CrossMark

Covariance tapering for multivariate Gaussian random fields estimation

M. Bevilacqua[1] · **A. Fassò**[2] · **C. Gaetan**[3] ·
E. Porcu[4] · **D. Velandia**[1]

Abstract In recent literature there has been a growing interest in the construction of covariance models for multivariate Gaussian random fields. However, effective estimation methods for these models are somehow unexplored. The maximum likelihood method has attractive features, but when we deal with large data sets this solution becomes impractical, so computationally efficient solutions have to be devised. In this paper we explore the use of the covariance tapering method for the estimation of multivariate covariance models. In particular, through a simulation study, we compare the use of simple separable tapers with more flexible multivariate tapers recently pro-

✉ M. Bevilacqua
moreno.bevilacqua@uv.cl

A. Fassò
alessandro.fasso@unibg.it

C. Gaetan
gaetan@unive.it

E. Porcu
georgepolya01@gmail.com

D. Velandia
dairavm@gmail.com

[1] Instituto de Estadística, Universidad de Valparaíso, Valparaiso, Chile

[2] Dipartimento di Ingegneria, Universitá degli Studi di Bergamo, Bergamo, Italy

[3] DAIS, Universitá Cá Foscari-Venezia, Venice, Italy

[4] Departamento de Matemática, Universidad Técnica Federico Santa María,
    Valparaiso, Chile

⚙ Springer

posed in the literature and we discuss the asymptotic properties of the method under increasing domain asymptotics.

Keywords Cross Covariance estimation · Large datasets · Multivariate compactly supported correlation function · Multivariate Gaussian process

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 models able to catch the simultaneous behavior of different variables observed in the same spatial region. The analysis of this kind of data requires not only the specification of the dependence within the phenomenon of interest but also the dependence between other phenomena observed in the same spatial domain. The related statistical modeling is often restrained by computational challenges when the data sets get larger. For example, modeling local air quality as in Arima et al. (2012) and Fontanella and Ippoliti (2003) or global carbon budget (Vetter et al. 2015) often expresses the need for tools capable to handle such complexity.

Multivariate Gaussian random fields (MGRF throughout) are important tools to describe such dependence (Wackernagel 2003). Nevertheless, modeling the covariance function of such MGRF is not an easy task. The linear model of corregionalization (LMC) has been the cornerstone in multivariate geostatistics for many years despite the drawbacks outlined in Porcu et al. (2013), Gneiting et al. (2010). For instance, the smoothness of any component of the MGRF is restricted to that of the roughest underlying univariate process. In order to overcome these problems, new multivariate covariance models have been proposed in the last years [for a recent review see Genton and Kleiber (2015)]. Among them, Gneiting et al. (2010) and Apanasovich et al. (2012) extend the Matérn model to the multivariate case. These models are very flexible and can capture both the marginal and the cross spatial dependence as well as the level of smoothness associated and the correlation between the components. Bevilacqua et al. (2015) compare the Matérn and the LMC models in terms of flexibility.

Effective estimation methods for multivariate 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 requires calculation of the inverse and determinant of the covariance matrix, and this evaluation is slow when the number of observations is large. Specifically, the computational problem is generally the Cholesky decomposition of the covariance matrix, whose computational cost in the multivariate case is of order $O(np)^3$ where n is the number of location sites for each of the p components of the MGRF. In the univariate case, different approaches have been proposed in order to find estimation methods with a good balance between computational complexity and statistical efficiency for large data sets. Some of these approaches proposed specific type of composite likelihood as in Vecchia (1988), Stein et al. (2004), Bevilacqua et al. (2012), Bevilacqua and Gaetan (2015), and Eidsvik et al. (2014), or stochastic approximations of the score function as in Stein et al. (2012, 2013).

Another possible approach is to consider an approximation of the covariance matrix, as in the covariance tapering (CT) method (Kaufman et al. 2008). With this approach, certain elements of the covariance matrix, that correspond to pairs with large distance, are set to zero by multiplying elementwise the covariance matrix with a taper matrix, this last being a positive definite sparse correlation matrix coming from a compactly supported correlation function. Then sparse matrix algorithms can be used to evaluate efficiently an approximate likelihood. Kaufman et al. (2008) propose two types of estimation based on CT. The first type, called one-taper, leads to a biased estimating equation and it is computationally more efficient with respect to the second type, called two-taper. Nervetheless the second type leads to an unbiased estimating equation and, under increasing domain, asymptotic properties of the method are well known (Shaby and Ruppert 2012).

For univariate Gaussian fields there is a large literature on compactly supported correlation functions [see Gneiting (2002) and the references therein] but the extension to the multivariate case is not trivial. Recently, Daley et al. (2015) have introduced a class of multivariate models with possibly different compact supports, extending the Wendland compactly supported correlation functions to the multivariate case.

In this paper we propose the multivariate two-taper CT method for the estimation of the MGRF. In particular, we make use of the multivariate tapers proposed in Daley et al. (2015). We compare, through a simulation study, the use of separable tapers with a common compact support, with more flexible tapers allowing different compact supports, when estimating a bivariate Matérn model and a LMC. Asymptotic properties of the proposed method (under increasing domain asymptotics) are discussed extending the results in Shaby and Ruppert (2012).

The paper is organized as follows. In Sect. 2 we review some possible models for building multivariate tapers, while Sect. 3 describes the multivariate CT method and its asymptotic properties. In Sect. 4, through a simulation study, we compare CT with ML from statistical and computational viewpoint and also we compare the CT using separable and nonseparable tapers. Finally, in Sect. 5 we give some conclusions.

2 Multivariate covariance models and tapers

For the remainder of the paper, we denote $\mathbf{Z}(\mathbf{s}) = \{(Z_1(\mathbf{s}), \ldots, Z_p(\mathbf{s}))^T\}$, a p-variate weakly stationary Gaussian field with continuous spatial index $\mathbf{s} \in \mathbb{R}^d$. The assumption of Gaussianity implies that the first and second order moments determine uniquely the finite dimensional distributions. In particular, we shall suppose weak stationarity throughout, so that the mean vector $\boldsymbol{\mu} = \mathbb{E}(\mathbf{Z})$ is constant and the covariance function between $\mathbf{Z}(\mathbf{s}_1)$ and $\mathbf{Z}(\mathbf{s}_2)$, for any pair $\mathbf{s}_1, \mathbf{s}_2$ in the spatial domain, is represented by a mapping $\mathbf{C} : \mathbb{R}^d \to M_{p \times p}$ defined through

$$\mathbf{C}(\mathbf{h}) = \left[C_{ij}(\mathbf{h}) \right]_{i,j=1}^p = \left[\text{cov} \left(Z_i(\mathbf{s}_1), Z_j(\mathbf{s}_2) \right) \right]_{i,j=1}^p, \quad \mathbf{h} = \mathbf{s}_1 - \mathbf{s}_2 \in \mathbb{R}^d. \quad (1)$$

The function $\mathbf{C}(\mathbf{h})$ is called matrix valued covariance function and when $p = 2$ we call it bivariate 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 $Z_i(s)$, $i = 1, \ldots, p$, whilst for $i \neq j$ the mapping C_{ij} is called cross-covariance between $Z_i(s)$ and $Z_j(s)$ at the spatial lag $h \in \mathbb{R}^d$. The matrix-valued mapping C must be positive definite, which means that, for a given realization $Z = (Z(s_1)^T, \ldots, Z(s_n)^T)^T$, the $(np) \times (np)$ covariance matrix $\Sigma := [C(s_i - s_j)]_{i,j=1}^n$ is positive definite.

We shall assume throughout that the mapping C comes from a parametric family of matrix valued covariances $\{C(\cdot; \theta), \theta \in \Theta \subseteq \mathbb{R}^k\}$, with Θ an arbitrary parametric space. Here, we shall list those parametric models that will be used throughout the paper. One of them is the linear model of coregionalization, that has been popular for over thirty 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:

$$C(h; \theta) = \left[\sum_{s=1}^{q} \alpha_{is} \alpha_{js} R_s(h, \psi_s) \right]_{i,j=1}^{p}, \tag{2}$$

with $A := [\alpha_{lm}]_{l,m=1}^{p,q}$ being a $p \times q$ dimensional matrix with full rank, and with $R_s(\cdot, \psi_s)$ with $s = 1, \ldots, q$, being a univariate parametric correlation model. Clearly, we have $\theta = (\text{vec}(A)^T, \psi_1^T, \ldots, \psi_q^T)^T$. Constructive criticism about this model has been expressed by Gneiting et al. (2010), Apanasovich et al. (2012) and Daley et al. (2015). For instance, if $\alpha_{lm} \neq 0$ for each l, m the smoothness of any component defaults to that of the roughest latent process.

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

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

where $R(\cdot, \psi)$ is a univariate parametric correlation model, $\sigma_i^2 > 0$, $i = 1, \ldots, p$, are the marginal variances and the ρ_{ij} expresses the marginal correlation between $Z_i(s)$ and $Z_j(s)$. In this case $\theta = (\rho, \psi^T, \sigma_1^2, \ldots, \sigma_p^2)^T$ where ρ is the vector containing all the pairwise correlations ρ_{ij}, $i = 1, \ldots p - 1$, $j = i + 1, \ldots, p$. This type of 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 characteristics of the components.

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

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

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 model (4) with $R(h, .)$ equal to the Matérn correlation model:

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

where $\boldsymbol{\psi} = (\alpha, \nu)$, $\alpha > 0$ is the scale parameter, $\nu > 0$ indexes differentiability at the origin and $\mathcal{K}_\nu(\cdot)$ is the modified Bessel function of the second kind of order ν. In the bivariate case, they find necessary and sufficient conditions on the parameters, whilst for the case $p \geq 3$ they only offered sufficient conditions. This kind of construction allows for a nice closed form, together with the possibility of different spatial scale and smoothness parameters and different variances.

2.1 Multivariate tapers

In the univariate case a taper is a correlation function being additionally compactly supported on a ball of $\mathbb{R}^d$ with given radius. Notable examples of tapers are the following.

We denote $(x)_+$ for the positive part of $x \in \mathbb{R}$. The Askey function (Askey 1973) defined here as

$$\mathcal{A}_\nu(\boldsymbol{h}; b) = \left(1 - \frac{\|\boldsymbol{h}\|}{b}\right)_+^\nu, \qquad \boldsymbol{h} \in \mathbb{R}^d, \quad b > 0, \tag{6}$$

is compactly supported over the ball of $\mathbb{R}^d$ with radius b. Positive definiteness in $\mathbb{R}^d$ is guaranteed whenever $\nu \geq \frac{(d+1)}{2}$ (Zastavnyi and Trigub 2002). Gneiting (2002) argues convincingly about the need for tapers being smooth at the origin, which is not the case for the Askey taper. The ingenious Wendland-Gneiting (Gneiting 2002) construction allows to overcome such a drawback through the use of the Montée operator (Matheron 1962). An example is the C^2 Wendland function, that we define as

$$\mathcal{W}_\nu(\boldsymbol{h}; b) = \left(1 + \nu \frac{\|\boldsymbol{h}\|}{b}\right) \mathcal{A}_\nu(\boldsymbol{h}; b), \qquad \boldsymbol{h} \in \mathbb{R}^d, \quad b > 0, \tag{7}$$

where $\nu > (d + 5)/2$ in order to preserve positive definiteness in $\mathbb{R}^d$.

For the remainder of the paper, we call a *multitaper* a matrix valued correlation function, having members R_{ij} being compactly supported over balls of arbitrary radius, and such that the functions on the diagonal ($i = j$) are univariate tapers. We denote with $\boldsymbol{R}_{\text{Tap}}$ a multitaper. Building a multitaper implies a nontrivial mathematical effort and the reader is referred to paper by Daley et al. (2015) with the details therein. It might be interesting to have multitapers that allow for

1. possibly different compact supports;
2. possibly different levels of differentiability at the origin;
3. a good balance between the number of parameters to be fixed and the multitaper flexibility.

The first feature is desirable since each compact support can be chosen in order to taper the marginal and the cross-covariance matrices in a different way, that is to fix the level of sparseness for each sub-matrix. The level of differentiability of the taper is an important feature since it is related to the level of differentiability of the covariance model to be tapered when considering results under infill asymptotics as in Furrer et al. (2013) and Du et al. (2009). Let us sketch some possible constructions:

1. *Separable multitapers*. This responds to the construction in Eq. (3), with emphasis on that R must be compactly supported (i.e., a univariate taper as those indicated above). In this case, the taper can be written as

$$\boldsymbol{R}_{\text{Tap}}(\boldsymbol{h}; \boldsymbol{d}) = \left[r_{ij} R_{\text{Tap}}(\boldsymbol{h}; b) \right]_{i,j=1}^{p}, \quad b > 0, \tag{8}$$

where the symmetric matrix $A := [r_{ij}]_{i,j=1}^{p}$, is a matrix of fixed coefficients such that $r_{ii} = 1$, $-1 \leq r_{ij} \leq 1$ and A is positive semidefinite. Here $\boldsymbol{d} = (\boldsymbol{r}^T, b)^T$ is the multitaper parameter vector where $\boldsymbol{r}$ is the vector containing all the pairwise coefficients r_{ij}, $i = 1, \ldots p - 1$, $j = i + 1, \ldots, p$. We write $R_{\text{Tap}}(\cdot; b)$ for a univariate taper, as for instance those in Sect. 2.1, where b determines the radius of the ball of $\mathbb{R}^d$ over which R is compactly supported. Thus, $\boldsymbol{R}_{\text{Tap}}(\cdot; \boldsymbol{d})$ inherits the compact support and level of differentiability of $R_{\text{Tap}}(\boldsymbol{h}; b)$. Note that the number of parameters to be fixed is $\frac{p(p-1)}{2} + 1$.

2. *Multitapers based on a linear model of coregionalization*, so that

$$\boldsymbol{R}_{\text{Tap}}(\boldsymbol{h}; \boldsymbol{d}) = \left[\sum_{k=1}^{q} r_{ik} r_{jk} R_{\text{Tap},k}(\boldsymbol{h}; b_k) \right]_{i,j=1}^{p}, \quad b_k > 0,$$

with $A := [r_{lm}]_{l,m=1}^{p,q}$ being a $p \times q$ dimensional matrix with full rank, and with $R_{\text{Tap},k}(\cdot; b_k)$ being a univariate taper. Note that since $\sum_{k=1}^{q} r_{ik}^2 R_{\text{Tap},k}(\boldsymbol{h}; b_k)$ must be a univariate taper for each i, then $\sum_{k=1}^{q} r_{ik}^2 = 1$ and the number of parameters to be fixed, i.e. the cardinality of $\boldsymbol{d}$, is $p(q - 1) + q$. Moreover, if $r_{lm} \neq 0$ for each l, m then the model shares the same compact support (the minimum of b_i, $i = 1, \ldots, k$) and the same level of differentiability (the minimum between the levels of differentiability associated to $R_{\text{Tap},k}(\boldsymbol{h}; b_k)$). However, it is possible to fix some coefficients equal to zero in order to attain p different compact supports and p different levels of differentiability.

3. *Nonseparable multitapers*, based on the following construction:

$$\boldsymbol{R}_{\text{Tap}}(\boldsymbol{h}; \boldsymbol{d}) = [r_{ij} R_{\text{Tap}}(\boldsymbol{h}, b_{ij})]_{i,j=1}^{p}, \quad b_{ij} > 0, \tag{9}$$

with $R_{\text{Tap}}(\cdot; b)$ being Wendland tapers (one of the entries in Sect. 2.1). Sufficient conditions for the validity of this construction (which depends on the coefficients r_{ij}) are given in Theorem 2 in Daley et al. (2015).

This last model allows to attain different compacts supports for each marginal and cross-covariance matrix, and the same level of differentiability at the origin. In this case, the number of parameters to be fixed is $p(p - 1) + p$.

Table 1 depicts the features of multitapers 1, 2 and 3 in the bivariate and trivariate case. As expected, there is a sort of trade off between the number of the multitaper parameters and its flexibility. The separable model is the simplest one, but it attains common compact supports and same level of differentiability at the origin. The LMC and the nonseparable multitaper share the same number of parameters but the last is more flexible from the number of compact supports viewpoint while the LMC

Table 1 Features of separable, nonseparable and LMC multitaper for $p = 2, 3$

	$p = 2$			$p = 3$		
	N. par.	N. comp.supp.	N. lev. diff.	N. par.	N. comp.supp.	N. lev. diff.
Separable	2	1	1	4	1	1
LMC	4	2	2	9	3	3
No separable	4	3	1	9	6	1

construction allows for different levels of differentiability at the origin. Since in this paper we are not interested in different level of differentiability, hereafter we consider multitapers of the type (9), including the separable multitaper as defined in Eq. (8).

3 Multivariate covariance tapering

Since we are assuming that the state of truth is represented by some parametric family of matrix–valued covariances $\{C(\cdot; \theta), \theta \in \Theta \subseteq R^k\}$, we may use the abuse of notation $\Sigma(\theta)$ for the covariance matrix Σ, in order to emphasize the dependence on the unknown parameters vector. For a realization Z from a p-variate Gaussian random field, the log likelihood can be written, up to an additive constant, as

$$l_n(\theta) = -\frac{1}{2} \log |\Sigma(\theta)| - \frac{1}{2} Z^T \Sigma(\theta).^{-1} Z, \tag{10}$$

The most time-consuming part when calculating (10) is to evaluate the determinant and inverse of $\Sigma(\theta)$. The most widely used algorithms, such as Cholesky decomposition, require up to $O((np)^3)$ steps. This can be prohibitive if n is large.

When extending the CT approach proposed in Kaufman et al. (2008) to the multivariate case, certain elements of the covariance matrix $\Sigma(\theta)$ are set to zero by multiplying $\Sigma(\theta)$ element by element with a sparse matrix coming from a multitaper model. Let us denote with $T(d)$ the $np \times np$ matrix associated to a multitaper $R_{\text{Tap}}(h; d)$ as for instance those described in Sect. 2.1. The 'tapered' covariance matrix is then obtained through

$$\Sigma_T(\theta) = \Sigma(\theta) \circ T(d),$$

where $\circ$ denotes the Schur product. The multitaper vector parameters d is fixed including the (possibly different) radiii and these are fixed in a way to determine the desired level of sparseness for the construction above. Observe that this construction guarantees positive definiteness of $\Sigma_T(\theta)$ when $T(d)$ is positive definite or positive semidefinite with positive diagonal elements (Horn and Johnson 1991). The multitapered likelihood is defined as:

$$l_{T,n}(\theta, d) = -\frac{1}{2} \log |\Sigma_T(\theta)| - \frac{1}{2} Z^T (\Sigma_T(\theta)^{-1} \circ T(d)) Z, \tag{11}$$

and algorithms for sparse matrices can be exploited in order to compute (11) efficiently.

Under increasing domain, it can be shown that the maximizer of (11) is consistent and asymptotically Gaussian, extending the result in Shaby and Ruppert (2012). Theorems 1 and 2 in the Supplementary material give conditions for the consistency and the asymptotically normality of the CT estimator. Let us define the Godambe information matrix

$$G_{T,n}(\boldsymbol{\theta}, \boldsymbol{d}) = H_{T,n}(\boldsymbol{\theta}, \boldsymbol{d}) J_{T,n}(\boldsymbol{\theta}, \boldsymbol{d})^{-1} H_{T,n}(\boldsymbol{\theta}, \boldsymbol{d})^{T}, \tag{12}$$

where

$$H_{T,n}(\boldsymbol{\theta}, \boldsymbol{d}) = -\mathrm{E}\left[\nabla^2 l_{T,n}(\boldsymbol{\theta}, \boldsymbol{d})\right], \qquad J_{T,n}(\boldsymbol{\theta}, \boldsymbol{d}) = \mathrm{E}\left[\nabla l_{T,n}(\boldsymbol{\theta}, \boldsymbol{d}) \nabla l_{T,n}(\boldsymbol{\theta}, \boldsymbol{d})^{T}\right]. \tag{13}$$

The generic entries of the $H_{T,n}(\boldsymbol{\theta}, \boldsymbol{d})$ and $J_{T,n}(\boldsymbol{\theta}, \boldsymbol{d})$ matrices are respectively:

$$[H_{T,n}(\boldsymbol{\theta}, \boldsymbol{d})]_{ij} = \frac{1}{2}\mathrm{tr}\left\{B_i\left(\frac{\partial \Sigma(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j} \circ \boldsymbol{T}(\boldsymbol{d})\right)\right\},$$

$$\text{and } [J_{T,n}(\boldsymbol{\theta}, \boldsymbol{d})]_{ij} = \frac{1}{2}\mathrm{tr}\left\{[B_i \circ \boldsymbol{T}(\boldsymbol{d})]\,\Sigma(\boldsymbol{\theta})\,[B_j \circ \boldsymbol{T}(\boldsymbol{d})]\,\Sigma(\boldsymbol{\theta})\right\},$$

$$\text{where } B_i = \Sigma_T(\boldsymbol{\theta})^{-1}\left(\frac{\partial \Sigma(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i} \circ \boldsymbol{T}(\boldsymbol{d})\right)\Sigma_T(\boldsymbol{\theta})^{-1}.$$

Then, under increasing domain it can be shown that the maximizer of (11) has Gaussian asymptotic distribution with variance covariance matrix equal to the inverse of the Godambe information (12). As outlined in Shaby and Ruppert (2012), the requirement of increasing domain is not stated explicitly but rather is implied by conditions on the eigenvalues of the covariance matrix and its derivatives that in general are not easy to check. In the Supplementary material, we have showed that a bivariate exponential separable model satisfies the condition of Theorem 1 and 2 when using a separable taper.

4 Numerical examples

This section is organized in order to answer, through numerical examples, to the following questions:

1. How does the CT method perform with respect to the the ML from statistical and computational efficiency point of view?
2. How does the choice of the parameters r_{ij} in the multitapers (8) and (9) affect the efficiency of the CT method?
3. Does the use of a flexible multitaper (9) allow to improve the statistical efficiency of the method with respect of a separable taper (8)?

We work under the bivariate case ($p = 2$) and in particular we consider 400 location sites, being uniformly distributed on the square $[0, 1]^2$. We consider two bivariate covariance models, the former being a special case of the model in Eq. (4), obtained by fixing $R(\boldsymbol{h}; .)$ equal to the exponential model, so that

$$C(\boldsymbol{h}, \boldsymbol{\theta}) = \left[\rho_{ij}\sigma_i\sigma_j e^{-\frac{\|\boldsymbol{h}\|}{\psi_{ij}}} \right]^2_{i,j}, \tag{14}$$

where $\rho_{ii} = 1$ and $\boldsymbol{\theta} = (\rho_{12}, \psi_{11}, \psi_{12}, \psi_{22}, \sigma_1^2, \sigma_2^2)^T$. The second model we consider is a special case of the bivariate LMC model in Eq. (2):

$$C(\boldsymbol{h}; \boldsymbol{\theta}) = \left[\sum_{k=1}^{2} \alpha_{ik}\alpha_{jk} e^{-\frac{\|\boldsymbol{h}\|}{\psi_i}} \right]^2_{i,j=1}, \tag{15}$$

where $\psi_i > 0$ and $[\alpha_{lm}]^2_{l,m=1}$ is a matrix of rank 2. Following Daley et al. (2015), as bivariate taper we consider :

$$\boldsymbol{R}_{\mathrm{Tap}}(\boldsymbol{h}; \boldsymbol{d}) = \left[r_{ij} \left(1 + 5\frac{\|\boldsymbol{h}\|}{b_{ij}} \right) \left(1 - \frac{\|\boldsymbol{h}\|}{b_{ij}} \right)^5_+ \right]^2_{i,j=1} \tag{16}$$

obtained from (9) fixing $R_{\mathrm{Tap}}(\boldsymbol{h}, b) = \mathcal{W}_5(\boldsymbol{h}; b)$, and $\boldsymbol{d} = (b_{11}, b_{12}, b_{22}, r_{12})^T$.

For statistical efficiency comparison and, in order to answer to the first question, we focus on the relative mean square error, defined through $RE(\boldsymbol{\theta}_i) = \dfrac{\mathrm{MSE}_{ML}(\boldsymbol{\theta}_i)}{\mathrm{MSE}_T(\boldsymbol{\theta}_i)}$ for $i = 1, \ldots, k$, whilst as global measure of relative efficiency we consider:

$$GRE(\boldsymbol{\theta}, k) = \left(\frac{\det(V_{ML}(\boldsymbol{\theta}))}{\det(V_T(\boldsymbol{\theta}))} \right)^{1/k}$$

where V_x, with $x = ML, T$, is the sample variance covariance matrix of the ML and CT estimates, k is the number of parameters involved in the estimation and $\det(A)$ is the determinant of the matrix A.

We consider three scenarios (denoted A, B and C), with increasing complexity of the bivariate exponential model in Eq. (14), and two additional scenarios (denoted D and E), with increasing complexity of the bivariate LMC model in Eq. (15). For each scenario, we simulate 1000 realizations from a bivariate zero mean GRF with a specified bivariate covariance model, and then we estimate using ML and CT, using both a separable and a nonseparable version of the bivariate taper as in Eq. (16). All the estimates have been carried out using un upcoming version of the R package CompRandFld (Padoan and Bevilacqua 2015), avalaible on CRAN. Let us describe further the five scenarios considered.

Scenario A. We use the model in Eq. (14) with $\psi_{12} = \psi_{11} = \psi_{22}$, i.e. a separable model. In this case, $\boldsymbol{\theta} = (\rho_{12}, \psi, \sigma_1^2, \sigma_2^2)^T$ and we fixed $\sigma_1^2 = \sigma_2^2 = 1$, $\psi_{11} = \psi_{12} = \psi_{22} = \psi = pr/3$ with $pr = 0.2$ (the common practical range) and $\rho_{12} = x$, $x = 0.1, 0.5, 0.9$. For this scenario, we consider a separable bivariate taper as in Eq. (16) with a common compact support b. Here, we fix $b = pr/2$ and r_{12} varying in its range of permissibility, that is $r_{12} = 0.1, 0.5, 1$.

Note that we fix the compact support as a fraction of the practical range, as suggested in Kaufman et al. (2008), and in order to answer to the second question, we consider

Table 2 $RE(\theta_i)$ and $GRE(\theta, 4)$ for the parameters of the bivariate exponential separable model using a separable taper with $r_{12} = 0.1, 0.5, 1$ when $\rho_{12} = 0.1, 0.5, 0.9$ (scenario A)

r_{12}	$\rho_{12} = 0.1$			$\rho_{12} = 0.5$			$\rho_{12} = 0.9$		
	0.1	0.5	1	0.1	0.5	1	0.1	0.5	1
σ_{11}^2	0.635	0.636	0.638	0.625	0.630	0.644	0.596	0.591	0.642
σ_{22}^2	0.681	0.682	0.685	0.635	0.640	0.653	0.598	0.599	0.653
ψ	0.546	0.550	0.556	0.437	0.476	0.557	0.289	0.306	0.560
ρ_{12}	0.314	0.348	0.350	0.367	0.370	0.371	0.341	0.376	0.378
$GRE(\theta, 4)$	0.420	0.431	0.432	0.406	0.413	0.430	0.343	0.369	0.426

different values of r_{12}. The percentage of non zero values in the tapered matrix is 3.2 %. Moreover, we considered different marginal correlations, i.e. different colocated correlation parameters, in order to look how the CT estimation is affected by the strength of the correlation between the components.

Table 2 shows $RE(\theta_i)$, $i = 1, 2, 3, 4$ and $GRE(\theta, 4)$ for $\rho_{12} = 0.1, 0.5, 0.9$ and $r_{12} = 0.1, 0.5, 1$. Some comments are in order. First, note that we consider only positive values of the taper parameter r_{12}. In fact it can be easily shown that, in the bivariate case, the CT estimating equations depend on r_{12}^2. This means that fixing positive or negative values of $r_{12} \in [-1, 1]$ leads to the same estimates. Second, the choice of r_{12} affects the statistical efficiency. In particular, the scale parameter is strongly affected by the choice of r_{12} especially when the correlation between the component is stronger ($\rho_{12} = 0.9$). As expected, the best efficiency is obtained when $r_{12} = 1$. The reason can be well explained in Fig. 1 where the tapered cross-covariance is shown for the scenario A when $\rho_{12} = 0.9$ and $r_{12} = 1, 0.5, 0.1$. It is apparent that when reducing r_{12} from 1 to 0.1 the resulting tapered cross-covariance function changes dramatically with respect to the non tapered cross-covariance function. For this reason, hereafter we consider $r_{12} = 1$ when using a separable taper.

Scenario B. Here, we use the model in Eq. (14) with constraints on the scale parameters. Specifically, we set $\sigma_1^2 = \sigma_2^2 = 1$, $\psi_{ii} = pr_{ii}/3$, $i = 1, 2$, $\psi_{12} = (\psi_{11} + \psi_{22})/2$ 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. Here, $\theta = (\rho_{12}, \psi_{11}, \psi_{22}, \sigma_1^2, \sigma_2^2)^T$. For this scenario, we consider two possible tapers:

- a separable bivariate taper as in Eq. (16), with a common compact support $b_{ij} = 0.0984$;
- a nonseparable bivariate taper as in Eq. (16), where we fixed $b_{ii} = 3 pr_{ii}/5$ $i = 1, 2$, $b_{12} = \min(b_{11}, b_{22})$ and $r_{12} = 0.65$.

We fix the common compact support in the first taper in such a way that the percentage of nonzero values in the tapered matrices is the same for both tapers (3.1 % approximatively). Since the tapered matrices have the same level of sparseness, we can compare the relative statistical efficiency of CT using both tapers.

Table 3 shows $RE(\theta_i)$, $i = 1, 2, 3, 4, 5$ and $GRE(\theta, 5)$ for both tapers when $\rho_{12} = 0.1$ and $\rho_{12} = 0.5$, while Fig. 2 depicts the boxplots for each parameter in the
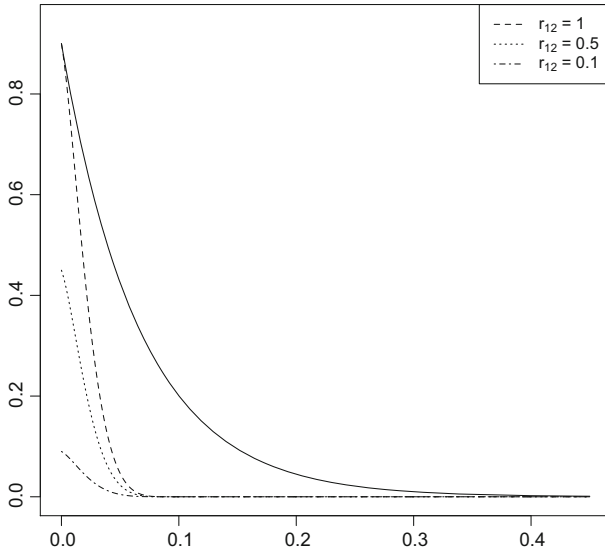
Fig. 1 Cross-covariance function with $\rho_{12} = 0.9$ (*solid line*) and tapered cross-covariance function when $r_{12} = 0.1, 0.5, 1$ for scenario A

Table 3 $RE(\theta_i)$ and $GRE(\theta, 5)$ for the parameters of the bivariate exponential model with contrasts using a separable and a nonseparable taper when $\rho_{12} = 0.1, 0.5, 0.9$ (scenario B)

	$\rho_{12} = 0.1$		$\rho_{12} = 0.5$		$\rho_{12} = 0.9$	
	Sep	Nosep	Sep	Nosep	Sep	Nosep
σ_{11}^2	0.890	0.900	0.858	0.866	0.849	0.840
σ_{22}^2	0.940	0.932	0.869	0.841	0.811	0.767
ρ_{12}	0.431	0.434	0.458	0.426	0.515	0.389
ψ_{11}	0.445	0.595	0.425	0.585	0.421	0.557
ψ_{22}	0.556	0.474	0.535	0.417	0.441	0.272
$GRE(\theta, 5)$	0.508	0.539	0.501	0.527	0.494	0.469

ML case and in the tapering case, using a separable and nonseparable taper, when the correlation is $\rho_{12} = 0.1$. Note that the scale parameters, as expected, are affected by the choice of the taper. In fact, the relative efficiency of ψ_{11} is much better in the case of the nonseparable taper, as expected, since the compact support of the first component in the nonseparable taper is 0.12, while it is 0.0984 in the separable case. On the other hand, the relative efficiency of ψ_{22} is slightly better in the separable case, since the compact support of the second component in the nonseparable taper is 0.09. Despite these discrepancies in the scale parameters estimation, the overall measure highlights basically no differences between the two types of tapers.

Scenario C. Here we use the model in Eq. (14), with the setting $\sigma_1^2 = \sigma_2^2 = 1$, $\psi_{ij} = pr_{ij}/3$, $i, j = 1, 2$ and $\rho_{12} = 0.5$, where $pr_{11} = 0.2$, $pr_{12} = 0.15$ and $pr_{22} = 0.175$
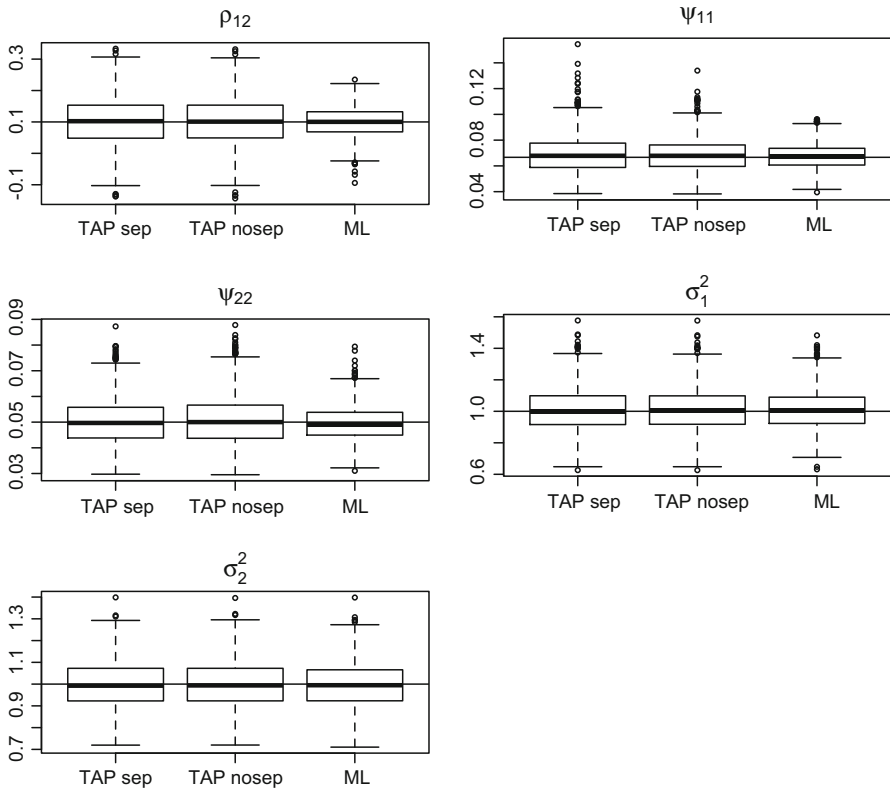
Fig. 2 Boxplots of the bivariate exponential model with constraints estimates using ML, and CT with a separable and a nonseparable taper when the correlation between the components is 0.1

are the practical ranges. In this case, $\boldsymbol{\theta} = (\rho_{12}, \psi_{11}, \psi_{12}, \psi_{22}, \sigma_1^2, \sigma_2^2)^T$. We consider the following tapers:

- a separable bivariate taper as in Eq. (16), with a common compact support $b_{ij} = 0.1133$;
- a nonseparable bivariate taper as in Eq. (16), with $b_{ii} = 3\, pr_{ii}/5$, $i = 1, 2$, and $b_{12} = \max(b_{11}, b_{22})$ and $r_{12} = 0.8$.

As before, we fix the common compact support b in the separable taper in such a way that the percentage of nonzero values in the tapered matrices is the same for the separable and non separable case (3.97 % approximatively).

Table 4 (left panel) shows $RE(\theta_i)$ $i = 1, 2, 3, 4, 5, 6$ and $GRE(\boldsymbol{\theta}, 6)$ for both tapers, when $\rho_{12} = 0.5$. As in the scenario B the choice of the taper affects the efficiency of the parameters. Specifically, the nonseparable taper outperforms the separable taper in the estimation of ψ_{11} and ψ_{12}, since the compact support of the first component and cross compact support are $b_{11} = b_{12} = 0.12$ while the common compact support in the separable taper is $b_{ij} = 0.1133$. On the other hand, the separable taper outperform the non separable taper in the estimation of ψ_{22} since $b_{22} = 0.105$.

Table 4 *Left panel* $RE(\theta_i)$ and $GRE(\theta, 6)$ for the parameters of the bivariate full exponential model using a separable and a nonseparable taper for the case $\rho_{12} = 0.5$ (scenario C)

	Sep	Nosep		Sep	Nosep
σ_{11}^2	0.988	0.991	α_{11}	0.733	0.846
σ_{22}^2	0.972	0.960	α_{22}	0.649	0.770
ρ_{12}	0.964	0.958	α_{12}	0.718	0.824
ψ_{11}	0.668	0.718	α_{21}	0.772	0.862
ψ_{12}	0.721	0.762	ψ_1	0.496	0.548
ψ_{22}	0.667	0.499	ψ_2	0.689	0.528
$GRE(\theta, 6)$	0.628	0.620		0.643	0.653

Right panel $RE(\theta_i)$ and $GRE(\theta, 6)$ for the parameters of the LMC model using the same separable and nonseparable tapers for the case when the correlation is 0.5 (scenario E)

Table 5 $RE(\theta_i)$ and $GRE(\theta, 5)$ for the parameters of the LMC model with constraints using a separable and a nonseparable taper when the correlation among the components is 0.1, 0.5, 0.9 (scenario D)

	$\rho_{12} = 0.1$		$\rho_{12} = 0.5$		$\rho_{12} = 0.9$	
	Sep	Nosep	Sep	Nosep	Sep	Nosep
α_{11}	0.899	0.909	0.876	0.882	0.791	0.754
α_{22}	0.949	0.942	0.920	0.900	0.588	0.515
α_{12}	0.417	0.416	0.468	0.457	0.828	0.769
ψ_1	0.451	0.594	0.433	0.550	0.431	0.407
ψ_2	0.562	0.476	0.548	0.423	0.535	0.307
$GRE(\theta, 5)$	0.510	0.536	0.503	0.517	0.502	0.456

Nevertheless, as in the previous scenario there are no substantial differences in the global measure of efficiency.

Scenario D. This scenario is based on the model in Eq. (15), with $\psi_1 = 0.2/3$, $\psi_2 = 0.15/3$, $\alpha_{ii} = \sqrt{1-x}$ $i = 1, 2$, and $\alpha_{12} = \alpha_{21} = \sqrt{x}$, with $x = 0.003, 0.067, 0.29$. Using this setting, the bivariate random field has unit variances, 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 and 0.9 when $x = 0.003, 0.067, 0.29$ respectively In this case, $\theta = (\alpha_{11}, \alpha_{12}, \alpha_{22}, \psi_1, \psi_2)^T$. We consider the same two tapers used for scenario B.

Table 5 show $RE(\theta_i)$, $i = 1, 2, 3, 4, 5$ and $GRE(\theta, 5)$ for both tapers. The scale parameters estimation, as expected, is more sensible to the type of taper. In particular the nonseparable taper outperforms in the estimation of ψ_1 and globally it slightly outperforms than the separable taper.

Scenario E. For this final scenario, we use the same model and the same setting of the scenario D, but without constraints on the parameters. In this case, $\theta = (\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}, \psi_1, \psi_2)^T$. As tapers, we use the same as in Scenario C. Table 4 (right panel) show $RE(\theta_i)$, $i = 1, 2, 3, 4, 5, 6$ and $GRE(\theta, 6)$ for both tapers

when the correlation is equal to 0.5 (case $x = 0.067$). The comments of the case D apply for this case.

In order to answer to the three questions at the begin of the section, we can resume our findings as follows:

– overall the CT relative efficiency is relatively good (between approximatively 0.4 and 0.6 of global relative efficiency with a percentage of non zero values between 3 and 4 %). Increasing the compact support(s) allows to improve the efficiency of the CT method as confirmed by simulation results not reported here.
– the sign of r_{12} in (16) does not affect the CT estimation in the bivariate case. Moreover for the separable taper, r_{12} should be fixed equal to 1, while in the nonseparable case it should be fixed equal to the upper bound of its range of validity, in order to improve the statistical efficiency.
– nonseparable tapers should be used in order to improve the efficiency of CT estimation for specific parameters. In particular, they can be useful if there is some preference in estimating the scale parameter of a specific component. Nevertheless, looking at the global efficiency, there are no substantial differences for CT using a separable and a nonseparable taper.

As further remark, the statistical efficiency of CT estimation depends on the strength of the marginal correlation between the components. Specifically, when decreasing the correlation, the efficiency improves. Finally, the choice of the marginal compact supports in the non separable taper should be related to the associated empirical practical ranges as suggested in Kaufman et al. (2008), while the "cross compact" support can be fixed as a function of the marginal ones. Possible choices are the minimum, the maximum or the mean of the marginal compact supports. Overall the choice of the compact supports should be driven by computational reasons that is the associated tapered matrix must be sufficiently sparse. In our examples the percentage of non zero values lies between 3 and 4 %.

We also consider an example in the trivariate case. Specifically we simulated 1000 realizations from a trivariate zero mean GRF, observed on 100 location sites uniformly distributed in $[0, 1]^2$, with covariance model obtained through Eqs. (4) and (5) with $p = 3$ and fixing $v_{ij} = 0.5 \, i, j = 1, 2, 3$ that is a trivariate exponential model. When $p \geq 3$, estimation can be troublesome because of the large number of parameters. A practical solution to this problem can be represented by decreasing the number of the parameters in some way such for instance some parsimonious versions of the covariance model. Here we consider a simple setting. We fix $\psi_{ij} = 0.1/3$, $\sigma_i^2 = 1$ for $i, j = 1, 2, 3$ and $\rho_{12} = -0.3$, $\rho_{13} = 0.3$, $\rho_{23} = 0.5$. We then estimate using a separable taper with common compact support equal to 0.13 (5.9 % of non zero values in the tapered matrix) and ML the three marginal scale parameters (considering the contrasts $\psi_{ij} = \psi_{ii} \, i, j = 1, 2, 3$), the three colocated correlation parameters and the three variances. Table 6 show the relative efficiency for the nine parameters.

We now compare the computational performance of the bivariate CT estimate with the ML one by using an increasing sequence of $n_w = 400 \cdot 2^w$ location sites, $w = 0, 1, 2, 3, 4$. The sites are uniformly distributed on the square $W_w = [0, 2^w] \times [0, 2^w]$. In Table 7 we report the mean of 5 evaluations of (10) and (11) in terms of elapsed time using the R function proc.time. Tapering is computed using a separable

Table 6 $RE(\boldsymbol{\theta}_i)$ and $GRE(\boldsymbol{\theta}, 9)$ for the parameters of a trivariate exponential model with constraints on the scale parameters

	σ_1^2	σ_2^2	σ_3^2	ψ_1	ψ_2	ψ_3	ρ_{12}	ρ_{13}	ρ_{23}	$GRE(\boldsymbol{\theta}, 9)$
Sep	0.903	0.873	0.790	0.268	0.291	0.300	0.899	0.911	0.877	0.616

Table 7 Time in seconds for evaluating l_n, $l_{T,n}$, in the bivariate case using a separable taper with $b = 0.2$ under increasing domain setup

w	n	l_n	$l_{T,n}$	%
0	800	0.26	0.26	11.500
1	1600	1.44	0.60	3.070
2	3200	9.10	1.36	0.800
3	6400	65.70	3.04	0.022
4	12800	579.20	16.90	0.006

The column (%) indicates the associated percentages of non zero values in the tapered covariance matrix

taper with common compact support $b = 0.2$ and $\rho_{12} = 1$. For evaluating (11) we use sparse matrix functions 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 only to be done once. In subsequent factorizations, one can pass in the structure and have spam only compute the numerical part. This can save a lot of time when the tapered likelihood function is evaluated repeatedly. For instance with 12,800 observations, the CT is approximatively 35 times faster than likelihood if the covariance matrix is very sparse.

5 Concluding remarks

For spatial Gaussian processes there are two types of covariance estimation methods based on CT (Kaufman et al. 2008). In this paper we focus on the one inducing unbiased estimating equations and we extend it to the multivariate case. The other version, obtained tapering the covariance matrix only, has been recently addressed in Furrer et al. (2015). In the same paper the authors use multivariate CT as prediction tool.

As outlined in Furrer et al. (2015) there is a clear trade off between statistical efficiency and computational complexity between the first and second version of the CT. Our proposal is much more efficient from statistical efficiency point of view but the price to pay is a severe loss of computational efficiency. On the other hand, the method proposed in Furrer et al. (2015) can lead to a substantial bias in the estimation, depending on the choice of the compact support in the multitaper. In our experience the second version of the multivariate CT can be an effective tool of estimation with a good balance between statistical efficiency and computational complexity, when

working with *large* data sets, where how *large* depends on the availability of memory and computing power. For instance in a computer with a 2.4 GHz processor and 8 GB of memory, CT estimation of a bivariate dataset with 5000 locations sites, i.e. 10,000 observations, starts to be troublesome. For very large data set, the approach proposed in Furrer et al. (2015) or extension to the multivariate case of methods proposed in the univariate case such as composite likelihood methods should be considered.

Tapering in a multivariate setting can be done with a simple separable taper or with more flexible tapers which allows to use different compact supports. In this paper, through a simulation study, we compare the performance of the CT using both tapers when estimating a multivariate Matern model and a LMC. Our numerical examples show that the colocated parameters of the multitaper must be chosen in the appropriate way in order to improve the efficiency. Moreover, the use of a non separable taper is recommended if there is an interest to improve the quality of the estimation of the parameters associated to a specific marginal component of the multivariate random fields. Nevertheless, globally there is no particular statistical gains when using a non separable tapers with respect to a separable one.

References

Apanasovich T, Genton M, Sun Y (2012) A valid Matérn class of cross-covariance functions for multivariate random fields with any number of components. J Am Stat Assoc 97:15–30

Arima S, Cretarola L, Lasinio GJ, Pollice A (2012) Bayesian univariate space-time hierarchical model for mapping pollutant concentrations in the municipal area of taranto. Stat Methods Appl 21:75–91

Askey R (1973) Radial characteristic functions. Technical report, Research Center, University of Wisconsin

Bevilacqua M, Gaetan C (2015) Comparing composite likelihood methods based on pairs for spatial gaussian random fields. Stat Comput 25:877–892

Bevilacqua M, Gaetan C, Mateu J, Porcu E (2012) Estimating space and space-time covariance functions for large data sets: a weighted composite likelihood approach. J Am Stat Assoc 107:268–280

Bevilacqua M, Hering A, Porcu E (2015) On the flexibility of multivariate covariance models: comment on the paper by Genton and Kleiber. Stat Sci 30:167–169

Daley D, Porcu E, Bevilacqua M (2015) Classes of compactly supported covariance functions for multi-variate random fields. Stoch Environ Res Risk Assess 29:1249–1263

Du J, Zhang H, Mandrekar VS (2009) Fixed-domain asymptotic properties of tapered maximum likelihood estimators. Ann Stat 37:3330–3361

Eidsvik J, Shaby BA, Reich BJ, Wheeler M, Niemi J (2014) Estimation and prediction in spatial models with block composite likelihoods. J Comput Graph Stat 23:295–315

Fontanella L, Ippoliti L (2003) Dynamic models for space-time prediction via karhunen-loeve expansion. Stat Methods Appl 12:61–78

Furrer R, Sain SR (2010) spam: a sparse matrix R package with emphasis on MCMC methods for Gaussian Markov random fields. J Stat Softw 36:1–25

Furrer R, Genton MG, Nychka D (2013) Covariance tapering for interpolation of large spatial datasets. J Comput Graph Stat 15:502–523

Furrer R, Bachoc F, Du J (2015) Asymptotic properties of multivariate tapering for estimation and prediction. ArXiv e-prints arXiv:1506.01833

Genton M, Kleiber W (2015) Cross-covariance functions for multivariate geostatistics. Stat Sci 30:147–163

Gneiting T (2002) Compactly supported correlation functions. J Multivar Anal 83:493–508

Gneiting T, Kleiber W, Schlather M (2010) Matérn cross-covariance functions for multivariate random fields. J Am Stat Assoc 105:1167–1177

Horn RA, Johnson CR (1991) Top matrix anal. Cambridge University Press, Cambridge

Kaufman CG, Schervish MJ, Nychka DW (2008) Covariance tapering for likelihood-based estimation in large spatial data sets. J Am Stat Assoc 103:1545–1555

Matheron G (1962) Traité de géostatistique appliquée, Tome 1. Mémoires du BRGM, n. 14, Technip, Paris

Padoan S, Bevilacqua M (2015) Analysis of random fields using CompRandFld. J Stat Softw 63:1–27

Porcu E, Daley D, Buhmann M, Bevilacqua M (2013) Radial basis functions with compact support for multivariate geostatistics. Stoch Environ Res Risk Assess 27:909–922

Shaby B, Ruppert D (2012) Tapered covariance: Bayesian estimation and asymptotics. J Comput Graph Stat 21:433–452

Stein M, Chi Z, Welty L (2004) Approximating likelihoods for large spatial data sets. J R Stat Soc B 66:275–296

Stein M, Chen J, Anitescu M (2012) Difference filter preconditioning for large covariance matrices. SIAM J Matrix Anal Appl 33:52–72

Stein M, Chen J, Anitescu M (2013) Stochastic approximation of score functions for gaussian processes. Ann Appl Stat 7:1162–1191

Vecchia A (1988) Estimation and model identification for continuous spatial processes. J R Stat Soc B 50:297–312

Vetter P, Schmid W, Schwarze R (2015) Spatio-temporal statistical analysis of the carbon budget of the terrestrial ecosystem. Stat Methods Appl

Wackernagel H (2003) Multivariate geostatistics: an introduction with applications, 3rd edn. Springer, New York

Zastavnyi V, Trigub R (2002) Positive definite splines of special form. Sbornik Math 193:1771–1800