Electronic Journal of Statistics
Vol. 17 (2023) 3287–3315
ISSN: 1935-7524
https://doi.org/10.1214/23-EJS2178

Assessing the estimation of nearly singular covariance matrices for modeling spatial variables

Javier Pérez[1], **Jonathan Acosta**[*2], **and Ronny Vallejos**[1]

[1] *Departamento de Matemática, Universidad Técnica Federico Santa María, Valparaíso, Chile*
e-mail: javier.perezo@usm.cl; ronny.vallejos@usm.cl

[2] *Departamento de Estadística, Pontificia Universidad Católica de Chile, Santiago, Chile*
e-mail: jonathan.acosta@mat.uc.cl

Abstract: Spatial analysis commonly relies on the estimation of a covariance matrix associated with a random field. This estimation strongly impacts the prediction where the process has not been observed, which in turn influences the construction of more sophisticated models. If some of the distances between all the possible pairs of observations in the plane are small, then we may have an ill-conditioned problem that results in a nearly singular covariance matrix. In this paper, we suggest a covariance matrix estimation method that works well even when there are very close pairs of locations on the plane. Our method is an extension to a spatial case of a method that is based on the estimation of eigenvalues of the unitary matrix decomposition of the covariance matrix. Several numerical examples are conducted to provide evidence of good performance in estimating the range parameter of the correlation structure of a spatial regression process. In addition, an application to macroalgae estimation in a restricted area of the Pacific Ocean is developed to determine a suitable estimation of the effective sample size associated with the transect sampling scheme.

MSC2020 subject classifications: Primary 62H11, 62J05; secondary 62P12.
Keywords and phrases: Spatial process, nearly singular covariance matrix, Vandermonde matrix, estimation algorithm.

Received October 2022.

1. Introduction and motivation

The analysis of spatial processes with singular or near singular covariance matrices has been a subject of study for decades (Penrose, 1955). Several approaches have been suggested to address nearly singular covariance matrices and their inverses. These matrices are characterized by a large condition number, which makes the estimation process unstable. In kriging analysis under the Gaussian assumption, Diamond and Armstrong (1984) found that the robustness of the kriging predictor depends on the condition number of the covariance matrix.

*Corresponding author.

Ababou et al. (1994) studied the condition number associated with several covariance functions. Salagame and Barton (1997) and Davis and Morris (1997) performed further analysis on the kriging stability and the condition number. Subsequently, there have been a number of papers undertaking a variety of approaches to deal with ill-conditioned covariance matrices. On the one hand, it is well known that the nugget effect can play a significant role in the estimation of the inverse of a covariance matrix (Cressie, 1993; Peng and Wu, 2014). On the other hand, methods such as covariance tapering (Furrer et al., 2006), fixed rank kriging (Cressie and Johannesson, 2008), penalized likelihood functions (Li and Sudjianto, 2005), matrix decomposition Ayyıldız et al. (2012), and compactly support covariance functions (Bevilacqua et al., 2019, 2022a) have been proposed to overcome the ill-conditioned problem.

The causes of instability of kriging have been classified and discussed by Peng and Wu (2014). The authors mention three possible causes of singularity of the correlation matrix: (i) sample size, (ii) dimension of the input vectors, and (iii) parameter values. In addition, we mention two more reasons that cause instability when doing spatial analysis: (iv) type of parametric correlation function and (v) spatial design. This classification makes it difficult to approach the problem from a unique perspective. For instance, Smith and Nicolik (2013) proposed an automatic method for handling nearly singular covariance structures by using a Cholesky decomposition. More recently, Zimmermann (2015b) in the context of (iv), addressed the ill-conditioning problem for the Gaussian correlation model from a theoretical perspective. In particular, it was proven that the condition number goes to infinity when the hyperparameter of the correlation approaches zero, thus distinguishing the Gaussian case from the remaining exponential correlation models in terms of how fast the condition number increases. In addition, Zimmermann (2015a) gives an improved growth rate estimate for the Gaussian condition number anomaly.

The main purpose of this paper is to study an estimation method of the inverse of a covariance matrix, also referred to as a precision matrix, associated with a spatial process when the covariance matrix is nearly singular due to the existence of pairs of points that are very close in space. Clearly, the closeness of two points in space is relative to the degree of correlation between the points, and can be characterized through the practical range of the process. We develop an estimation method for the parameters of a spatial process when the covariance matrix is nearly singular, due to the sampling scheme of the locations, or when there are repeated observations yielding a singular covariance matrix. Our suggested method is a generalization of the method proposed by Marzetta et al. (2011) in an image processing context, in which the eigenvalues close to zero of a nearly singular matrix are estimated.

Their method provides an alternative to numerical regularization (Cressie, 1993) used to estimate the covariance matrix and its inverse for a Gaussian regression process. Our method is developed in an iterative fashion, so that the eigenvalues, the mean of the process, and the covariance parameters are estimated in each iteration. We present the estimation algorithms and show how to perform spatial prediction (kriging). Eigenvalue estimation and parameter

estimation of a Gaussian process are explored numerically through Monte Carlo simulation. An application with a real dataset that has been collected using transect sampling is analyzed to estimate the effective sample size when a spatial linear regression model is assumed.

2. Preliminaries and background

In this section, we describe an approach that allows us to estimate a singular covariance matrix when having N independent identically distributed realizations of a random vector. Marzetta et al. (2011) considered a data matrix $\boldsymbol{X}$ of size $M \times N$ with $N < M$ that contains N independent and identically distributed realizations of a M-dimensional zero mean random vector with a covariance matrix $\boldsymbol{\Sigma}$. In this framework, $\boldsymbol{K}$ is the sample covariance matrix of $\boldsymbol{\Sigma}$, which is singular, because $\text{rank}(\boldsymbol{K}) = N < M$. Let $\Omega_{L,N} = \{\boldsymbol{\Phi} \in \mathbb{R}^{L\times M} : \boldsymbol{\Phi}\boldsymbol{\Phi}^* = \boldsymbol{I}_L\}$ be the set of all $L \times M$ one-sided unitary matrices, with manifold structure called the Stiefel manifold. Marzetta et al. (2011) endow the Stiefel manifold with the *Haar measure* and define the operators $\text{cov}_L(\boldsymbol{K}) = \mathbb{E}_{\boldsymbol{\Phi}}(\boldsymbol{\Phi}^*(\boldsymbol{\Phi}\boldsymbol{K}\boldsymbol{\Phi}^*)\boldsymbol{\Phi})$ and $\text{InvCov}_L(\boldsymbol{K}) = \mathbb{E}_{\boldsymbol{\Phi}}(\boldsymbol{\Phi}^*(\boldsymbol{\Phi}\boldsymbol{K}\boldsymbol{\Phi}^*)^{-1}\boldsymbol{\Phi})$ as estimators of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}^{-1}$, respectively. In addition, the authors found that $\text{cov}_L(\boldsymbol{K})$ coincides with nugget regularization (see for instance, Ranjan et al., 2011). However, $\text{InvCov}_L(\boldsymbol{K})$ does not necessarily coincide with $\text{cov}_L(\boldsymbol{K})^{-1}$. For the InvCov estimator, consider the spectral decomposition of $\boldsymbol{K}$ ($\boldsymbol{K} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^*$) with $\boldsymbol{D} = \text{diag}(d_1, \ldots, d_N, 0, \ldots, 0)$ containing the eigenvalues of $\boldsymbol{K}$ and $\boldsymbol{U}$ being the corresponding eigenvector matrix. Marzetta et al. (2011) proved that $\text{InvCov}_L(\boldsymbol{K}) = \boldsymbol{U}\ \text{InvCov}_L(\boldsymbol{D})\ \boldsymbol{U}^*$, where

$$\text{InvCov}_L(\boldsymbol{D}) = \text{diag}(\lambda_1, \ldots, \lambda_N, \mu, \ldots, \mu),$$

and provided formulas to compute the values of λ_i and μ, and investigated their asymptotic behavior (see Appendix A). Notice that λ_i is an approximation of d_i^{-1}.

The developments presented in this section were designed for a stochastic process with N replicates. In such scenarios, the aim of Marzetta et al. (2011) is to address the problem of an ill-conditioned covariance matrix by introducing dimension reduction techniques, which effectively reduce the dimensionality of the problem to a lower-dimensional space. The focus of this paper is on a spatial process defined in $\mathbb{R}^2$, where $N = 1$. Even in this case, the problem of an ill-conditioned covariance matrix persists due to the presence of spatial autocorrelation, which leads to redundant information in the observed data Acosta and Vallejos (2018). The next section will cover the extension of Marzetta et al. (2011)'s approach to the spatial context.

3. Estimation of the covariance function in a spatial process

Consider a second-order stationary spatial process $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^2\}$, with the locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_M \in \mathbb{R}^2$ and the parameter vector $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^p$. Assume that

$\text{var}[Z(\boldsymbol{s})] = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ is a nearly singular matrix of order $M \times M$. For the sake of simplicity we denote $\boldsymbol{\Sigma} := \boldsymbol{\Sigma}(\boldsymbol{\theta})$. Given $Z(\boldsymbol{s}_i)$, $i = 1, \ldots, M$, the parametric estimation of the covariance matrix is denoted by $\widehat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})$.

Let $d_1, d_2, \ldots, d_N, d_{N+1}, \ldots, d_M$ be the eigenvalues of $\widehat{\boldsymbol{\Sigma}}$, then the spectral decomposition of $\widehat{\boldsymbol{\Sigma}}$ is $\widehat{\boldsymbol{\Sigma}} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^\top$, where $\boldsymbol{D}$ is a diagonal matrix with entries equal to the eigenvalues of $\widehat{\boldsymbol{\Sigma}}$ and $\boldsymbol{U}$ is an orthogonal matrix with the respective eigenvectors. Assume that $d_{N+1}, \ldots, d_M \approx 0$ and define $\boldsymbol{D}_N = \text{diag}(d_1, d_2, \ldots, d_N)$, where N corresponds to the number of nonzero eigenvalues, then $\boldsymbol{D} \approx \text{diag}(\boldsymbol{D}_N, \boldsymbol{0}_{M-N})$. Consequently, the estimation method developed by Marzetta et al. (2011) can be applied in this framework to obtain estimations of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}^{-1}$ as

$$\text{cov}_L(\widehat{\boldsymbol{\Sigma}}) = \frac{L}{(M^2-1)M}\big[(ML-1)\widehat{\boldsymbol{\Sigma}} + (M-L)\,\text{tr}(\widehat{\boldsymbol{\Sigma}})\boldsymbol{I}_M\big], \tag{3.1}$$

$$\text{InvCov}_L(\boldsymbol{D}) = \text{diag}(\lambda_1, \ldots, \lambda_N, \mu, \ldots, \mu), \tag{3.2}$$

$$\text{InvCov}_L(\widehat{\boldsymbol{\Sigma}}) = \boldsymbol{U}\ \text{InvCov}_L(\boldsymbol{D})\boldsymbol{U}^{-1}, \tag{3.3}$$

with $L < N$,

$$\mu = \frac{\det(\boldsymbol{G})}{\det(\boldsymbol{\Delta}(\boldsymbol{D}_N))}, \text{ and}$$

$$\lambda_i = \frac{\partial}{\partial d_i}\left(\sum_{k=0}^{L-1}\frac{(N-(k+1))!}{(L-(k+1))!}\frac{\det(\boldsymbol{G}_k)}{\det(\boldsymbol{\Delta}(\boldsymbol{D}_N))}\right),$$

where $\boldsymbol{G}$, $\boldsymbol{G}_k$, and $\boldsymbol{\Delta}(\cdot)$ are defined in Appendix A. Given the structures of $\boldsymbol{G}_k$ and $\boldsymbol{\Delta}(\boldsymbol{D}_N)$, it is possible to provide an explicit expression for λ_i. The derivatives of matrices $\boldsymbol{G}_k$ and $\boldsymbol{\Delta}(\boldsymbol{D}_N)$ with respect to d_i are straightforward to obtain. Then, the expression depends only on a power of d_i, one element from the inverse of $\boldsymbol{G}_k$, and on another element from the inverse of $\boldsymbol{\Delta}(\boldsymbol{D}_N)$. The computation of the inverse of matrices $\boldsymbol{G}_k$ and $\boldsymbol{\Delta}(\boldsymbol{D}_N)$ is done by using a QR decomposition. In a similar fashion, the determinants of these matrices are obtained. The following result gives an expression for λ_i.

Theorem 3.1. *In the framework of Equations* (3.1)–(3.3),

$$\lambda_i = \sum_{k=0}^{L-1}\frac{(N-(k+1))!}{(L-(k+1))!}\frac{|\boldsymbol{G}_k|}{|\boldsymbol{\Delta}(\boldsymbol{D}_N)|}\left(\sum_{\substack{l=1\\ l\neq k+1}}^{N-1}\big((g_{il}-b_{il})(N-l)d_i^{N-(l+1)}\big) + g_{i,k+1}G_i - b_{i,k+1}\big(N-(k+1)\big)d_i^{N-(k+2)}\right), \tag{3.4}$$

where $\boldsymbol{G}_k^{-1} = [g_{ij}]_{i,j=1}^N$, $\boldsymbol{\Delta}(\boldsymbol{D}_N)^{-1} = [b_{ij}]_{i,j=1}^N$ *and* G_i *is the* $(N-L-1)$*-th integral of* $x^{L-(k+1)}\log(x)$, *evaluated at* $x = d_i$.

Proof. See Appendix B. □

In practice, eigenvalues equal to zero are not possible in the nearly singular case, hence the final estimates rely on the selection of N. A sensitivity study about N was carried and is described at the end of section 6.1. Alternative criteria for the selection of N can be constructed considering eigenvalues greater than a fixed tolerance; or considering a fixed value of the total variance.

4. Estimation algorithm

In this section, we present an iterative estimation algorithm for the parameters of $\boldsymbol{\Sigma}$ for a spatial regression model with constant mean. The InvCov method is used because in this case, it is possible to obtain a closed form for the estimation based on the estimation of the eigenvalues of $\boldsymbol{\Sigma}$. The usual methods of estimation, such as maximum likelihood (ML), cannot be applied here because $\boldsymbol{\Sigma}$ is nearly singular. Hence, it is not practically possible to compute $\boldsymbol{\Sigma}^{-1}$ or $\ln|\boldsymbol{\Sigma}|$. Let $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^2\}$ be a spatial process observed in the locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_M \in \mathbb{R}^2$, such that the vector of observations is $\boldsymbol{Z} = (Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_M))^\top$. In addition, assume that

$$\boldsymbol{Z} \sim \mathcal{N}(\mathbf{1} \cdot \beta, \boldsymbol{\Sigma}), \tag{4.1}$$

where β is a real parameter, $\mathbf{1}$ is a vector of ones of size M, $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{R}(\phi)$, where $\boldsymbol{R}(\phi) = \boldsymbol{R}(\|\boldsymbol{h}\|, \phi)$ is a parametric correlation matrix, ϕ is a correlation scale component, $\|\boldsymbol{h}\| = \|\boldsymbol{s}_i - \boldsymbol{s}_j\|$ and $\boldsymbol{\theta} = (\sigma^2, \phi)^\top$. We denote the derivative of $\boldsymbol{R}(\phi)$ with respect to ϕ as $\boldsymbol{R}_\phi(\phi)$.

The log-likelihood for model (4.1) is

$$\ell(\beta, \sigma^2, \phi) = -\frac{M}{2}\ln(2\pi) - \frac{M}{2}\ln(\sigma^2) - \frac{1}{2}\ln|\boldsymbol{R}(\phi)| - \frac{1}{2\sigma^2}(\boldsymbol{Z} - \mathbf{1}\cdot\beta)^\top \boldsymbol{R}^{-1}(\phi)(\boldsymbol{Z} - \mathbf{1}\cdot\beta), \tag{4.2}$$

from which the ML estimators of β, σ^2 and ϕ are obtained by solving the system $\nabla\ell(\beta, \sigma^2, \phi) = \mathbf{0}$. In this system, β and σ^2 have closed forms in terms of ϕ, which can be substituted into the log-likelihood (4.2) to obtain a so-called concentrated log-likelihood. Consequently, the ML estimate of β can be computed as

$$\widehat{\beta} = \left(\mathbf{1}^\top \boldsymbol{R}^{-1}(\widehat{\phi})\mathbf{1}\right)^{-1} \mathbf{1}^\top \boldsymbol{R}^{-1}(\widehat{\phi})\boldsymbol{Z}. \tag{4.3}$$

Let $\boldsymbol{e} = \boldsymbol{Z} - \mathbf{1}\cdot\widehat{\beta}$ be the error estimates; then, the ML estimate for σ^2 is

$$\widehat{\sigma}^2 = \frac{1}{M}\boldsymbol{e}^\top \boldsymbol{R}^{-1}(\widehat{\phi})\boldsymbol{e}. \tag{4.4}$$

Substituting the expressions from Equations (4.3) and (4.4) into the log-likelihood (Equation (4.2)) yields the concentrated log-likelihood

$$\ell_C(\phi) = \text{const.} - \frac{1}{2}\ln|\boldsymbol{R}(\phi)| - \frac{M}{2}\ln\left(\boldsymbol{e}^\top \boldsymbol{R}^{-1}(\phi)\boldsymbol{e}\right), \tag{4.5}$$

a nonlinear function in a single parameter ϕ, where const. $= -M\ln(2\pi)/2 - M/2 + M\ln(M)/2$. Let $d_1, d_2, \ldots, d_M$ be the eigenvalues of $\boldsymbol{R}(\phi)$, and let $\boldsymbol{v}_1, \boldsymbol{v}_2$,

$\ldots, \boldsymbol{v}_M$ be the corresponding eigenvectors of $\boldsymbol{R}(\phi)$. Note that $|\boldsymbol{R}(\phi)| = \prod_{i=1}^{M} d_i$, and the quadratic form $\boldsymbol{e}^\top \boldsymbol{R}^{-1}(\phi)\boldsymbol{e}$ may also be written in terms of the eigenvalues and eigenvectors as follows:

$$\boldsymbol{R}^{-1}(\phi) = \boldsymbol{P}\boldsymbol{D}^{-1}\boldsymbol{P}^\top \quad \Longrightarrow \quad \boldsymbol{e}^\top \boldsymbol{R}^{-1}(\phi)\boldsymbol{e} = \left(\boldsymbol{P}^\top \boldsymbol{e}\right)^\top \boldsymbol{D}^{-1}\left(\boldsymbol{P}^\top \boldsymbol{e}\right) = \sum_{i=1}^{M} \frac{q_i^2}{d_i},$$

where $\boldsymbol{D}$ and $\boldsymbol{P}$ are the matrix of eigenvalues and eigenvectors of $\boldsymbol{R}(\phi)$, respectively, and $q_i = \boldsymbol{v}_i^\top \boldsymbol{e}$, i.e., $\boldsymbol{q} = \boldsymbol{P}^\top \boldsymbol{e}$. Furthermore, Equation (4.5) can be rewritten as

$$\ell_C(\phi) = \text{const.} - \frac{1}{2}\sum_{i=1}^{M} \ln(d_i) - \frac{M}{2} \ln\left(\sum_{i=1}^{M} \frac{q_i^2}{d_i}\right). \tag{4.6}$$

Let $\lambda_1, \lambda_2, \ldots, \lambda_N, \mu, \ldots, \mu$ be the estimated eigenvalues of $\boldsymbol{R}^{-1}(\phi)$ obtained by using the InvCov method. That is, $d_i^{-1} \approx \lambda_i$ for $i = 1, \ldots, N$, and $d_i^{-1} \approx \mu$ for $i = N+1, \ldots, M$. To obtain an ML estimate for ϕ, we suggest approximating the likelihood by replacing the eigenvalues provided by the InvCov method with (4.6), which will yield the approximate log-likelihood:

$$\begin{aligned} \ell_{\text{InvCov}}(\phi) = \text{const.} &+ \frac{1}{2}\sum_{i=1}^{N} \ln(\lambda_i) \\ &+ \frac{(M-N)}{2} \ln(\mu) - \frac{M}{2} \ln\left(\sum_{i=1}^{N} q_i^2 \lambda_i + \mu \sum_{i=N+1}^{M} q_i^2\right). \end{aligned} \tag{4.7}$$

Note that (4.7) is a nonlinear function of ϕ, which can be maximized by numerical optimization. Additionally, note that the dependence of ϕ is implicit through μ, λ_i and q_i. Thus, an approximation of the ML estimate of ϕ is

$$\widehat{\phi}_{\text{ML-InvCov}} = \arg\max_{\phi} \ell_{\text{InvCov}}(\phi).$$

The motivation to use the estimated eigenvalues provided by the InvCov method is because some of the eigenvalues are approximately zero. It is worth mentioning that the approximate likelihood (4.7) results in a biased estimation equation. Nevertheless, the estimations provided by this method are computationally plausible, as will be shown in the numerical experiments presented in Section 6. One way to overcome the biased of an estimation equation is by considering a bias correction, similar to how it is achieved in robustness (Stefanski and Boos, 2002).

The estimations $\widehat{\beta}$ and $\widehat{\sigma}^2$ given in Equations (4.3) and (4.4) cannot be obtained due to their dependence on the inverse of $\boldsymbol{R}$. Thus, we approach the estimation of $\boldsymbol{R}^{-1}$ via the InvCov method to obtain

$$\boldsymbol{R}_{\text{InvCov}}^{-1} = \boldsymbol{R}^{-1}(\widehat{\phi}_{\text{ML-InvCov}}) = \boldsymbol{P}\boldsymbol{D}_{\text{InvCov}}^{-1}\boldsymbol{P}^\top, \tag{4.8}$$

where $\boldsymbol{D}_{\text{InvCov}}^{-1} = \text{diag}\{\lambda_1, \ldots, \lambda_N, \mu, \ldots, \mu\}$ (the diagonal matrix with approximated eigenvalues using the InvCov method with $\lambda_1, \ldots, \lambda_N, \mu, \ldots, \mu$ estimates from $\boldsymbol{R}(\widehat{\phi}_{\text{ML-InvCov}})$). As a consequence, β and σ^2 are estimated by

$$\widehat{\beta}_{\text{ML-InvCov}} = \left(\mathbf{1}^\top \boldsymbol{R}_{\text{InvCov}}^{-1}\mathbf{1}\right)^{-1}\mathbf{1}^\top \boldsymbol{R}_{\text{InvCov}}^{-1}\boldsymbol{Z}. \tag{4.9}$$

$$\widehat{\sigma}^2_{\text{ML-InvCov}} = \frac{1}{M}(\boldsymbol{Z} - \mathbf{1}\cdot\widehat{\beta}_{\text{ML-InvCov}})^\top \boldsymbol{R}_{\text{InvCov}}^{-1}(\boldsymbol{Z} - \mathbf{1}\cdot\widehat{\beta}_{\text{ML-InvCov}}). \tag{4.10}$$

Note that $\text{tr}(\boldsymbol{R}) = \sum_{i=1}^{M} d_i = M$; however, $\sum_{i=1}^{N}\lambda_i^{-1} + (M-N)\mu^{-1}$ does not necessarily add up to M. This clearly affects the estimator of σ^2 in Equation (4.10), but it does not affect the estimator of β in Equation (4.9) because it contains the inverse correlation matrix in both the numerator and denominator. A corrected version of $\widehat{\sigma}^2_{\text{ML-InvCov}}$ described in Equation (4.10) is obtained by scaling the eigenvalues yielded by the InvCov method; thus, the corrected estimator of σ^2 is

$$\widetilde{\sigma}^2_{\text{ML-InvCov}} = \left(\frac{1}{M}\sum_{i=1}^{N}\frac{1}{\lambda_i} + \frac{(M-N)}{M\mu}\right)\widehat{\sigma}^2_{\text{ML-InvCov}}. \tag{4.11}$$

This procedure is summarized in the following algorithm, which will be used in the Monte Carlo simulations developed in Section 6 and in the application developed in Section 7.

Algorithm 1: InvCov estimation.

input : $\boldsymbol{Z}$, N, *tolerance*

1 Maximize $\ell_{\text{InvCov}}(\phi)$ given in Equation (4.7), and obtain $\widehat{\phi}_{\text{ML-InvCov}}$;

2 Compute $\boldsymbol{R}_{\text{InvCov}}^{-1}$ through Equation (4.8);

3 Compute $\widehat{\beta}_{\text{ML-InvCov}}$ through Equation (4.9);

4 Compute $\widehat{\sigma}^2_{\text{ML-InvCov}}$ through Equation (4.11);

output: $\widehat{\phi}_{\text{ML-InvCov}}$, $\widehat{\beta}_{\text{ML-InvCov}}$, $\widehat{\sigma}^2_{\text{ML-InvCov}}$

Sometimes Algorithm 1 does not converge because $\ell_{\text{InvCov}}(\phi)$ in (4.7) is not bounded (see the first row of Figure 1). One way to overcome this drawback is to introduce a penalty to the objective function $\ell_{\text{InvCov}}(\phi)$ (in order to yield a bounded log-likelihood) through

$$G_\eta(\phi) = \ell_{\text{InvCov}}(\phi) - \eta H(\phi), \tag{4.12}$$

where $H(\phi)$ is a concave positive and differentiable function. In such a case, an approximation of the ML estimate of ϕ is

$$\widehat{\phi}_{\text{ML-InvCov}} = \arg\max_{\phi} G_\eta(\phi).$$

From the first-order conditions $\partial G_\eta/\partial\phi = 0$, it follows that

$$\eta = \max\left\{\frac{\ell'_{\text{InvCov}}(\phi)}{H'(\phi)}, 0\right\}. \tag{4.13}$$

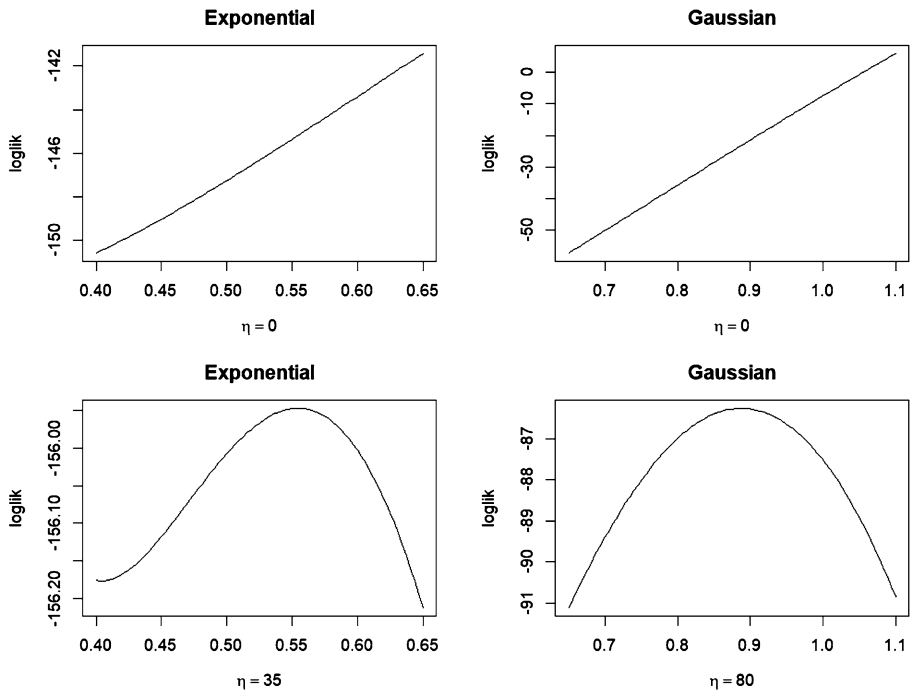


FIG 1. *Plot of* (4.7) *(first row) and* (4.12) *(second row). These plots correspond to the log-likelihood evaluated at a realization from a Gaussian random field with exponential and Gaussian correlation function, with parameters* $\beta = 0$, $\sigma^2 = 1$, *and the same practical range equal to 1.5 (*$\phi \approx 0.50$ *for exponential, and* $\phi \approx 0.87$ *for Gaussian). For the penalized loglik case* $\eta = 35$ *for the exponential covariance function and* $\eta = 80$ *for the Gaussian covariance function.*

A typical choice of H is $H(\phi) = \phi^2$ (Fu, 1998), which implies that $H'(\phi) = 2\phi$. Figure 1 shows the penalization effect described in Equation (4.12) when the penalization is a quadratic function. A bounded objective function is obtained for both covariance structures, exponential and Gaussian with a different value of η in each case.

Note that $H'(\phi) > 0$ for $\phi > 0$, but $\ell'_{\text{InvCov}}(\phi)$ is not necessarily positive. However, if $\ell'_{\text{InvCov}}(\phi) < 0$ the objective function is bounded (reaches a maximum), so the penalty should not apply.

In addition, an approximation of $\ell'_{\text{InvCov}}(\phi)$ is

$$\ell'_{\text{InvCov}}(\phi) = \frac{\ell_{\text{InvCov}}(\phi + a) - \ell_{\text{InvCov}}(\phi - a)}{2a},$$

where a is a tolerance level that could be fixed (for instance) as $a = 10^{-4}$. Algorithm 2 summarizes the procedure describe above:

Algorithm 2: InvCov penalized estimation.

input : $\boldsymbol{Z}$, N, *tolerance*, max.iter, $\widehat{\phi}_0$

Set j=0;

Compute η through Equation (4.13) evaluated at $\widehat{\phi}_j$;

Maximize $G_\eta(\phi)$ given by Equation (4.12), and obtain $\widehat{\phi}_{j+1}$;

Set j=j+1;

Repeat steps 2-4 until $j = max.iter$ or $\frac{|\widehat{\phi}_{j+1} - \widehat{\phi}_j|}{\widehat{\phi}_j} < tolerance$;

Set $\widehat{\phi}_{\text{ML-InvCov}} = \widehat{\phi}_j$;

Compute $\boldsymbol{R}^{-1}_{\text{InvCov}}$ through Equation (4.8);

Compute $\widehat{\beta}_{\text{ML-InvCov}}$ through Equation (4.9);

Compute $\widehat{\sigma}^2_{\text{ML-InvCov}}$ through Equation (4.11);

output: $\widehat{\phi}_{\text{ML-InvCov}}$, $\widehat{\beta}_{\text{ML-InvCov}}$, $\widehat{\sigma}^2_{\text{ML-InvCov}}$

5. Spatial prediction

Let $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}\}$ be a spatial process. Define $\boldsymbol{Z} = (Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_n))^\top$ for $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n \in \mathbb{R}$, denote $\boldsymbol{s}_0$ as a new location where the process has not been observed ($Z(\boldsymbol{s}_0)$), and denote the predictor of $Z(\boldsymbol{s})$ at $\boldsymbol{s}_0$ as $p(\boldsymbol{Z}, \boldsymbol{s}_0)$. We consider the ordinary kriging scenario (Cressie, 1993), with a spatial regression process of the form

$$Z(\boldsymbol{s}) = \beta + \delta(\boldsymbol{s}), \quad \boldsymbol{s} \in D, \beta \in \mathbb{R},$$

with unknown β. The InvCov predictor in such a case is

$$\widehat{p}_{\text{InvCov}}(\boldsymbol{Z}, \boldsymbol{s}_0) = \boldsymbol{w}^\top \boldsymbol{Z}, \quad \text{subject to } \sum_{i=1}^{n} w_i = 1,$$

where $\boldsymbol{w} = (w_1, \ldots, w_n)^\top$ and m (parameter associated with the restriction $\sum_{i=1}^n w_i = 1$) is obtained by minimizing the quadratic prediction error

$$\sigma_e^2 = \mathbb{E}\big[\big(Z(\boldsymbol{s}_0) - p(\boldsymbol{Z}, \boldsymbol{s}_0)\big)^2\big].$$

As a result,

$$\boldsymbol{w}^\top = \left(\boldsymbol{c} + \boldsymbol{1}\frac{(1 - \boldsymbol{1}^\top \text{InvCov}_L(\boldsymbol{\Sigma})\boldsymbol{c})}{\boldsymbol{1}^\top \text{InvCov}_L(\boldsymbol{\Sigma})\boldsymbol{1}}\right)^\top \text{InvCov}_L(\boldsymbol{\Sigma}), \tag{5.1}$$

$$m = \frac{1 - \boldsymbol{1}^\top \text{InvCov}_L(\boldsymbol{\Sigma})\boldsymbol{c}}{\boldsymbol{1}^\top \text{InvCov}_L(\boldsymbol{\Sigma})\boldsymbol{1}}, \tag{5.2}$$

where $\boldsymbol{c} = \text{cov}(\boldsymbol{Z}, Z(\boldsymbol{s}_0))$. In this case, $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{R}$ where $\boldsymbol{R}$ is a correlation matrix, then $\text{InvCov}_L(\boldsymbol{\Sigma}) = \sigma^{-2}\boldsymbol{R}^{-1}_{\text{InvCov}}$ and $\boldsymbol{c} = \sigma^2 \boldsymbol{r}$ ($\boldsymbol{r} = \text{corr}(\boldsymbol{Z}, Z(\boldsymbol{s}_0))$. Hence the kriging predictor can be written as

$$\widehat{p}_{\text{InvCov}}(\boldsymbol{Z}, \boldsymbol{s}_0) = \widehat{\beta}_{\text{ML-InvCov}} + \boldsymbol{r}^\top \boldsymbol{R}^{-1}_{\text{InvCov}}(\boldsymbol{Z} - \widehat{\beta}_{\text{ML-InvCov}}\boldsymbol{1}).$$

The kriging variance is given by $\sigma_k^2(\boldsymbol{s}_0) = C(\mathbf{0}) - \boldsymbol{w}^\top \boldsymbol{c} + m$; then, given $\boldsymbol{w}$, we have

$$\begin{aligned}\sigma_k^2(\boldsymbol{s}_0) &= C(\mathbf{0}) - \boldsymbol{c}^\top \mathrm{InvCov}_L(\boldsymbol{\Sigma})\boldsymbol{c} + \frac{(1 - \mathbf{1}^\top \mathrm{InvCov}_L(\boldsymbol{\Sigma})\boldsymbol{c})^2}{\mathbf{1}^\top \mathrm{InvCov}_L(\boldsymbol{\Sigma})\mathbf{1}},\\ &= \sigma^2\left(1 - \boldsymbol{r}^\top \boldsymbol{R}_{\mathrm{InvCov}}^{-1}\boldsymbol{r} + \frac{(1 - \mathbf{1}^\top \boldsymbol{R}_{\mathrm{InvCov}}^{-1}\boldsymbol{r})^2}{\mathbf{1}^\top \boldsymbol{R}_{\mathrm{InvCov}}^{-1}\mathbf{1}}\right),\end{aligned}$$

under the condition $C(\mathbf{0}) - \boldsymbol{c}^\top \mathrm{InvCov}_L(\boldsymbol{\Sigma})\boldsymbol{c} \geq 0$. The last equation was obtained as a particular case when $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{R}$. Then $C(\mathbf{0}) = \sigma^2$. Because the estimation of $\boldsymbol{\Sigma}^{-1}$ is carried out through $\mathrm{InvCov}_L(\boldsymbol{\Sigma})$, we propose an estimation based on the cov estimator for the augmented matrix $\tilde{\boldsymbol{\Sigma}}$ by including the covariance between $\boldsymbol{Z}$ and the process at the new location $\boldsymbol{s}_0$. More precisely,

$$\tilde{\boldsymbol{\Sigma}} = \begin{pmatrix} C(\mathbf{0}) & \boldsymbol{c} \\ \boldsymbol{c}^\top & \boldsymbol{\Sigma} \end{pmatrix}.$$

Then,

$$\mathrm{cov}(\tilde{\boldsymbol{\Sigma}}) = \frac{\tilde{L}[((M+1)\tilde{L} - 1)\tilde{\boldsymbol{\Sigma}} + ((M+1) - \tilde{L})\,\mathrm{tr}(\tilde{\boldsymbol{\Sigma}})\boldsymbol{I}_{M+1}]}{(M^2 + 2M)(M+1)}.$$

Normalizing $\mathrm{cov}(\tilde{\boldsymbol{\Sigma}})$ by $(M+1)/\tilde{L}$ such that $\tilde{\boldsymbol{\Sigma}}$ and $\mathrm{cov}(\boldsymbol{\Sigma})$ have the same trace, and using the fact that $\mathrm{tr}(\tilde{\boldsymbol{\Sigma}}) = (M+1)C(0)$, the estimators for $\boldsymbol{c}$ and $C(\mathbf{0})$ can be written as

$$\begin{aligned}\tilde{\boldsymbol{c}} &= \frac{(M+1)\tilde{L} - 1}{M^2 + 2M}\boldsymbol{c} = k_1 \boldsymbol{c},\\ \tilde{C}(\mathbf{0}) &= \frac{[((M+1)\tilde{L} - 1)C(\mathbf{0}) + ((M+1) - \tilde{L})\,\mathrm{tr}(\tilde{\boldsymbol{\Sigma}})]}{M^2 + 2M} = C(\mathbf{0}),\end{aligned}$$

where $k_1 = \frac{(M+1)\tilde{L}-1}{M^2+2M}$. Finally, the corrected kriging predictor is

$$\widehat{p}_{\mathrm{InvCov}}(\boldsymbol{Z}, \boldsymbol{s}_0) = \widehat{\beta}_{\text{ML-InvCov}} + k_1 \boldsymbol{r}^\top \boldsymbol{R}_{\mathrm{InvCov}}^{-1}(\boldsymbol{Z} - \widehat{\beta}_{\text{ML-InvCov}}\mathbf{1}),$$

and its variance is given by

$$\begin{aligned}\sigma_k^2(\boldsymbol{s}_0) &= \tilde{C}(\mathbf{0}) - \tilde{\boldsymbol{c}}^\top \mathrm{InvCov}_L(\boldsymbol{\Sigma})\tilde{\boldsymbol{c}} + \frac{(1 - \mathbf{1}^\top \mathrm{InvCov}_L(\boldsymbol{\Sigma})\tilde{\boldsymbol{c}})^2}{\mathbf{1}^\top \mathrm{InvCov}_L(\boldsymbol{\Sigma})\mathbf{1}},\\ &= \sigma^2\left(1 - k_1^2 \boldsymbol{r}^\top \boldsymbol{R}_{\mathrm{InvCov}}^{-1}\boldsymbol{r} + \frac{(1 - k_1 \mathbf{1}^\top \boldsymbol{R}_{\mathrm{InvCov}}^{-1}\boldsymbol{r})^2}{\mathbf{1}^\top \boldsymbol{R}_{\mathrm{InvCov}}^{-1}\mathbf{1}}\right).\end{aligned}$$

6. Numerical experiments

This section introduces numerical examples that allow us to have a better understanding of the problems discussed in this work. The material is divided into

three parts. First, we present an example to explore the estimations of the eigenvalues of $\mathrm{cov}_L(\widehat{\boldsymbol{\Sigma}})$. Second, a simulation experiment is carried out to gain more insights into the estimation of a nearly singular covariance structure. Finally, a Monte Carlo simulation study addresses spatial prediction (kriging) when the covariance matrix is nearly singular.

6.1. Eigenvalue estimation

Here, we consider a spatial process defined on the grid $[0,1]^2$, with a parametric and isotropic covariance model. The InvCov method is used to compare the estimates of the theoretical eigenvalues of $\boldsymbol{\Sigma}$, obtained using the `R` function `eigen`, with the estimates of the inverse of $\mathrm{InvCov}_L(\boldsymbol{\Sigma})$, normalized by L/N. In this study, three covariance structures were considered. The Gaussian covariance function is defined as $C(h) = \exp[-(h/\phi)^2]$, where $h = \|\boldsymbol{h}\|$ and $\phi > 0$. The Matérn covariance given by

$$C(h) = \frac{1}{2^{\nu-1}\Gamma(\nu)}\left(\frac{h}{\phi}\right)^{\nu} K_\nu\left(\frac{h}{\phi}\right),$$

where $\Gamma(\cdot)$ is the gamma function, $K(\cdot)$ is the modified Bessel function of the second kind, and $\phi > 0$ and $\nu > 0$. The third covariance is the exponential one, obtained as a particular of the Matérn, when $\nu = 1/2$. The following three sets of parameters were used in the experiment:

- Case 1: Exponential covariance with practical ranges (PR) 0.5, 1.0, 1.5, where $\phi = PR/2.996$.
- Case 2: Gaussian covariance with practical ranges 0.5, 1.0, 1.5, where $\phi = PR/1.731$.
- Case 3: Matérn covariance with practical ranges 0.5, 1.0, 1.5, where $\phi = PR/4.744$, and smoothness parameter $\nu = 1.5$.

The results of the eigenvalue estimations and the theoretical values are plotted in Figure 2 for the sets of covariance functions and parameters listed in Cases 1–3. The estimations are visually close to the true values, and it can be observed that the error of the estimated eigenvalues is reduced as practical range increases. Among the three approximate covariance models, the Gaussian covariance outperforms the other two, which is in agreement with Zimmermann (2015b) who argued that the Gaussian covariance is an anomalous model within the class of exponential covariances, showing a larger condition number than the other models for a fixed practical range. To quantify the discrepancy between them, we used the root mean square error (RMSE) to obtain a nonnegative number associated with every pair of curves. The formula used for the comparison is

$$\mathrm{RMSE} = \sqrt{\frac{1}{T}\sum_{i=1}^{T}(\widehat{\lambda}_i - \lambda_i)^2},$$

where $\widehat{\lambda}_i$ is the estimated value, λ_i is the true value and T is the total number of eigenvalues in each case. These values are summarized in Table 1. Among all cases, the largest values of the RMSE are attained for a practical range equal to 0.5. The RMSE also decreases as the practical range increases. For small practical ranges the Gaussian covariance has the worse performance, however, for large practical ranges the Gaussian covariance outperforms the other two models.

TABLE 1
RMSE for the eigenvalue estimations for Cases 1–3 displayed in Figures 2.

	Practical Range		
Covariance function	0.5	1.0	1.5
Exponential	0.461	0.345	0.267
Gaussian	0.637	0.151	0.028
Matérn ($\nu = 1, 5$)	0.584	0.352	0.190

In addition, Figure 2 also shows the sensitivity of the estimated eigenvalues by the proposed method described in Section 3, for different values of N (number of non-zero eigenvalues) and for the exponential, Gaussian, and Matérn covariance functions with different levels of dependence in each case.

Table 2 shows the standardized difference of the spectral radius between matrices $\boldsymbol{R}$ and $\text{InvCov}(\boldsymbol{R})^{-1}$ given by $(\lambda_1 - \widehat{\lambda}_1)/\lambda_1$, where λ_1 is the largest eigenvalue of $\boldsymbol{R}$ and $\widehat{\lambda}_1$ is its estimator. We observe that all the values are negative, indicating an underestimation of the largest eigenvalue.

TABLE 2
Difference between the largest correlation eigenvalue and the largest estimated eigenvalue for different values of N and covariance structures with different degrees of spatial dependence.

		Covariance function		
Practical Range	N	Exponential	Gaussian	Matérn ($\nu = 1.5$)
0.5	9	−0.0495	−0.0437	−0.0493
	10	−0.0400	−0.0328	−0.0389
	11	−0.0335	−0.0257	−0.0318
1.0	9	−0.0201	−0.0053	−0.0145
	10	−0.0153	−0.0032	−0.0104
	11	−0.0123	−0.0019	−0.0077
1.5	9	−0.0112	−0.0008	−0.0056
	10	−0.0084	−0.0004	−0.0038
	11	−0.0066	−0.0002	−0.0027

6.2. Parameter estimation

A Monte Carlo simulation study is carried out to observe the performance of the estimation method described in (3.1) and (3.3). We consider 100 fixed locations randomly sampled from $[0, 1] \times [0, 1]$, as displayed in Figure 3, where it can be seen how close some pairs of locations are.

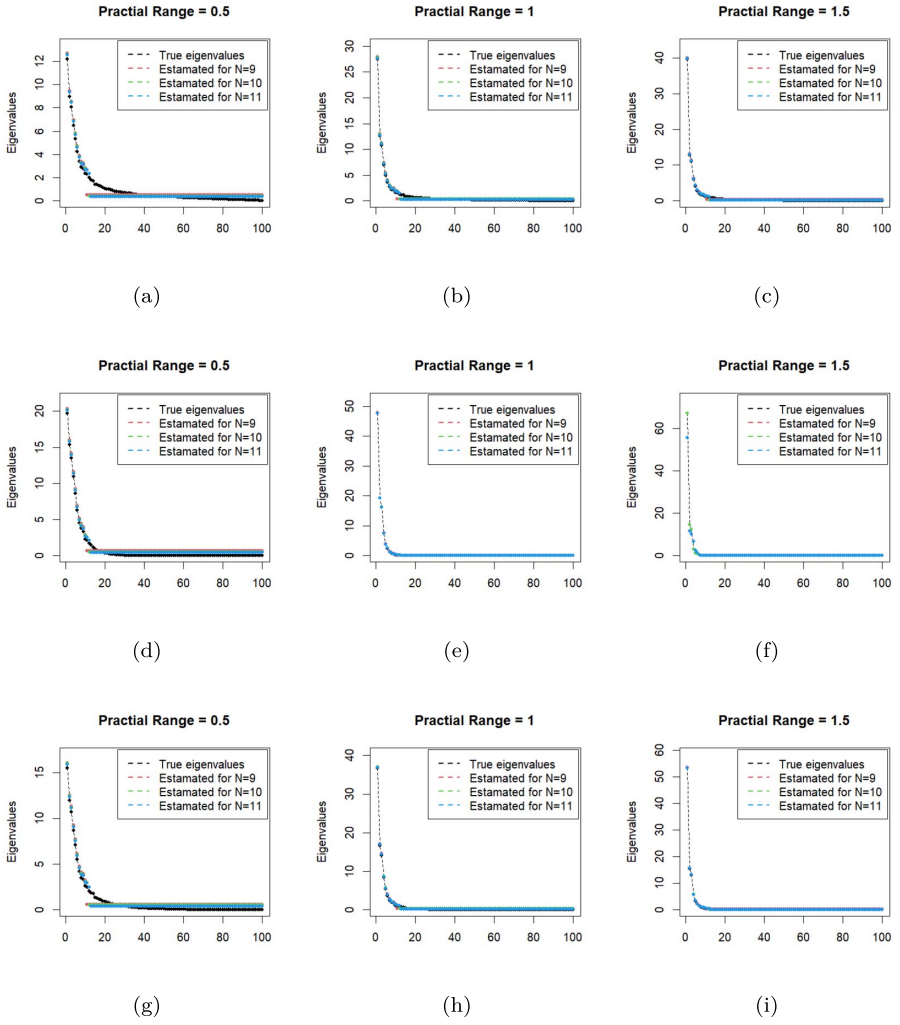


FIG 2. *Eigenvalue estimations and theoretical values for three covariance models with practical ranges equal to 0.5, 1.0, and 1.5. (a)–(c) Exponential covariance; (d)–(f) Gaussian covariance; (g)–(i) Matérn covariance with* $\nu = 1.5$

Exponential and Gaussian correlation models with different ranges (1.0, 1.5, 2.5, and 5.0) were used. For each case, we simulated 1000 independent random fields with zero mean and unit variance, and then the ML and ML-InvCov estimates of the range, β, and σ^2, were computed. In most cases, it was not feasible to obtain the ML estimator because a computationally singular system was obtained. Then, for comparison purposes, a fixed nugget equal to 0.01 was used to regularize the system and compared with the InvCov method. Figure 4 shows the empirical variograms (black dots) for the 1000 realizations of the generated

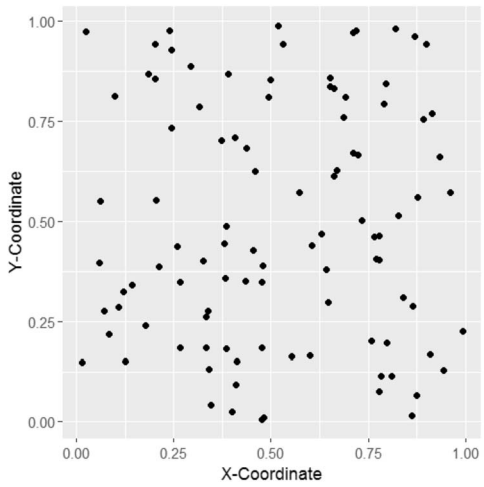


FIG 3. *Coordinates of the locations to be used in the Monte Carlo simulation study.*

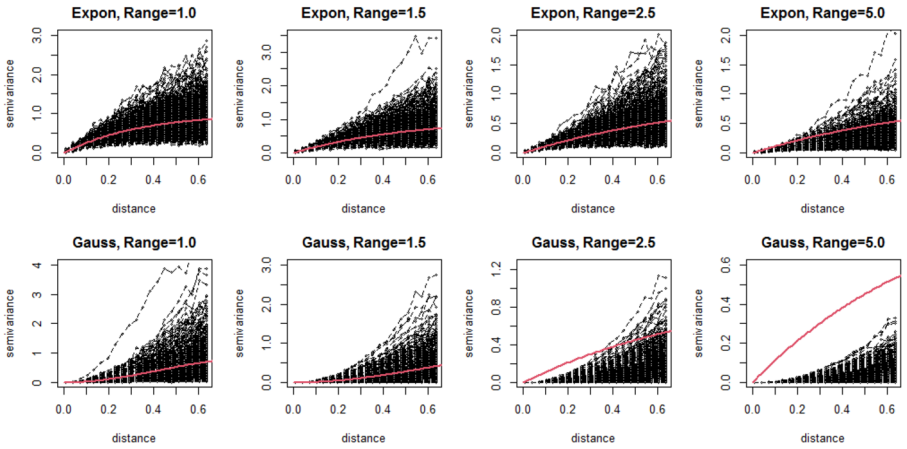


FIG 4. *Empirical variogram (black dots) for the 1000 realizations of the Gaussian random field. The red line represents the true model from which the data were simulated.*

Gaussian random field with exponential and Gaussian correlations. The red line represents the true model from which the observations were simulated. Note that as the spatial dependence increases (increasing range), the empirical variogram becomes increasingly distant from the theoretical variogram; therefore, using the least squares type technique is not very appropriate. We conjecture that the empirical variogram differs from the theoretical one as the practical range increases. One potential reason for this could be that the simulation process relies on the inverse of the square root of the covariance matrix, which, in this case, is computed using Cholesky decomposition. This can introduce insta-

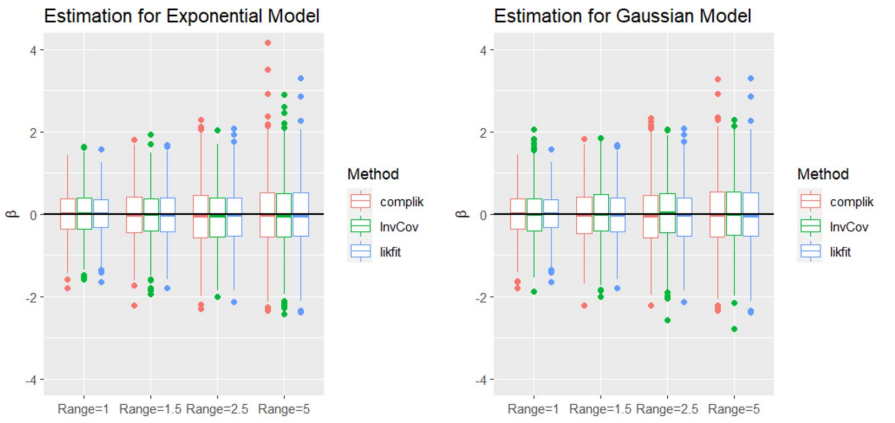
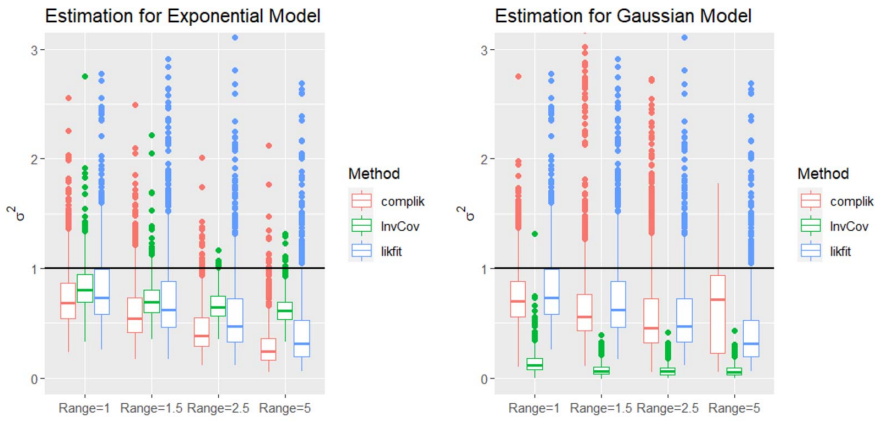


FIG 5. *Estimation of β using the pairwise composite likelihood (complik), the InvCov algorithm, and the ML method (likfit), with a fixed nugget equal to* 0.01 *for the complik anf likfit methods.*

bility that affects the randomly generated numbers, In this study, the maximum simulation range was selected to ensure that Cholesky decomposition could be performed computationally.

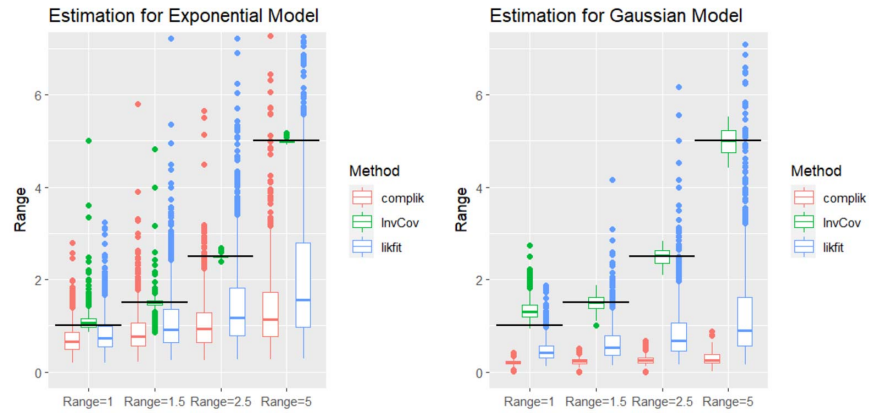
The application of the methodology described in Algorithm 2 can be seen in Figures 5–7. The schemes shown on the x-axis indicate the range for which the observations were simulated, where schemes 1, 2, 3, and 4 have practical ranges equal to 1, 1.5, 2.5, and 5.0, respectively. The estimation process was carried out using the ML method through the likfit function of Ribeiro Jr et al. (2020), the pairwise composite likelihood (CL) method, through the GeoFit2 function of Bevilacqua et al. (2022b), and the InvCov method. To implement the pairwise composite likelihood method, the number of neighbors was set to a fixed value of 3, and a conditional likelihood approach was employed. In particular, Figure 5 shows the mean estimator. The performance of the all methods is very similar for both the exponential and Gaussian covariance models for the different ranges simulated, highlighting that the uncertainty associated with the estimation of the β parameter increases as the range increases, although this increase is very slight. In schemes 1 and 2, it can be observed that the likfit method performs slightly better than the rest, while in scheme 4 the InvCov method ouperforms the others.

Figure 6 shows the behavior of the estimator of the variance. Only in the scheme 1 for exponential correlation the InvCov have good result. As the practical range increases, the true variance is underestimated regardless of the estimation method used. For the exponential correlation, the InvCov method performs better than the others, but in Gaussian case has the worst behavior. The composite likelihood method has a slightly better performance in Gaussian case, while the likfit has a similar behavior for both correlation functions. This



(a) (b)

FIG 6. *Estimation of* σ^2 *using the complik, InvCov, and likfit methods, with a fixed nugget equal to* 0.01 *for the complik anf likfit methods.*



(a) (b)

FIG 7. *Estimation of the practical range using the complik, InvCov, likfit methods, with a fixed nugget equal to* 0.01 *for the complik anf likfit methods.*

fact may be may be related to to the discrepancy between the empirical and theoretical variograms in those cases where the practical range is large.

Although the performance for the estimator of σ^2 is not as expected with regards to the Gaussian case, it can be seen from Figure 7 that the behavior for the range is quite different. Here, ϕ is estimated and then transformed into the practical range for the exponential model range $= 3\phi$ and for the Gaussian model range $= \sqrt{3}\phi$. In all cases, the InvCov method outperforms likfit method in terms of variance and bias (which has been marked with a black line in Fig-

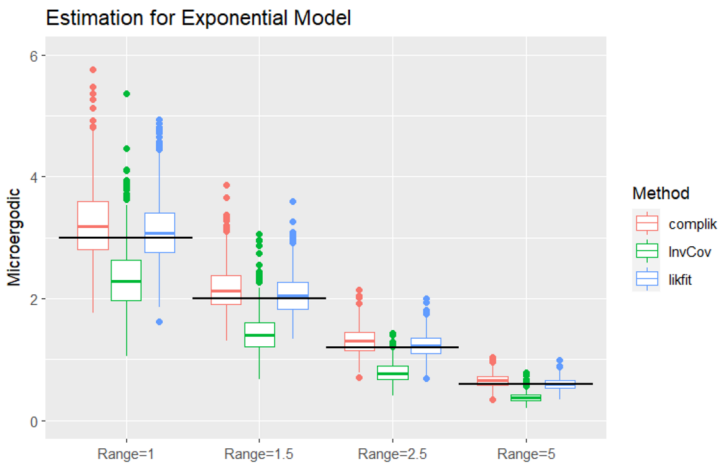


FIG 8. *Estimation of the microergodic parameter using the complik, InvCov, and ML likfit method, with a fixed nugget equal to* 0.01 *for the complik anf likfit methods.*

ure 7). In the Gaussian covariance model the pairwise composite likelihood has the smallest variance for each scheme but a higher bias. Moreover, for this case 1.8%, 5.2%, 17.6%, and 56.4% of the times the estimations of ϕ were negative for schemes 1-4 respectively. Also notice that in all cases the likfit and pairwise composite likelihood methods underestimate the true spatial dependence. Perhaps this positively impacts the likfit estimates of the variance (see Figure 6(b)). Surprisingly, the InvCov estimates for schemes 3 and 4 in the exponential model, present a very tiny dispersion.

Note that the poor performance of the ML and CL methods in estimating the parameters σ^2 and ϕ is not surprising, because when the practical range significantly increases, it becomes challenging to consistently estimate both parameters separately, when using an increasing domain sampling scheme. However, by using the asymptotic infill approach, it is possible to consistently estimate the microergodic parameter, which, in the case of the Matérn model is $\sigma^2/\phi^{2\nu}$. This can be seen in Figure 8 for the exponential model ($\nu = 0.5$). Here, the complik and likfit methods exhibit the expected behavior, as the ratio cancels out the biases stemming from the imperfect estimation of σ^2 and ϕ, respectively. For the InvCov method, an underestimation of the microergodic parameter is observed, which can be explained by the fact that the estimator of ϕ is unbiased while the estimator of σ^2 is biased, therefore, the ratio maintains the bias of the estimator of σ^2. It is worth noting that, in all cases, as the practical range increases, the uncertainty surrounding the microergodic parameter decreases.

The proposed methodology allows spatial prediction (kriging) to be carried out when the inverse of $\boldsymbol{\Sigma}$ is not computationally feasible to obtain. In this case, the Gaussian covariance is the most affected, especially when the practical range is large. Figure 9 displays the kriging estimates and their variance for a

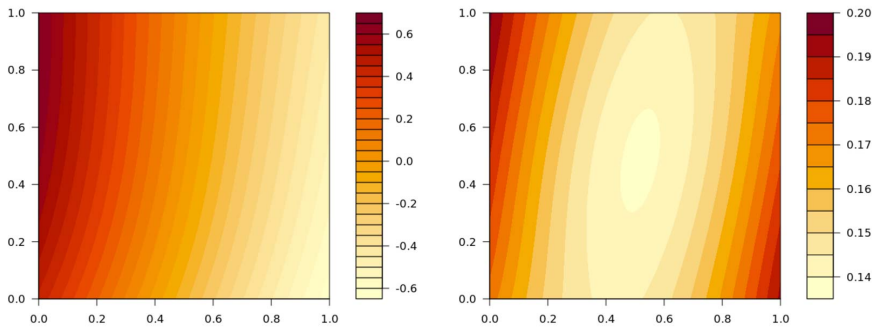


FIG 9. *Kriging predictor (a) and its standard deviation (b) for a Gaussian covmodel with practical range equal to 5.0.*

rectangular grid (101×101) in the unit square. Figure 9(a) shows the average of 1000 simulations using ordinary kriging, while Figure 9(b) shows the square root of the average of the respective variances. To explore the kriging behavior, the same parameters considered for the Monte Carlo simulation were used to generate the plots. i.e., $\beta = 0$, $\sigma^2 = 1$, $\tau^2 = 0$ and practical range equal to 5.0, that is, $\phi = 2.8888$. The observed trend from left to right that is observe in Figure 9(a) is apparent due to the short range of the scale. The surface shown corresponds to just one realization of the process. Other patterns can be observed when running the experiments many times. The standard deviation of the kriging predictor shown in Figure 9(b) displays largest values for those areas with small number of observations.

Finally, Figure 10 shows the scalability of the estimation process proposed in Algorithm 2, measuring the computational time of the estimation process for different sample sizes, setting the covariance models to Exponential and Gaussian, and the practical range equal to 2.0 when the data are simulated from a zero mean Gaussian random field over the square $[0, 1]^2$. An exponential trend for the computational is observed for both covariance models when they are plotted as a function of the sample size.

7. An application

The macroalgae dataset was analyzed in Acosta et al. (2016) whithin the context of effective sample size (ESS) for a spatial regression process with a constant mean. Subsequently, the same dataset was considered as an illustration for the generalization of the ESS for a general spatial regression process (Acosta and Vallejos, 2018). In this study, we revisit the macroalgae dataset, which consists of 427 observations pertaining to the density of *Lessonia trabeculata* (scientific name of macroalgae in the study) per 20 m^2. These observations were collected by the IFOP[1] institute in a protected area near Quintero, Chile. A transect

[1]Instituto de Fomento Pesquero (Chilean Fisheries Research).

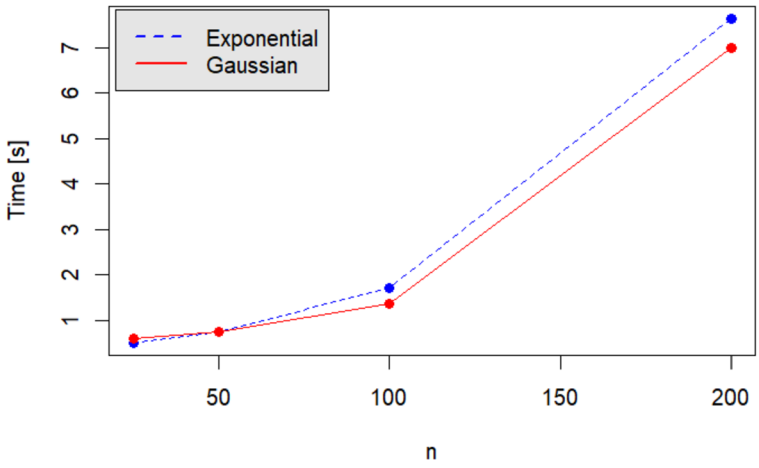


FIG 10. *Execution time (in seconds) for the InvCov method proposed in Algorithm 2, for the Exponential and Gaussian covariance models as a function of the sample size n. The estimations were obtained using the R function proc.time(·). The computations were run on a i7-8565U processor, CPU 1.99 GHz with 8 GB in RAM.*

sampling scheme consisting of 26 perpendicular line transects was employed to study species found at depths no greater than 20 m. The locations, transects and study area can be seen in Figure 11.

In order to achieve symmetry in the distribution of the entire raw dataset, Acosta et al. (2016) applied a logarithmic transformation of the form $W(\boldsymbol{s}_i) = \ln(Z(\boldsymbol{s}_i) + 2.586)$, where $Z(\boldsymbol{s}_i)$ represents the original process observed at location $\boldsymbol{s}_i$ and $W(\boldsymbol{s}_i)$ denotes the transformed process at the same location. This transformation also has an impact on the estimation of the covariance components of the process. As a result, multiple covariance matrices were employed to estimate the ESS, which indicates the number of independent and identically distributed variables associated with a spatial sample of size n. The ESS is particularly important in this context as it quantifies the decrease in sample size caused by the presence of spatial correlation among the georeferenced observations. This has implications, for example, for determining the number of observations to be collected in a subsequent study assuming that the correlation structure of the underlying process remains unchanged.

Here, we use the raw data (density of the macroalgae) and assume that the original spatial regression process is as in (4.1). The relationship between density and the UTM coordinates (north and east) is depicted in Figure 12. No distinct patterns or trends can be observed between density and the spatial variables. However, since the study encompasses two sampled zones—one area unaffected by human intervention and another open area—the means of these two zones exhibit differences (as evident in Figure 12(a), particularly for values of North less than 6375000). A similar effect is observed in Figure 12(b).

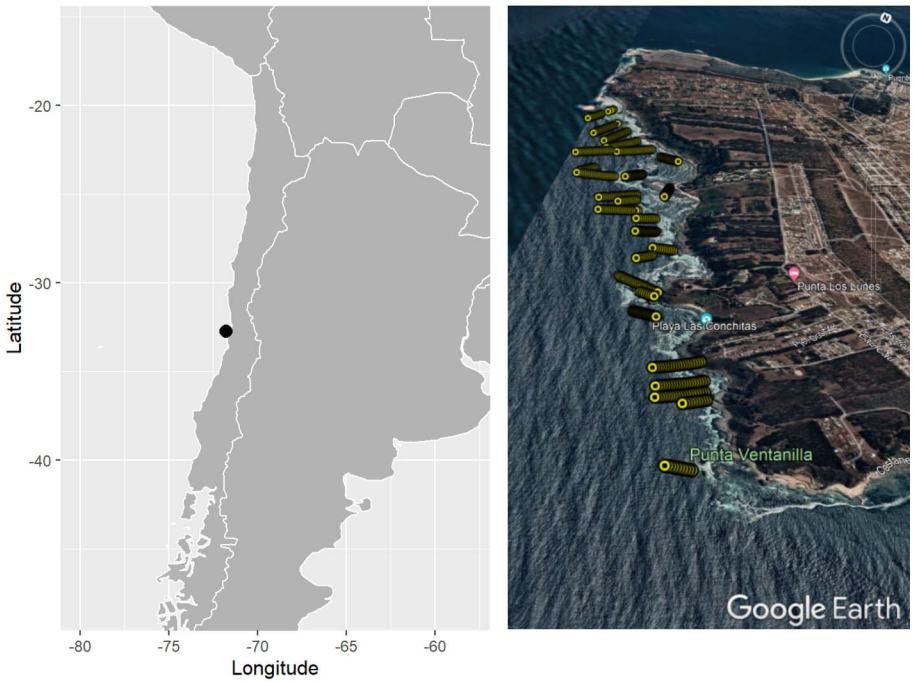
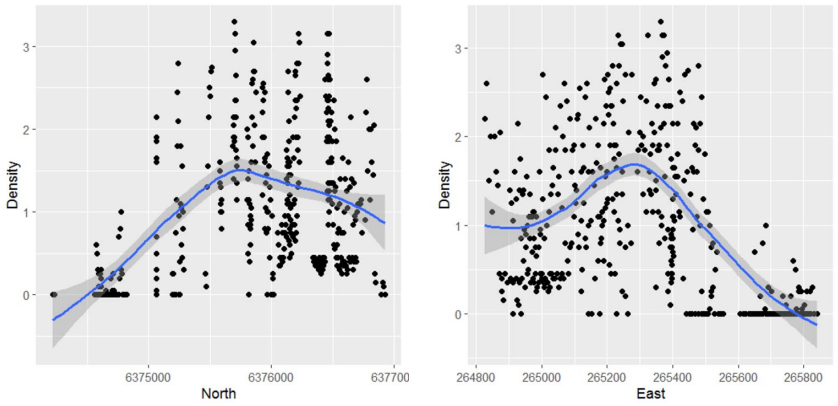


FIG 11. *Location of study in central Chile (left); Transect lines (in yellow) where the observations were taken (right). This area is near Quintero in the fifth region of Chile.*



(a) (b)

FIG 12. *Density versus the coordinates north and east for the macroalgae dataset. The coordidnates were measured in the UTM metric system, thus one unity corresponds to one meter.*

Direct maximum likelihood estimation of a spatial regression process with a constant mean and a Gaussian covariance function is not feasible with the macroalgae dataset. For instance, when using the `geoR` package, an error message is encountered, indicating a reciprocal condition number of 4.34×10^{-21}. If a fixed nonzero nugget effect is added to the covariance structure, the remaining parameters of the model can be estimated successfully. On the other hand, if the nugget effect is not fixed and treated as an additional parameter to be estimated, the estimation algorithm produces a sequence of nugget effect values that converge to zero. In such cases, the InvCov method can be applied, particularly when the nugget effect is equal to zero. Observing the empirical variogram depicted in Figure 7, it is evident that the practical range is approximately 450. Therefore, a suitable value for ϕ_0 in the Gaussian model would be 250. Evaluating the correlation matrix associated with the Gaussian model at $\phi = 250$, it is found that the 16 largest eigenvalues account for 95% of the total variance. Consequently, a fixed value of 16 was chosen for N, with $L = N - 1 = 15$. Table 3 presents the results of the maximum likelihood (ML) estimates obtained with a fixed nugget effect that is nonzero, as well as the ML estimates using the InvCov methodology. The most significant distinction between the InvCov and regular ML estimations is observed for the parameter ϕ. In accordance with the performance displayed in Figure 7, the InvCov method exhibits a small bias, while the ML method consistently underestimates the true value. Consequently, this characteristic also impacts the range, which in this case is equal to $\sqrt{3}\phi$.

TABLE 3
Parameter estimation of the Gaussian covariance model.

	Fixed			Estimated				
Method	τ_0^2	σ_0^2	ϕ_0	σ^2	ϕ	β	range	η
InvCov	0.00	–	250	0.789	264.00	0.673	457.26	0.01
ML1	0.01	0.80	250	0.632	7.17	0.962	12.42	–
ML2	0.10	0.80	250	0.520	43.73	0.944	75.68	–
ML3	0.20	0.80	250	0.536	70.95	0.943	122.79	–
Average	–	–	–	0.769	–	1.026	0	–

Figure 13 shows the outcome of parameter estimation for the Gaussian correlation model in combination with the empirical semivariogram. The maximum likelihood (ML) estimates, when considering a model with a fixed and nonzero nugget effect, underestimate the spatial dependence of the process. Conversely, the InvCov method accurately captures the range as indicated by the empirical variogram. For distances less than 200 meters, the patterns of all other methods closely resemble that of a white noise sequence.

To compare the estimation methods, leave-one-out cross-validation was used, with the metrics performance the RMSE (defined in Section 6.1), and the DMSE, defined through

$$DMSE = \sqrt{\frac{1}{M}\sum_{j=1}^{M}\left(\frac{\widehat{Z}_{(-j)}(\boldsymbol{s}_j) - Z(\boldsymbol{s}_j)}{\widehat{\sigma}_{(-j)}}\right)^2},$$

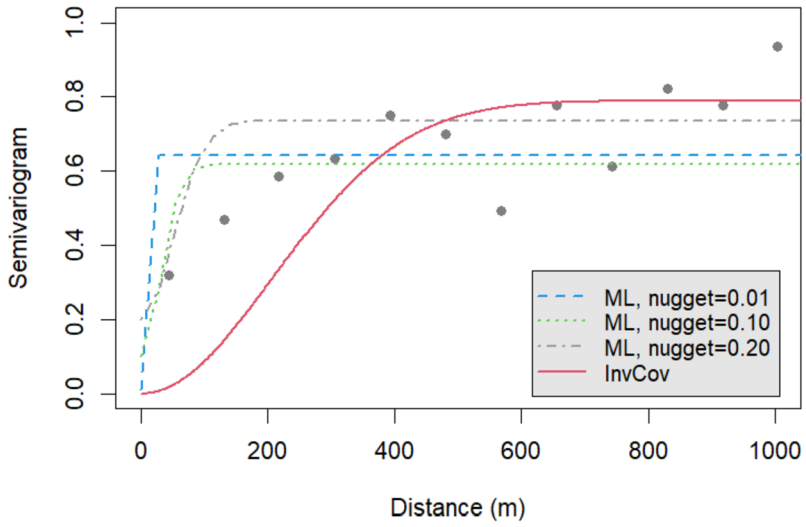


FIG 13. *Empirical variogram (dots) and ML estimations of the Gaussian variogram for different nugget effects. Distance has been measured in meters.*

where $\widehat{Z}_{(-j)}(\boldsymbol{s}_j)$ is the kriging estimator of $Z(\boldsymbol{s}_j)$, and $\widehat{\sigma}^2_{(-j)}$ is the kriging variance with all data points except j. The results are shown in Table 4. The best performance, according with RMSE, was obtained for the maximum likelihood estimates with a fixed nugget equal to 0.10. The performance of our method is comparable with the maximum likelihood approach. However, the RMSE does not take into account the spatial location where the prediction was made, so it is strongly dependent on the observations at the boundary of the region. According with DMSE, our proposal (invcov method) has the best performance. Figure 14 shows the kriging variance associated with the all estimation methods.

TABLE 4

*First and second row contain root mean square error (RMSE) and dimensionless mean squared error (DMSE) for the leave-one-out cross validation for the invcov method, maximum likelihood with fixed nugget (*0.01*,* 0.10*, and* 0.20*, respectively), and the average estimator. The third row shows the coverage rate for a 95% confidence interval.*

	InvCov	ML1	ML2	ML3	Average
RMSE	0.608	0.659	0.411	0.430	0.876
DMSE	1.010	1.106	1.115	0.878	0.998
Coverage Prob.	0.918	0.780	0.895	0.965	0.967

Therefore, we can conclude that incorporating a nugget improves the point estimator but not the Kriging variance. This is evident in the coverage of the 95% confidence intervals, where the invcov method is closer to the expected value compared to the maximum likelihood method with a nugget value of 0.10 (refer to Table 4 and Figure 14 for more details).

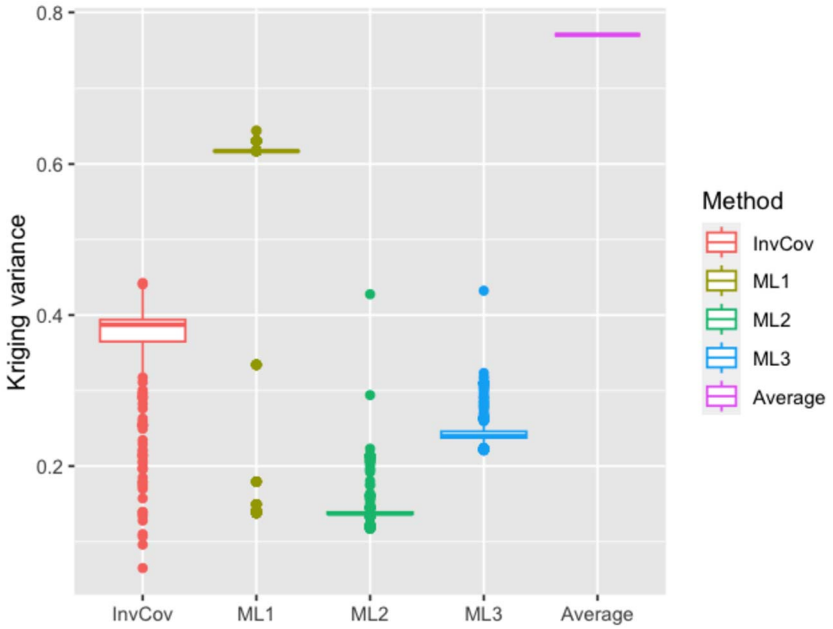


FIG 14. *Kriging variance for leave-one-out cross-validation using the transect data employed in the application.*

As a direct result of estimating the covariance parameters, the ESS can be estimated, which is defined as

$$\text{ESS} = \mathbf{1}^\top \boldsymbol{R}(\boldsymbol{\theta})^{-1}\mathbf{1}, \tag{7.1}$$

where $\boldsymbol{R}(\boldsymbol{\theta})$ is as in model (4.1). For simplicity let us define $\boldsymbol{R} = \boldsymbol{R}(\boldsymbol{\theta})$. When $\boldsymbol{R}$ is singular or nearly singular, the ESS cannot be obtained. There are at least two alternatives to overcome this inconvenience. The first alternative, inspired by Vallejos and Osorio (2014), involves setting the correlation between all pairs of observations to one, resulting in $\boldsymbol{R} = \mathbf{1}\mathbf{1}^\top$, which is not invertible. In this case, the Moore-Penrose pseudoinverse $\boldsymbol{R}^+ = \mathbf{1}\mathbf{1}^\top/M^2$ is used to obtain ESS = 1. The second alternative is to use the InvCov approximation of $\boldsymbol{R}^{-1}$, $\boldsymbol{R}^{-1}_{\text{InvCov}}$, and thus obtain the ESS as

$$\text{ESS} = \mathbf{1}^\top \boldsymbol{R}^{-1}_{\text{InvCov}}\mathbf{1}.$$

In the estimation process, the sum of eigenvalues, $\sum_{i=1}^{N}\lambda_i + (M-N)\mu$, is not necessarily equal to M. To account for this, a correction factor defined in Equation (4.11) is applied. The ESS can be expressed as

$$\text{ESS} = \left(\frac{1}{M}\sum_{i=1}^{N}\frac{1}{\lambda_i} + \frac{(M-N)}{M\mu}\right)\mathbf{1}^\top \boldsymbol{R}^{-1}_{\text{InvCov}}\mathbf{1}. \tag{7.2}$$

By using the InvCov-ML estimate for ϕ and applying the correction factor used for σ, the ESS in (7.2) amounts to 16 sample units. When using the Moore-Penrose pseudoinverse, the ESS is calculated as 14 sample units.

8. Discussion

In this paper, we suggest an estimation method to overcome the problem of having a nearly singular covariance matrix in a spatial regression model. It is a generalization of a methodology first proposed by Marzetta et al. (2011) for processes in which there are replicates. Our generalization includes the treatment of processes for which only one trajectory is available and for those datasets with one or more distances between pairs of locations that are close in space. The estimation relies on previous knowledge of the parameters N and L such that $L < N$, and where both parameters depend on M (the size of the spatial dataset). The novelty of our paper lies in the fact that we provide an estimation method for spatial data and covariance models where traditional ML estimations cannot be computed due to dealing with an ill-conditioned problem. Numerical experiments point out that the computation of the InvCov estimator is demanding and it highly depends on the spatial sample size. Hence, the computation of the proposed estimator for large spatial datasets is a challenge.

Tucci and Wang (2019) show that $\mathrm{InvCov}_L(\boldsymbol{D})$ has a surprisingly simple algebraic structure, i.e., it is a polynomial of the diagonal matrix $\boldsymbol{D}$. They also provide formulas to compute the coefficients of the polynomial, and extend these ideas by replacing random unitary matrices with random permutation matrices and by using the Ewens measure. Then, they define two new operators $K_{\vartheta,M,L} = \mathbb{E}[V_\sigma^\top (V_\sigma \boldsymbol{K} V_\sigma^\top) V_\sigma]$, and $\widetilde{K}_{\vartheta,M,L} = \mathbb{E}[V_\sigma^\top (V_\sigma \boldsymbol{K} V_\sigma^\top)^+ V_\sigma]$ to estimate $\boldsymbol{\Sigma}$ and its inverse $\boldsymbol{\Sigma}^{-1}$, respectively. Here, V_σ is a unitary permutation matrix, and $\boldsymbol{A}^+$ is the Moore Penrose pseudoinverse of the $\boldsymbol{A}$ (Penrose, 1955). Additionally, Tucci and Wang (2019) provide an explicit formula for $K_{\vartheta,M,L}$ and an inductive formula to compute $\widetilde{K}_{\vartheta,M,L}$. Moreover, they studied the asymptotic behavior for certain matrices with the mean conjugate estimator under the Ewens measure. The extension of their work to a spatial statistics framework is an interesting open problem that will be addressed in future research.

In a future article, we will also report on further numerical experiments, in which different sampling schemes and sample sizes will be considered, as well as the practical issues that arise.

Appendix A: Definition and results for Section 2

Definition A.1. Given a matrix $\boldsymbol{A} = \mathrm{diag}(a_1, a_2, \ldots, a_n)$, the Vandermonde matrix associated with $\boldsymbol{A}$, denoted by $\boldsymbol{\Delta}(\boldsymbol{A})$, is defined as

$$\boldsymbol{\Delta}(\boldsymbol{A}) := \begin{pmatrix} a_1^{n-1} & a_2^{n-1} & \cdots & a_n^{n-1} \\ \vdots & \vdots & & \vdots \\ a_1 & a_2 & \cdots & a_n \\ 1 & 1 & \cdots & 1 \end{pmatrix}, \tag{A.1}$$

whose determinant is given by $\det(\boldsymbol{\Delta}(\boldsymbol{A})) = \prod_{1\le i\le j\le n}(a_j - a_i)$.

Theorem A.2 (Marzetta et al., 2011)**.** *Let $\boldsymbol{D}_N$ be a full rank $N \times N$ diagonal matrix. For $f \in \mathcal{C}[d_{\min}, d_{\max}]$, we have that*

$$\int_{\Omega_{L,N}} tr\big(f\big(\boldsymbol{\Phi}^*\boldsymbol{D}_N\boldsymbol{\Phi}\big)\big)\,d\phi = \sum_{k=0}^{L-1}\frac{(N-(k+1))!}{(L-(k+1))!}\frac{det(\boldsymbol{G}_k)}{det(\boldsymbol{\Delta}(\boldsymbol{D}_N))}, \tag{A.2}$$

where $\Omega_{L,N}$ is defined in Section 2, $\boldsymbol{G}_k$ is the matrix defined by replacing the $(k+1)$-th row of $\boldsymbol{\Delta}(\boldsymbol{D}_N)$, $\{d_i^{N-(k+1)}\}_{i=1}^n$, by $\{\boldsymbol{I}^{(N-L)}(x^{(L-(k+1))}f(x))|_{x=d_i}\}_{i=1}^N$, and $\mathrm{tr}(\cdot)$ is the trace operator.

Proposition A.3 (Marzetta et al., 2011)**.** *Let $\boldsymbol{D}_N$ be a $N\times N$ diagonal matrix. Consider $1 \le L < N$ and $\boldsymbol{Z}$ as a Gaussian random matrix with zero mean entries and variances equal to one such that all entries are iid of size $N \times L$. Then*

$$\int_{\Omega_{L,N}} \mathrm{tr}\big(\big(\boldsymbol{\Phi}^*\boldsymbol{D}_N\boldsymbol{\Phi}\big)^{-1}\big)\ d\phi = (N-L)\frac{det(\boldsymbol{G})}{det(\boldsymbol{\Delta}(\boldsymbol{D}_N))}, \tag{A.3}$$

and

$$\mu := \mathbb{E}\big[\mathrm{tr}\big(\big(\boldsymbol{Z}^*\boldsymbol{D}_N\boldsymbol{Z}\big)^{-1}\big)\big] = \frac{det(\boldsymbol{G})}{det(\boldsymbol{\Delta}(\boldsymbol{D}_N))}, \tag{A.4}$$

where $\boldsymbol{G}$ is the matrix constructed in Theorem A.2 by replacing the L-th row of the Vandermonde $\boldsymbol{\Delta}(\boldsymbol{D}_N)$ by the row $(d_1^{N-(L+1)}\log(d_1), \ldots, d_N^{N-(L+1)}\log(d_N))$.

The matrix $\mathrm{InvCov}_L(\boldsymbol{D})$ is diagonal and can be decomposed as

$$\mathrm{InvCov}_L(\boldsymbol{D}) = \mathrm{diag}\big(\boldsymbol{\Lambda}_L(\boldsymbol{D}_N), \mu\boldsymbol{I}_{M-N}\big),$$

where $\boldsymbol{\Lambda}_L(\boldsymbol{D}_N) = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$. Using Lemma 1 in Marzetta et al. (2011), it follows that

$$\lambda_k = \frac{\partial}{\partial d_k}\int_{\Omega_{L,N}} \mathrm{tr}\log\big(\boldsymbol{\Phi}^*\boldsymbol{D}_N\boldsymbol{\Phi}\big)\ \mathrm{d}\phi. \tag{A.5}$$

As a result, the components d_k^{-1}, $k = 1, \ldots, N$, of $\boldsymbol{D}$ are approximated by λ_k given in Equation (A.5), while the components equal to zero in $\boldsymbol{D}$ are replaced by μ given in Equation (A.4) in the approximation of the inverse of $\boldsymbol{D}$.

Appendix B: Proof of Theorem 3.1

We note that using the linearity of the derivative operator and pulling out the constants we have that

$$\begin{aligned}\lambda_i &= \frac{\partial}{\partial d_i}\left(\sum_{k=0}^{L-1}\frac{(N-(k+1))!}{(L-(k+1))!}\frac{|\boldsymbol{G}_k|}{|\boldsymbol{\Delta}(\boldsymbol{D}_N)|}\right)\\ &= \sum_{k=0}^{L-1}\frac{(N-(k+1))!}{(L-(k+1))!}\frac{\partial}{\partial d_i}\left(\frac{|\boldsymbol{G}_k|}{|\boldsymbol{\Delta}(\boldsymbol{D}_N)|}\right).\end{aligned}$$

By application of the quotient rule it follows that

$$\lambda_i = \sum_{k=0}^{L-1} \frac{(N-(k+1))!}{(L-(k+1))!} \left(\frac{\frac{\partial |\boldsymbol{G}_k|}{\partial d_i} |\boldsymbol{\Delta}(\boldsymbol{D}_N)| - |\boldsymbol{G}_k| \frac{\partial |\boldsymbol{\Delta}(\boldsymbol{D}_N)|}{\partial d_i}}{|\boldsymbol{\Delta}(\boldsymbol{D}_N)|^2} \right).$$

Using the fact that $\partial(\det(X)) = \det(X)\mathrm{trace}(X^{-1}\partial X)$ (Magnus and Neudecker, 2007) we obtain

$$\lambda_i = \sum_{k=0}^{L-1} \frac{(N-(k+1))!}{(L-(k+1))!} \left(\frac{|\boldsymbol{G}_k||\boldsymbol{\Delta}(\boldsymbol{D}_N)|(\mathrm{tr}[\boldsymbol{G}_k^{-1} \frac{\partial(\boldsymbol{G}_k)}{\partial d_i}] - \mathrm{tr}[\boldsymbol{\Delta}(\boldsymbol{D}_N)^{-1} \frac{\partial(\boldsymbol{\Delta}(\boldsymbol{D}_N))}{\partial d_i}])}{|\boldsymbol{\Delta}(\boldsymbol{D}_N)|^2} \right).$$

Canceling $|\Delta(D_N)|$ we have that

$$\lambda_i = \sum_{k=0}^{L-1} \frac{(N-(k+1))!}{(L-(k+1))!} \left(\frac{|\boldsymbol{G}_k|(\mathrm{tr}[\boldsymbol{G}_k^{-1} \frac{\partial(\boldsymbol{G}_k)}{\partial d_i}] - \mathrm{tr}[\boldsymbol{\Delta}(\boldsymbol{D}_N)^{-1} \frac{\partial(\boldsymbol{\Delta}(\boldsymbol{D}_N))}{\partial d_i}])}{|\boldsymbol{\Delta}(\boldsymbol{D}_N)|} \right).$$

Let $\boldsymbol{V}_N$ be a Vandermonde matrix defined as

$$\boldsymbol{V}_N = \boldsymbol{V}(\boldsymbol{D}_N) := \begin{pmatrix} 1 & d_1 & \cdots & d_1^{N-1} \\ 1 & d_2 & \cdots & d_2^{N-1} \\ \vdots & \vdots & & \vdots \\ 1 & d_N & \cdots & d_N^{N-1} \end{pmatrix}.$$

Then, for a row permutation matrix $\boldsymbol{P}_N$, we have that

$$\boldsymbol{\Delta}(\boldsymbol{D}_N) = (\boldsymbol{V}_N \boldsymbol{P}_N)^\top.$$

Furthermore,

$$\frac{\partial}{\partial d_i}(\boldsymbol{\Delta}(\boldsymbol{D}_N)) = \begin{pmatrix} 0 & 0 & \cdots & (N-1)d_i^{N-2} & \cdots & 0 & 0 \\ 0 & 0 & \cdots & (N-2)d_i^{N-3} & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 2d_i & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 1 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & \cdots & 0 & 0 \end{pmatrix},$$

and denoting the inverse of $\boldsymbol{\Delta}(\boldsymbol{D}_N)$ by $\boldsymbol{\Delta}(\boldsymbol{D}_N)^{-1} = [b_{ij}]_{i,j=1}^N$, we obtain

$$\mathrm{tr}\left[\boldsymbol{\Delta}(\boldsymbol{D}_N)^{-1} \frac{\partial(\boldsymbol{\Delta}(\boldsymbol{D}_N))}{\partial d_i}\right] = \sum_{l=1}^{N-1} b_{il}(N-l)d_i^{N-(l+1)}.$$

For $\boldsymbol{G}_k$, we notice that

$$\frac{\partial}{\partial d_i}(\boldsymbol{G}_k) = \begin{pmatrix} 0 & 0 & \cdots & (N-1)d_i^{N-2} & \cdots & 0 & 0 \\ 0 & 0 & \cdots & (N-2)d_i^{N-3} & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & G_i & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 2d_i & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 1 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & \cdots & 0 & 0 \end{pmatrix},$$

where $G_i = \frac{\partial}{\partial d_i}(\boldsymbol{I}^{(N-L)}(x^{(L-(k+1))}\log(x))\big|_{x=d_i}) = \boldsymbol{I}^{(N-L-1)}(x^{(L-(k+1))} \times \log(x))\big|_{x=d_i}$. Defining the inverse of $\boldsymbol{G}_k$ by $\boldsymbol{G}_k^{-1} = [g_{ij}]_{i,j=1}^{N}$, one gets

$$\operatorname{tr}\left[\boldsymbol{G}_k^{-1}\frac{\partial(\boldsymbol{G}_k)}{\partial d_i}\right] = \sum_{\substack{l=1 \\ l\neq k+1}}^{N-1} g_{il}(N-l)d_i^{N-(l+1)} + g_{i,k+1}G_i.$$

Hence, we obtain that

$$\begin{aligned} &\operatorname{tr}\left[\boldsymbol{G}_k^{-1}\frac{\partial(\boldsymbol{G}_k)}{\partial d_i}\right] - \operatorname{tr}\left[\boldsymbol{\Delta}(\boldsymbol{D}_N)^{-1}\frac{\partial(\boldsymbol{\Delta}(\boldsymbol{D}_N))}{\partial d_i}\right] \\ &= \sum_{\substack{l=1 \\ l\neq k+1}}^{N-1} (g_{il}-b_{il})(N-l)d_i^{N-(l+1)} + g_{i,k+1}G_i - b_{i,k+1}\big(N-(k+1)\big)d_i^{N-(k+2)}. \end{aligned}$$

Thus,

$$\begin{aligned} \lambda_i = \sum_{k=0}^{L-1} \frac{(N-(k+1))!}{(L-(k+1))!}\frac{|\boldsymbol{G}_k|}{|\boldsymbol{\Delta}(\boldsymbol{D}_N)|}\Bigg(&\sum_{\substack{l=1 \\ l\neq k+1}}^{N-1} \big((g_{il}-b_{il})(N-l)d_i^{N-(l+1)}\big) \\ &+ g_{i,k+1}G_i - b_{i,k+1}\big(N-(k+1)\big)d_i^{N-(k+2)}\Bigg), \end{aligned}$$

completing the proof.

Acknowledgments

The authors are grateful to Felipe Osorio for helpful discussions and support.

Funding

R. Vallejos acknowledges financial support from CONICYT through the MATH-AMSUD program, grant 20-MATH-03, from UTFSM, grant P_LIR_2020_20, from AC3E, grant FB-0008, and from ANID, Fondecyt grant 1230012. J. Acosta acknowledges financial support from ANID, Fondecyt grant 11230502.

References

Ababou, R., Bagtzoglou, A. C., and Wood, E. F. (1994). On the condition number of covariance matrices in kriging, estimation, and simulation of random fields. *Mathematical Geology*, 26:99–133. MR1256401

Acosta, J. and Vallejos, R. (2018). Effective sample size for spatial regression models. *Electronic Journal of Statistics*, 12:1935–1524. MR3858695

Acosta, J., Vallejos, R., and Osorio, F. (2016). Effective sample size for line transect models with an application to marine macroalgae. *Journal of Agricultural, Biological and Environmental Statistics*, 21:407–425. MR3542079

Ayyıldız, E., Purutçuoglu, V., and Wit, E. (2012). A short note on resolving singularity problems in covariance matrices. *International Journal of Statistics and Probability*, 1:113–118.

Bevilacqua, M., Caamaño-Carrillo, C., and Porcu, E. (2022a). Unifying compactly supported and matérn covariance functions in spatial statistics. *Journal of Multivariate Analysis*, 189(104949). MR4375672

Bevilacqua, M., Furrer, R., Faouzi, T., and Porcu, E. (2019). Estimation and prediction using generalized wendland covariance functions under fixed domain asymptotics. *Annals of Statistics*, 47:828–856. MR3909952

Bevilacqua, M., Morales-Oñate, V., and Caamaño-Carrillo, C. (2022b). *GeoModels: Procedures for Gaussian and Non Gaussian Geostatistical (Large) Data Analysis*. R package version 1.0.2.

Cressie, N. (1993). *Statistics for spatial data*. Wiley, New York. MR1239641

Cressie, N. and Johannesson, G. (2008). Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society: Series B*, 70:209–226. MR2412639

Davis, G. J. and Morris, M. D. (1997). Six factors which affect the condition number of matrices associated with kriging. *Mathematical Geology*, 29:669–683.

Diamond, P. and Armstrong, M. (1984). Robustness of variograms and conditioning of kriging matrices. *Mathematical Geology*, 16:809–822. MR0774542

Fu, W. J. (1998). Penalized regressions: The bridge versus the lasso. *Journal of Computational and Graphical Statistics*, 7(3):397–416. MR1646710

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. MR2291261

Li, R. and Sudjianto, A. (2005). Analysis of computer experiments using penalized likelihood in gaussian kriging models. *Technometrics*, 47:111–120. MR2188073

Magnus, J. and Neudecker, H. (2007). *Matrix Differential Calculus with Applications in Statistics and Econometrics. Third Edition*. John Wiley, New York. MR1698873

Marzetta, T. L., Tucci, G. H., and Simon, S. H. (2011). A random matrix-theoretic approach to handling singular covariance estimates. *IEEE Transactions on Information Theory*, 57:6256–6271. MR2857971

Peng, C.-Y. and Wu, C. F. (2014). On the choice of nugget in kriging model-

ing for deterministic computer experiments. *Journal of Computational and Graphical Statistics*, 23:151–168. MR3173765

Penrose, R. (1955). A generalized inverse for matrices. *Mathematical Proceedings of the Cambridge Philosophical Society*, 51:406–413. MR0069793

Ranjan, P., Haynes, R., and Karsten, R. (2011). A computationally stable approach to gaussian process interpolation of deterministic computer simulation data. *Technometrics*, 53(4):366–378. MR2850469

Ribeiro Jr, P. J., Diggle, P. J., Schlather, M., Bivand, R., and Ripley, B. (2020). *geoR: Analysis of Geostatistical Data*. R package version 1.8-1. MR2293378

Salagame, R. R. and Barton, R. R. (1997). Factorial hypercube designs for spatial correlation regression. *Journal of Applied Statistics*, 24:453–473.

Smith, J. and Nicolik, M. (2013). Automatic methods for handling nearly singular covariance structures using the Cholesky decomposition of an indefinite matrix. arXiv:1201.1699.

Stefanski, L. and Boos, D. (2002). The calculus of m-estimation. *The American Statistician*, 56:29–38. MR1939394

Tucci, G. and Wang, K. (2019). New methods for handling singular sample covariance matrices. *IEEE Transactions on Information Theory*, 65:770–786. MR3904912

Vallejos, R. and Osorio, F. (2014). Effective sample size of spatial process models. *Spatial Statistics*, 9(3):66–92. MR3326832

Zimmermann, R. (2015a). An improved estimate for the condition number anomaly of univariate gaussian correlation matrices. *Electronic Journal of Linear Algebra*, 30:592–598. MR3414315

Zimmermann, R. (2015b). On the condition number anomaly of gaussian correlation matrices. *Linear Algebra and its Applications*, 466:512–526. MR3278266