Blockwise Euclidean likelihood for spatio-temporal covariance models

Journal Pre-proof

Blockwise Euclidean likelihood for spatio-temporal covariance models

Víctor Morales-Oñate, Federico Crudu, Moreno Bevilacqua

PII: S2452-3062(21)00003-4
DOI: https://doi.org/10.1016/j.ecosta.2021.01.001
Reference: ECOSTA 207

To appear in: *Econometrics and Statistics*

Received date: 11 February 2020
Revised date: 11 January 2021
Accepted date: 12 January 2021

Please cite this article as: Víctor Morales-Oñate, Federico Crudu, Moreno Bevilacqua, Blockwise Euclidean likelihood for spatio-temporal covariance models, *Econometrics and Statistics* (2021), doi: https://doi.org/10.1016/j.ecosta.2021.01.001

Blockwise Euclidean likelihood for spatio-temporal covariance models

Víctor Morales-Oñate[a,*], Federico Crudu[b], Moreno Bevilacqua[c]

[a] *Colegio de Administración y Economía, Universidad San Francisco de Quito, Ecuador*
Banco Solidario, Risk Division, Data Analytics, Quito, Ecuador
Data Science Research Group - CIDED, Escuela Superior Politécnica de Chimborazo
Territorial Development, Business and Innovation Research Group - DeTEI, Universidad Técnica de Ambato
[b]*Department of Economics and Statistics, University of Siena, Piazza San Francesco, 7/8 53100 Siena, Italy*
Università di Siena
CRENoS
[c]*Facultad de Ingeniería y Ciencias, Universidad Adolfo Ibáñez, Viña del Mar, Chile*
ANID-Millennium Science Initiative Program-Millennium Nucleus Center for the Discovery of Structures in Complex Data

Abstract

A spatio-temporal blockwise Euclidean likelihood method for the estimation of covariance models when dealing with large spatio-temporal Gaussian data is proposed. 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, a general purpose graphics processing unit implementation using OpenCL is implemented. The asymptotic properties of the proposed estimator are derived and the finite sample properties of this methodology by means of a simulation study highlighting the computational gains of the OpenCL graphics processing unit implementation. Finally, there is an application of the estimation method to a wind component data set.

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

1. Introduction

With the advent and expansion of Geographical Information Systems (GIS) along with related software, statisticians today routinely encounter large spatial or spatio-

☆For reproducible research purposes, we developed the R package `STBEU` (Morales-Oñate et al., 2019) that includes the full code.

*Corresponding author

Email addresses: `victor.morales@uv.cl` (Víctor Morales-Oñate), `federico.crudu@unisi.it` (Federico Crudu), `moreno.bevilacqua@uai.cl` (Moreno Bevilacqua)

URL: `https://sites.google.com/site/moralesonatevictor/` (Víctor Morales-Oñate), `https://sites.google.com/site/federicocrudu/` (Federico Crudu), `https://sites.google.com/a/uv.cl/moreno-bevilacqua/` (Moreno Bevilacqua)

Preprint submitted to Econometrics and Statistics *January 26, 2021*

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, Cressie & Wikle (2015) and Sherman (2011).

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, Bevilacqua et al. (2020), De Oliveira et al. (1997) and Xu & Genton (2017)). 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 (Cressie & Johannesson, 2008; Furrer et al., 2006; Kaufman et al., 2008; Litvinenko et al., 2017), stochastic approximations of the score function (Stein et al., 2013), approximations based on Markov RFs (Lindgren et al., 2011; Rue & Held, 2005; Rue & Tjelmeland, 2002), Gaussian predictive process (Banerjee et al., 2008) or on the composite likelihood idea (Bai et al., 2012; Bevilacqua & Gaetan, 2015; Bevilacqua et al., 2012; Eidsvik et al., 2014) and the so-called Vecchia approximations (Katzfuss & Guinness, 2020; Stein et al., 2004) among others. Another interesting proposal merging a parametric and non parametric approach can be found in Ma & Kang (2020). 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.: Kitamura, 1997; Newey & Smith, 2004; Nordman & Caragea, 2008; 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 (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 $\rho(\boldsymbol{0}, 0) = 1$ that only depends on $\boldsymbol{h} = \boldsymbol{s}' - \boldsymbol{s}$ and $u = t' - t$. Additionally, in the remainder of the paper we assume a zero nugget effect. 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 $\rho(\boldsymbol{h}, u) = \phi(\|\boldsymbol{h}\|, |u|)$.[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 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 $\widehat{\boldsymbol{\theta}}_{PL} = \mathrm{argmax}_{\boldsymbol{\theta} \in \Theta}\, pl(\boldsymbol{\theta})$. 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).

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

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

[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.

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,

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

where $\boldsymbol{\theta}_0$ is unique.[2] The moment condition in Equation (4) is one of the building blocks of our approach.

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 Bevilacqua & Gaetan (2015), Davis & Yau (2011) and Joe & Lee (2009), 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. A theoretical guideline on how to choose d_s and d_t is given in Bevilacqua et al. (2012) but its implementation is computationally demanding. In practice, the choice of d_s or d_t depends on the problem at hand and on the size of the dataset. A rule of thumb is to fix d_s or d_t as a small proportion of the maximum spatial and temporal distances (Bevilacqua et al., 2012).

The recent literature on the topic has put forward alternative and more statistically efficient weighting schemes (see, e.g., Pace et al., 2019). However, also in this case, their practical implementation is computationally demanding.

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 $\mathcal{U} = \left(-\frac{1}{2}, \frac{1}{2}\right]^d \times (0, 1]$ (see e.g. Nordman & Caragea, 2008). Then, a $(1+d)$-dimensional block is defined as

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

[2] The assumption that $\boldsymbol{\theta}_0$ be unique is rather standard in the literature (see e.g. Bevilacqua et al., 2012), it may be, though, problematic to maintain when dealing with complex models such as those treated in this paper. In this case a researcher may invoke some modifications that accommodate for the presence of multiple optima (see for example Van der Vaart, 2007, Section 5.2.1). It is possible to adapt the standard proof for the consistency of M-estimators to the presence of multiple optima. In particular, one can define a set of population optima, say, $\Theta_0 \in \Theta$ and show, under fairly standard assumptions, that, for every $\epsilon > 0$ and every compact set $\mathcal{K} \subset \Theta$, $P(\mathsf{d}(\widehat{\boldsymbol{\theta}}, \Theta_0) \geq \epsilon \wedge \widehat{\boldsymbol{\theta}} \in \mathcal{K}) \to 0$ where $\widehat{\boldsymbol{\theta}}$ is an M-estimator and $\mathsf{d}(\cdot, \cdot)$ measures the distance between a point and a set. Further details can be found in Theorem 5.14 in Van der Vaart (2007).

Notice that the set $\mathcal{U}$ is a $(1+d)$-dimensional square and can be seen as the prototypical space for the construction of the generic block $\mathcal{B}_{b_n}(\boldsymbol{\kappa})$. The size of the block depends on b_n while its position depends on the point of coordinates $\boldsymbol{\kappa}$. 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 $b_n^{1+d} = |\mathcal{D}_{b_n}|$,

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

The STBEU objective function is defined as

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

(see Antoine et al., 2007).[3] 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

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

and

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

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]The auxiliary parameter $\boldsymbol{\lambda}$ comes from the fact that the EU estimator is a member of the generalized empirical likelihood family of estimators. These estimators admit a dual representation as the solution of a Lagrangian optimization problem. The parameter vector $\boldsymbol{\lambda}$ is related to the corresponding Lagrange multipliers. See Newey & Smith (2004) for some general results on generalized empirical likelihood, Bevilacqua et al. (2015) for an application of EU to the spatial case and Owen (2001) for a textbook treatment of the problem.

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 $\varsigma(\cdot,\cdot)$, we have $\varsigma(\boldsymbol{l},\boldsymbol{k}) \geq d_0$ with $d_0 > 0$.

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 $\nabla_{\theta}\boldsymbol{m}(\boldsymbol{\theta}_0)$ is full column rank;

(iv) $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\theta}_0) \to_p \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$, 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 $\varsigma(\boldsymbol{l},\boldsymbol{k}) = \max_{1\leq i\leq 1+d}|l_i - k_i|$. In addition to that define the set distance as $\varsigma(\mathcal{V},\mathcal{W}) = \inf\{\varsigma(\boldsymbol{l},\boldsymbol{k}) : \boldsymbol{l} \in \mathcal{V}, \boldsymbol{k} \in \mathcal{W}\}$ for any subset $\mathcal{V},\mathcal{W} \subset \mathbb{R}^d \times \mathbb{R}^+$. Then, the α-mixing coefficient for the RF is given by

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

where

$$\alpha(\mathcal{V},\mathcal{W}) = \sup\left(|P(\mathcal{A}\cap\mathcal{B}) - P(\mathcal{A})P(\mathcal{B})|; \mathcal{A} \in \sigma(\mathcal{V}), \mathcal{B} \in \sigma(\mathcal{W})\right).$$

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 $\varepsilon > 0$.

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. Assumption A4 together with assumptions A1, A2 and the α-mixing condition A6 allows 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.

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

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 $[-a, a]^2$ in both directions 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 RF 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 Bevilacqua et al. (2015) and Lee & Lahiri (2002), 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.

As outlined in Section 2, the efficiency of the method depends on the choice of d_s and d_t. Here we follow Bevilacqua et al. (2012) for the choice of these two parameters and we fix the distances d_s and d_t in the weight function (5) to be 25% of their 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.

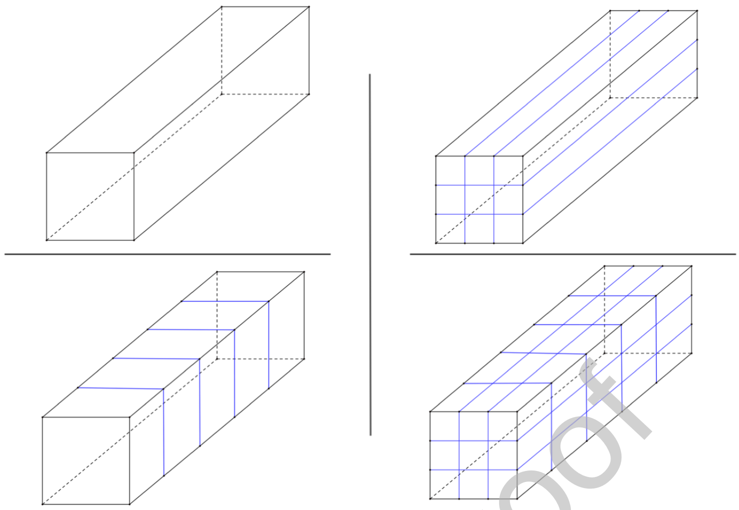

Figure 1: Intuition behind the spatio-temporal blocking procedure

	Double exponential				Gneiting			
	Regular		Irregular		Regular		Irregular	
	$b=2$	$b=4$	$b=2$	$b=4$	$b=2$	$b=4$	$b=2$	$b=4$
	$\alpha_s = 1.2/3$ $\alpha_t = 1.2/3$				$\alpha_s = 1.2/3$ $\alpha_t = 1.2/3$			
α_s	1.015	0.961	0.765	0.824	1.133	1.009	0.799	0.865
	(1.035)	(0.988)	(0.666)	(0.720)	(1.051)	(1.094)	(0.706)	(0.768)
α_t	0.949	0.912	0.651	0.751	1.101	1.051	0.714	0.800
	(0.916)	(0.954)	(0.570)	(0.667)	(1.131)	(1.115)	(0.571)	(0.715)
σ^2	0.997	0.957	0.651	0.846	1.006	0.878	0.665	0.825
	(1.071)	(0.975)	(0.554)	(0.764)	(0.930)	(0.983)	(0.548)	(0.756)
$STRE$	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)
	$\alpha_s = 1.8/3$ $\alpha_t = 1.8/3$				$\alpha_s = 1.8/3$ $\alpha_t = 1.8/19$			
α_s	1.008	1.005	0.832	0.801	1.188	1.087	0.897	0.902
	(1.021)	(1.084)	(0.625)	(0.706)	(1.212)	(1.224)	(0.759)	(0.835)
α_t	1.012	0.975	0.730	0.824	1.195	1.074	0.830	0.891
	(1.038)	(1.020)	(0.617)	(0.723)	(1.292)	(1.156)	(0.647)	(0.885)
σ^2	0.993	0.893	0.688	0.853	0.980	0.923	0.682	0.851
	(1.001)	(0.987)	(0.588)	(0.763)	(0.996)	(0.981)	(0.595)	(0.795)
$STRE$	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, i.e. $SRE = \frac{mad_{PL}}{mad_{STBEU}}$) 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.

	Double exponential				Gneiting			
	Regular		Irregular		Regular		Irregular	
	$b=2$	$b=3$	$b=2$	$b=3$	$b=2$	$b=3$	$b=2$	$b=3$
	$\alpha_s = 3.1/3$ $\alpha_t = 3.1/3$				$\alpha_s = 3.1/3$ $\alpha_t = 3.1/19$			
α_s	1.111	0.827	1.072	0.787	1.038	0.842	1.035	0.811
	(0.752)	(0.619)	(0.674)	(0.593)	(0.676)	(0.591)	(0.690)	(0.555)
α_t	1.134	1.010	1.218	0.945	1.323	1.205	1.332	0.972
	(0.888)	(0.752)	(0.818)	(0.709)	(1.088)	(0.932)	(0.820)	(0.763)
σ^2	0.966	0.739	1.055	0.741	0.984	0.749	1.018	0.746
	(0.610)	(0.504)	(0.657)	(0.544)	(0.670)	(0.568)	(0.655)	(0.537)
$STRE$	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)
	$\alpha_s = 4/3$ $\alpha_t = 4/3$				$\alpha_s = 4/3$ $\alpha_t = 4/19$			
α_s	1.119	0.825	1.074	0.759	0.981	0.769	1.033	0.798
	(0.786)	(0.623)	(0.690)	(0.584)	(0.634)	(0.548)	(0.690)	(0.559)
α_t	1.184	1.045	1.264	0.956	1.400	1.377	1.357	1.074
	(0.946)	(0.804)	(0.852)	(0.702)	(1.166)	(1.020)	(0.894)	(0.836)
σ^2	0.955	0.756	1.090	0.743	1.008	0.775	1.014	0.768
	(0.627)	(0.542)	(0.670)	(0.540)	(0.675)	(0.559)	(0.651)	(0.518)
$STRE$	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, i.e. $SRE = \frac{mad_{PL}}{mad_{STBEU}}$) 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.

	Double exponential				Gneiting			
	Regular		Irregular		Regular		Irregular	
	$b_{st}=4$	$b_{st}=9$	$b_{st}=4$	$b_{st}=9$	$b_{st}=4$	$b_{st}=9$	$b_{st}=4$	$b_{st}=9$
	$\alpha_s=3.1/3$ $\alpha_t=3.1/3$				$\alpha_s=3/3$ $\alpha_t=3/19$			
α_s	1.283	0.918	1.034	0.842	1.187	0.890	0.970	0.784
	(1.054)	(0.692)	(0.584)	(0.625)	(0.899)	(0.674)	(0.574)	(0.684)
α_t	1.494	0.914	1.042	0.913	1.786	1.177	1.166	1.039
	(1.030)	(0.723)	(0.632)	(0.707)	(1.297)	(0.941)	(0.729)	(0.809)
σ^2	0.951	0.721	0.759	0.711	1.024	0.782	0.746	0.699
	(0.773)	(0.642)	(0.464)	(0.569)	(0.815)	(0.689)	(0.462)	(0.566)
$STRE$	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)
	$\alpha_s=4/3$ $\alpha_t=4/3$				$\alpha_s=4/3$ $\alpha_t=4/19$			
α_s	1.244	0.876	1.112	0.891	0.897	0.708	1.022	0.853
	(1.034)	(0.631)	(0.638)	(0.701)	(0.659)	(0.501)	(0.565)	(0.655)
α_t	1.507	0.976	1.176	1.020	1.223	1.055	1.318	1.121
	(1.122)	(0.794)	(0.715)	(0.785)	(1.009)	(0.744)	(0.818)	(0.928)
σ^2	0.991	0.722	0.793	0.690	0.738	0.706	0.749	0.686
	(0.814)	(0.652)	(0.527)	(0.600)	(0.650)	(0.533)	(0.529)	(0.610)
$STRE$	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, i.e. $SRE = \frac{mad_{PL}}{mad_{STBEU}}$) 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.

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 $SRE = \frac{mad_{PL}}{mad_{STBEU}}$, where mad_{PL} and mad_{STBEU} are the median absolute deviations associated with PL and STBEU estimators respectively. SRE is reported for every parameter and scenario. The choice of the mad as a measure of statistical efficiency is due to the fact that the STBEU estimator may display fatter tails than its competitor (see, e.g., Hansen et al., 1996, for a similar situation in a different context).[4] 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 $STRE = \left(\frac{D_{PL}}{D_{STBEU}}\right)^{1/d_{\boldsymbol{\theta}}}$ where $d_{\boldsymbol{\theta}} = 3$ 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 specifications STBEU tends to outperform 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

[4]In Appendix C we also provide results for performance measures based on the mean square error and on the nine decile range.

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), 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.

We additionally consider a simulation study using a special case of the spatio-temporal Wendland correlation function proposed in Porcu et al. (2020):

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

where $\boldsymbol{\theta} = (\sigma^2, \alpha_s, \alpha_t, \beta)^\top$ (see Appendix E for the corresponding code). This covariance model is compactly supported in time and has some computational benefits with respect to the covariance models (1) and (2) since the associated covariance matrix is sparse. We use this model in the application in Section 5. The case $\beta = 0$ implies a separable spatio-temporal covariance and the case $0 < \beta \leq 1$ leads to a non separable covariance function We opt for a non overlapping regular spatial blocking setting ($b_s = 0.2$) with $n_s = 400$ and $n_t = 10$, a total of $n_{st} = 4000$. The distances in the weight function are set to $d_s = 0.06$ and $d_t = 3$. Figure 2 shows the boxplots of the estimated parameters. As a general comment, the distribution of the estimates for the four parameters tends to be symmetric and with very few outliers.

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 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)[5]. 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 process data.

[5]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.

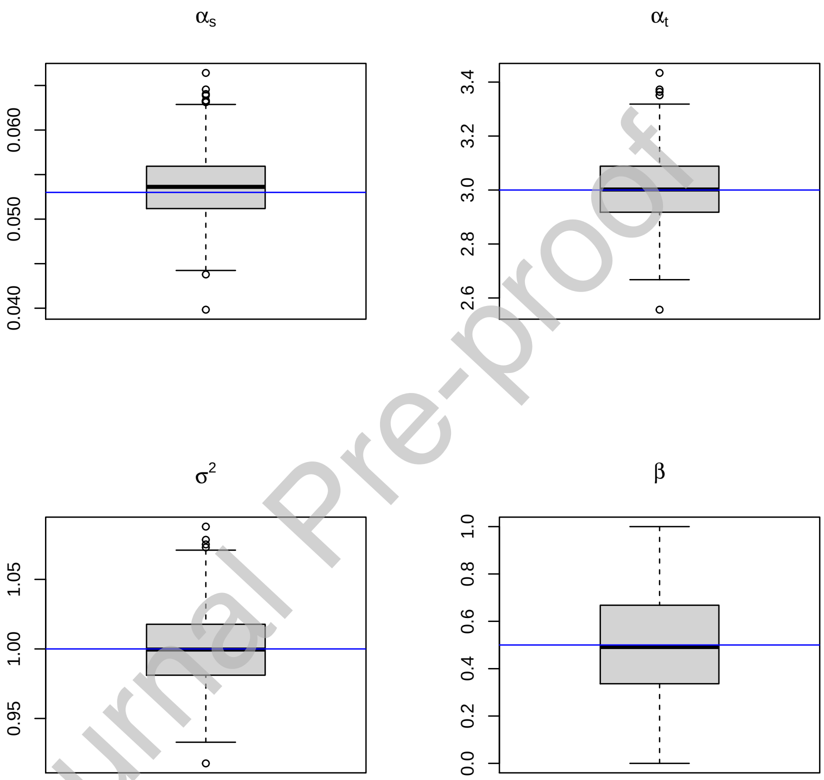

Figure 2: Boxplots of the parameters of the space time Wendland covariance model in equation (11), using a non overlapping regular spatial blocking setting.

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 $n_{st} = 32768$ 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.

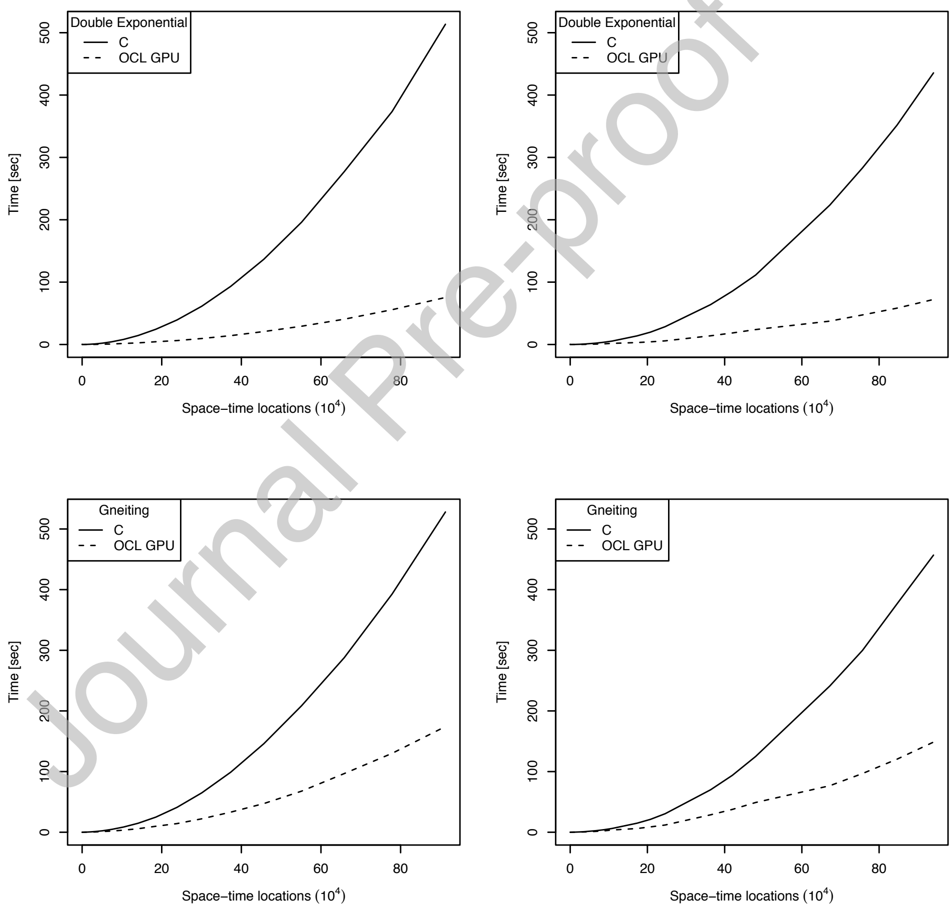

Figure 3: Gradient ($\boldsymbol{g}_{ij}$) evaluation time performance comparison C vs OpenCL (denoted OCL) for Double Exponential and Gneiting covariance functions. Space locations vary from 4 to 9409 and time locations from 2 to 97 on the left panel, the opposite in the right panel.

Figure 3 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 specifications, 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.

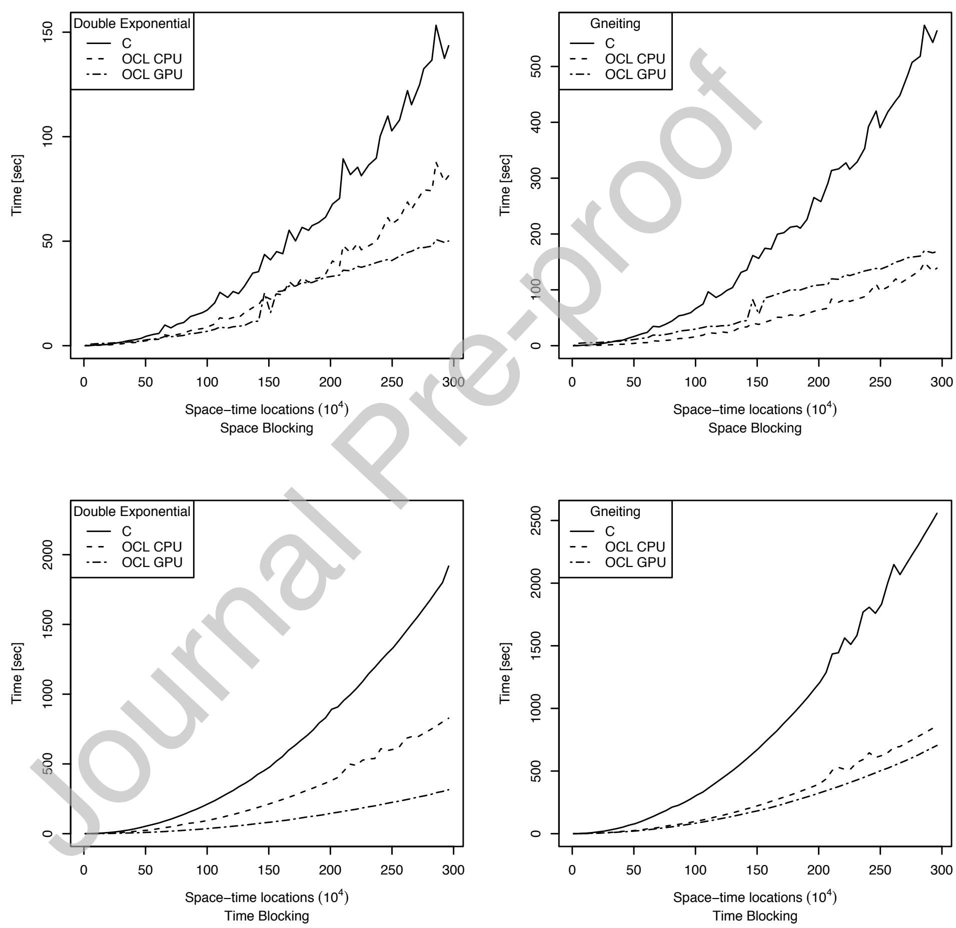

Figure 4: Blockwise time performance comparison for C vs OpenCL (denoted OCL with CPU and GPU). The x axis is divided to $10e4$. Rows compare spatial vs temporal blocking and columns compare the correlation model.

Rows from Figure 4 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 4 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 time performance. Rows from Figure 4 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 4) 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 four panels.

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 5 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 for this application.

Figure 5: 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.

STBEU			
Parameters	α_s	α_t	σ^2
$\beta = 0$	385.73	16.93	13.45
	(2.90)	(0.26)	(0.27)
$\beta = 0.5$	386.69	17.07	13.38
	(4.07)	(0.27)	(0.38)
$\beta = 1$	399.37	16.62	13.16
	(26.45)	(0.32)	(0.27)
PL			
$\beta = 0$	351.79	18.44	12.03
	(19.75)	(1.47)	(0.87)
$\beta = 0.5$	352.91	18.45	12.03
	(19.82)	(1.47)	(0.87)
$\beta = 1$	354.04	18.47	12.03
	(19.91)	(1.48)	(0.87)

Table 5: Estimation results of the spatio-temporal Gaussian process with Wendland covariance model (11) using Mediterranean wind data with STBEU and PL for $\beta = 0, 0.5, 1$. Standard errors are shown in parenthesis.

Scenario	Elapsed time	Time Gain (with 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).

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 intrduced Equation (11).

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 are reported in Table 5 while standard errors are shown in parenthesis. In terms of magnitude, the estimated coefficients for STBEU are not too susceptible to the choice of β. With respect to PL, we notice some sizable difference in the estimates of α_t both in comparison with STBEU and in relation with the choice of

β. Furthermore, the estimates of the other parameters, α_t and σ^2, tend to be similar for the two estimation methods and with respect to the choice of β. In addition to that, we observe that the standard errors for σ^2 are systematically smaller for PL. However, identifying a similar pattern for the other parameters seems to be more complicated. We notice, though, that for $\beta = 0.5$ the standard errors for STBEU are much smaller than those produced by PL.

Additionally, in Figure 6, 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. The shaded area between the solid lines represents the confidence band for the STBEU while that between the dotted lines the PL. Additional information about the standard errors are presented in Appendix B.

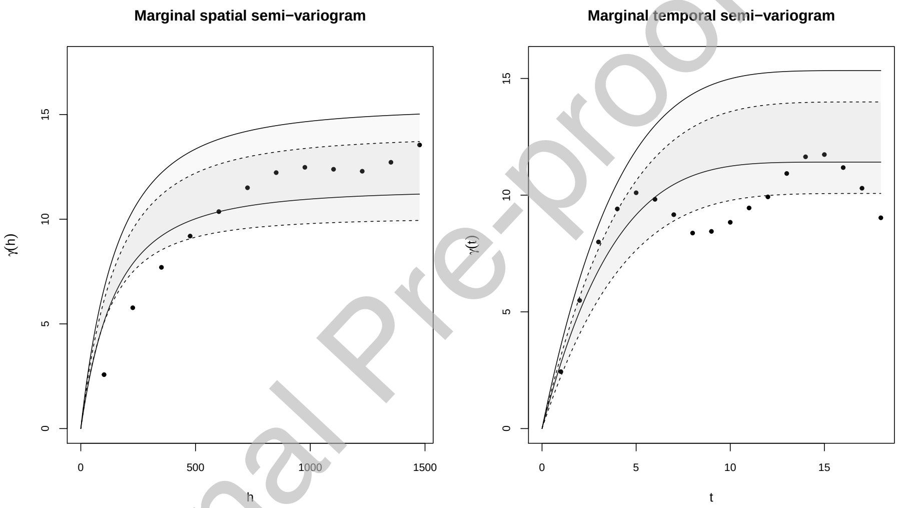

Figure 6: Confidence bands for the empirical spatial and temporal marginal semi-variogram versus the estimated semi-variograms for model (11) with $\beta = 0.5$ using STBEU (solid line) and PL (dotted line) estimates.

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 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 RFs. 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 RFs, the proposed methodology can be easily extended to the case of the estimation of spatio-temporal non-Gaussian RFs 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 ANID-Millennium Science Initiative Program-NCN17_059 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 and Territorial Development, Business and Innovation Research Group -DeTEI at Universidad Técnica de Ambato.

Appendix

Appendix 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 $\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}}$, $\nabla_{\boldsymbol{\lambda}\boldsymbol{\lambda}}$ and $\nabla_{\boldsymbol{\theta}\boldsymbol{\lambda}}$ 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

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

Hence, by taking the norm and by triangle inequality

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

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

$$
\sup_{\boldsymbol{\theta}\in\Theta}\mid\widehat{Q}_n(\boldsymbol{\theta})-Q(\boldsymbol{\theta})\mid\ \to_p 0. \tag{A.1}
$$

Therefore,

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

where the latter inequality follows from the triangular inequality and the uniform convergence condition (A.1). 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

$$
0=R_{n,\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\lambda}})=\nabla_{\boldsymbol{\theta}}\widehat{\boldsymbol{m}}(\widehat{\boldsymbol{\theta}})\widehat{\boldsymbol{\lambda}}+\frac{\boldsymbol{\lambda}^\top}{Nb_n^{1+d}}\sum_{\boldsymbol{i}\in\mathcal{I}_{b_n}}\boldsymbol{m_i}(\widehat{\boldsymbol{\theta}})\nabla_{\boldsymbol{\theta}}\boldsymbol{m_i}(\widehat{\boldsymbol{\theta}})\widehat{\boldsymbol{\lambda}}, \tag{A.2}
$$

$$
0=R_{n,\boldsymbol{\lambda}}(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\lambda}})=\widehat{\boldsymbol{m}}(\widehat{\boldsymbol{\theta}})+\frac{1}{b_n^{1+d}}\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\theta}})\widehat{\boldsymbol{\lambda}}. \tag{A.3}
$$

Let us now take a mean value expansion of the first order conditions (A.2) and (A.3) about the true values $(\boldsymbol{\theta}^\top, \boldsymbol{\lambda}^\top)^\top = (\boldsymbol{\theta}_0^\top, \mathbf{0}^\top)^\top$

$$\begin{aligned}\mathbf{0} = R_{n,\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\lambda}}) &= R_{n,\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \mathbf{0}) + R_{n,\boldsymbol{\theta\lambda}}(\dot{\boldsymbol{\theta}}, \dot{\boldsymbol{\lambda}})\widehat{\boldsymbol{\lambda}} + R_{n,\boldsymbol{\theta\theta}}(\dot{\boldsymbol{\theta}}, \dot{\boldsymbol{\lambda}})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\ &= R_{n,\boldsymbol{\theta\lambda}}(\dot{\boldsymbol{\theta}}, \dot{\boldsymbol{\lambda}})\frac{\sqrt{n}}{b_n^{1+d}}\widehat{\boldsymbol{\lambda}} + \frac{1}{b_n^{1+d}}R_{n,\boldsymbol{\theta\theta}}(\dot{\boldsymbol{\theta}}, \dot{\boldsymbol{\lambda}})\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0),\end{aligned} \tag{A.4}$$

$$\begin{aligned}\mathbf{0} = R_{n,\boldsymbol{\lambda}}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\lambda}}) &= R_{n,\boldsymbol{\lambda}}(\boldsymbol{\theta}_0, \mathbf{0}) + R_{n,\boldsymbol{\lambda\lambda}}(\dot{\boldsymbol{\theta}}, \dot{\boldsymbol{\lambda}})\widehat{\boldsymbol{\lambda}} + R_{n,\boldsymbol{\lambda\theta}}(\dot{\boldsymbol{\theta}}, \dot{\boldsymbol{\lambda}})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\ &= \sqrt{n}R_{n,\boldsymbol{\lambda}}(\boldsymbol{\theta}_0, \mathbf{0}) + b_n^{1+d}R_{n,\boldsymbol{\lambda\lambda}}(\dot{\boldsymbol{\theta}}, \dot{\boldsymbol{\lambda}})\frac{\sqrt{n}}{b_n^{1+d}}\widehat{\boldsymbol{\lambda}} + R_{n,\boldsymbol{\lambda\theta}}(\dot{\boldsymbol{\theta}}, \dot{\boldsymbol{\lambda}})\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0).\end{aligned} \tag{A.5}$$

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,

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

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.

□

Appendix B. Standard errors

In this section we show via simulation the performance of the STBEU estimator for a space-time Gaussian RF with Double Exponential covariance function in terms of confidence intervals. Theorem 1 gives us an expression for the covariance matrix of $\widehat{\boldsymbol{\theta}}$:

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

Using this formula we compute the corresponding standard errors. Table B.7 shows the coverage rates for the parameters of interest obtained by the simulation experiment detailed below.

α_s	α_s	σ^2
95.6%	95.8%	93.8%

Table B.7: Coverage rates for the Monte Carlo experiment using a space time Double Exponential covariance function. The number of Monte Carlo replications is set to 1000.

In what follows we first show as an example how to simulate a realization of a space-time Gaussian RF with Double Exponential covariance function and how to calculate the 95% confidence interval. We start by creating the grid and the data:

```
# ST: Creating grid & Data:
rm(list = ls())
graphics.off()
library(GeoModels)
library(STBEU)

#######################################################
type_dist=1 ### type of distance 1:euclidean
type_subs=1 ### type of subsampling 1=in space 2= in time

scale_t=0.6
scale_s=0.6

sill=1
nugget=0
mean=0.05

start=list(scale_s=scale_s,scale_t=scale_t,sill=sill)
fix=c( nugget=nugget,mean = mean)

####location sites ####################################
lambda=6
xx=seq(-lambda,lambda);
coords=as.matrix(expand.grid(xx,xx))    ###regular

####temporal instants #################################
nt = 4
times=seq(1,nt,1)
(NT <- nrow(coords)*nt)

param=list(scale_s=scale_s,scale_t=scale_t,
sill=sill,nugget=nugget,mean = mean)
maxdist1=max(dist(coords))*.25
maxtime1=ceiling(max(dist(times))*.25)

winc=c(0,0)   ### length of temporal window
winstp=1 ###   0.5 half overlapping  1 "no" overlapping
```

```
cc = 1 #Exp_Exp

datos <- GeoSim(coordx=coords,coordt=times,
corrmodel="Exp_Exp", param=param)$data
weighted = 0
# END: Creating grid & Data
```

We are now ready to estimate the model and calculate the 95% confidence interval:

```
# ST: Starting values & estimation:
res1=STBEUFit(start,fix,coords,times,cc,datos,
              type_dist,maxdist1 ,maxtime1,
              winc,winstp,0,0,type_subs,weighted,
              GPU =0, local = c(1,1),varest = TRUE)

NT = nrow(coords)*length(times)
qq <- qnorm(0.975)

# ******** SCALE S
c(res1$par[1] - qq*res1$stderr[1],
res1$par[1] + qq*res1$stderr[1])
# ******** SCALE T
c(res1$par[2] - qq*res1$stderr[2],
res1$par[2] + qq*res1$stderr[2])
# ******** SILL
c(res1$par[3] - qq*res1$stderr[3],
res1$par[3] + qq*res1$stderr[3])
# END: Starting values & estimation:
```

In order to determine how close the simulated coverage is to the nominal 95% coverage, we now simulate this process a 1000 times.

```
#### ST: Simulation:

SolPar <- NULL
SolSD <- NULL

semilla = 1537
set.seed(semilla)
i = 1
nsim = 1000
while(i <=nsim)
{
  dd <- GeoSim(coordx=coords,coordt=times,
  corrmodel="Exp_Exp", param=param)$data

  tryCatch({aux=STBEUFit(start,fix,coords,times,cc,dd,
        type_dist,maxdist1 ,maxtime1,
        winc,winstp,0,0,type_subs,weighted,
```

```
        GPU =0,  local = c(1,1), varest = TRUE)},
  error=function(e){cat("ERROR :",
  conditionMessage(e), "\n")})

  SolPar <- rbind(SolPar,aux$par)
  SolSD <- rbind(SolSD,aux$stderr)
  cat("Iter: ",i,"de: ",nsim,"\n")
  i = i+1
}
#### END Simulation:
```

The final step is to evaluate the coverage rate.

```
mm <- SolPar
se <- SolSD

qq <- qnorm(0.975)

# ******** SCALE S
lo.conf.scale_s <- mm[,1] - qq*se[,1]
up.conf.scale_s <- mm[,1] + qq*se[,1]

bl.scale_s <- sum(start$scale_s<lo.conf.scale_s) # bad lower
bu.scale_s <- sum(up.conf.scale_s<start$scale_s) # bad upper

1-(bl.scale_s+bu.scale_s)/nsim # should be close to 1

# ******** SCALE T
lo.conf.scale_t <- mm[,2] - qq*se[,2]
up.conf.scale_t <- mm[,2] + qq*se[,2]

bl.scale_t <- sum(start$scale_t<lo.conf.scale_t) # bad lower
bu.scale_t <- sum(up.conf.scale_t<start$scale_t) # bad upper
1-(bl.scale_t+bu.scale_t)/nsim # should be close to 1



# ******** SILL
lo.conf.sill <- mm[,3] - qq*se[,3]
up.conf.sill <- mm[,3] + qq*se[,3]

bl.sill <- sum(start$sill<lo.conf.sill) # bad lower
bu.sill <- sum(up.conf.sill<start$sill) # bad upper
1-(bl.sill+bu.sill)/nsim # should be close to 1
```

For reproducible research purposes, we developed the R package `STBEU` (Morales-Oñate et al., 2019) that includes the full code for this application.

Appendix C. Simulated relative efficiency

This section shows the relative efficiency results for STBEU and PL considering as statistical efficiency measures $SRE = \frac{mse_{PL}}{mse_{STBEU}}$ and $SRE = \frac{ndr_{PL}}{ndr_{STBEU}}$, where mse and ndr stand for mean square error and nine decile range respectively. Like the mad, the ndr is robust to the presence of outliers (see, for example, Bekker & Crudu, 2015; Hausman et al., 2012). Comparing the results in Tables C.8, C.9 and C.10 against those in Tables C.11, C.12 and C.13 we may reasonably conjecture that extreme values affect the results when the relative performance measure is based on the mse. On the other hand, there seems to be no substantial qualitative difference when we use the ndr versus the mad.

	Double exponential				Gneiting			
	Regular		Irregular		Regular		Irregular	
	$b=2$	$b=4$	$b=2$	$b=4$	$b=2$	$b=4$	$b=2$	$b=4$
	$\alpha_s = 1.2/3$ $\alpha_t = 1.2/3$				$\alpha_s = 1.2/3$ $\alpha_t = 1.2/3$			
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)
	$\alpha_s = 1.8/3$ $\alpha_t = 1.8/3$				$\alpha_s = 1.8/3$ $\alpha_t = 1.8/19$			
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 C.8: Simulated relative efficiency (with respect to the PL, i.e. $SRE = \frac{mse_{PL}}{mse_{STBEU}}$) 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.

	Double exponential				Gneiting			
	Regular		Irregular		Regular		Irregular	
	$b=2$	$b=3$	$b=2$	$b=3$	$b=2$	$b=3$	$b=2$	$b=3$
	$\alpha_s=3.1/3$ $\alpha_t=3.1/3$				$\alpha_s=3.1/3$ $\alpha_t=3.1/19$			
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)
	$\alpha_s=4/3$ $\alpha_t=4/3$				$\alpha_s=4/3$ $\alpha_t=4/19$			
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 C.9: Simulated relative efficiency (with respect to the PL, i.e. $SRE = \frac{mse_{PL}}{mse_{STBEU}}$) 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.

	Double exponential				Gneiting			
	Regular		Irregular		Regular		Irregular	
	$b_{st}=4$	$b_{st}=9$	$b_{st}=4$	$b_{st}=9$	$b_{st}=4$	$b_{st}=9$	$b_{st}=4$	$b_{st}=9$
	$\alpha_s=3.1/3$ $\alpha_t=3.1/3$				$\alpha_s=3/3$ $\alpha_t=3/19$			
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)
	$\alpha_s=4/3$ $\alpha_t=4/3$				$\alpha_s=4/3$ $\alpha_t=4/19$			
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 C.10: Simulated relative efficiency (with respect to the PL, i.e. $SRE = \frac{mse_{PL}}{mse_{STBEU}}$) 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.

	Double exponential				Gneiting			
	Regular		Irregular		Regular		Irregular	
	$b=2$	$b=4$	$b=2$	$b=4$	$b=2$	$b=4$	$b=2$	$b=4$
	$\alpha_s = 1.2/3$ $\alpha_t = 1.2/3$				$\alpha_s = 1.2/3$ $\alpha_t = 1.2/3$			
α_s	0.929	0.892	0.716	0.789	1.085	1.006	0.779	0.875
	(0.968)	(0.979)	(0.567)	(0.722)	(1.082)	(1.120)	(0.617)	(0.797)
α_t	0.944	0.939	0.620	0.795	1.096	0.996	0.777	0.885
	(0.993)	(0.941)	(0.491)	(0.677)	(1.075)	(1.029)	(0.641)	(0.819)
σ^2	0.903	0.969	0.628	0.845	0.950	0.926	0.614	0.843
	(0.918)	(0.977)	(0.559)	(0.753)	(0.939)	(0.974)	(0.548)	(0.733)
$STRE$	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)
	$\alpha_s = 1.8/3$ $\alpha_t = 1.8/3$				$\alpha_s = 1.8/3$ $\alpha_t = 1.8/19$			
α_s	1.070	0.939	0.721	0.784	1.294	1.120	0.844	0.904
	(1.095)	(1.031)	(0.558)	(0.765)	(1.316)	(1.203)	(0.687)	(0.885)
α_t	0.993	0.911	0.737	0.832	1.294	1.116	0.896	0.953
	(0.996)	(0.974)	(0.571)	(0.753)	(1.314)	(1.209)	(0.712)	(0.897)
σ^2	0.975	0.949	0.590	0.753	0.949	0.925	0.577	0.775
	(0.953)	(0.990)	(0.518)	(0.709)	(0.912)	(0.916)	(0.487)	(0.725)
$STRE$	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 C.11: Simulated relative efficiency (with respect to the PL, i.e. $SRE = \frac{ndr_{PL}}{ndr_{STBEU}}$) 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.

	Double exponential				Gneiting			
	Regular		Irregular		Regular		Irregular	
	$b = 2$	$b = 3$	$b = 2$	$b = 3$	$b = 2$	$b = 3$	$b = 2$	$b = 3$
	$\alpha_s = 3.1/3$ $\alpha_t = 3.1/3$				$\alpha_s = 3.1/3$ $\alpha_t = 3.1/19$			
α_s	1.109	0.815	1.059	0.843	1.058	0.802	0.976	0.782
	(0.721)	(0.616)	(0.723)	(0.594)	(0.728)	(0.647)	(0.695)	(0.584)
α_t	1.193	0.965	1.162	0.953	1.503	1.297	1.270	1.081
	(0.886)	(0.672)	(0.841)	(0.725)	(1.177)	(0.916)	(0.901)	(0.782)
σ^2	1.013	0.844	1.003	0.819	1.037	0.842	0.989	0.809
	(0.680)	(0.572)	(0.678)	(0.587)	(0.711)	(0.595)	(0.685)	(0.569)
$STRE$	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)
	$\alpha_s = 4/3$ $\alpha_t = 4/3$				$\alpha_s = 4/3$ $\alpha_t = 4/19$			
α_s	1.139	0.857	1.085	0.814	1.017	0.772	0.997	0.796
	(0.765)	(0.628)	(0.732)	(0.597)	(0.708)	(0.604)	(0.698)	(0.587)
α_t	1.263	1.034	1.222	1.009	1.579	1.391	1.360	1.164
	(0.955)	(0.746)	(0.867)	(0.727)	(1.241)	(0.985)	(0.989)	(0.850)
σ^2	1.013	0.825	0.990	0.781	1.032	0.822	0.986	0.782
	(0.684)	(0.575)	(0.662)	(0.587)	(0.703)	(0.575)	(0.684)	(0.568)
$STRE$	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 C.12: Simulated relative efficiency (with respect to the PL, i.e. $SRE = \frac{ndr_{PL}}{ndr_{STBEU}}$) 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.

	Double exponential				Gneiting			
	Regular		Irregular		Regular		Irregular	
	$b_{st}=4$	$b_{st}=9$	$b_{st}=4$	$b_{st}=9$	$b_{st}=4$	$b_{st}=9$	$b_{st}=4$	$b_{st}=9$
	$\alpha_s=3.1/3$ $\alpha_t=3.1/3$				$\alpha_s=3/3$ $\alpha_t=3/19$			
α_s	1.183	0.754	0.989	0.836	1.351	0.925	1.030	0.875
	(0.861)	(0.628)	(0.600)	(0.627)	(0.995)	(0.734)	(0.595)	(0.659)
α_t	1.379	0.987	1.080	1.017	1.756	1.167	1.188	1.124
	(1.073)	(0.742)	(0.627)	(0.730)	(1.246)	(0.846)	(0.712)	(0.827)
σ^2	0.947	0.729	0.754	0.724	0.972	0.737	0.802	0.698
	(0.737)	(0.584)	(0.461)	(0.567)	(0.736)	(0.595)	(0.474)	(0.553)
$STRE$	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)
	$\alpha_s=4/3$ $\alpha_t=4/3$				$\alpha_s=4/3$ $\alpha_t=4/19$			
α_s	1.219	0.777	0.990	0.824	0.908	0.743	1.018	0.863
	(0.910)	(0.653)	(0.624)	(0.613)	(0.691)	(0.549)	(0.591)	(0.663)
α_t	1.635	1.112	1.247	1.079	1.191	0.974	1.323	1.147
	(1.214)	(0.807)	(0.767)	(0.810)	(0.821)	(0.708)	(0.827)	(0.834)
σ^2	0.952	0.743	0.763	0.726	0.732	0.695	0.802	0.726
	(0.737)	(0.569)	(0.494)	(0.582)	(0.590)	(0.475)	(0.498)	(0.553)
$STRE$	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 C.13: Simulated relative efficiency (with respect to the PL, i.e. $SRE = \frac{ndr_{PL}}{ndr_{STBEU}}$) 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.

Appendix D. Statistical efficiency

The following code is for the simulation study using a special case of the spatio-temporal Wendland correlation function proposed in Porcu et al. (2020):

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

where $\boldsymbol{\theta} = (\sigma^2, \alpha_s, \alpha_t, \beta)^\top$.

```
# Separability parameter
rm(list = ls())
graphics.off()
cat("\014")
library(GeoModels)
library(STBEU)

# ST: Creating grid & Data:
tt = 10
N = 20

times <- 1:tt
```

```
x <- seq(0,1,length.out = N)
y <- x
coords <- expand.grid(x,y)
(NT = length(times)*nrow(coords))



smooth_t=0 # k or kappa
scale_t=time_comp_supp=3  # compact supportt scale_t
scale_s=.053
power2_t=3.5+smooth_t #nu
power_s=2
power2_s=2.5+2*smooth_t# tau

sep=0.5  ## 0 0.5  1
sill=1
mean=0
nugget=0



param = c(scale_s = scale_s,scale_t=scale_t, sill = sill ,
          nugget = nugget, power_s=power_s, mean =mean,
          power2_s = power2_s,
          power2_t =power2_t,  smooth_t  =smooth_t, sep =sep)



set.seed(2)
datos <- GeoSim(coordx=coords,coordt=times,sparse=TRUE,
        corrmodel="Wen_time", param=as.list(param))$data
# END: Creating grid & Data


# ST: Starting values & estimation:
start <- NULL
start$scale_s <- as.numeric(param[1])
start$scale_t <-  as.numeric(param[2])
start$sill <-  as.numeric(param[3])
# start$mean <-  as.numeric(param[6])
start$sep <-  as.numeric(param[10])

fix <- as.list(param[c(4,5,6,7,8,9)])

l <- .2
winc=l
winstp=1
```

```
(maxdist1=(1*winc))

maxdist1=0.06
maxtime1=3

type_subs=1
type_dist=1
weighted=0

#### Simluation:

SolPar <- NULL

semilla = 1537
set.seed(semilla)
i = 1
nsim = 1000
while(i <=nsim){
  dd <- GeoSim(coordx=coords,coordt=times,
               corrmodel="Wen_time", param=as.list(param),
               model = "Gaussian",sparse = TRUE)$data

  aux=STBEUFit(theta =start,fixed = unlist(fix),
               coords = coords,times=times,cc=3,datos=dd,
               type_dist=type_dist,
               maxdist=maxdist1 ,maxtime=maxtime1,
               winc_s=winc,winstp_s=winstp,
               winc_t=NULL,
        winstp_t=NULL,subs=type_subs,weighted=weighted)

  SolPar <- rbind(SolPar,aux$par)
  cat("Iter: ",i,"de: ",nsim,"\n")
  i = i+1
}
solSTBEU <- SolPar
apply(solSTBEU,2,mean)
par(mfrow = c(2,2))
boxplot(solSTBEU[,1], main = "scale_s");
abline(h = scale_s, col = "blue")
boxplot(solSTBEU[,2], main = "scale_t");
abline(h = scale_t, col = "blue")
boxplot(solSTBEU[,3], main = "sill");
abline(h = sill, col = "blue")
boxplot(solSTBEU[,4], main = "sep");
abline(h = sep, col = "blue")
par(mfrow = c(1,1))
```

Appendix E. Application including nugget

The following code is for the estimation of the parameters in the application data including the nugget.

```
rm(list = ls())
graphics.off()
cat("\014")
library(GeoModels)
library(STBEU)
library(scatterplot3d)
# devtools::install_github("andrewzm/STRbook")
data("Medwind_data",package = "STRbook")

coords = Edat$ECMWFxylocs

datos = matrix(unlist(Edat$EUdat),
ncol = ncol(Edat$EUdat),nrow = nrow(Edat$EUdat))
datos = t(datos)

coords_ll = coords #lon-lat coords

prj = mapproj::mapproject(coords[ ,1],coords[ ,2],
projection ="sinusoidal")
coords = cbind(prj$x,prj$y) # Projected coords
coords = coords*6371

time <- 1:nrow(datos)

#************** sep = 0.5 **********#

#### ************************** Estimation

################################################
#   parameters for the subsampling ####
###############################################
coordx=coords[,1]
coordy=coords[,2]
LX=abs(range(coordx)[1]-range(coordx)[2])
LY=abs(range(coordy)[1]-range(coordy)[2])
lato_fin=400 #changing window size
lx=lato_fin          #lunghezza lato x quadrato subfinestra
ly=lato_fin          #lunghezza lato y quadrato subfinestra
winc=c(lx/sqrt(LX),ly/sqrt(LY))
###  1/lato_fin complete overlapping
## in space 1 "no" overlapping in space
winstp= 1
###############################################
```

```
winc_t=6    ### length of temporal window
winstp_t=1 ###  0.5 half overlapping  1 "no" overlapping

(maxdist <-  40)
(maxtime <- 1*winc_t)
################################################
weighted=0
type_dist=1 ### type of distance 1:euclidean
type_subs=1 ### type of subsampling  1=in space 2= in time


smooth_t=0
scale_t=20
scale_s=350
power2_t=3.5+smooth_t   +1
power_s=2
power2_s=2.5+2*smooth_t
sep=0.5

 sill=var(c(datos),na.rm = TRUE)
nugget=0.01
mean=mean(datos,na.rm=TRUE)

start=list(scale_s=scale_s,scale_t=scale_t,sill=sill,
nugget=nugget)

fix=c(  power_s=power_s,mean = mean,
        power2_s=power2_s,
        power2_t=power2_t,smooth_t=smooth_t,
        sep=sep)
param <- c(start,fix)

summary(dist(coords))
max(dist(coords))/maxdist
max(dist(time))/maxtime
summary(dist(time))
cc = 3


#2 : STBEU in OpenCL framework with CPU
res1_0=STBEUFit(start,fix,coords,time,cc,datos,
                  type_dist,maxdist  ,maxtime,
                  winc,winstp,NULL,NULL,type_subs,weighted
                  ,varest = TRUE
)
```

```
###### pairwise likelihood ########
fixed=as.list(fix)
res2_0=GeoFit(data=datos,coordx=coords,coordt=time,
                corrmodel="Wen_time",
                start=start,fixed=fixed,
                maxdist=maxdist ,maxtime=maxtime,varest = TRUE)


res1_0$par
res2_0$par
```

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*, 825–848.

Bekker, P. A., & Crudu, F. (2015). Jackknife instrumental variable estimation with heteroskedasticity. *Journal of Econometrics*, *185*, 332–342.

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*, 335–346.

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

Bevilacqua, M., Gaetan, C., Mateu, J., & Porcu, E. (2012). Estimating space and space-time covariance functions for large data sets: a weighted composite likelihood approach. *Journal of the American Statistical Association*, *107*, 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*, 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*, 295–315.

Furrer, R., Genton, M. G., & Nychka, D. (2006). Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*, *15*, 502–523.

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

Hansen, L. P., Heaton, J., & Yaron, A. (1996). Finite-sample properties of some alternative GMM estimators. *Journal of Business & Economic Statistics*, *14*, 262–280.

Hausman, J. A., Newey, W. K., Woutersen, T., Chao, J. C., & Swanson, N. R. (2012). Instrumental variable estimation with heteroskedasticity and many instruments. *Quantitative Economics*, *3*, 211–255.

Heaton, M. J., Datta, A., Finley, A. O., Furrer, R., Guinness, J., Guhaniyogi, R., Gerber, F., Gramacy, R. B., Hammerling, D., Katzfuss, M., Lindgren, F., Nychka, D. W., Sun, F., & Zammit-Mangion,

A. (2019). A case study competition among methods for analyzing large spatial data. *Journal of Agricultural, Biological and Environmental Statistics*, *24*, 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*, (pp. 86–98).
Joe, H., & Lee, Y. (2009). On weighting of bivariate margins in pairwise likelihood. *Journal of Multivariate Analysis*, *100*, 670–685.
Katzfuss, M., & Guinness, J. (2020). A general framework for Vecchia approximations of Gaussian processes. *Statistical Science*, *to appear*.
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*, 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*, 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*, 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*, 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*, .
Ma, P., & Kang, E. L. (2020). A fused Gaussian process model for very large spatial data. *Journal of Computational and Graphical Statistics*, (pp. 1–11).
Morales-Oñate, V., Bevilacqua, M., & Crudu, F. (2019). *STBEU: SpaceTime Blockwise Euclidean Likelihood for Gaussian Models in Geostatistics*. 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*, 350–361.
Owen, A. B. (2001). *Empirical Likelihood*. Chapman & Hall/CRC, London.
Pace, L., Salvan, A., & Sartori, N. (2019). Efficient composite likelihood for a scalar parameter of interest. *Stat*, *8*, e222.
Porcu, E., Bevilacqua, M., & 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*, 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*, 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*, 419–438.
Van der Vaart, A. W. (2007). *Asymptotic Statistics*. Cambridge University Press.
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*, *112*, 1236–1249.

Conflicts of Interest Statement

Manuscript title:

Blockwise Euclidean likelihood for spatio-temporal covariance models

The authors whose names are listed immediately below certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

Author names:

Víctor Morales-Oñate
Federico Crudu
Moreno Bevilacqua

The authors whose names are listed immediately below report the following details of affiliation or involvement in an organization or entity with a financial or non-financial interest in the subject matter or materials discussed in this manuscript. Please specify the nature of the conflict on a separate sheet of paper if the space below is inadequate.

Author names:

Víctor Morales-Oñate Banco Solidario Risk Division, Data Analytics, Quito, Ecuador, victor.morales@uv.cl. Facultad Latinoamericana de Ciencias Sociales Departamento de Desarrollo Ambiente y Territorio, Quito, Ecuador
Federico Crudu Università di Siena CRENoS Department of Economics and Statistics, University of Siena, Piazza San Francesco, 7/8 53100 Siena, Italy, federico.crudu@unisi.it
Moreno Bevilacqua Universidad de Valparaíso Millennium Nucleus Center for the Discovery of Structures in Complex Data Department of Statistics, Avenida Gran Breta~na 1111 Playa Ancha, Valparaso, Chile, moreno.bevilacqua@uv.cl

This statement is signed by all the authors to indicate agreement that the above information is true and correct (a photocopy of this form may be used if there are more than 10 authors):

Author's name (typed)	Author's signature	Date
Víctor Morales-Oñate		11-feb-2020
Federico Crudu		11-feb-2020
Moreno Bevilacqua		11-feb-2020