. Our CPU (called the *host* in OpenCL) has access to 16 Gb of the main memory, while the GPU has 2 Gb of memory from which it can directly

[2]All GPU vendors have some fundamental building block they scale up/down to hit various performance/power/price targets. AMD calls theirs a Compute Unit, NVIDIA's is known as an SMX, and Intel's is called a sub-slice.

process data.

Now, in order to evaluate the correlation functions, we need to compute $n_{st}(n_{st}-1)/2$ distances for the upper triangular matrix formed by all possible pairs of n_{st} spatio-temporal locations. At first glance, this would mean that the problem size (called *NDrange* in OpenCL where ND stands for $N-dimensional$, $N=1,2,3$) is $n_{st}(n_{st}-1)/2$ too. Say, for example, we have $n_s = 1024$ locations in space and $n_t = 32$ in time, that makes $nt = 32768$ spatio-temporal locations. Double precision requires 8 bytes per location, that means that our host and device memory requirement would be $8 \times (32768 \times (32767)/2) \approx 4.3Gb$. To overcome this memory requirement issue, we set the NDrange to have two dimensions with sizes n_s and n_t. It means that our device memory requirement is now $8 \times 1024 \times 32 \approx 33kB$, roughly 0.0007% of the initial requirement in our example. The latter was possible due to the *workgroup* concept in OpenCL.

Figure 2 compares C and OpenCL performance of equations (1) and (2) as specified before. Space locations vary from 4 to 9409 and time locations from 2 to 97 on the left panel, the opposite in the right panel. These results are dependent on the characteristics of the computer, such as the graphic card, OpenCL version, hardware specs, and so on. Nonetheless, it provides a relative sense of the computational improvement potential. We used AMD in this case, local size is 16 work-items in each dimension, which makes our total max Work Group Size (256). In both panels, OpenCL GPU timing outperforms C from roughly $n_{st} \approx 10000$ reaching approximately 6 and 3 times faster for the double exponential and Gneiting case respectively.

Rows from Figure 3 compare spatial blocking against temporal blocking and columns compare Double Exponential (1) and Gneiting (2) correlation models. In the spatial blocking procedure, n_t is fixed to 100 and n_s maximum is 29584, meaning $n_{st} = 2958400$, and n_s is fixed to 100 and the maximum value of n_t is 29600 ($n_{st} = 2960000$) in the temporal blocking case. We can see that OpenCL outperforms C in all cases. An important conclusion from Figure 3 is that OpenCL should be used when having more locations per block. In the blockwise context, this implies that having a denser block improves the

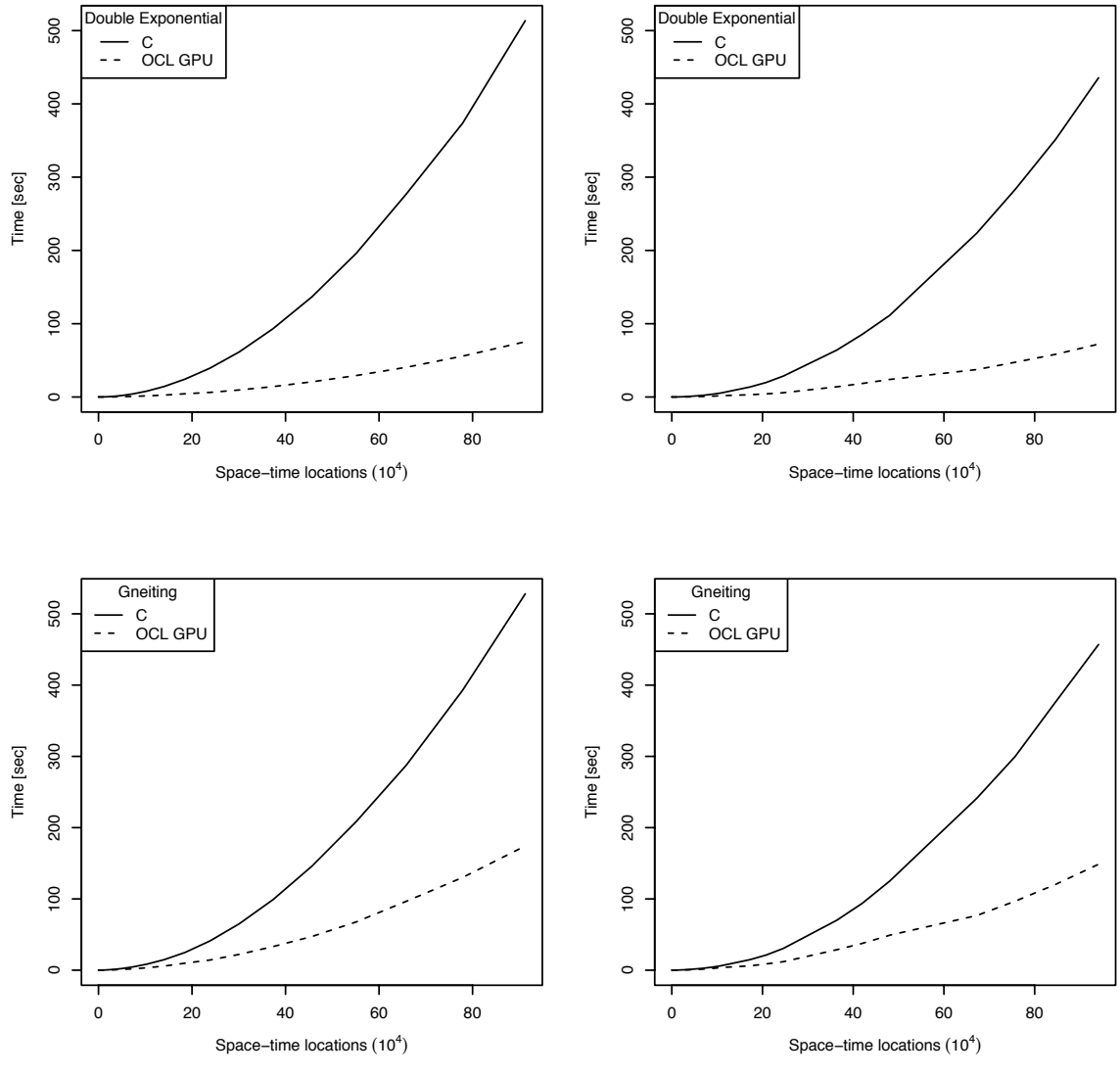

Figure 2: Gradient ($\boldsymbol{g}_{ij}$) evaluation time performance comparison C vs OpenCL (denoted OCL) for Double Exponential and Gneiting covariance functions. Space locations vary from 4 to 9409 and time locations from 2 to 97 on the left panel, the opposite in the right panel.

time performance. Rows from Figure 3 reinforce this conclusion as we set 50 temporal blocks and approximately 11 spatial blocks. Comparing the correlation function used in the blockwise procedure (i.e. the columns from Figure 3) suggests that using the Double Exponential covariance function outperforms the Gneiting covariance function. Finally, note that OpenCL GPU outperforms OpenCL CPU in three out of for panels. The upper right panel is the exception, but their difference seems to converge around the maximum.

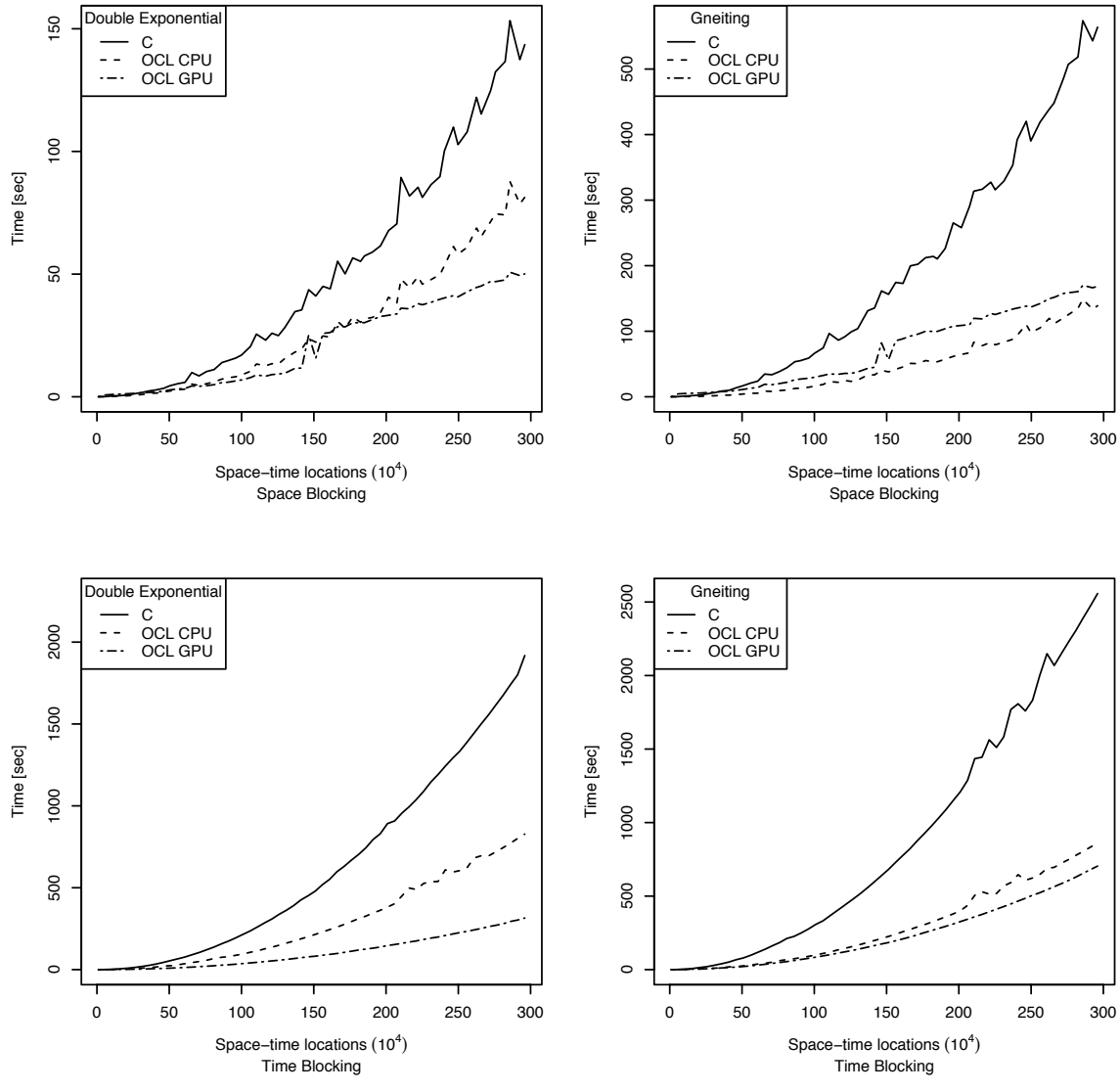

Figure 3: Blockwise time performance comparison for C vs OpenCL (denoted OCL with CPU and GPU). The x axis is divided to $10e4$. Rows compare spatial vs temporal blocking and columns compare the correlation model.

5 Application: Mediterranean winds

The Mediterranean winds data set contains wind component observations (east-west) for 1175 space locations and 28 time periods taken every 6 hours from 00:00 UTC on 29 January 2005 to 18:00 UTC on 04 February 2005. These data are available in Wikle et al. (2019). Figure 4 shows a map of the spatial locations. For reproducible research purposes, we developed the R package `STBEU` (Morales-Oñate et al., 2019) that includes the full code

STBEU				
Parameters	α_s	α_t	σ^2	Objective
$\beta = 0$	364.19	29.58	11.65	$5.813082e-16$
$\beta = 0.5$	373.97	38.06	12.47	$2.503008e-16$
$\beta = 1$	372.03	36.94	12.28	$5.642065e-16$
PL				
$\beta = 0$	338.54	18.45	13.01	-2816935.05
$\beta = 0.5$	338.75	18.54	13.02	-2816938.93
$\beta = 1$	339.00	18.63	13.02	-2816944.73

Table 5: Estimation results of the spatio-temporal Gaussian process with Wendland model (11) to Mediterranean winds data with SBEU and PL for $\beta = 0, 0.5, 1$.

for this application.

Figure 4: Mediterranean region. The light blue dots are the space locations where the wind component data are recorded in the region from 6.5° W-16.5° E and 33.5° N-45.5° N.

We assume data to be a realization of an isotropic in space and symmetric in time spatio-temporal Gaussian RF with spatio-temporal Wendland correlation function (Porcu

Scenario	Elapsed time	Time Gain (respect to i))
i)	16.6696	1.0000
ii)	2.5202	6.6144
iii)	1.0604	15.7201
iv)	0.4764	34.9908
v)	0.2237	74.5177

Table 6: Estimation elapsed times (minutes) of the spatio-temporal Gaussian process with Wendland covariance model (11) to Mediterranean winds data. Scenarios are i) PL using CPU one core (default in R), ii) PL using OpenCL framework with CPU (Intel(R) Core(TM) i7-4980HQ), iii) STBEU using CPU one core (default in R), iv) STBEU using OpenCL framework with GPU (AMD Radeon R9 M370X) and v) STBEU using OpenCL framework with CPU (Intel(R) Core(TM) i7-4980HQ).

et al., 2020; Bevilacqua et al., 2019):

$$\phi(\boldsymbol{h}, u, \boldsymbol{\theta}) = \frac{\sigma^2}{(1 + ||\boldsymbol{h}||/\alpha_s)^{2.5}} \left(1 - \frac{|u|}{\alpha_t (1 + ||\boldsymbol{h}||/\alpha_s)^{-\beta}}\right)_+^{4.5}, \tag{11}$$

where $\boldsymbol{\theta} = (\sigma^2, \alpha_s, \alpha_t, \beta)^\top$. Here $\beta \in [0, 1]$ is a separability parameter. The case $\beta = 0$ implies a separable spatio-temporal covariance and the case $0 < \beta \leq 1$ leads to a non separable parameter. Since the data set has more space than time locations, spatial (non overlapping) blocks are constructed in the following manner: $[0, 400]^2$ and $n_t = 28$, that is $n_s = 1175$ and $n_{st} = 32900$. We estimate the model with STBEU considering the cases $\beta = 0, 0.5, 1$ and with weights such that only pairs with spatial and temporal distances lower than 50 and 6 respectively are considered for each block, that is $d_s = 50$ and $d_t = 6$ in the weight function (5). The results reported in Table 5 show that the objective function of the STBEU is minimized for $\beta = 0.5$. Additionally, in Figure 5, the empirical marginal spatial and temporal semi-variograms are compared with their estimated theoretical counterparts using STBEU and PL estimates with $\beta = 0.5$ and they show a satisfactory fitting in particular for the STBEU estimation.

Finally we show the computational benefits of the STBEU method. Results in Table 6 show the elapsed time (in minutes) of the entire optimization process (we use the simplex method proposed in Nelder & Mead (1965) as implemented in the R function `optim`) for

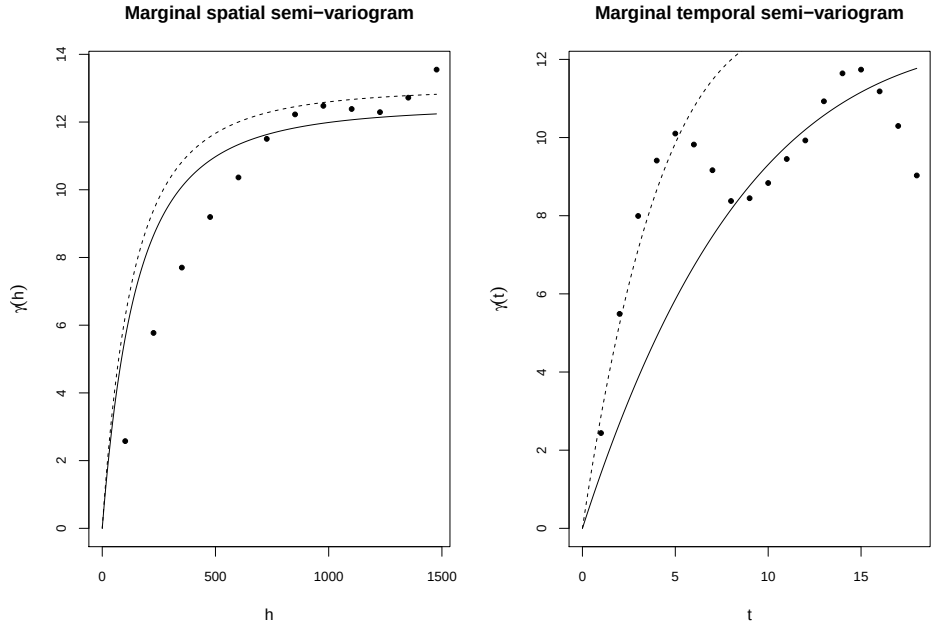

Figure 5: Empirical spatial and temporal marginal semi-variogram versus the estimated semi-variograms using model (11) with $\beta = 0.5$ using STBEU (solid line) and PL (dotted line) estimates.

five setups:

i) PL using CPU one core (default in R),

ii) PL using OpenCL framework with CPU (Intel(R) Core(TM) i7-4980HQ),

iii) STBEU using CPU one core (default in R),

iv) STBEU using OpenCL framework with GPU (AMD Radeon R9 M370X) and

v) STBEU using OpenCL framework with CPU (Intel(R) Core(TM) i7-4980HQ).

Using the Wendland covariance function and comparing against the PL (CPU-only) setup, the STBEU method is approximately 35 and 75 times faster in setups iv) and v) respectively.

6 Conclusions

In this paper we introduce a blockwise Euclidean likelihood method based on the score of the pairwise likelihood objective function for the estimation of spatio-temporal covariance models of Gaussian random fields. This approach is particularly useful when dealing with large data sets. We show that the proposed estimator, denoted as STBEU, is consistent and asymptotically normal. Furthermore, a set of simulation results and an application on a wind speed data set suggest that the STBEU works well in finite samples. The blockwise approach guarantees considerable computational gains over the standard pairwise composite likelihood method and our implementation in OpenCL allows us to obtain further improvements in the computation of the estimates. Although in this paper we only considered spatio-temporal Gaussian random fields, the proposed methodology can be easily extended to the case of the estimation of spatio-temporal non-Gaussian random fields with known bivariate distribution as, for example, in Alegría et al. (2017) and Bevilacqua et al. (2020).

Acknowledgements

Partial support was provided by FONDECYT grant 1200068, Chile and by Millennium Science Initiative of the Ministry of Economy, Development, and Tourism, grant "Millenium Nucleus Center for the Discovery of Structures in Complex Data" and by regional MATH-AmSud program, grant number 20-MATH-03 for Moreno Bevilacqua. Federico Crudu's research was supported by the FONDECYT grant 11140433 and the Regione Autonoma della Sardegna Master and Back grant PRR-MABA2011-24192. Víctor Morales-Oñate's research was partially supported by the Data Science Research Group at Escuela Superior Politécnica de Chimborazo - Ecuador.

A Proofs

In this section we collect the proof of the asymptotic results described in Theorem 1. Let us introduce some useful notation: $\nabla_{\boldsymbol{\theta}}$ and $\nabla_{\boldsymbol{\lambda}}$ are the first derivative operators for $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$ respectively, while $\nabla_{\boldsymbol{\theta\theta}}$, $\nabla_{\boldsymbol{\lambda\lambda}}$ and $\nabla_{\boldsymbol{\theta\lambda}}$ indicate second and cross derivatives and are defined accordingly. Similarly, for a certain function $R_n(\boldsymbol{\theta}, \boldsymbol{\lambda})$ defined below, $R_{n,\boldsymbol{\theta}}(\boldsymbol{\theta}, \boldsymbol{\lambda})$ is its first derivative with respect to $\boldsymbol{\theta}$. Derivatives with respect to $\boldsymbol{\lambda}$, second derivatives and cross derivatives are defined in a similar manner. Let us also define $Q(\boldsymbol{\theta}) = \boldsymbol{m}(\boldsymbol{\theta})^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \boldsymbol{m}(\boldsymbol{\theta})$, the population version of our objective function.

Proof. We first prove part 1. We have to show that, for some $\delta > 0$, $P(\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| > \delta) \to 0$ as $n \to \infty$. By continuity of $Q(\boldsymbol{\theta})$ and the assumption that $\boldsymbol{\theta}_0$ is the unique minimizer, we have that, for some $\varepsilon > 0$, $\{\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| > \delta\} \implies \{ \mid Q(\widehat{\boldsymbol{\theta}}) - Q(\boldsymbol{\theta}_0) \mid > \varepsilon\}$. This is, the latter set contains the former. Hence, $P(\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| > \delta) \leq P(\mid Q(\widehat{\boldsymbol{\theta}}) - Q(\boldsymbol{\theta}_0) \mid > \varepsilon)$. By some simple algebraic manipulation we have

$$
\begin{aligned}
\widehat{Q}_n(\boldsymbol{\theta}) - Q(\boldsymbol{\theta}) &= \widehat{\boldsymbol{m}}(\boldsymbol{\theta})^\top \widehat{\boldsymbol{\Sigma}}(\boldsymbol{\theta})^{-1} \widehat{\boldsymbol{m}}(\boldsymbol{\theta}) - \boldsymbol{m}(\boldsymbol{\theta})^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \boldsymbol{m}(\boldsymbol{\theta}) \\
&= (\widehat{\boldsymbol{m}}(\boldsymbol{\theta}) - \boldsymbol{m}(\boldsymbol{\theta}))^\top \widehat{\boldsymbol{\Sigma}}(\boldsymbol{\theta})^{-1} (\widehat{\boldsymbol{m}}(\boldsymbol{\theta}) - \boldsymbol{m}(\boldsymbol{\theta})) + 2 (\widehat{\boldsymbol{m}}(\boldsymbol{\theta}) - \boldsymbol{m}(\boldsymbol{\theta}))^\top \widehat{\boldsymbol{\Sigma}}(\boldsymbol{\theta})^{-1} \boldsymbol{m}(\boldsymbol{\theta}) \\
&\quad - \boldsymbol{m}(\boldsymbol{\theta})^\top \left(\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} - \widehat{\boldsymbol{\Sigma}}(\boldsymbol{\theta})^{-1} \right) \boldsymbol{m}(\boldsymbol{\theta}).
\end{aligned}
$$

Hence, by taking the norm and by triangle inequality

$$
\begin{aligned}
|\widehat{Q}_n(\boldsymbol{\theta}) - Q(\boldsymbol{\theta})| &\leq \|\widehat{\boldsymbol{m}}(\boldsymbol{\theta}) - \boldsymbol{m}(\boldsymbol{\theta})\|^2 \left\| \widehat{\boldsymbol{\Sigma}}(\boldsymbol{\theta})^{-1} \right\| + 2 \|\widehat{\boldsymbol{m}}(\boldsymbol{\theta}) - \boldsymbol{m}(\boldsymbol{\theta})\| \left\| \widehat{\boldsymbol{\Sigma}}(\boldsymbol{\theta})^{-1} \right\| \|\boldsymbol{m}(\boldsymbol{\theta})\| \\
&\quad - \|\boldsymbol{m}(\boldsymbol{\theta})\|^2 \left\| \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} - \widehat{\boldsymbol{\Sigma}}(\boldsymbol{\theta})^{-1} \right\|.
\end{aligned}
$$

By assumptions A5 and A6 and the continuous mapping theorem we get the following uniform convergence result

$$
\sup_{\boldsymbol{\theta} \in \Theta} \mid \widehat{Q}_n(\boldsymbol{\theta}) - Q(\boldsymbol{\theta}) \mid \to_p 0. \tag{12}
$$

Therefore,

$$\begin{aligned}\varepsilon < \mid\ Q(\widehat{\boldsymbol{\theta}}) - Q(\boldsymbol{\theta}_0)\ \mid\ &= \mid\ Q(\widehat{\boldsymbol{\theta}}) - \widehat{Q}_n(\boldsymbol{\theta}_0) + \widehat{Q}_n(\boldsymbol{\theta}_0) - Q(\boldsymbol{\theta}_0)\ \mid \\ &\leq 2 \sup_{\boldsymbol{\theta}\in\Theta} \mid \widehat{Q}_n(\boldsymbol{\theta}) - Q(\boldsymbol{\theta})\ \mid\ \to_p 0\end{aligned}$$

where the latter inequality follows from the triangular inequality and the uniform convergence condition (12). This implies that $P(\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| > \delta) \leq P(\mid\ Q(\widehat{\boldsymbol{\theta}}) - Q(\boldsymbol{\theta}_0)\ \mid\ > \varepsilon) \to 0$ as $n \to \infty$. Hence, $\widehat{\boldsymbol{\theta}} \to_p \boldsymbol{\theta}_0$. Before showing asymptotic normality we show that the estimate of the Lagrange multiplier $\frac{\widehat{\boldsymbol{\lambda}}}{b_n^{1+d}}$ converges to zero in probability. By a mean value argument, the uniform convergence results in part 1 and the continuous mapping theorem we get

$$\frac{\widehat{\boldsymbol{\lambda}}}{b_n^{1+d}} \to_p \mathbf{0}.$$

Let us now prove part 2 and define

$$2R_n(\boldsymbol{\theta}, \boldsymbol{\lambda}) = 1 + 2\boldsymbol{\lambda}^\top \widehat{\boldsymbol{m}}(\boldsymbol{\theta}) + \frac{1}{b_n^{1+d}} \boldsymbol{\lambda}^\top \widehat{\boldsymbol{\Sigma}}(\boldsymbol{\theta})^\top \boldsymbol{\lambda}.$$

The first order conditions of $\widehat{R}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\lambda}})$ with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$ are

$$0 = R_{n,\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\lambda}}) = \nabla_{\boldsymbol{\theta}} \widehat{\boldsymbol{m}}(\widehat{\boldsymbol{\theta}}) \widehat{\boldsymbol{\lambda}} + \frac{\boldsymbol{\lambda}^\top}{N b_n^{1+d}} \sum_{\boldsymbol{i}\in\mathcal{I}_{b_n}} \boldsymbol{m_i}(\widehat{\boldsymbol{\theta}}) \nabla_{\boldsymbol{\theta}} \boldsymbol{m_i}(\widehat{\boldsymbol{\theta}}) \widehat{\boldsymbol{\lambda}} \tag{13}$$

$$0 = R_{n,\boldsymbol{\lambda}}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\lambda}}) = \widehat{\boldsymbol{m}}(\widehat{\boldsymbol{\theta}}) + \frac{1}{b_n^{1+d}} \widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\theta}}) \widehat{\boldsymbol{\lambda}}. \tag{14}$$

Let us now take a mean value expansion of the first order conditions (13) and (14) about

the true values $(\boldsymbol{\theta}^\top, \boldsymbol{\lambda}^\top)^\top = (\boldsymbol{\theta}_0^\top, \mathbf{0}^\top)^\top$

$$\begin{aligned}\mathbf{0} = R_{n,\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\lambda}}) &= R_{n,\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \mathbf{0}) + R_{n,\boldsymbol{\theta\lambda}}(\dot{\boldsymbol{\theta}}, \dot{\boldsymbol{\lambda}})\widehat{\boldsymbol{\lambda}} + R_{n,\boldsymbol{\theta\theta}}(\dot{\boldsymbol{\theta}}, \dot{\boldsymbol{\lambda}})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\ &= R_{n,\boldsymbol{\theta\lambda}}(\dot{\boldsymbol{\theta}}, \dot{\boldsymbol{\lambda}})\frac{\sqrt{n}}{b_n^{1+d}}\widehat{\boldsymbol{\lambda}} + \frac{1}{b_n^{1+d}} R_{n,\boldsymbol{\theta\theta}}(\dot{\boldsymbol{\theta}}, \dot{\boldsymbol{\lambda}})\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\end{aligned} \tag{15}$$

$$\begin{aligned}\mathbf{0} = R_{n,\boldsymbol{\lambda}}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\lambda}}) &= R_{n,\boldsymbol{\lambda}}(\boldsymbol{\theta}_0, \mathbf{0}) + R_{n,\boldsymbol{\lambda\lambda}}(\dot{\boldsymbol{\theta}}, \dot{\boldsymbol{\lambda}})\widehat{\boldsymbol{\lambda}} + R_{n,\boldsymbol{\lambda\theta}}(\dot{\boldsymbol{\theta}}, \dot{\boldsymbol{\lambda}})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\ &= \sqrt{n}R_{n,\boldsymbol{\lambda}}(\boldsymbol{\theta}_0, \mathbf{0}) + b_n^{1+d} R_{n,\boldsymbol{\lambda\lambda}}(\dot{\boldsymbol{\theta}}, \dot{\boldsymbol{\lambda}})\frac{\sqrt{n}}{b_n^{1+d}}\widehat{\boldsymbol{\lambda}} + R_{n,\boldsymbol{\lambda\theta}}(\dot{\boldsymbol{\theta}}, \dot{\boldsymbol{\lambda}})\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0).\end{aligned} \tag{16}$$

More compactly,

$$\begin{pmatrix} \mathbf{0} \\ \sqrt{n}\widehat{R}_{\boldsymbol{\lambda}}(\boldsymbol{\theta}_0, \mathbf{0}) \end{pmatrix} = - \begin{pmatrix} \frac{1}{b_n^{1+d}}\widehat{R}_{\boldsymbol{\theta\theta}}(\dot{\boldsymbol{\theta}}, \dot{\boldsymbol{\lambda}}) & \widehat{R}_{\boldsymbol{\theta\lambda}}(\dot{\boldsymbol{\theta}}, \dot{\boldsymbol{\lambda}}) \\ \widehat{R}_{\boldsymbol{\lambda\theta}}(\dot{\boldsymbol{\theta}}, \dot{\boldsymbol{\lambda}}) & b_n^{1+d}\widehat{R}_{\boldsymbol{\lambda\lambda}}(\dot{\boldsymbol{\theta}}, \dot{\boldsymbol{\lambda}}) \end{pmatrix} \begin{pmatrix} \sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\ \frac{\sqrt{n}}{b_n^{1+d}}\widehat{\boldsymbol{\lambda}} \end{pmatrix}.$$

By the unifrom weak law of large numbers we get $\frac{1}{b_n^{1+d}}\widehat{R}_{\boldsymbol{\theta\theta}}(\dot{\boldsymbol{\theta}}, \dot{\boldsymbol{\lambda}}) \to_p \mathbf{0}$, $b_n^{1+d}\widehat{R}_{\boldsymbol{\lambda\lambda}}(\dot{\boldsymbol{\theta}}, \dot{\boldsymbol{\lambda}}) \to_p \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ and $\widehat{R}_{\boldsymbol{\lambda\theta}}(\dot{\boldsymbol{\theta}}, \dot{\boldsymbol{\lambda}}) \to_p \nabla_{\boldsymbol{\theta}}\boldsymbol{m}(\boldsymbol{\theta}_0)$. Hence,

$$\begin{aligned}\begin{pmatrix} \sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\ \frac{\sqrt{n}}{b_n^{1+d}}\widehat{\boldsymbol{\lambda}} \end{pmatrix} = &- \begin{pmatrix} \boldsymbol{\Omega}(\boldsymbol{\theta}_0) & \boldsymbol{\Omega}(\boldsymbol{\theta}_0)\nabla_{\boldsymbol{\theta}}\boldsymbol{m}(\boldsymbol{\theta}_0)^\top \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)^{-1} \\ \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)^{-1}\nabla_{\boldsymbol{\theta}}\boldsymbol{m}(\boldsymbol{\theta}_0)\boldsymbol{\Omega}(\boldsymbol{\theta}_0) & \boldsymbol{\Lambda}(\boldsymbol{\theta}_0) \end{pmatrix} \begin{pmatrix} \mathbf{0} \\ \sqrt{n}\widehat{\boldsymbol{m}}(\boldsymbol{\theta}_0) \end{pmatrix} \\ &+ o_p(1)\end{aligned}$$

where

$$\boldsymbol{\Omega}(\boldsymbol{\theta}_0) = (\nabla_{\boldsymbol{\theta}}\boldsymbol{m}(\boldsymbol{\theta}_0)^\top \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)^{-1}\nabla_{\boldsymbol{\theta}}\boldsymbol{m}(\boldsymbol{\theta}_0))^{-1}$$

and

$$\boldsymbol{\Lambda}(\boldsymbol{\theta}_0) = \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)^{-1} - \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)^{-1}\nabla_{\boldsymbol{\theta}}\boldsymbol{m}(\boldsymbol{\theta}_0)\boldsymbol{\Omega}(\boldsymbol{\theta}_0)\nabla_{\boldsymbol{\theta}}\boldsymbol{m}(\boldsymbol{\theta}_0)^\top \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)^{-1}.$$

The result follows from an application of the central limit theorem and the continuous mapping theorem. □

References

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

Antoine, B., Bonnal, H. & Renault, E. (2007). On the efficient use of the informational content of estimating equations: Implied probabilities and Euclidean empirical likelihood. *Journal of Econometrics*, *138*, 461–487.

Bai, Y., Song, P.-K. & Raghunathan, T. E. (2012). Joint composite estimating functions in spatiotemporal models. *Journal of the Royal Statistical Society, B*, *74*, 799–824.

Banerjee, S., Gelfand, A. E., Finley, A. O. & Sang, H. (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *70*(4), 825–848.

Bevilacqua, M., Caamaño, C., Arellano Valle, R. & Víctor Morales-Oñate, V. (2020). Non-gaussian geostatistical modeling using (skew) t processes. *Scandinavian Journal of Statistics*. (Forthcoming)

Bevilacqua, M., Crudu, F. & Porcu, E. (2015). Combining euclidean and composite likelihood for binary spatial data estimation. *Stochastic environmental research and risk assessment*, *29*(2), 335–346.

Bevilacqua, M., Faouzi, T., Furrer, R. & Porcu, E. (2019). Estimation and prediction using generalized wendland functions under fixed domain asymptotics. *The Annals of Statistics*, *47*, 828-856.

Bevilacqua, M. & Gaetan, C. (2015). Comparing composite likelihood methods based on pairs for spatial Gaussian random fields. *Statistics and Computing*, *25*(5), 877–892.

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

Cressie, N. & Johannesson, G. (2008). Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *70*(1), 209–226.

Cressie, N. & Wikle, C. K. (2015). *Statistics for spatio-temporal data*. John Wiley & Sons.

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

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

Eidsvik, J., Shaby, B. A., Reich, B. J., Wheeler, M. & Niemi, J. (2014). Estimation and prediction in spatial models with block composite likelihoods. *Journal of Computational and Graphical Statistics*, *23*(2), 295–315.

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

Heaton, M. J., Datta, A., Finley, A. O., Furrer, R., Guinness, J., Guhaniyogi, R., ... Zammit-Mangion, A. (2019, 1st Sep). A case study competition among methods for analyzing large spatial data. *Journal of Agricultural, Biological and Environmental Statistics*, *24*(3), 398–425.

Jenish, N. & Prucha, I. R. (2009). Central limit theorems and uniform laws of large numbers for arrays of random fields. *Journal of Econometrics*, 86–98.

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

Kaufman, C. G., Schervish, M. J. & Nychka, D. W. (2008). Covariance tapering for likelihood-based estimation in large spatial data sets. *Journal of the American Statistical Association*, *103*(484), 1545–1555.

Kitamura, Y. (1997). Empirical likelihood methods with weakly dependent processes. *Annals of Statistics*, *25*, 2084-2102.

Lee, A., Yau, C., Giles, M. B., Doucet, A. & Holmes, C. C. (2010). On the utility of graphics cards to perform massively parallel simulation of advanced monte carlo methods. *Journal of computational and graphical statistics*, *19*(4), 769–789.

Lee, Y. D. & Lahiri, S. N. (2002). Least squares variogram fitting by spatial subsampling. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(4), 837–854.

Lindgren, F., Rue, H. & Lindstrom, J. (2011). An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *73*(4), 423-498.

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

Litvinenko, A., Sun, Y., Genton, M. G. & Keyes, D. (2017). Likelihood approximation with hierarchical matrices for large spatial datasets. *arXiv preprint arXiv:1709.04419*.

Matloff, N. (2011). The art of r programming. *No Starch Pess*, *3*.

Morales-Oñate, V., Bevilacqua, M. & Crudu, F. (2019). Stbeu: Spacetime blockwise euclidean likelihood for gaussian models in geostatistics [Computer software manual]. Retrieved from `https://github.com/vmoprojs/STBEU` (R package version 1.0.0)

Nelder, J. & Mead, R. (1965). A simplex method for function minimization. *The Computer Journal*, *7*, 308-313.

Newey, W. & Smith, R. J. (2004). Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica*, *72*, 219–255.

Nordman, D. J. & Caragea, P. C. (2008). Point and interval estimation of variogram models using spatial empirical likelihood. *Journal of the American Statistical Association*, *103*(481), 350–361.

Owen, A. B. (2001). *Empirical Likelihood*. Chapman & Hall/CRC, London.

Porcu, E., M., B. & Genton, M. (2020). Nonseparable, space-time covariance functions with dynamical compact supports. *Statistica Sinica*, *to appear*.

Qin, J. & Lawless, J. (1994). Empirical likelihood and general estimating equations. *Annals of Statistics*, *22*, 300–325.

Rue, H. & Held, L. (2005). *Gaussian markov random fields: theory and applications*. CRC Press.

Rue, H. & Tjelmeland, H. (2002). Fitting Gaussian Markov random fields to Gaussian fields. *Scandinavian Journal of Statistics*, *29*, 31–49.

Sherman, M. (2011). *Spatial statistics and spatio-temporal data: covariance functions and directional properties*. John Wiley & Sons.

Stein, M., Chen, J., Anitescu, M. et al. (2013). Stochastic approximation of score functions for gaussian processes. *The Annals of Applied Statistics*, *7*(2), 1162–1191.

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

Stone, J. E., Gohara, D. & Shi, G. (2010). Opencl: A parallel programming standard for heterogeneous computing systems. *Computing in science & engineering*, *12*(3), 66–73.

Suchard, M. A., Wang, Q., Chan, C., Frelinger, J., Cron, A. & West, M. (2010). Understanding gpu programming for statistical computation: Studies in massively parallel massive mixtures. *Journal of computational and graphical statistics*, *19*(2), 419–438.

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

Wikle, C. K., Zammit-Mangion, A. & Cressie, N. (2019). *Spatio-temporal statistics with r*. CRC Press.

Xu, G. & Genton, M. G. (2017). Tukey g-and-h random fields. *Journal of the American Statistical Association*, 1–14.