CrossMark

Composite Likelihood Inference for Multivariate Gaussian Random Fields

Moreno BEVILACQUA, Alfredo ALEGRIA, and Daira VELANDIA, and Emilio PORCU

In the recent years, there has been a growing interest in proposing covariance models for multivariate Gaussian random fields. Some of these covariance models are very flexible and can capture both the marginal and the cross-spatial dependence of the components of the associated multivariate Gaussian random field. However, effective estimation methods for these models are somehow unexplored. Maximum likelihood is certainly a useful tool, but it is impractical in all the circumstances where the number of observations is very large. In this work, we consider two possible approaches based on composite likelihood for multivariate covariance model estimation. We illustrate, through simulation experiments, that our methods offer a good balance between statistical efficiency and computational complexity. Asymptotic properties of the proposed estimators are assessed under increasing domain asymptotics. Finally, we apply the method for the analysis of a bivariate dataset on chlorophyll concentration and sea surface temperature in the Chilean coast.

Key Words: Cross-covariance; Large datasets; Geostatistics.

1. INTRODUCTION

In the last years, there has been a great demand for models describing the evolution of environmental or geophysical spatial processes. In particular, there is a considerable need for modeling the simultaneous behavior of different variables observed in the same spatial region and for providing accurate prediction maps with associated uncertainties.

Multivariate Gaussian random fields (MGRF throughout) are important tools to model and predict these kind of data. Since optimal prediction at unknown locations requires the knowledge of the covariance function, a common practice in geostatistics is to select some parametric classes of covariance functions, so that the covariance is known up to a parameter vector, being estimated through some estimation method.

Moreno Bevilacqua, Daira Velandia (✉) Instituto de Estadística, Universidad de Valparaíso, Valparaíso, Chile (E-mail: *moreno.bevilacqua@uv.cl*). Alfredo Alegria, Emilio Porcu Departamento de Matemática, Universidad Técnica Federico Santa María, Valparaíso, Chile. Daira Velandia Facultad de Ciencias Básicas, Universidad Tecnológica de Bolívar, Cartagena, Colombia.

Nevertheless, modeling the covariance function of such MGRF is not an easy task. For instance, the linear model of coregionalization [LMC , see Wackernagel (2003)] has been criticized for having a number of drawbacks, and we refer the reader to Porcu et al. (2013), Gneiting et al. (2010) and Genton and Kleiber (2015) among others.

The lack of flexible multivariate models has justified recent efforts (see Genton and Kleiber (2015) for a review). Among them, Gneiting et al. (2010) and Apanasovich et al. (2012) extend the Matérn model to the multivariate case and Li and Zhang (2011) propose a class of asymmetric models obtained from stationary multivariate symmetric covariance models. These models are very flexible and can capture both the marginal and the cross-spatial dependence, as well as different levels of smoothness associated, the colocated correlation between the components and the possible presence of asymmetry in the data.

Literature concerning estimation of MGRF has been focussed on the LMC model. Goulard and Voltz (1992) and Pelletier et al. (2004) proposed to estimate it extending least square-type estimators, while Zhang (2007) proposed a ML estimation method capable of handling high-dimensional data exploiting the EM algorithm.

Effective estimation methods, outside the LMC model, for covariance models are somehow unexplored. Maximum likelihood (ML) is probably the best method of estimation when dealing with MGRF, but the exact computation of the likelihood, for N irregularly spaced data, requires $O((Np)^3)$ operations and $O((Np)^2)$ memory, with p being the number of components of the MGRF.

In the case of scalar-valued fields, different approaches have been proposed in order to find estimation methods with a good balance between computational complexity and statistical efficiency for large data set. Some of these approaches are based on the composite likelihood (CL) method as, for instance, in Bevilacqua et al. (2012), Bevilacqua and Gaetan (2015), Stein et al. (2004) and Eidsvik et al. (2014).

Composite likelihood (Lindsay 1988) is a general class of estimation methods based on the likelihood of marginal or conditional events (see Varin et al. (2011) for a review), useful for performing statistical inference in complex problems where standard likelihood methods are difficult to apply. In this paper, we propose two estimation methods for multivariate covariance models, based on CL idea: the former being based on weighted cross- pairwise likelihood (pl_1) and the latter being based on weighted pairs of pairwise likelihoods (pl_2).

We discuss pros and cons of the two methods, including the type of weights that can improve the statistical and computational efficiency, the computation of the standard errors and the asymptotic properties of the estimates under increasing domain asymptotics. We show that the pl_1 and pl_2 estimation methods are useful tools when estimating the covariance model of a MGRF. This is done through simulations experiments, considering different covariance models, such as the multivariate Matérn, the LMC and the asymmetric model proposed in Li and Zhang (2011), using the multivariate covariance tapering (CT) (Bevilacqua et al. 2016) as benchmark.

Finally, we apply the proposed method to the analysis of a large bivariate dataset (approximatively 70000 observations) on chlorophyll concentration and sea surface temperature in the Chilean coast. In particular, we fit a bivariate Matérn model, and consider the problem of model selection between three different versions of the bivariate Matérn (separable, constrained and full).

The paper is organized as follows. Section 2 briefly reviews some multivariate covariance models and their estimation through ML method. Section 3 describes our proposal. In Sect. 4, through a simulation study, we compare pl_1 and pl_2 from statistical and computational efficiency point of view, using ML and CT methods as benchmark. The real data example is described in Sect. 5. Finally, in Sect. 6, we give some conclusions. Supplementary material contains the asymptotic results of the proposed methods.

2. MULTIVARIATE COVARIANCE MODELS AND MAXIMUM LIKELIHOOD

Let $\mathbf{Z}(s) = \{(Z_1(s), \ldots, Z_p(s))^T\}$ be a p-variate Gaussian field with continuous spatial index $s \in \mathbb{R}^d$. The assumption of Gaussianity implies that the first- and second-order moments characterize completely the finite-dimensional distributions. In particular, we assume weak stationarity throughout, so that the mean vector $\boldsymbol{\mu} = \mathbb{E}(\mathbf{Z})$ is constant and the covariance function between $\mathbf{Z}(s_1)$ and $\mathbf{Z}(s_2)$, for any pair s_1, s_2 in the spatial domain, is represented by a mapping $\boldsymbol{C} : \mathbb{R}^d \to M_{p \times p}$ defined through

$$\boldsymbol{C}(\boldsymbol{h}) = \left[C_{ij}(\boldsymbol{h})\right]_{i,j=1}^p = \left[\mathrm{cov}\left(Z_i(s_1), Z_j(s_2)\right)\right]_{i,j=1}^p, \quad \boldsymbol{h} = s_1 - s_2 \in \mathbb{R}^d. \quad (1)$$

The function $\boldsymbol{C}(\boldsymbol{h})$ is called matrix-valued covariance function. Here, $M_{p \times p}$ is the set of squared, symmetric and positive definite matrices. For $i = j$, the functions C_{ii} are called autocovariances or marginal covariances of $Z_i(\cdot), i = 1, \ldots, p$, while for $i \neq j$ the mapping C_{ij} is called cross-covariance between $Z_i(\cdot)$ and $Z_j(\cdot)$. The matrix- valued mapping $\boldsymbol{C}$ must be positive definite, which means that, for a given realization $\mathbf{Z} = (\mathbf{Z}(s_1)^T, \ldots, \mathbf{Z}(s_N)^T)^T$, the $(Np) \times (Np)$ covariance matrix $\Gamma := [\boldsymbol{C}(s_m - s_n)]_{m,n=1}^N$ is positive definite, where $\mathbf{Z}(s_i) = (Z_1(s_i), \ldots Z_p(s_i))^T$ and $\boldsymbol{C}(s_m - s_n) = [C_{ij}(s_m - s_n)]_{i,j=1}^p$, for $m, n = 1, ..., N$, is the generic submatrix. Hereafter for convenience of notation, we set $C_{ij}(s_m - s_n) = c_{ijmn}$.

On the other hand, the second-order properties of the MGRF can be represented by the cross-variogram, defined as

$$2\gamma_{ij}(s_1 - s_2) = \mathbb{E}[\{Z_i(s_1) - Z_i(s_2)\}\{Z_j(s_1) - Z_j(s_2)\}].$$

Note that, under weak stationarity assumptions, the linear identity

$$2\gamma_{ij}(\boldsymbol{h}) = 2C_{ij}(\mathbf{0}) - C_{ij}(\boldsymbol{h}) - C_{ji}(\boldsymbol{h})$$

shows that covariance and variogram are equivalent in terms of modeling.

In order to illustrate the estimation methods proposed in this paper, we shall assume throughout that the mapping $\boldsymbol{C}$ comes from a parametric family of matrix-valued covariances $\{\boldsymbol{C}(\cdot; \boldsymbol{\theta}), \boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq R^k\}$, with $\boldsymbol{\Theta}$ an arbitrary parametric space. Recent literature has been preoccupied on offering new models for matrix-valued covariances. For a thorough review, the reader is referred to Genton and Kleiber (2015), with their exhaustive list of references. Here, we shall merely list those parametric models that will be used through the paper. We have already mentioned the linear model of coregionalization (LMC) that has been popular for over 30 years (Wackernagel 2003). It consists of representing the p-variate Gaussian

field as a linear combination of q-independent univariate fields, with $q = 1, \ldots, p$. The resulting matrix-valued covariance function takes the form:

$$C(\boldsymbol{h}; \boldsymbol{\theta}) = \left[\sum_{k=1}^{q} \alpha_{ik} \alpha_{jk} R_k(\boldsymbol{h}, \boldsymbol{\beta}_k) \right]_{i,j=1}^{p}, \tag{2}$$

with $A := [\alpha_{lk}]_{l,k=1}^{p,q}$ being a $p \times q$-dimensional matrix with full rank and with R_k being a univariate parametric correlation model with parameter vector $\boldsymbol{\beta}_k$. Clearly, we have $\boldsymbol{\theta} = (\text{vec}(A)^T, \boldsymbol{\beta_1}^T, \ldots, \boldsymbol{\beta_q}^T)^T$.

Constructive criticism about this model has been expressed, for instance, by Gneiting et al. (2010), Apanasovich et al. (2012) and Daley et al. (2015). For instance, if $\alpha_{ik} \neq 0$ for each i, k the smoothness of any component defaults to that of the roughest latent process.

Another popular construction, called *separable*, is obtained through:

$$C(\boldsymbol{h}, \boldsymbol{\theta}) = \left[\rho_{ij} \sigma_i \sigma_j R(\boldsymbol{h}, \boldsymbol{\psi}) \right]_{i,j=1}^{p}, \quad \rho_{ii} = 1, \tag{3}$$

with $R(\cdot, \boldsymbol{\psi})$ a univariate parametric correlation model, $\sigma_i^2 > 0$, $i = 1, \ldots, p$, are the marginal variances, and the ρ_{ij} is the colocated correlation coefficient between $Z_i(\boldsymbol{s})$ and $Z_j(\boldsymbol{s})$. In this case, $\boldsymbol{\theta} = (\boldsymbol{\rho}, \boldsymbol{\psi}^T, \sigma_1^2, \ldots, \sigma_p^2)^T$ where $\boldsymbol{\rho}$ is the vector containing all the pairwise colocated correlation coefficients ρ_{ij}, $i = 1, \ldots, p-1$, $j = i+1, \ldots, p$. This construction assumes that the components of the multivariate random field have the same spatial correlation structure. Therefore, the model is not able to capture the different scales and smoothness of the components. Note that the separable model is a special case of LMC model in Eq. (2).

A generalization of the model (3), which allows to overcome this drawback, is:

$$C(\boldsymbol{h}, \boldsymbol{\theta}) = \left[\rho_{ij} \sigma_i \sigma_j R(\boldsymbol{h}; \boldsymbol{\psi}_{ij}) \right]_{i,j=1}^{p}, \quad \rho_{ii} = 1. \tag{4}$$

Bevilacqua et al. (2015) use this general class in order to test the significance of the correlation between the components of a bivariate Gaussian random field. In this general approach, the difficulty lies in deriving conditions on the model parameters that result in a valid multivariate covariance model. For instance, Gneiting et al. (2010) proposed a special case of the construction in Eq. (4), with R belonging to the Matérn family, being defined through:

$$R(\boldsymbol{h}; \boldsymbol{\psi}) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\|\boldsymbol{h}\|}{\alpha} \right)^{\nu} \mathcal{K}_{\nu} \left(\frac{\|\boldsymbol{h}\|}{\alpha} \right), \tag{5}$$

where $\boldsymbol{\psi} = (\alpha, \nu), \alpha > 0$ is the scale parameter, and $\nu > 0$ is a smoothness parameter. In the bivariate case, the authors find necessary and sufficient conditions based on the parameters for positive definiteness, while for the case $p \geq 3$ they only offer sufficient conditions. This kind of construction allows for a nice closed form, together with the possibility of different spatial scales and smoothness, different variances and colocated correlation parameters.

The previous models are symmetric, i.e., they assume that $C_{ij}(\boldsymbol{h}) = C_{ji}(\boldsymbol{h})$, but in some circumstances, such hypothesis can be restrictive. In this case, asymmetric models must be taken into account, as proposed in Li and Zhang (2011):

$$C(\boldsymbol{h}, \boldsymbol{\theta}) = [C_{ij}^*(\boldsymbol{h} + \boldsymbol{a}_i - \boldsymbol{a}_j; \boldsymbol{\vartheta})]_{i,j=1}^p, \tag{6}$$

where $C_{ij}^*(\boldsymbol{h}, \boldsymbol{\vartheta})$ is a valid symmetric parametric covariance model. Here $\boldsymbol{\theta} = (\boldsymbol{\vartheta}, \boldsymbol{a}_1, \ldots, \boldsymbol{a}_p)^T$ where $\boldsymbol{a}_i \in \mathbb{R}^d$, for $i = 1, \ldots, p$, are the vector parameters which allows to introduce asymmetry in the symmetric model.

Since we are assuming that the state of truth is represented by some parametric family of matrix-valued covariances $\{C(\cdot; \boldsymbol{\theta}), \boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^k\}$, we may use the abuse of notation $\Gamma(\boldsymbol{\theta})$ for the covariance matrix Γ, in order to emphasize the dependence on the unknown parameter vector. For a N-dimensional realization from a p-variate Gaussian field, the log-likelihood can be written as

$$l(\boldsymbol{\theta}) = -\frac{1}{2} \log |\Gamma(\boldsymbol{\theta})| - \frac{1}{2} \boldsymbol{Z}^T [\Gamma(\boldsymbol{\theta})]^{-1} \boldsymbol{Z}. \tag{7}$$

The most time-consuming part when calculating (7) is to evaluate the determinant and inverse of $\Gamma(\boldsymbol{\theta})$. The most widely used algorithms, such as Cholesky decomposition, require up to $\mathcal{O}((Np)^3)$ operations and $O((Np)^2)$ memory, and this can be prohibitive if N is large. Various approximation methods have been developed to address this computational problem. For instance, the CT method has been proposed in order to reduce the number of operations and memory requirement in the case of scalar- valued Gaussian fields, for both estimation (see Kaufman et al. (2008), Shaby and Ruppert (2012), Bevilacqua et al. (2016) in the multivariate case) and prediction (Furrer et al. 2006). The method consists in setting to zero certain elements of the covariance matrix $\Gamma(\boldsymbol{\theta})$, by multiplying $\Gamma(\boldsymbol{\theta})$ element by element with a sparse matrix coming from a multitaper function. Let us denote with $\boldsymbol{T}(\mathbf{d})$ the $Np \times Np$ matrix associated with a multitaper function as, for instance, those described in Bevilacqua et al. (2016).

The "tapered" covariance matrix is then obtained through $\Gamma_{CT}(\boldsymbol{\theta}) = \Gamma(\boldsymbol{\theta}) \circ \boldsymbol{T}(\mathbf{d})$, where $\circ$ denotes the Schur product. The multitaper vector parameters $\mathbf{d}$ include the (possibly different) compact supports, these being fixed in a way to determine the desired level of sparseness for the construction above. Then, the tapered likelihood is defined as:

$$l_{CT}(\boldsymbol{\theta}, \mathbf{d}) = -\frac{1}{2} \log |\Gamma_{CT}(\boldsymbol{\theta})| - \frac{1}{2} \boldsymbol{Z}^T (\Gamma_{CT}(\boldsymbol{\theta})^{-1} \circ \boldsymbol{T}(\mathbf{d})) \boldsymbol{Z}, \tag{8}$$

and algorithms for sparse matrices can be exploited in order to compute (8) efficiently.

3. COMPOSITE LIKELIHOOD METHODS

We propose two alternative versions of the CL approach in the multivariate context:

(a) First, let us consider the bivariate random vector based on the cross-pairs $\boldsymbol{Y}_{ijmn} = [Z_i(\boldsymbol{s}_m), Z_j(\boldsymbol{s}_n)]^T$, for $i, j = 1, \ldots, p$ and $m, n = 1, \ldots, N$, with associated log-likelihood

$$l_{ijmn}(\boldsymbol{\theta}) = -\frac{1}{2} \left(\log |Q_{ijmn}| + \boldsymbol{Y}_{ijmn}^T Q_{ijmn}^{-1} \boldsymbol{Y}_{ijmn} \right), \boldsymbol{\theta} \subset \boldsymbol{\Theta},$$

where

$$Q_{ijmn} = \begin{pmatrix} c_{iimm} & c_{ijmn} \\ - & c_{jjnn} \end{pmatrix}.$$

(9)

Here for convenience of notation, we drop out the dependence on $\boldsymbol{\theta}$ of the generic element c_{ijmn}. We define the index set $\Lambda_1 = \Lambda_1^{(1)} \cup \Lambda_1^{(2)} \cup \Lambda_1^{(3)}$, where

$$\Lambda_1^{(1)} = \{(i, j, m, n) : m = 1, \ldots, N-1, \ n = m+1, \ldots, N, \ i{=}j{=}1, \ldots, p\},$$
$$\Lambda_1^{(2)} = \{(i, j, m, n) : m = 1, \ldots, N, \ n = m, \ldots, N, \ i{=}1, \ldots, p-1, \ j{=}i{+}1, \ldots, p\},$$
$$\Lambda_1^{(3)} = \{(i, j, m, n) : m{=}1, \ldots, N-1, \ n{=}m+1, \ldots, N,$$
$$j{=}1, \ldots, p-1, \ i{=}j+1, \ldots, p\}.$$

Then, the first objective function, based on cross-pairs, is defined as

$$pl_1(\boldsymbol{\theta}) = \sum_{(i,j,m,n) \in \Lambda_1} l_{ijmn}(\boldsymbol{\theta}) w_{ijmn},$$

(10)

where w_{ijmn} are positive suitable weights.

(b) The second version of CL we propose is based on the four- dimensional random vector $X_{ijmn} = \left[Z_i(s_m), Z_i(s_n), Z_j(s_m), Z_j(s_n) \right]^T$, for $i, j = 1, \ldots, p$ and $m, n = 1, \ldots, N$, with associated log-likelihood

$$g_{ijmn}(\boldsymbol{\theta}) = -\frac{1}{2} \left(\log |\Sigma_{ijmn}| + X_{ijmn}^T \Sigma_{ijmn}^{-1} X_{ijmn} \right),$$

where

$$\Sigma_{ijmn} = \begin{pmatrix} c_{iimm} & c_{iimn} & c_{ijmm} & c_{ijmn} \\ - & c_{iinn} & c_{ijnm} & c_{ijnn} \\ - & - & c_{jjmm} & c_{jjmn} \\ - & - & - & c_{jjnn} \end{pmatrix}.$$

(11)

Thus, the second approach of the CL we propose is defined as:

$$pl_2(\boldsymbol{\theta}) = \sum_{(i,j,m,n) \in \Lambda_2} g_{ijmn}(\boldsymbol{\theta}) w_{ijmn},$$

(12)

where the index set Λ_2 is defined as:

$$\Lambda_2 = \{(i, j, m, n) : i = 1, \ldots, p-1, \ j = i+1, \ldots, p,$$
$$m = 1, \ldots, N-1, \ n = m+1, \ldots, N\}.$$

Therefore, the maximum CL estimator for both methods is given by $\widehat{\boldsymbol{\theta}}_a{=}\mathrm{argmax}_{\boldsymbol{\theta}} \ pl_a(\boldsymbol{\theta})$, $a = 1, 2$. Some comments are in order:

- The function $pl_1(\boldsymbol{\theta})$ involves the marginal pairwise log-likelihood associated with the $i - th$ component for $i = 1, \ldots, p$ and the cross-pairwise log-likelihood. In fact,

$pl_1(\boldsymbol{\theta})$ can be rewritten as

$$pl_1(\boldsymbol{\theta}) = \sum_{i=1}^{p}\sum_{m=1}^{N-1}\sum_{n=m+1}^{N} l_{iimn}(\boldsymbol{\theta})w_{iimn} + \sum_{i=1}^{p-1}\sum_{j=i+1}^{p}\sum_{m=1}^{N}\sum_{n=m}^{N} l_{ijmn}(\boldsymbol{\theta})w_{ijmn}$$
$$+\sum_{j=1}^{p-1}\sum_{i=j+1}^{p}\sum_{m=1}^{N-1}\sum_{n=m+1}^{N} l_{ijmn}(\boldsymbol{\theta})w_{ijmn}.$$

The first term is the contribution of each pairwise likelihood associated with the $i-th$ component. The second and the third terms, instead, involve the cross-pairwise log-likelihoods. The reason to split the contribution of the cross-pairwise log-likelihood into two parts is that $l_{ijmn}(\boldsymbol{\theta}) \neq l_{jimn}(\boldsymbol{\theta})$ because $c_{ijmn}(\boldsymbol{\theta}) \neq c_{jimn}(\boldsymbol{\theta})$ under asymmetry. In the special case of multivariate symmetric models, $pl_1(\boldsymbol{\theta})$ simplifies to

$$pl_1(\boldsymbol{\theta}) = \sum_{i=1}^{p}\sum_{m=1}^{N-1}\sum_{n=m+1}^{N} l_{iimn}(\boldsymbol{\theta})w_{iimn} + \sum_{i=1}^{p-1}\sum_{j=i+1}^{p}\sum_{m=1}^{N}\sum_{n=1}^{N} l_{ijmn}(\boldsymbol{\theta})w_{ijmn}. \tag{13}$$

Genton et al. (2015) use (13) in order to estimate multivariate max stable processes considering only multivariate symmetric covariances of a certain type.

- The role of the weights w_{ijmn} in $pl_a(\boldsymbol{\theta})$, $a = 1, 2$, is to save computational time and to improve statistical efficiency. Weight functions with compact support have clear computational advantages and can improve the statistical efficiency, as shown in Joe and Lee (2009), Davis and Yau (2011) and Bevilacqua et al. (2012). A possible choice for the weights w_{ijmn} in Eqs. (10) and (12) is the simple cutoff function:

$$u_{mn}(\kappa_{ij}) = \begin{cases} 1, & h_{ijmn} \leq 1 \\ 0, & h_{ijmn} > 1 \end{cases}, \tag{14}$$

where $h_{ijmn} = \dfrac{\|\boldsymbol{s}_m - \boldsymbol{s}_n\|}{\kappa_{ij}}$, or alternatively, a smoother function such as a correlation function with compact support can be used. An example is the Bohman function (Gneiting 2002):

$$v_{mn}(\kappa_{ij}) = \begin{cases} (1 - h_{ijmn})\left(\dfrac{\sin(2\pi h_{ijmn})}{2\pi h_{ijmn}}\right) + \left(\dfrac{1 - \cos(2\pi h_{ijmn})}{2\pi^2 h_{ijmn}}\right), & h_{ijmn} \leq 1 \\ 0, & h_{ijmn} > 1 \end{cases}. \tag{15}$$

The choice of the compact supports, κ_{ij}'s, in the weight function is not in general an easy task. Bevilacqua et al. (2012) proposed, in the univariate case, a formal criterion based on the optimization of the trace of the Godambe information matrix, but this method can be computationally hard. An ad hoc strategy is to fix the compact support as a fraction of the (empirically estimated) practical range of the covariance function (Bevilacqua and Gaetan 2015). We follow this strategy in the simulation study.

Note that with the proposal denoted pl_1, it is possible to use different types of weights. In particular, Eq. (10) shows that it is possible to use different weights for the marginal and cross-pairwise likelihoods. Instead, for the pl_2 case, only weights of the second type are allowed. For instance, if $p = 2$, then we can use w_{11mn}, w_{22mn} and w_{12mn} for pl_1, while for pl_2 we can just use w_{12mn}. In this respect, pl_1 is more flexible since it allows to consider different weight functions.

- For the pl_1 case, a bivariate Gaussian distribution must be evaluated at each sum and a 4-variate Gaussian distribution in the second case. The order of computation is, in the first case, $\mathcal{O}\left(\frac{Np(Np-1)}{2}\right)$, while in the second case, it is $\mathcal{O}\left(\frac{Np(N-1)(p-1)}{4}\right)$, when considering the unweighted version of the two methods. When considering the weighted versions, the order of computation depends also on the choice of the (compactly supported) weight functions. In the simulation study, we compare the computational performance of the two methods using both the weighted and unweighted versions.

- One of the benefits of pl_1 and pl_2 is that, as outlined in Varin et al. (2011), composite likelihood requires only model assumptions on the lower-dimensional marginal densities and not detailed specification of the full joint distribution. Varin et al. (2011) call this feature robustness of composite likelihood. In our case, only a two- dimensional or a four-dimensional Gaussian assumption is required for pl_1 and pl_2, respectively. In this sense, pl_1 is more robust than pl_2.

- Throughout the study, we have supposed to work in the isotopic case (Wackernagel 2003). Under this setting, each component of the MGRF is observed at the same location sites. In the heterotopic case, such random field $Z_i(\cdot), i = 1, \ldots, p$ is observed at points $s_{ni}, n = 1, \ldots, N_i$, at possibly different location sites. In this case, a little modification of the index set Λ_1 in pl_1 is needed in order to take into account the different location sites and the different number of locations. Note that pl_2 is not defined in this case unless $N_i = N_j$ for $i \neq j$.

In the Supplementary material, we show that the maximum CL estimators are consistent and asymptotically normal under increasing domain asymptotics. Under these results, the inverse of the Godambe information matrix G_a, defined through

$$G_a(\boldsymbol{\theta}) = H_a(\boldsymbol{\theta})J_a(\boldsymbol{\theta})^{-1}H_a(\boldsymbol{\theta})^T, \quad a = 1, 2, \tag{16}$$

where

$$H_a(\boldsymbol{\theta}) = -\mathrm{E}[\nabla^2 pl_a(\boldsymbol{\theta})], \quad J_a(\boldsymbol{\theta}) = \mathrm{E}[\nabla pl_a(\boldsymbol{\theta})\nabla pl_a(\boldsymbol{\theta})^T], \quad a = 1, 2, \tag{17}$$

is the asymptotic variance of the pl_a, $a = 1, 2$ estimators. In the Supplementary material, we offer closed-form expressions for the matrices $H_a(\boldsymbol{\theta})$ and $J_a(\boldsymbol{\theta})$, for $a = 1, 2$.

Computation of the standard errors requires consistent estimation of the matrices $H_a(\boldsymbol{\theta})$ and $J_a(\boldsymbol{\theta})$, for $a = 1, 2$. This can be achieved through the plug-in estimates $H_a(\widehat{\boldsymbol{\theta}}_a)$ and $J_a(\widehat{\boldsymbol{\theta}}_a)$, $a = 1, 2$. Nevertheless, the latter becomes computationally unfeasible for large data sets since it is of order $O((Np)^4)$. In order to estimate $J_a(\widehat{\boldsymbol{\theta}}_a)$, we extend to the multivariate

case the subsampling method as described in Heagerty and Lumley (2000). Provided that $W^{-1}J_a(\widehat{\boldsymbol{\theta}}_a)$ converges to a matrix J_a^* as $n \to \infty$, where W is the sum of the weights involved in the estimation, we use the subsampling method in order to obtain an estimate $\widehat{J}_a^*$ of J_a^* and then estimate $J_a(\widehat{\boldsymbol{\theta}}_a)$ through $W\widehat{J}_a^*$. Given $S_1, \ldots, S_M$ subsets of $\{1, 2, \ldots, n\}$, associated with the observation points $\{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_n\}$, the estimator is defined as

$$\widehat{J}_a^* = \frac{1}{M} \sum_{k=1}^{M} \frac{1}{W_a^{(k)}} \sum_{\substack{(i,j,m,n)\in\Lambda_a \\ (i',j',m',n')\in\Lambda_a}} [\nabla pl_a]_{ijmn}[\nabla pl_a]_{i'j'm'n'}^\mathsf{T}, \, w_{ijmn}w_{i'j'm'n'}, \quad a = 1, 2,$$

(18)

where $W_a^{(k)} = \sum_{(i,j,m,n)\in\Lambda_a} w_{ijmn}$. Here, the pairs (i, j) and (i', j') belong to S_k and the analytic expression of $[\nabla pl_a]_{ijmn}$, $a = 1, 2$ is offered in the Supplementary material. The subsets are derived gathering the points that fall in a collection of overlapping subregions of the same shape of the region of observations but of smaller volume (Lee and Lahiri 2002). Then, the asymptotic covariance matrix of $\widehat{\boldsymbol{\theta}}_a$ can then be estimated by using the subsampling approximation

$$\widehat{G_a^{-1}}(\widehat{\boldsymbol{\theta}}_a) = H_a^{-1}(\widehat{\boldsymbol{\theta}}_a)(W\widehat{J}_a^*)H_a^{-1}(\widehat{\boldsymbol{\theta}}_a).$$

(19)

and standard error estimation of each parameter is computed by taking the square root of the diagonal elements of $\widehat{G_a^{-1}}(\widehat{\boldsymbol{\theta}}_a)$. As outlined in Bevilacqua et al. (2012), computational reasons can drive the choice of the number of subsets M. For instance, if M grows at a rate $\mathcal{O}((Np)^{2/3})$, evaluation of J_a^* implies an order of computation $\mathcal{O}((Np)^2)$ that is the same of H_a and pl_a.

Finally, as in the classical likelihood framework, model selection based on CL is feasible. Model selection is a major problem in multivariate spatial statistics, since when increasing the number of components of the vector random field, the complexity of the associated covariance models grows up quickly, and the choice of a parsimonious model can be troublesome. Following Varin and Vidoni (2005), the CL information criterion (CLIC) selects the model maximizing

$$\text{CLIC}(\widehat{\boldsymbol{\theta}}_a) = pl_a(\widehat{\boldsymbol{\theta}}_a) + \text{tr}(J_a(\widehat{\boldsymbol{\theta}}_a)H_a(\widehat{\boldsymbol{\theta}}_a)^{-1}),$$

(20)

where the estimation of $J_a(\widehat{\boldsymbol{\theta}}_a)$ has been discussed previously.

4. SIMULATION STUDY

The aim of this section is to compare the performance of pl_1 and pl_2, with ML estimation method, from statistical and computational efficiency point of view. Multivariate CT, described in Sect. 2, is used as benchmark. We work under the bivariate case ($p = 2$), and in particular, we consider N location sites, being uniformly distributed on the square $[0, 1]^2$. In the simulation study, since we want to compare pl_a, $a = 1, 2$ with ML, we choose a relatively small number of observations by fixing $N = 300$ so that we are considering 600 observations. The maximum distance between the location sites is $\kappa_{max} = 1.339$.

We consider two bivariate covariance models, the former being obtained through Eqs. (4) and (5), fixing $v_{ij} = 0.5$; this choice relates to bivariate exponential model, defined as

$$C(\boldsymbol{h}, \boldsymbol{\theta}) = \left[\rho_{ij} \sigma_i \sigma_i e^{-\frac{\|\boldsymbol{h}\|}{\psi_{ij}}} \right]_{i,j}^2, \tag{21}$$

where $\rho_{ii} = 1$ and $\boldsymbol{\theta} = (\rho_{12}, \psi_{11}, \psi_{12}, \psi_{22}, \sigma_1^2, \sigma_2^2)^T$. Here, $\sigma_m^2 > 0$ and $\alpha_{ii} > 0$, $i = 1, \ldots 2$ are, respectively, the variance and scale parameters, while $\psi_{12} > 0$ and ρ_{12} are, respectively, the cross-scale and the correlation parameter that describes the marginal correlation between the components.

The second model is a special case of the bivariate LMC model in Eq. (2):

$$C(\boldsymbol{h}; \boldsymbol{\theta}) = \left[\sum_{k=1}^{2} \alpha_{ik} \alpha_{jk} e^{-\frac{\|\boldsymbol{h}\|}{\beta_i}} \right]_{i,j=1}^p, \tag{22}$$

where $\beta_i > 0$ and $[\alpha_{lm}]_{l,m=1}^2$ is a matrix of rank 2.

Multivariate CT is performed considering a separable bivariate taper function, given by $[T(\boldsymbol{h}, b)]_{i,j=1}^2$ (Bevilacqua et al. 2016), where $T(\boldsymbol{h}, b)$ is a univariate compactly supported correlation function of the Wendland type $i.e$, $T(\boldsymbol{h}; b) = \left(1 + 4\frac{\|\boldsymbol{h}\|}{b}\right)\left(1 - \frac{\|\boldsymbol{h}\|}{b}\right)_+^4$. This is a special case of a general class of tapers proposed in Daley et al. (2015). In this case, the taper matrix is given by $\boldsymbol{T}(b) = \boldsymbol{1} \otimes W$ where $\boldsymbol{1}$ is a 2×2 unit matrix and W is the $N \times N$ sparse correlation matrix associated with the Wendland function. In the simulation study, the compact support b has been fixed in order to achieve 2.5% of nonzero values in the tapered matrices.

4.1. STATISTICAL EFFICIENCY COMPARISON

Comparison in terms of statistical efficiency is taken through the sample relative mean squared error, defined through

$$\text{RE}_a(\boldsymbol{\theta}_i) = \frac{\text{MSE}_{\text{ML}}(\boldsymbol{\theta}_i)}{\text{MSE}_a(\boldsymbol{\theta}_i)}, \quad a = 1, 2, CT \tag{23}$$

for $i = 1, \ldots, k$ where $\text{MSE}_{\text{ML}}(\boldsymbol{\theta}_i)$, $\text{MSE}_{CT}(\boldsymbol{\theta}_i)$ and $\text{MSE}_a(\boldsymbol{\theta}_i)$ are the sample mean squared error of ML, CT and pl_a, $a = 1, 2$, respectively, associated with the generic parameter $\boldsymbol{\theta}_i$.

As global measure of relative efficiency, we consider

$$\text{GRE}_a(\boldsymbol{\theta}, k) = \left(\frac{\det(V_{\text{ML}}(\boldsymbol{\theta}))}{\det(V_a(\boldsymbol{\theta}))} \right)^{1/k}, \quad a = 1, 2, CT$$

where V_{ML}, V_{CT} and V_a, $a = 1, 2$ are the sample variance–covariance matrices of ML, CT and pl_a, respectively, k is the number of parameters involved in the estimation, and $\det(F)$ is the determinant of the matrix F.

Table 1. Scenario A: $RE_a(\theta_i)$ and $GRE_a(\theta, 5)$ for the parameters of the bivariate exponential model with constraints with increasing marginal correlation between the two random fields when considering pl_a and its weighted version pl_a^w, $a = 1, 2$ and covariance tapering (l_{CT}) estimation methods.

| ρ_{12}	0.1					0.5					0.9				
a	pl_1	pl_1^w	pl_2	pl_2^w	l_{CT}	pl_1	pl_1^w	pl_2	pl_2^w	l_{CT}	pl_1	pl_1^w	pl_2	pl_2^w	l_{CT}
$RE_a(\sigma_1^2)$	0.910	0.732	0.917	0.702	0.954	0.863	0.694	0.870	0.686	0.913	0.931	0.751	0.942	0.766	0.994
$RE_a(\sigma_2^2)$	0.886	0.642	0.892	0.681	0.944	0.821	0.600	0.825	0.654	0.889	0.791	0.613	0.795	0.670	0.861
$RE_a(\psi_{11})$	0.402	0.582	0.435	0.584	0.449	0.350	0.540	0.432	0.595	0.446	0.391	0.596	0.448	0.731	0.500
$RE_a(\psi_{22})$	0.367	0.535	0.475	0.638	0.518	0.227	0.251	0.499	0.650	0.531	0.220	0.189	0.517	0.702	0.545
$RE_a(\rho_{12})$	0.127	0.273	0.428	0.444	0.499	0.174	0.306	0.426	0.448	0.493	0.371	0.393	0.472	0.475	0.542
$GRE_a(\theta, 5)$	0.354	0.517	0.489	0.600	0.523	0.342	0.459	0.489	0.603	0.523	0.393	0.388	0.516	0.629	0.537
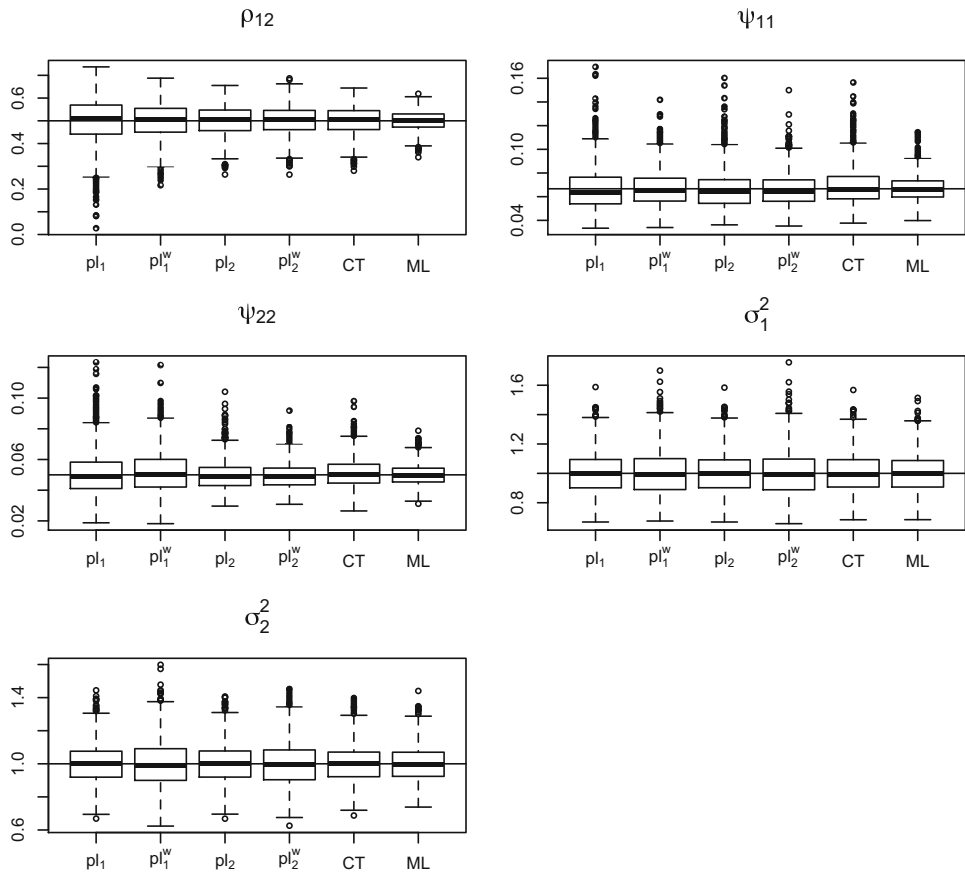
Figure 1. *Boxplots* of the bivariate exponential model (scenario A) estimates using maximum likelihood (ML) covariance tapering (CT) and CL using (pl_a^w) and not using (pl_a) ($a = 1, 2$) compact support weight functions when the correlation between the components is 0.5.

We consider three scenarios (denoted "A," "B," "C"), and for each scenario, we simulate 1000 realizations from a bivariate zero mean GRF, using Cholesky decomposition. Then, we estimate using ML, CT and both weighted and unweighted pl_a, $a = 1, 2$. Let us give more details about the indicated scenarios:

(A) The model in Eq. (21) with constraints on the scales parameters. Specifically, we set $\sigma_1^2 = \sigma_2^2 = 1$, $\psi_{ii} = \frac{pr_{ii}}{3}$, $i = 1, 2$, $\psi_{12} = \frac{\psi_{11}+\psi_{22}}{2}$ and $\rho_{12} = 0.1, 0.5, 0.9$, where $pr_{11} = 0.2$ and $pr_{22} = 0.15$ are the practical ranges for the first and second components. In this case, $\boldsymbol{\theta} = (\rho_{12}, \psi_{11}, \psi_{22}, \sigma_1^2, \sigma_2^2)^T$ and we estimate through CT and pl_a, for $a = 1, 2$ with the following weight function :

- For pl_1, we choose $w_{iimn} = u_{mn}(\kappa_{ii})$ with $\kappa_{ii} = \frac{pr_{ii}}{2}$, $i = 1, 2$ and $w_{12mn} = v_{mn}(\kappa_{12})$ with $\kappa_{12} = 0.5(\kappa_{11} + \kappa_{22})$

- For pl_2, we choose $w_{12mn} = u_{mn}(\kappa)$ with $\kappa = \frac{\kappa_{11}+\kappa_{22}+\kappa_{12}}{3}$ and for CT we fix $b = \kappa$.

Note that in pl_1 we choose the marginal compact supports as a fraction of the practical ranges, and for pl_2 and CT, we fix the compact support equal to the practical range. Table 1 shows $RE_a(\boldsymbol{\theta}_i)$, $i = 1, \ldots, 5$ and $GRE_a(\boldsymbol{\theta}, 5)$ for $a = 1, 2, CT$ when $\rho_{12} = 0.1, 0.5, 0.9$. In the same Table, we report the results for unweighted pl_1 and pl_2.

Note that, when increasing the correlation between the random fields from 0.1 to 0.9, the relative efficiency of the colocated and scale parameters changes substantially for pl_1, but not for pl_2. By contrast, the efficiencies of the variance parameters are not affected by the correlation. Table 1 also shows that using compactly supported weight functions is a good choice for both methods when looking at the global measure of relative efficiency. Nevertheless, the relative efficiency of variances parameters does not improve when considering the weighted version. Overall pl_2 improves the global measure of relative efficiency of pl_1 of 10 percent approximatively. CT performs satisfactorily, in particular for the ρ_{12} and variances parameters, but pl_2 outperforms CT in terms of global measure of efficiency. Figure 1 depicts the boxplots of each parameter for ML, CT and weighted and unweighted pl_a $a = 1, 2$.

(B) The model in Eq. (22), with $\beta_1 = 0.2/3$, $\beta_2 = 0.15/3$, $\alpha_{ii} = \sqrt{1-x}$ $i = 1, 2$, and the constraint $\alpha_{12} = \alpha_{21} = \sqrt{x}$, with $x = 0.003, 0.067, 0.29$. Under this setting, the bivariate random field has unit variance, with marginal practical ranges equal to $pr_{11} = 0.2$ and $pr_{22} = 0.15$ and correlation between the components approximatively equal to $0.1, 0.5, 0.9$ when $x = 0.003, 0.067, 0.29$, respectively. In this case, $\boldsymbol{\theta} = (\alpha_{11}, \alpha_{12}, \alpha_{22}, \beta_1, \beta_2)^T$. We estimate with CT and pl_a $a = 1, 2$, using the same weights and the same compact support b in CT method, as in scenario A. Table 2 shows $RE_a(\boldsymbol{\theta}_i)$, $i = 1, 2, 3, 4, 5$ and $GRE_a(\boldsymbol{\theta}, 5)$ for $a = 1, 2, CT$ when $\rho_{12} = 0.1, 0.5, 0.9$.

Note that, when increasing the correlation between the random fields from 0.1 to 0.9, the relative efficiency of α_{12} and the scale parameters β_1 and β_2 change substantially for pl_1, but not for pl_2. By contrast, the efficiencies α_{11} and α_{22} parameters are not affected by the correlation. Table 2 shows that also in this scenario, using compactly supported weight functions is a good choice for both methods from global relative efficiency point of view. As in scenario A, pl_2 improves the global relative efficiency of pl_1 of 10 percent approximatively. CT relative efficiency is overall good, in particular when estimating α_{ii}, $i = 1, 2$. Also in this scenario, pl_2 outperforms CT in terms of global relative efficiency.

We replicate the simulation study using unconstrained versions of the covariance models in Eqs. (21) and (22), and the conclusions are very similar to those associated with scenarios A and B.

(C) The asymmetric model defined in Eq. (6) with $p = 2$ and $C_{ij}^*(\boldsymbol{h}, \cdot)$ equal to the separable bivariate exponential model was obtained by considering $\psi = \psi_{ij}$ for $i, j = 1, 2$ in the scenario (A). As outlined in Li and Zhang (2011), $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ are not jointly identifiable, so we set $\boldsymbol{a}_1 = [0, 0]^T$ and $\boldsymbol{a}_2 = [k, k]^T$. We set $\sigma_1^2 = \sigma_2^2 = 1$, $\psi = 0.1/3$, $k = -0.1$ and consider different colocated correlation that is $\rho_{12} = 0.15, 0.25, 0.35$. In this case, $\boldsymbol{\theta} = (k, \rho_{12}, \psi, \sigma_1^2, \sigma_2^2)^T$, but we note some numerical evidence of multimodality of the CL surface when estimating $\boldsymbol{\theta}$. For this reason, we

Table 2. Scenario B: $RE_a(\theta_i)$ and $GRE_a(\theta, 5)$ for the parameters of the LMC model with constraints with increasing marginal correlation between the two random fields, when considering pl_a and its weighted version pl_a^w, $a = 1, 2$ and covariance tapering (l_{CT}) estimation methods.

ρ_{12}	0.1					0.5					0.9				
a	pl_1	pl_1^w	pl_2	pl_2^w	l_{CT}	pl_1	pl_1^w	pl_2	pl_2^w	l_{CT}	pl_1	pl_1^w	pl_2	pl_2^w	l_{CT}
$RE_a(\alpha_{11})$	0.947	0.768	0.955	0.733	0.994	0.872	0.718	0.887	0.691	0.925	0.771	0.633	0.780	0.652	0.842
$RE_a(\alpha_{12})$	0.113	0.240	0.379	0.398	0.445	0.265	0.326	0.445	0.459	0.525	0.531	0.434	0.554	0.533	0.628
$RE_a(\alpha_{22})$	0.904	0.701	0.951	0.732	0.999	0.851	0.660	0.890	0.684	0.944	0.767	0.628	0.796	0.648	0.856
$RE_a(\beta_1)$	0.417	0.603	0.452	0.616	0.477	0.325	0.488	0.440	0.590	0.281	0.165	0.247	0.405	0.619	0.466
$RE_a(\beta_2)$	0.046	0.602	0.523	0.725	0.579	0.032	0.219	0.478	0.651	0.365	0.067	0.182	0.513	0.650	0.529
$GRE_a(\theta, 5)$	0.233	0.530	0.503	0.624	0.544	0.221	0.422	0.489	0.602	0.442	0.254	0.353	0.491	0.606	0.530

fix the colocated parameter and then we estimate with ML, CT and pl_a $a = 1, 2$, using the same weights and the same compact support b of CT as in scenario (A) but with $d = \kappa_{max}/3$. Table 3 shows $RE(\theta_i)$, $i = 1, 2, 3, 4$ and $GRE(\theta, 4)$ for CT, weighted and unweighted pl_a $a = 1, 2$. Note that when increasing the correlation, pl_1 performs better than pl_2 when estimating the scale parameter and the relative efficiency associated with the asymmetry parameter is better for pl_2. The global measure of efficiency suggests that pl_1 performs very similar to pl_2 for this specific model. Finally, CT is very inefficient when estimating the asymmetry parameter. This suggests that the use of symmetric bivariate tapers, as proposed in Daley et al. (2015), can be very inefficient when estimating asymmetric models.

We can summarize our findings as follows:

- The use of compactly supported weight functions allows to improve significantly the global statistical relative efficiency of pl_a, $a = 1, 2$.

- pl_2 generally outperforms pl_1 in terms of global statistical relative efficiency.

- CT shows in general good statistical relative efficiency, in particular for some specific parameters. Nevertheless, pl_2, in its weighted version, performs slightly better than CT.

4.2. Computational efficiency comparison

In order to give an idea of the computational gains of pl_1 and pl_2 with respect to ML and CT, we consider $n_w = N \times 2^w$ location sites, being uniformly distributed on the square $W_w = [0, 2^w] \times [0, 2^w]$, $w = 0, 1, \ldots, 7$. The case $w = 0$ has been used as simulation setting in the previous section. In Table 4, we report times in seconds (in terms of elapsed time using the R function $proc.time$) for the evaluations of the ML, CT, pl_a, $a = 1, 2$ functions, by considering the unweighted and weighted versions of both methods. Results in Table 4 have been obtained using an upcoming version of the package *CompRandFld* (Padoan and Bevilacqua 2015), where ML, CT and pl_a, $a = 1, 2$ have been implemented. While carrying out the experiment, we have used a 2.4 GHz processor with 16 GB of memory.

The compact support of the weight functions and the bivariate taper has been fixed equal to 0.3. Computational advantages are clear for both types of CL when increasing the number of location sites. In particular, when increasing the number of observations, pl_1 slightly outperforms pl_2, especially in the unweighted version. Considerable computational gains are obtained when pl_1 and pl_2 are evaluated with compactly supported weights. For instance, when $w = 5$, pl_1 and pl_2 are, respectively, 5800 and 20600 times faster than the standard likelihood. For CT, we use the sparse matrix implementation in the R package *spam* (Furrer and Sain 2010). The spam package allows users to separate structural and numerical computations needed for Cholesky factorization. The result is that for a given sparsity structure, the full factorization needs to be done once only. This can save a lot of time and memory requirement when the tapered likelihood function is evaluated repeatedly. The time in Table 4 is associated with the computation of the Cholesky factor, the log determinant

Table 3. Scenario C: $RE_a(\theta_i)$ and $GRE_a(\theta, 4)$ for the parameters of the bivariate separable asymmetric exponential separable model with marginal correlation between the two random fields equal to 0.15, 0.25, 0.35 when considering pl_a and its weighted version pl_a^w, $a = 1, 2$ and covariance tapering (l_{CT}) estimation methods.

ρ_{12}	0.15					0.25					0.35				
a	pl_1	pl_1^w	pl_2	pl_2^w	l_{CT}	pl_1	pl_1^w	pl_2	pl_2^w	l_{CT}	pl_1	pl_1^w	pl_2	pl_2^w	l_{CT}
$RE_a(\sigma_{11}^2)$	0.682	0.628	0.680	0.625	0.889	0.681	0.626	0.679	0.623	0.885	0.686	0.632	0.685	0.630	0.887
$RE_a(\sigma_{22}^2)$	0.702	0.650	0.701	0.647	0.894	0.703	0.652	0.702	0.649	0.891	0.705	0.653	0.703	0.651	0.890
$RE_a(\psi)$	0.447	0.450	0.506	0.502	0.792	0.356	0.363	0.450	0.450	0.795	0.284	0.292	0.406	0.403	0.788
$RE_a(k)$	0.420	0.460	0.624	0.441	0.102	0.416	0.440	0.525	0.466	0.091	0.297	0.310	0.376	0.350	0.087
$GRE_a(\theta, 4)$	0.484	0.477	0.554	0.485	0.423	0.451	0.443	0.512	0.478	0.365	0.381	0.390	0.456	0.430	0.398

Table 4. Time in seconds for evaluating the standard likelihood $l(.)$, pl_1 and pl_2 using and not using compactly supported weights functions with compact support equal to 0.3 and $l_{CT}(0.3)$ with associated percentage of nonzero values in the covariance-tapered matrices for $w = 0, 1, \ldots, 7$. When $w = 6, 7$, time in seconds for $l(.)$ and $l_{CT}(0.3)$ is not reported for computational reasons.

w	$2n_w$	$l(.)$	$l_{CT}(0.3)$	%	$pl_1(0.3)$	pl_1	$pl_2(0.3)$	pl_2
0	600	0.3	0.3	21.85	0.03	0.1	0.03	0.1
1	1200	1.1	0.6	6.51	0.04	0.3	0.03	0.3
2	2400	4.3	1.1	1.78	0.05	1.1	0.04	1.2
3	4800	24.5	2.3	0.43	0.10	3.9	0.05	4.7
4	9600	203.0	28.6	0.15	0.28	15.1	0.10	18.1
5	19, 200	5996.1	259.6	0.03	1.02	58.4	0.29	71.2
6	38, 400	–	–	–	3.87	192.1	1.04	241.2
7	76, 800	–	–	.–	15.4	617.3	4.08	808.0

of the covariance matrix and the quadratic form in (5), given a fixed sparsity structure. In Table 4, we also report the percentage of nonzero values in the covariance-tapered matrices. Although in our experiment the covariance-tapered matrices are highly sparse (0.03% of non zero elements when $w = 5$ for instance), pl_a $a = 1, 2$ methods are clearly computationally preferable with respect to the CT approach when increasing the number of location sites. Further, computational gains in the CT approach could be obtained tapering the covariance matrix only (Furrer et al. 2016), this choice leading to a biased estimator. Note the times for ML and CT when $w = 6, 7$ are not reported, since evaluations of ML and CT functions in these cases take a very long time in our machine. Our results are consistent with those of Bevilacqua and Gaetan (2015), where a comparison between CL based on pairs and CT is performed in the univariate case.

Since $pl_a, a = 1, 2$ are highly amenable to parallelization, further computational gains can be obtained using parallel computing as, for instance, in Eidsvik et al. (2014). The package *CompRandFld* uses a sequential implementation of pl_a, $a = 1, 2$ so, potentially, results in Table 4 could be considerably improved. Finally, a clear advantage of $pl_a, a = 1, 2$ with respect to ML and CT is in terms of memory storage, as it requires only 2×2, or 4×4 covariance matrices.

5. A REAL DATA EXAMPLE

In marine ecosystems, rising atmospheric CO2 and climate change are associated with concurrent shifts in temperature, circulation, stratification, nutrient input, oxygen content and ocean acidification, with potentially wide-ranging biological effects (Doney et al. 2012). An association of great interest is the concurrent shifts in temperature and the relationship with the presence of nutrients in the ocean, because of identifying changing trends in global marine phytoplankton.

This section analyzes satellite ocean data from MODIS/NASA. Specifically, we analyze the monthly average of chlorophyll concentration (Microgram per liter (μg/L)) and sea surface temperature (°K) during march of 2011 in the north of Chile observed in a regularly
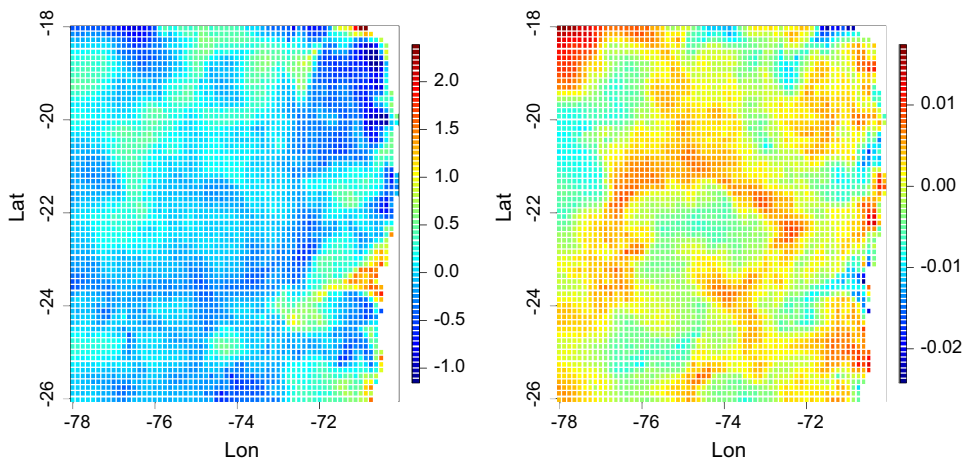
Figure 2.   Maps of residuals of chlorophyll concentration (*left*) and sea surface temperature (*right*) in the Chilean coast.

spaced grid with a resolution of 4 km, for a total of 35295 locations sites, i.e., 70590 observations. ML and CT estimations are clearly unfeasible in this case. It is well known that there is a negative correlation between the two variables (Behrenfeld et al (2006) and Boyce et al. (2010)).

We keep out the location sites where the measurement of at least one of the two variable has not been recorded (598 location sites). We then use spline regression, as implemented in the R package *mgcv* (Wood 2006), with a fixed number of knots, in order to remove the cyclic pattern of both variables along the longitude and latitude directions. A preliminary analysis (see the, for instance, the empirical cross-variograms in Fig. 2) suggests that the residuals can be considered as a realization of a zero mean weakly stationary and isotropic bivariate Gaussian random field. As outlined by an anonymous Referee, the empirical cross-variogram estimates the average variance of the first-order contrast, but it is not informative of possible local nonstationary behaviors. Figure 2 shows the maps of the residuals, and as expected, a local nonstationary behavior near the coast is apparent for both chlorophyll concentration and sea surface temperature in particular at latitude $[-26, 22]$ and longitude $[-72, 71]$.

We specify the bivariate covariance model in Eq. (21) in an increasing order of complexity (separable version, constrained version as in the setting A of the simulation study and unconstrained version). We then estimate through pl_1, using the weight functions $w_{11mn} = u_{11mn}(70)$, $w_{22mn} = u_{22mn}(50)$ and $w_{12mn} = k_{12mn}(60)$, as well as through pl_2, using $w_{12mn} = u_{12mn}(60)$. The choice of the different compact supports in the weight functions for pl_1 is due to the different spatial dependence (see Fig. 3). Table 5 reports pl_a, $a = 1, 2$ estimates with the associated standard errors, computed using subsampling technique as described in Sect. 3. Note that the estimation of colocated parameter, variances and scale(s) parameters is very similar for both methods in each considered model. As expected, the estimation of the parameter ρ_{12}, expressing the marginal correlation between chlorophyll concentration and sea surface temperature, is negative.
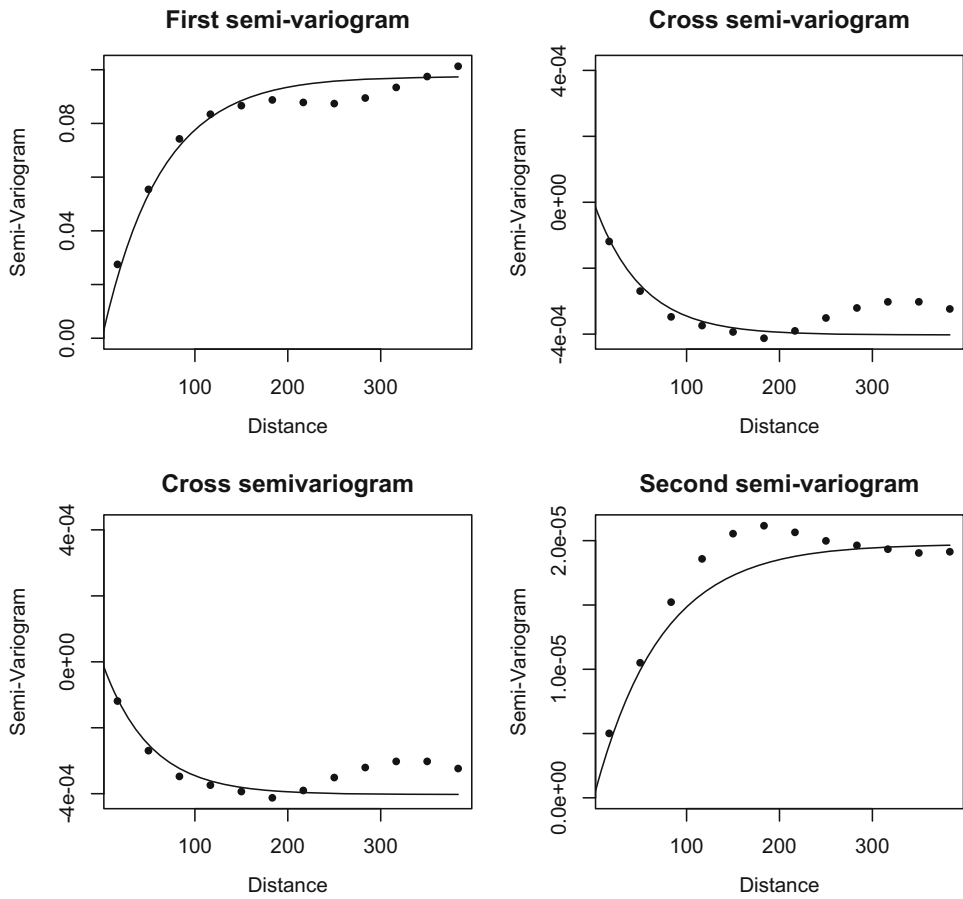
Figure 3. Comparison of the empirical (cross) variograms estimation with full exponential variogram model (model 3) estimated with pl_1. First semivariogram is associated with chlorophyll, and second semivariogram is associated with temperature.

Table 5 reports the value of the CL information criteria (CLIC) for pl_a $a = 1, 2$ for each covariance model considered. In both cases, the model selected is Model 3. Note that the flexibility of this model allows to estimate a smaller cross-scale parameter α_{12} with respect the marginal scale parameters α_{ii}, $i = 1, 2$.

Finally, Fig. 3 shows the comparison between the empirical semi(cross)-variograms and the estimated semi(cross)- variograms associated with the model C estimated using pl_2. The comparison shows a good fitting of the estimated model C with the empirical (cross) variogram.

6. CONCLUDING REMARKS

CL is an appealing method of estimation when dealing with large datasets. As outlined in Bevilacqua and Gaetan (2015), CL is a large class of estimating functions, and for a given estimation problem, it is not clear how to choose within this class. Some insights are

Table 5. $pl_a, a = 1, 2$ estimates with associated standard errors for models 1, 2, 3 and associated CL information criteria (CLIC).

	1		2		3	
	pl_1	pl_2	pl_1	pl_2	pl_1	pl_2
σ_{11}^2	9.80e-02	9.94e-02	9.76e-02	9.782e-02	9.75e-02	9.77e-02
	(1.4e-02)	(1.3e-02)	(1.e-02)	(1.5e-02)	(1.5e-02)	(1.4e-02)
σ_{22}^2	1.94e-05	1.94e-05	1.97e-05	1.96e-05	1.97e-05	1.96e-05
	(1.6e-06)	(1.5e-06)	(1.7e-06)	(1.5e-06)	(1.7e-06)	(1.5e-06)
ρ_{12}	−0.288	−0.298	−0.290	−0.290	−0.290	−0.290
	(9.4e-02)	(5.3e-02)	(9.3e-02)	(5.4e-02)	(1.0e-01)	(5.9e-02)
α_{11}	66.63	68.67	62.98	65.42	62.98	63.77
	(5.21)	(5.52)	(6.42)	(6.29)	(6.42)	(6.53)
α_{12}	-	-	-	-	51.20	52.94
					(13.99)	(11.63)
α_{22}	-	-	71.71	71.96	71.74	70.39
			(7.69)	(7.84)	(7.68)	(7.25)
CLIC	−140876799	−212867004	−140853167	−212834694	−140852194	−212095990

given in Castruccio et al. (2016), where different types of marginal CLs are compared from statistical efficiency viewpoint. In the Gaussian case, if the choice of the CL is driven by computational reasons only, then the CL based on pairs has clear computational advantages with respect to other types of CLs. For this reason, we consider this type of CL when estimating multivariate Gaussian random fields. In particular, in this paper we have proposed two possible approaches, based on weighted composite likelihood methods, for the estimation of MGRF. The first is based on cross-pairwise likelihood, while the second is based on pairs of marginal pairwise likelihoods. Through numerical examples, we have compared the two methods using the maximum likelihood and covariance tapering as benchmarks when estimating different types of multivariate covariance models. The second version generally outperforms the first from statistical efficiency point of view, but it cannot be used in the heterotopic case.

The numerical examples show that the weighted versions have better performance from statistical efficiency point of view and considerable computational gains with respect to the unweighted versions. CT is a good competitor of both methods; nevertheless, our methods are clearly preferable in terms of computational complexity when considering very large datasets. The benefit of our methods in terms of memory requirement with respect to ML and CT is apparent since it requires only very small covariance matrices storage. Moreover, since our methods are highly amenable to parallelization, further computational gains can be achieved, using parallel computing techniques as described, for instance, in Lee et al. (2010) and Suchard et al. (2010).

Our methods can be easily extended to the estimation of multivariate space time covariance models when considering fully symmetric covariance models. For estimation of asymmetric covariance models (Stein 2005; Gneiting et al. 2007), asymmetry in time should be taken into account in CL following the lines of section 3. Finally, since our methods require model assumptions on the lower-dimensional marginal densities, they can be very useful for multivariate non Gaussian random field estimation where, typically, the likelihood function does not have a simple form, as in Genton et al. (2015).

ACKNOWLEDGEMENTS

REFERENCES

Apanasovich, T., Genton, M., and Sun, Y. (2012), "A valid Matérn class of cross-covariance functions for multivariate random fields with any number of components," *Journal of the American Statistical Association*, 97, 15–30.

Bevilacqua, M., Fassò, A., Gaetan, C., Porcu, E., and Velandia, D. (2016), "Covariance tapering for multivariate Gaussian random fields estimation," *Statistical Methods & Applications*, 25(1), 21–37.

Bevilacqua, M., and 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., and 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.

Bevilacqua, M., Vallejos, R., and Velandia, D. (2015), "Assessing the significance of the correlation between the components of a bivariate Gaussian random field," *Environmetrics*, 26, 545–556.

Boyce, D. G., Lewis, M. R., and Worm, B. (2010), "Global phytoplankton decline over the past century," *Nature. International weekly journal of science*, 466. doi:10.1038/nature09268.

Castruccio, S., Huser, R., and Genton, M. G. (2016), "High-order composite likelihood inference for max-stable distributions and processes," *Journal of Computational and Graphical Statistics*. To appear.

Daley, D., Porcu, E., and Bevilacqua, M. (2015), "Classes of compactly supported covariance functions for multivariate random fields," *Stoch Environ Res Risk Assess*, 29, 1249–1263.

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

Doney, S. C., Ruckelshaus, M., Duffy, J. E., Barry, J. P., Chan, F., English, C. A., Galindo, H. M., Grebmeier, J. M., Hollowed, A. B., Knowlton, N., Polovina, J., Rabalais, N. N., Sydeman, W. J., and Talley, L. D. (2012), "Annual Review of Marine Science," *Nature. International weekly journal of science*, 4, 11–37.

Eidsvik, J., Shaby, B., Reich, B., Wheeler, M., and Niemi, J. (2014), "Estimation and prediction in spatial models with block composite likelihoods," *Journal of Computational and Graphical Statistics*, 29, 295–315.

Furrer, R., Bachoc, F., and Du, J. (2016), "Asymptotic properties of multivariate tapering for estimation and prediction," *Journal of Multivariate Analysis*, In press.

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

Genton, M. G., Padoan, S., and Sang, H. (2015), "Multivariate max-stable spatial processes," *Biometrika*, 102, 215–230.

Genton, M., and Kleiber, W. (2015), "Cross-Covariance Functions for Multivariate Geostatistics," *Statistical Science*, in press.

Gneiting, T. (2002), "Compactly supported correlation functions," *Journal of Multivariate Analysis*, 83, 493–508.

Gneiting, T., Genton, M. G., and Guttorp, P. (2007), "Geostatistical space-time models, stationarity, separability and full symmetry," in *Statistical Methods for Spatio-Temporal Systems*, eds. B. Finkenstadt, L. Held, and V. Isham, Boca Raton: FL: Chapman & Hall/CRC, pp. 151–175.

Gneiting, T., Kleiber, W., and Schlather, M. (2010), "Matérn Cross-Covariance Functions for Multivariate Random Fields," *Journal of the American Statistical Association*, 105, 1167–1177.

Goulard, M., and Voltz, M. (1992), "Linear coregionalization model: Tools for estimation and choice of cross-variogram matrix," *Mathematical Geology*, 24, 269–286.

Heagerty, P., and Lumley, T. (2000), "Window subsampling of estimating functions with application to regression models," *Journal of the American Statistical Association*, 95, 197–211.

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

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

Lee, A., Yau, C., Giles, M., Doucet, A., and Holmes, 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., and Lahiri, S. (2002), "Least squares variogram fitting by spatial subsampling," *Journal of the Royal Statistical Society B*, 64, 837–854.

Li, B., and Zhang, H. (2011), "An approach to modeling asymmetric multivariate spatial covariance structures," *Journal of Multivariate Analysis*, 102, 1445–1453.

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

Padoan, S. A., and Bevilacqua, M. (2015), "Analysis of Random Fields Using CompRandFld," *Journal of Statistical Software*, 63, 1–27.

Pelletier, B., Dutilleul, P., Larocque, G., and Fyles, J. (2004), "Fitting the linear model of coregionalization by generalized least squares," *Mathematical Geology*, 36(3), 323–343.

Porcu, E., Daley, D., Buhmann, M., and Bevilacqua, M. (2013), "Radial basis functions with compact support for multivariate geostatistics," *Stochastic Environmental Research and Risk Assessment*, 27, 909–922.

Shaby, B., and Ruppert, D. (2012), "Tapered covariance: Bayesian estimation and asymptotics," *Journal of Computational and Graphical Statistics*, 21, 433–452.

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

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

Suchard, M., Wang, Q., anf J. Frelinger, C. C., Cron, A., and West, M. (2010), "Understanding GPU programming for statistical computation: studies in massively parallel massive mixtures," *Journal of Computational and Graphical Statistics*, 19, 419 –438.

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

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

Wackernagel, H. (2003), *Multivariate Geostatistics: An Introduction with Applications*, 3rd edn, New York: Springer.

Wood, S. (2006), *Generalized Additive Models: An Introduction with R*, : Chapman and Hall CRC.

Zhang, H. (2007), "Maximum-likelihood estimation for multivariate spatial linear coregionalization models," *Environmetrics*, 18, 125–139.