PROPOSED RESEARCH:

In this section, you must include the following aspects:

a) Theoretical foundations and state of the art
b) Objectives and hypothesis or research questions
c) Scientific or technological novelty of your proposal
d) Methodology
e) Work plan, or Gantt chart

In accordance with point 5.5 of the guidelines that govern this contest, you must not include in this section information that identifies you, such as: names, initials, institution of membership or sponsor, links or other background.

The maximum length of this file is **10 pages** (please, use Verdana, or similar, 10 pt. font).

Aspects related to the proposed research that are included in annexes will not be considered in the evaluation process.

Citations to own works (self-citations) or previous works must be made in a neutral way, that is, they must not allow the applicant to be identified.

1 Introduction

Spatial Statistics, a branch of Statistics dedicated to the study of georeferenced data, is currently immersed in a wealth of applications, including agriculture, ecology, social sciences, medicine, fisheries, climatology and mining engineering, among many others (Cressie, 1993; Goovaerts, 1997; Kitanidis, 1997; Diggle and Ribeiro, 2007; Chilès and Delfiner, 2009; Gelfand et al., 2010; Banerjee et al., 2015; Plant, 2019; Oyana, 2021). In recent decades, substantial advances have been made in the development of computational tools that allow the collection of real-time data from a variety of sources (e.g., orbiting satellites), entailing several research challenges for statisticians. Random field models provide a helpful framework to capture the spatial uncertainty of the observations. In this context, the covariance function or variogram (the key ingredient for the simulation, estimation and prediction problems (Stein, 2012)) is typically modeled through a mapping that only depends on the spatial separation between observations. This principle obeys Tobler's First Law of Geography; "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970, 2004).

This project focuses on the effect that produces a strong spatial dependence on estimation and prediction, as well as on the quantification of redundant information within a sample of spatially correlated data. The general idea is that strongly correlated data lead to nearly singular correlation matrices (ill-condition problem) and provide duplicated information (effective sample size problem). On the one hand, the ill-conditioning problem and its consequences on estimation and prediction have been extensively discussed in the spatial literature, particularly related to kriging analysis (see Diamond and Armstrong (1984); Ababou et al. (1994); Salagame and Barton (1997); Davis and Morris (1997); Peng and Wu (2014)). Both estimation and prediction require the inverse of the covariance matrix, i.e. precision matrix. Lindgren et al. (2011); Simpson et al. (2012) proposed estimating the precision matrix using the relationship between Gaussian fields with the Matérn covariance function, Gaussian Markov random field and stochastic partial differential equations (SPDE), entailing a very sparse precision matrix; however, this sparsity decreases with regularity. We find applications of the SPDE approach to Gaussian fields in Bakka et al. (2019); Bolin and Wallin (2021); Moses et al. (2021); Lindgren et al. (2022). On the other hand, the effective sample size (ESS), initially introduced in spatial data by Cressie (1993), measures the actual information contained within a sample. Griffith et al. (2021) mentions ESS as "a complicating factor", indicating that their calculation is somewhat sensitive to the assumed spatial statistical model. In recent years, several alternative definitions of the ESS as well as exciting applications have been studied by Griffith (2005, 2008); Faes et al. (2009); Vallejos and Osorio (2014); Berger et al. (2014); Acosta et al. (2016); Li et al. (2016); Acosta and Vallejos (2018); Acosta et al. (2018); Chatterjee and Diaconis (2018); Bayarri et al. (2019); Chen et al. (2019); Neuenschwander et al. (2020); Egidi (2021); Vallejos and Acosta (2021).

To address the problems of ill conditions and ESS. First, we plan to investigate how to obtain a suitable approximation of the inverse of an ill-conditioned matrix, which is equivalent to approximating the precision matrix. For this purpose, we propose to use the Marzetta et al. (2011) and Tucci and Wang (2019) approximations (which in turn use the Haar and Ewens measures, respectively) of the correlation matrix and its inverse. We propose to use these approximations in an iterative parameter estimation algorithm, where the objective function is the penalized likelihood or quasi-likelihood. In addition, a nonparametric estimation of the variogram is analyzed; here, the ideas of Shapiro and Botha (1991) and Gorsich and Genton (2004) are relevant, and in this context, we propose incorporating a quadratic penalization in the minimum squares objective function. Second, intending to properly quantify the redundant information caused by the spatial dependency, we have noticed that previous

notions of the ESS can be unified into a single definition. This generalized definition allows one to analyze random fields with either constant or nonconstant means. Additionally, it can be used to prove, among others, the desirable property, which indicates that the ESS must belong to the interval $[1, n]$ (Vallejos and Osorio, 2014). In addition, a new application of ESS is proposed, which establishes the number of neighbors needed in a linear predictor, including a selection technique. Finally, the study of the ESS can be carried out without assuming a predefined distribution for the random field. As a result, a non-parametric ESS could be derived. Finally, several open problems are left open that can be addressed once all the objectives set out in this proposal are achieved.

2 Background: Theoretical foundations and state-of-the-art

A random field, denoted by $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in D \subset \mathbb{R}^d\}$ or $\{Y(\boldsymbol{s})\}$, is a collection of random variables indexed by D, where each $\boldsymbol{s} \in D$ is a location (Cressie, 1993). For a random field, the mean function $\mu : \mathbb{R}^d \to \mathbb{R}$ is given by $\mu(\boldsymbol{s}) = \mathbb{E}[Y(\boldsymbol{s})]$, whereas the covariance function $C : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is given by $C(\boldsymbol{s}_1, \boldsymbol{s}_2) = \operatorname{cov}(Y(\boldsymbol{s}_1), Y(\boldsymbol{s}_2))$ (Schabenberger and Gotway, 2005). $\{Y(\boldsymbol{s})\}$ is called a stationary random field if the mean function is constant, i.e., $\mu(\boldsymbol{s}) = \mu \in \mathbb{R}$, and if the covariance function only depends on the difference between the sites, $C(\boldsymbol{s}_1, \boldsymbol{s}_2) = C(\boldsymbol{s}_1 - \boldsymbol{s}_2, \boldsymbol{0}) = C_0(\boldsymbol{s}_1 - \boldsymbol{s}_2)$. With abuse of notation, the covariance function of a stationary process will be denoted as $C(\boldsymbol{h})$, where $\boldsymbol{h} = \boldsymbol{s}_1 - \boldsymbol{s}_2$. Thus, $\operatorname{var}(Y(\boldsymbol{s})) = C(\boldsymbol{0}) = \sigma^2 > 0$, and $\rho(\boldsymbol{h}) = C(\boldsymbol{h})/C(\boldsymbol{0})$ is the correlation function. $\{Y(\boldsymbol{s})\}$ is said to be an intrinsically stationary random field if $\{Y(\boldsymbol{s} + \boldsymbol{h}) - Y(\boldsymbol{s})\}$ is stationary. For intrinsically stationary random fields, the mapping $2\gamma(\boldsymbol{h}) = \operatorname{var}(Y(\boldsymbol{s} + \boldsymbol{h}) - Y(\boldsymbol{s}))$ is known as the variogram function, and $\gamma(\boldsymbol{h})$ is the semivariogram function. The random field will be said to be isotropic if the covariance function (or variogram) depends exclusively on the distance between the spatial locations and not on the direction. Again, with abuse of notation, we write $C(h)$ or $\gamma(h)$, where $h = \|\boldsymbol{h}\|$. For a stationary random field, the semivariogram and covariance function relationship is $\gamma(\boldsymbol{h}) = C(\boldsymbol{0}) - C(\boldsymbol{h})$. To estimate the second-order dependency structure of the random field, we can assume a specific admissible parametric model, or we can adopt an appropriate nonparametric framework.

In spatial data analysis, the random field is often assumed to be Gaussian (Rasmussen and Williams, 2006), that is, the finite-dimensional distributions of $\{Y(\boldsymbol{s})\}$ are Gaussian. Under the Gaussian assumption, a spatial regression model is defined as follows. Suppose that we observe $\{Y(\boldsymbol{s})\}$ at n spatial locations and $\boldsymbol{s} = (\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n)^\top \subset D^d$ be a set of mutually distinct sample points. Let $\boldsymbol{Y}(\boldsymbol{s}) = (Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n))^\top$, $\boldsymbol{\mu}(\boldsymbol{s}) = (\mu(\boldsymbol{s}_1), \ldots, \mu(\boldsymbol{s}_n))^\top$, and $\boldsymbol{e}(\boldsymbol{s}) = (e(\boldsymbol{s}_1), \ldots, e(\boldsymbol{s}_n))^\top$; then,

$$\boldsymbol{Y}(\boldsymbol{s}) = \boldsymbol{\mu}(\boldsymbol{s}) + \boldsymbol{e}(\boldsymbol{s}), \qquad \boldsymbol{e}(\boldsymbol{s}) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}), \qquad \boldsymbol{\Sigma} = \sigma^2 \boldsymbol{R}, \tag{1}$$

where $\boldsymbol{e}(\boldsymbol{s})$ is an error term with a zero mean and covariance matrix $\boldsymbol{\Sigma}$ (an $n \times n$ semi-positive definite matrix) with the (i, j)-th entry given by $\sigma_{ij} = C(\boldsymbol{s}_i - \boldsymbol{s}_j) = \sigma^2 \rho(\boldsymbol{s}_i - \boldsymbol{s}_j)$, $\boldsymbol{R}$ is a correlation matrix, $\mathcal{N}$ denotes the Gaussian distribution, and $(\cdot)^\top$ denotes the transpose operator. The spatial linear model in the classical setting (Mardia and Marshall, 1984) is defined as $\mu(\boldsymbol{s}) = \boldsymbol{f}(\boldsymbol{s})^\top \boldsymbol{\beta}$, where $\boldsymbol{f}(\boldsymbol{s}) = (f_1(\boldsymbol{s}), \ldots, f_p(\boldsymbol{s}))^\top$ is the regressor vector and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ is the vector of regression coefficients. Let the matrix $\boldsymbol{F}$ feature the regressor vectors $f(\boldsymbol{s}_i)^\top$ as rows and $\widehat{\boldsymbol{\beta}} = (\boldsymbol{F}^\top \boldsymbol{R}^{-1} \boldsymbol{F})^{-1} \boldsymbol{F}^\top \boldsymbol{R}^{-1} \boldsymbol{Y}$ be the generalized least-squares solution to the regression problem. In this context, the natural problem is the prediction in $\boldsymbol{s}_0 \in D$ an unobserved location, i.e., $\boldsymbol{s}_0 \neq \boldsymbol{s}_i$ for all $i = 1, \ldots, n$. Let $Y_0 = Y(\boldsymbol{s}_0)$ be the true, but unknown, value of random at site $\boldsymbol{s}_0$, and let $\widehat{Y}_0$ be a predictor of Y_0. $\boldsymbol{r}_0 = \rho(\boldsymbol{s}, s_0)$ denotes the correlation vector between $\boldsymbol{Y}$ and Y_0. Then, the best linear unbiased predictor (Journel and Huijbregts, 1978) is $\widehat{Y}_0 = \boldsymbol{f}(\boldsymbol{s})^\top \widehat{\boldsymbol{\beta}} + \boldsymbol{r}_0^\top \boldsymbol{R}^{-1} \left(\boldsymbol{Y} - \boldsymbol{F} \widehat{\boldsymbol{\beta}} \right)$.

The best linear predictor is a particular case of statistical tools typically known as **kriging** (Krige, 1951; Matheron, 1963), where it requires predicting $g(Y(\boldsymbol{s}_0))$ from a set of observed data for a given function $g(\cdot)$. These techniques, in addition to obtaining the punctual estimator, also obtain the kriging variance. For example, if $g(x) = x$ (identity function) and $\mu(\boldsymbol{s}) = \beta$ is an unknown constant (i.e., $f(\boldsymbol{s}) := 1$, $\boldsymbol{F} = \boldsymbol{1}$ be a vector of ones), we have obtained **ordinary kriging**, where $\widehat{Y}_0 = \widehat{\beta} + \boldsymbol{r}_0^\top \boldsymbol{R}^{-1} \left(\boldsymbol{Y} - \boldsymbol{1}\widehat{\beta} \right)$, and $\sigma^2_{ok}(\boldsymbol{s}_0) = \sigma^2 \left(1 - \boldsymbol{r}_0^\top \boldsymbol{R}^{-1} \boldsymbol{r}_0 + \left(\boldsymbol{1}^\top \boldsymbol{R}^{-1} \boldsymbol{1} \right)^{-1} \left(1 - \boldsymbol{1}^\top \boldsymbol{R}^{-1} \boldsymbol{r}_0 \right)^2 \right)$ is the kriging variance, with $\widehat{\beta} = (\boldsymbol{1}^\top \boldsymbol{R}^{-1} \boldsymbol{1})^{-1} \boldsymbol{1}^\top \boldsymbol{R}^{-1} \boldsymbol{Y}$ (Wakernagel, 2003; Schabenberger and Gotway, 2005; Chilès and Delfiner, 2009). The predictor function requires computing the inverse of the correlation matrix; consequently, kriging with large datasets has computational inconveniences. In practice, one has to restrict the data to a subset of neighbors (Stein, 2002; Emery, 2009; Madini, 2019). The justification for such an approach is the so-called screening effect (Stein, 2002). For linear predictors, when observations are on a regular grid, Stein (2002) proves that there generally is a screening effect as the grid becomes increasingly dense. The selection of the neighborhood about $\boldsymbol{s}_0$ has been addressed in Emery (2009); Kleijnen and van Beers (2020), although it is still subject to ongoing investigations. For a discussion of the evolution of kriging over time, see Chilès and Desassis (2018).

2.1 Estimation of covariance and variogram functions

The covariance (or variogram) estimation must be done carefully because the covariance function must be a positive semidefinite, and the variogram must be a conditionally negative definite function (Cressie, 1993).

Parametric Covariance Estimation: Suppose that the stationary covariance structure is modeled via a positive definite covariance function $C(\boldsymbol{\theta}, \boldsymbol{h}) = \sigma^2 \rho(\boldsymbol{\theta}, \boldsymbol{h})$, by convention parametrized such that for $\nu \in \mathbb{R}^d \setminus \{\mathbf{0}\}$, $\rho(\nu, \boldsymbol{\theta}) \to 1$ for $\|\boldsymbol{\theta}\| \to 0$, and $\rho(\nu, \boldsymbol{\theta}) \to 1$ for $\|\boldsymbol{\theta}\| \to \infty$, with coordinate-wise range-parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$, also referred to as the model's hyperparameters. The reciprocal values $1/\theta_k$ are called the correlation lengths (Zimmermann, 2015). Therefore, $\boldsymbol{R} := (\rho(\boldsymbol{\theta}, \boldsymbol{s}_i - \boldsymbol{s}_j)_{ij} \in \mathbb{R}^{n \times n}$ is the corresponding correlation matrix. The maximum likelihood estimation (MLE) method is generally the best option for estimating model parameters. Under the Gaussian assumption, the log likelihood is, up to an additive constant, $\ell(\boldsymbol{\theta}, \sigma^2, \boldsymbol{\beta}) = -0.5(\log|\sigma^2 \boldsymbol{R}| + \sigma^{-2}(\boldsymbol{Y} - \boldsymbol{F}\boldsymbol{\beta})^\top \boldsymbol{R}^{-1}(\boldsymbol{Y} - \boldsymbol{F}\boldsymbol{\beta}))$. Maximizing over $\boldsymbol{\beta}$ and σ^2 gives $\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}) = (\boldsymbol{F}^\top \boldsymbol{R}^{-1} \boldsymbol{F})^{-1} \boldsymbol{F}^\top \boldsymbol{R}^{-1} \boldsymbol{Y}$, and $\sigma^2(\boldsymbol{\theta}) = (\boldsymbol{Y} - \boldsymbol{F}\boldsymbol{\beta})^\top \boldsymbol{R}^{-1}(\boldsymbol{\theta})(\boldsymbol{Y} - \boldsymbol{F}\boldsymbol{\beta})/n$. Then, the profile log likelihood is

$$\ell_P(\boldsymbol{\theta}) = -0.5 \left(n \log(\sigma^2(\boldsymbol{\theta})) + \log|\boldsymbol{R}(\boldsymbol{\theta})| \right). \tag{2}$$

Hence, $\widehat{\boldsymbol{\theta}}_{\text{ML}} = \text{argmax}\{\ell_P(\boldsymbol{\theta})\}$. Similar to the linear predictor, the likelihood function also requires calculating the inverse of the correlation matrix. In addition, as is well known in spatial statistics (Warnes and Ripley, 1987), the log-likelihood function for the covariance parameters of a Gaussian random field (GRF) is nonconcave, which leads to numerical difficulties in solving the optimization problem (computing a global optimal solution) for MLE. One way of overcoming this drawback has been to use penalized likelihood functions (Li and Sudjianto, 2005; Chu et al., 2011; Tajbakhsh et al., 2020). Furthermore, each evaluation of the negative loglikelihood function requires $O(n^3)$ operations due to covariance matrix inversions (Tajbakhsh et al., 2020). Since n is typically large in GRF modeling, the computational issues due to large n are called the "big n problem" (Banerjee et al., 2008). This fact motivates the search for estimation methods with a good balance between computational complexity and statistical efficiency, such as quasi-likelihood methods, e.g., composited likelihood (CL) (Lindsay, 1988; Lele, 1997; Heagerty and Lele, 1998; Varin et al., 2011; Bevilacqua and Gaetan, 2015; Bachoc et al., 2019), including block likelihood (BL) (Caragea and Smith, 2007).

Nevertheless, the "big n problem" is not the only problem because when the hyperparameters approach zero, then for every admissible choice of parametrized correlation functions (Zimmermann, 2015), the correlation matrix converges to the singular matrix with every entry equal to 1, i.e., $\lim_{\|\boldsymbol{\theta}\| \to 0} \boldsymbol{R}(\boldsymbol{\theta}) = \mathbf{1}\mathbf{1}^\top$, and when $\|\boldsymbol{\theta}\|$ is close to 0, we have a nearly singular covariance matrix. These matrices are characterized by a large condition number which makes the estimation process unstable. The condition number $\kappa_2(\boldsymbol{R})$ may be expressed as the ratio between the maximum and the minimum eigenvalues. Hence, $\kappa_2(\boldsymbol{R}(\boldsymbol{\theta})) \to \infty$ when $\|\boldsymbol{\theta}\| \to 0$ (Zimmermann, 2015). Between the parametric correlation functions, it has been observed in several applications that the Gaussian correlation model is particularly prone to suffer from ill conditioning. The Gaussian model is the special case of $p = 2$ for the family of exponential correlation functions $\rho_p(\boldsymbol{h}; \boldsymbol{\theta}) = \exp\left\{ -\sum_{k=1}^d \theta_k |h_k|^p \right\}$, $1 \le p \le 2$, and as the limit case, when $\nu \to \infty$, in the Matérn correlation (Matern, 1986), which is given by $\rho_\nu(\boldsymbol{h}; \phi) = 2^{1-\nu}\Gamma(\nu)^{-1}(\phi\|\boldsymbol{h}\|)^\nu \mathcal{K}_\nu(\phi\|\boldsymbol{h}\|)$, where $\Gamma(\cdot)$ is the gamma function, $\mathcal{K}_\nu(\cdot)$ is the modified Bessel function of the second kind of order $\nu > 0$, and $\phi > 0$. For example, consider a regular grid of 2×2, and a single value θ. Then, if $\theta = 0.1$, we have that $\kappa_2(\boldsymbol{R})$ is 63, 103, and 401 when p is 1, 1.5, and 2, respectively. However, if $\theta = 0.01$, we have that $\kappa_2(\boldsymbol{R})$ is 677, 1230, and 40001 when p is 1, 1.5, and 2, respectively. Therefore, ill-conditioning does not only occur when n is large.

The exceptionally high conditioning of the Gaussian correlation model was termed the Gaussian anomaly in Kostinski and Koivunen (1999). A theoretical explanation of the Gaussian condition number anomaly is within (Theorem 1 and 2, Zimmermann, 2015). They prove that the conditioning of the Gaussian correlation matrix grows at least as fast as $1/\|\boldsymbol{\theta}\|^2$, while it grows generally only as fast as $1/\|\boldsymbol{\theta}\|$ for the remaining members of the exponential correlation family. In other contexts, Gaussian correlation matrices arise frequently as interpolation matrices in radial basis function (RBF) interpolation (Buhmann, 2003).

The problem of ill-conditioning related to kriging analysis has been extensively discussed in the literature (Diamond and Armstrong, 1984; Ababou et al., 1994; Salagame and Barton, 1997; Davis and Morris, 1997). The causes of kriging instability have been classified and discussed by Peng and Wu (2014). The authors mentioned three possible causes of singularity of the correlation matrix: (i) sample size, (ii) dimension of the input vectors, and (iii) parameter value. To overcome this problem, some alternatives have been developed. On the one hand, nugget regularization means replacing $\boldsymbol{R}$ with $\boldsymbol{R} + \delta \boldsymbol{I}$ for a small number $\delta > 0$ (Cressie, 1993). Nugget selection and its effects on the predictor and on the MLE process are investigated in Ranjan et al. (2011); Andrianakis and Challenor (2012); Won et al. (2013). On the other hand, methods such as covariance tapering (Furrer et al., 2006; Kaufman et al., 2008) and fixed rank kriging (Cressie and Johannesson, 2008), SPDE approach (Chilès and Desassis, 2018), among others have been proposed and used to overcome the ill-conditioned problem.

Outside the field of spatial statistics, Marzetta et al. (2011) use the idea of random dimension reduction to handle singular covariance matrices. In this framework, $\boldsymbol{K}$ is the sample covariance matrix, and $\text{rank}(\boldsymbol{K}) = N < M$,

which is a singular matrix. Let $\Omega_{L,M} = \{\boldsymbol{\Phi} : \boldsymbol{\Phi}\boldsymbol{\Phi}^* = \boldsymbol{I}_L\}$ 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 $\text{cov}_L(\boldsymbol{K}) = \mathbb{E}(\boldsymbol{\Phi}^*(\boldsymbol{\Phi}\boldsymbol{K}\boldsymbol{\Phi}^*)\boldsymbol{\Phi})$ and $\text{invcov}_L(\boldsymbol{K}) = \mathbb{E}(\boldsymbol{\Phi}^*(\boldsymbol{\Phi}\boldsymbol{K}\boldsymbol{\Phi}^*)^{-1}\boldsymbol{\Phi})$ as estimators of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}^{-1}$, respectively. In addition, they found that $\text{cov}_L(\boldsymbol{K})$ coincides with nugget regularization. However, $\text{invcov}_L(\boldsymbol{K})$ does not necessarily coincide with $\text{cov}_L(\boldsymbol{K})^{-1}$. Let $\boldsymbol{K} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^*$ (spectral decomposition of $\boldsymbol{K}$) with $\boldsymbol{D} = \text{diag}(d_1, \ldots, d_N, 0, \ldots, 0)$ the eigenvalues matrix and $\boldsymbol{U}$ the eigenvector matrix. Marzetta et al. (2011) show that $\text{invcov}_L(\boldsymbol{K}) = \boldsymbol{U}\,\text{invcov}_L(\boldsymbol{D})\,\boldsymbol{U}^*$, where $\text{invcov}_L(\boldsymbol{D}) = \text{diag}(\lambda_1, \ldots, \lambda_N, \mu, \ldots, \mu)$. The formulas to compute the values of λ_i and μ, joint with their asymptotic behavior using techniques from free probability, are found in Marzetta et al. (2011). Later, Tucci and Wang (2019) show that $\text{invcov}_L(\boldsymbol{D})$ has a surprisingly simple algebraic structure. it is a polynomial of the diagonal matrix $\boldsymbol{D}$. They also provide formulas to compute the coefficients of the polynomial. Tucci and Wang (2019) extended these ideas by replacing random unitary matrices with random permutation matrices and used 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 pseudo inverse of the $\boldsymbol{A}$ (Moore, 1920; Penrose, 1955). Additionally, Tucci and Wang (2019) provide an explicit formula for $K_{\vartheta,M,L}$ and an inductive formula to compute $\widetilde{K}_{\vartheta,M,L}$ and include a study of the asymptotic behavior for certain matrices with the mean conjugate estimator under the Ewens measure. Showing that, in some cases, the Ewens approach performs better than Haar.

NonParametric Variogram Estimation: Choosing a valid parametric semivariogram and fitting it to data ensures that the predicted variogram has the needed properties. However, one is restricted to a relatively small number of semivariogram models. Often, the empirical semivariogram appears erratic or wavy. What is termed the "nonparametric" approach to semivariogram modeling consists of choosing a family of semivariogram models that is sufficiently flexible to accommodate a wider range of shapes than the model parametrics (Schabenberger and Gotway, 2005).

Matheron (1965) proposed estimating the empirical variogram, $2\widehat{\gamma}(\boldsymbol{h})$, based on the empirical estimator of the second moment of $Y(\boldsymbol{s}_i) - Y(\boldsymbol{s}_j)$. Cressie and Hawkins (1980) proposed a robust variogram estimator based on the empirical estimator of the fourth moment of the transformation $|Y(\boldsymbol{s}_i) - Y(\boldsymbol{s}_j)|^{1/2}$. This estimator attenuates the effect of outliers compared to Matheron's estimator, but it is not sufficiently flexible because even a single outlier could still influence the estimation procedure. In this sense, Genton (1998) suggested, based on a highly robust estimator of scale $Q_{N(\boldsymbol{h})}$ (Rousseeuw and Croux, 1993), an estimator of the variogram as $2\widehat{\gamma}(\boldsymbol{h}) = \left(Q_{N(\boldsymbol{h})}\right)^2$. This variogram estimator is more robust than previous competitors. Unfortunately, these variogram estimators did not guarantee a conditionally negative definite function. Therefore, none of them is valid. Fortunately, for an isotropic random field, Shapiro and Botha (1991) proposed a general class of functions where a valid variogram is achievable. Consider $\gamma(h) = c_0 - f(h)$, where f is a positive function. From the Hankel Transform (Schoenberg , 1938), we have

$$f(r) = \int_0^\infty \Omega_d(tr)F(dt), \quad \text{where} \quad \Omega_d(x) = \left(\frac{2}{x}\right)^{(d-2)/2} \Gamma\left(\frac{d}{2}\right) J_{(d-2)/2}(x), \tag{3}$$

F is a bounded function, which is monotonically increasing on $(0, +\infty)$, $\Gamma(\cdot)$ stands for the gamma function, and $J_\nu(\cdot)$ is the Bessel function of the first kind of order ν. Let $\widehat{\gamma}_i = \widehat{\gamma}(h_i)$, $i = 1, \ldots, l$, be several estimates of the semivariogram at l different distances (any variogram introduced above). By discretizing (3) at m points ($m < l - 1$), or equivalently, by using a function F with m jumps at the points $t_1, \ldots, t_m$. The problem of finding $\gamma(h)$ is equal to finding $\boldsymbol{z} = (p_1, \ldots, p_m, c_0)^\top$, which minimizes the function

$$Q(\boldsymbol{z}) = \sum_{i=1}^{l} \omega_i \Big(\widehat{\gamma}_i - c_0 + \sum_{j=1}^{m} p_j \Omega_d(t_j h_i)\Big)^2, \quad \text{subject to} \quad p_j \geq 0,\ j = 1, \ldots, m, \quad \text{and} \quad c_0 - \sum_{j=1}^{m} p_j \geq 0. \tag{4}$$

These restrictions guarantee that $\widehat{\gamma}(h) = \widehat{c}_0 - \sum_{j=1}^m \widehat{p}_j \Omega_d(t_j h)$ be a conditionally negative definite function (Shapiro and Botha, 1991). Let $\boldsymbol{\gamma} = (\widehat{\gamma}_1, \ldots, \widehat{\gamma}_m)^\top$, $\boldsymbol{W} = \text{diag}(\omega_1, \ldots, \omega_m)$, and $\boldsymbol{A} = (a_{ij})$ with $a_{m+1,m+1} = c_0$, $a_{i,j} = -\Omega_d(t_j h_i)$. $i, j = 1, \ldots, m$, and $a_{i,m+1} = a_{m+1,j} = 0$ for $i, j = 1, \ldots, m$. Then, the matrix version is given by $Q(\boldsymbol{z}) = (\boldsymbol{\gamma} - \boldsymbol{A}\boldsymbol{z})^\top \boldsymbol{W}(\boldsymbol{\gamma} - \boldsymbol{A}\boldsymbol{z})$. Later, Gorsich and Genton (2004) provided a solution to the discontinuity problem of Shapiro and Botha (1991); although the methodology is similar, they estimated the covariance function instead of the variogram. Additionally, they propose an alternative to select the nodes in an optimal way based on the zeros of the Bessel-Fourier function.

2.2 The Effective Sample Size

Strong spatial dependence leads to ill-conditioning problems in the underlying correlation matrix. It can also affect statistical properties that are commonly used in practice. For example, suppose that $Y_i \sim \mathcal{N}(\mu_y, \sigma^2)$, $i = 1, \ldots, n$, and $\text{cov}(Y_i, Y_j) = \sigma^2\rho$, $(i \neq j)$, and assume that $\rho > 0$. Ignoring the fact that the Y_i are correlated, one might consider the sample mean $\overline{Y}$ as the "natural" estimator of μ_y. Thus, $\mathbb{E}[\overline{Y}] = \mu_y$, regardless of the

correlations. Some straightforward manipulations yield $\text{var}[\overline{Y}] = \sigma^2/n\{1+(n-1)\rho\}$; we note that $\text{var}[\overline{Y}] > \sigma^2/n$, and more importantly, $\lim_{n\to\infty} \text{var}[\overline{Y}] = \sigma^2\rho$. The sample mean is not a consistent estimator of the population mean μ_y. That is bad news. For other examples about the effect of correlation, see (Schabenberger and Gotway, 2005, Section 1.5). The effect of positive autocorrelation is that n correlated observations do not provide the same amount of information as n uncorrelated observations. (Cressie, 1993, p.15) approaches this problem by asking "How many samples of the uncorrelated kind provide the same precision as a sample of correlated observations?" If n denotes the number of correlated samples and n^* the number of uncorrelated samples, the effective sample size (ESS), for the case of the previous example, is calculated $\text{ESS} = n^* = n/(1+(n-1)\rho)$.

The problem of how uncorrelated samples provide the same precision as correlated observations is mentioned in several statistics books that describe the methodology for spatial modeling Haining (1990); Cressie (1993); Schabenberger and Gotway (2005); Lantuejoul (2010); Griffith and Paelinck (2011); Plant (2019); Vallejos et al. (2020). Furthermore, adequate calculation of the ESS is essential due to the possibility of duplicate information in a sample, which has an impact on subsequent modeling and inferences (Clifford et al., 1989; Dutilleul, 1993; Dutilleul et al., 2008). For example, Clifford et al. (1989) assessed the spatial association between two positively correlated processes through a modified t-test using an "effective sample size". Other applications of the ESS on spatial statistics are available in Cogley (1999); Madden and Hughes (1999); Griffith (2008); Dale and Fortin (2009); Griffith (2013); Acosta et al. (2016). The ESS also arises in the context of survey analysis by Kish (1965); although the motivation is totally different, their definition has a similar structure. In another example, the ESS, has been defined in a linear mixed model with replicas by Faes et al. (2009). The search for unambiguous definitions of the ESS in the fields of network modeling and point processes has also attracted recent scientific interest Krivitsky and Kolacyk (2015); Renner et al. (2021).

The first study of this concept, issued by Bayley and Hammersley (1946) in times series, denoted the ESS as the "Effective" number of independent observations. Therefore, the ESS is given by $\text{ESS} = n/\left(1+2\sum_{k=1}^{n-1}(1-k/n)\,\rho_k\right)$, where ρ_k is the lag-k serial correlation coefficient. If $\rho_k = \rho$, for all k, we also obtain $\text{ESS} = n/(1+(n-1)\rho)$; in particular, for $\rho_k = 0$ (uncorrelated observations) and $\rho_k = 1$ (perfect positive correlation), we obtain $\text{ESS} = n$, and $\text{ESS} = 1$, respectively. Later, Thiébaux and Zwiers (1984) mention that the ESS is not uniquely defined and that this has led to some confusion about its correct interpretation. In a time series context, they examine various methods of estimating and conclude that this quantity is quite difficult to estimate reliably. In addition, it reasoned that the concept could be extended to spatial data, and afterward was Cressie (1993), who defined the ESS in this context, exploring the variance of the sample average $\overline{Y}$. In fact, for noncorrelated observations, one has $\text{var}(\overline{Y}) = \sigma^2/n$. For dependent observations, instead, we have that $\text{var}(\overline{Y}) = (\sigma^2/n^2)\sum_{i=1}^{n}\sum_{j=1}^{n}\rho(\boldsymbol{h}_{ij})$, $\boldsymbol{h}_{ij} = \boldsymbol{s}_i - \boldsymbol{s}_j$. Assuming that $\text{var}(\overline{Y}) = \sigma^2/\text{ESS}$ (Cressie, 1993), then the ESS is given by

$$\text{ESS} = n^2 \Bigg/ \sum_{i=1}^{n}\sum_{j=1}^{n}\rho(\boldsymbol{h}_{ij}) \tag{5}$$

For instance, suppose an intra-class correlation structure, i.e. $\rho(\boldsymbol{h}_{ij}) = \rho$ is constant, then the definitions given by Bayley and Hammersley (1946) and Cressie (1993) agree. Subsequently, Yue and Wang (2004) analyzed the ability of ESS to eliminate the influence of serial correlation AR(1) on the nonparametric Mann-Kendall (MK) statistical test. However, in the presence of a linear trend, they failed to correct the type I error. Alternatively, Griffith (2005) based their quantification of the ESS on the variance inflation factor of the sample mean. The model used is $\boldsymbol{Y} = \mu\boldsymbol{1} + \boldsymbol{\epsilon},\ \boldsymbol{\epsilon} \sim \mathcal{N}\left(\boldsymbol{0}, \sigma_\epsilon^2\boldsymbol{V}\right)$, where $\boldsymbol{V}$ is a positive definite matrix but not necessarily a correlation matrix since the diagonal elements of $\boldsymbol{V}$ do not necessarily coincide; therefore, the process is not stationary. Thus, the ESS is given by

$$\text{ESS} = \frac{n\cdot\text{tr}(\boldsymbol{V})}{\boldsymbol{1}^\top\boldsymbol{V}\boldsymbol{1}}, \tag{6}$$

where $\text{tr}(\cdot)$ denotes the trace operator. Suppose again that $\boldsymbol{V}$ is a correlation matrix, then the formula (6) reduces to equation (5). Vallejos and Osorio (2014) consider the realization of a Gaussian random field as in (1), such that $\boldsymbol{\mu}(\boldsymbol{s}) = \mu\boldsymbol{1}$ and $\boldsymbol{\Sigma} = \boldsymbol{R}(\boldsymbol{\theta})$ (a nonsingular correlation matrix) are parameterized by the vector $\boldsymbol{\theta}$. Their definition of the ESS comes from the Fisher information about the mean μ, that is,

$$\text{ESS} = \boldsymbol{1}^\top\boldsymbol{R}^{-1}(\boldsymbol{\theta})\boldsymbol{1}. \tag{7}$$

If the sample is uncorrelated, then $\text{ESS} = n$. On the other hand, if a perfect positive spatial correlation prevails, then $\boldsymbol{R}(\boldsymbol{\theta}) = \boldsymbol{1}\boldsymbol{1}^\top$ is a singular matrix. In this case, Definition (7) is extended by considering the Moore-Penrose pseudoinverse of $\boldsymbol{R}(\boldsymbol{\theta})$; then, $\text{ESS} = 1$. Again, for an intraclass structure, the ESS (using Equation (7)) is $n/(1+(n-1)\rho)$. On the other hand, Acosta and Vallejos (2018) extended the definition in (7) to spatial regression processes with linear means. Assume model (1), with $\boldsymbol{\mu}(\boldsymbol{s}) = \boldsymbol{X}\boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is a p-dimensional vector of parameters, and $\boldsymbol{X}$ is a design full-rank matrix. They used the trace as a generalized variance of $\widehat{\boldsymbol{\beta}}$ to define the ESS. Acosta et al. (2021) used the block likelihood approach to define the ESS through the Godambe information matrix in a BigData context. The ESS has also been developed in Bayesian analysis (Berger et al., 2014; Bayarri et al., 2019), where this quantity is defined to generalize the BIC for non-iid samples.

3 Objectives and hypothesis

This research proposal is divided into two problems: The former focuses on the estimation of covariance functions and variograms in the presence of a nearly singular matrix and for large datasets, and the second is devoted to a unified definition of ESS, including a nonparametric version and applications at kriging.

3.1 Hypothesis

Proposal 1: On Estimation Methods.

(**H1**) It is possible to estimate the parameters of a covariance model in the presence of strong spatial dependence.

(**H2**) It is possible to construct a valid nonparametric variogram estimator that includes an overfitting regularizing function and is used in big data.

Proposal 2: On Effective Sample Size.

(**H3**) There is a general definition of the ESS that frames the different versions circulating in the literature.

(**H4**) The ESS may determine the size of the local neighborhood about s_0, a location where the prediction is required, when a linear predictor is used.

(**H5**) There is a definition of the distribution-free ESS and its nonparametric estimation.

3.2 Objectives

General Goal: Proposals 1 and 2 provide adequate tools for managing georeferenced data with strong spatial dependence, including estimation, prediction and correct definition of the rejection region in hypothesis testing.

We now describe the **Specific Goals** associated with each hypothesis:

Hyp.	Specif.	Description
(H1)	**Goal 1**	Adapting the methods proposed by Marzetta et al. (2011); Tucci and Wang (2019) for use the correlation matrix. Compare the different approximations of the correlation matrix when the dependency is strong using the Frobenius norm in a Monte-Carlo study.
	Goal 2	Proposed an algorithm of parameter estimation of the covariance function using the Haar and Ewens measure when the objective function is a penalized-likelihood or quasi-likelihood.
(H2)	**Goal 3**	Provide guidelines for selecting the penalty parameter. For this, we will use the Cross-validation method. Compare through a Monte-Carlo study the different methods of the non-parametric estimation of the variogram, and identify the conditions under which our proposed method is better than others.
(H3)	**Goal 4**	Deduce an expression of ESS that unified all definitions in the literature spatial and other areas. Under this expression, we found assumptions to ensure that the ESS belongs to $[1, n]$.
	Goal 5	Deduce the ESS using the composite-likelihood method and evaluated of asymptotic normality for the ESS found.
	Goal 6	For a linear model, we will use the determinant, as generalized variance, for deducing the ESS. Compare the different ESS in a selection scheme of the covariate.
(H4)	**Goal 7**	Deduce the ESS using the variance of a linear predictor (as ordinary kriging) and use this ESS to define the size of the local neighborhood of prediction. Besides, we proposed an algorithm for the selection of the local neighborhood and compared it with other selection methods in a cross-validation context.
(H5)	**Goal 8**	Define the non-parametric ESS and deduce the asymptotic distribution of the non-parametric ESS estimator, and determine a confidence interval for the ESS estimation.

BIBLIOGRAPHIC REFERENCES:

In this section, include the complete list of cited references in the Proposed Research section. **Maximum extension 5 pages**. (Must use letter size, Verdana size 10 or similar)

In accordance with point 5.5 of the guidelines that govern this contest, you must not include in this section information that identifies you, such as: names, initials, institution of membership or sponsor, links or other background.

Citations to own works (self-citations) or previous works must be made in a neutral way, that is, they must not allow the applicant to be identified.

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

Acosta, J., and Vallejos, R. (2018) Effective sample size for spatial regression models. *Electronic Journal of Statistics*, **12**, 3147-3180.

Acosta, J., Vallejos, R., and Griffith, D. (2018) On the effective geographic sample size. *Journal of Statistical Computation and Simulation*, **88**, 1958-1975.

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

Acosta, J., Alegria, A., Osorio, F., and Vallejos, R. (2021) Assessing the effective sample size for large spatial datasets: A block likelihood approach *Computational Statistics & Data Analysis*, **162**, October 2021, 107282.

Andrianakis I., and Challenor, P.G. (2012) The effect of the nugget on Gaussian process emulators of computer models. *Comput. Statist. Data Anal.*, **56**, 4215–4228.

Bachoc, F., Bevilacqua, M. and Velandia, D. (2019) Composite likelihood estimation for a Gaussian process under fixed domain asymptotics *Journal of Multivariate Analysis*, **174**, online.

Bakka H., Vanhatalo J., Illian J.B., Simpson D., Rue H. (2019) Non-stationary Gaussian models with physical barriers *Spatial Statistics*, **29**, 268–288

Banerjee, S., Gelfand, A., Finely, A., and Sang, H. (2008) Gaussian predictive process models for large spatial data sets. *Journal of Royal Statistical Society, Series B*, **70**, 825–848.

Banerjee S., Carlin B., and Gelfand A. (2015) *Hierarchical Modeling and Analysis for Spatial Data*, Boca Raton: Chapman Hall/CRC.

Bayarri, M. J., Berger, J. O., Jang, W., Ray, S., Pericchi, L. R., and Visser, I. (2019) Prior-based Bayesian information criterion *Statistical Theory and Related Fields*, **3**(1), 2–13.

Bayley, G. V., and Hammersley, J. M. (1946) The "Effective" Number of Independent Observations in an Autocorrelated Time Series *Journal of the Royal Statistical Society*, **8**(2), 184–197.

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

Berger, J., Bayarri, M. J., and Pericchi, L. R. (2014) The effective sample size. *Econometric Letters*, **33**, 197–217.

Bolin D., and Wallin J. (2021) Efficient methods for Gaussian Markov random fields under constraints *Advances In Neural Information Processing Systems*, **34**, Curran Associates, 1–13.

Buhmann, M. D. (2003) *Radial Basis Functions*. Cambridge Monographs on Applied and Computational Mathematics, vol. 12, Cambridge University Press, Cambridge, UK.

Caragea, P. C., Smith, R. L., (2007) Asymptotic properties of computationally efficient alternative estimators for a class of multivariate normal models. *Journal of Multivariate Analysis*, **98**(7), 1417–1440.

Chatterjee, S. and Diaconis, P.(2018) The sample size required in importance sampling *The Annals of Applied Probability*, **28**, 1099–1135.

Chen, F., Crow, W.T., Cosh, M.H., Colliander, A., Asanuma, J., Berg, A., Bosch, D.D., Caldwell, T., Holifield, C., Høgh, K., Martínez-Fernández, J., Mcnairn, H., Starks, P.J., Su, Z., And Walker, J.P. (2019) Uncertainty of Reference Pixel Soil Moisture Averages Sampled at SMAP Core Validation Sites. *American Meteorological Society*. Published online, DOI: 10.1175/JHM-D-19-0049.1

Chilès, J.P. and Delfiner, P.(2009) *Geostatistics: Modeling Spatial Uncertainty*. John Wiley & Sons.

Chilès, J.-P., and Desassis, N. (2018) *Fifty years of Kriging. In B. S. Daya Sagar, Q. Cheng, & F. Agterberg (Eds.), Handbook of mathematical geosciences: Fifty years of IAMG*. Cham, Switzerland: Springer International Publishing AG.

Chu, T., Zhu, J., and Wang, H. (2011) Penalized maximum likelihood estimation and variable selection in geostatistics. *The Annals of Statistics*, **39**(5), 2607–2625.

Clifford, P., Richardson, S. and Hémon, D. (1989) Assessing the significance of the correlation between two spatial processes. *Biometrics*, **45**, 123–134.

Cogley, J. G. (1999) Effective sample size for glacier mass balance. *Geografiska Annaler*, **81 A**, 497–507.

Cressie, N. (1993) *Statistics for Spatial Data*. New York: Wiley.

Cressie, N., and Hawkins, M. (1980) Robust estimation of the variogram I. *Mathematical Geology*, **12**, 115–125.

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

Dale, M. R. T., and Fortin, M-J. (2009) Spatial autocorrelation and statistical tests: some solutions. *Journal of Agricultural, Biological, and Environmental Statistics*, **14**, 188–206.

Davis, G. J., and Morris, M. D. (1997) Six factors which affect the condition numberof 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.

Diggle. P.J., Ribeiro, P.J., (2007). *Model-based Geostatistics*. Springer, New York, NY

Dutilleul, P. (1993) Modifying the t test for assessing the correlation between two spatial processes. *Biometrics*, **49**, 305–314.

Duttilleul, P., Pelletier, B., and Alpargu, G. (2008) Modified F tests for assessing the multiple correlation between one spatial process and several others. *Journal of Statistical Planning and Inference*, **138**, 1402–1415

Egidi, L. (2022) Effective sample size for a mixture prior *Statistics & Probability Letters*, 183:109335.

Emery, X. (2009) The kriging update equations and their application to the selection of neighboring data *Computational Geosciences*, **13**, 269–280.

Faes, C., Molenberghs, G., Aerts, M., Verbeke, G., and Kenward M. (2009) The effective sample size and an alternative small-sample degrees-of-freedom method. *The American Statistician*, **63**, 389–399.

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.

Gelfand, A. E., Diggle, P. J., Fuentes, M., and Guttorp, P. (2010) *Handbook of Spatial Statistics*, Boca Raton, Fl: Chapman & Hall/CRC.

Genton, M., (1998). Highly robust variogram estimation. *Mathematical Geology*, **30**, 213–221.

Godambi, V. P., Kale, B. K., (1991) *Estimating functions: an overview*. In V.P. Godambe (Ed.) Estimating Functions, Claredon Press: Oxford.

Goovaerts, P. (1997). *Geostatistics for natural resource evaluation* Oxford University Press

Gorsich, D., Genton, M. (2004). On the discretization of nonparametric isotropic covariogram estimators. *Statistics and Computing*, **14**, 99–108.

Griffith D. (2005) Effective geographic sample size in the presence of spatial autorrelation. *Ann. Assoc. Amer. Geogr.*, **95**, 740–760.

Griffith D. (2008) Geographic sampling of urban soils for contaminant mapping: how many samples and from where. *Environ Geochem Health*, **30**, 495–509.

Griffith, D., and Paelinck, J. H. P. (2011) *Non-Stardard Spatial Statistics*. New York: Springer.

Griffith, D. A. (2013) Establishing qualitative geographic sample size in the presence of spatial autocorrelation. *Journal Annals of the Association of American Geographers* **103**(5) 1107–1122.

Griffith, D. A., Chun, Y., and Lee, M. (2021) Deeper Spatial Statistical Insights into Small Geographic Area Data Uncertainty. *International Journal of Environmental Research and Public Health* **18**(1) 231; https://doi.org/10.3390/ijerph18010231

Haining, R. (1990) *Spatial Data Analysis in the Social Environmental Sciences*. Cambridge: Cambridge University Press.

Heagerty, P. and Lele, S. (1998) A composite likelihood approach to binary spatial data. *Journal of the Amerian Statistical Association* , **93**, 1099–1111.

Huang, C. Hsing, T,. Cressie, N. (2011). Nonparametric estimation of the variogram and its spectrum. *Biometrika*, **98**, 775–789.

Johnson, R. A., Wichern, D. W.(1992). *Applied Multivariate Statistical Analysis*. Prentice Hall, Englewood Cliffs, New Jersey.

Journel, A., Huijbregts, C. (1978) *Minig Geostatistics*. Academic Press: New York

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

Kish, L., (1965). *Survey Sampling*. John Wiley and Sons: New York

Kitanidis, P. K. (1997) *Introduction to Geostatistics: Applications to Hydrogeology*. Cambridge University Press, UK.

Kleijnen, J., and van Beers, W. (2019) Prediction for Big Data Through Kriging: Small Sequential and One-Shot Designs *American Journal of Mathematical and Management Sciences*, **39**, 199–213.

Kostinski, A.B., and Koivunen, A.C. (1999) The feasibility of data whitening to improve performance of weather radar *J. Appl. Math.*, **38**, 741–749.

Krige, D.G. (1951) A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of Chemical, Metallurgical, and Mining Society of South Africa*, **52**, 119–139.

Krivitsky, P. V., Kolaczyk, E. D. (2015). On the question of effective sample size in network modeling: An asymptotic inquiry. *Statistical Science* 30, 184–198.

Lantuejoul, C. (2010). *Geostatistical Simulation Models and Algorithms*. Springer.

Lele S. R. (1997) "Estimating functions for semivariogram estimation," in selected proceedings of the conference on Estimating Functions, eds. I. V. Basawa, V. P. Godambe, and R. L. Taylor, Haywood, CA: Institute of Mathematical Statistics, pp. 381–396.

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

Li, B., Griffith, D. A., and Becker, B. (2016) Spatially simplified scatterplots for large raster datasets. Geo-spatial Information Science, 19(2), 81-93.

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

Lindgren, F., Bolin, D., and Rue, H. (2022) The SPDE approach for Gaussian and non-Gaussian fields: 10 years and still running *Spatial Statistics*, In Press, doi: https://doi.org/10.1016/j.spasta.2022.100599

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

Madden, L. V., Hughes, G. (1999). An effective sample size for predicting plant disease incidence in a spatial hierarchy. *Analytical and Theoretical Plant Pathology*, **89**, 770–781.

Madini, N., and Emery, X. (2019) A comparison of search strategies to design the cokriging neighborhood for predicting coregionalized variables *Stochastic Environmental Research and Risk Assessment*, **33**, 183–199.

Mardia, K., and Marshall, R. (1984) Maximum likelihood of models for residual covariance in spatial regression. *Biometrika*, **71**, 135–146.

Martino, L., Elvira, V., and Louzada, F. (2017) Effective sample size for importance sampling based on discrepancy measures *Signal Processing*, 131:386–401.

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

Matérn, B., (1986) *Spatial Variation*, 2nd ed. Lecture Notes in Statistics, SpringerVerlag, New York..

Matheron, G., (1963) Principles of geostatistics. *Economic Geology*, **58**, 1246–1266.

Matheron, G., (1965) *Les variables régionalisés et leur estimation*. Masson, Paris.

Moore, E. H. (1920) On the Reciprocal of the General Algebraic Matrix. *Bulletin of American Mathematical Society*, **26**, 394–395.

Morales-Oñate, V., Crudu, F., and Bevilacqua M. (2021) Blockwise Euclidean likelihood for spatio-temporal covariance models. *Econometrics and Statistics*. Available online https://doi.org/10.1016/j.ecosta.2021.01.001.

Moses J., Aheto K., and Dagne G.A. (2021) Geostatistical analysis, web-based mapping, and environmental determinants of under-5 stunting: evidence from the 2014 Ghana Demographic and Health Survey *The Lancet Planetary Health*, **5**(6), e347–e355.

Neuenschwander, B, Weber, S, Schmidli, H, O'Hagan, A. Predictively consistent prior effective sample sizes. *Biometrics*, **76**, 578–587. https://doi.org/10.1111/biom.13252

Oyana, T.J. (2021) *Spatial Analysis with R Statistics, Visualization, and Computational Methods*. CRC Press; 2nd edición

Peng, C.Y., and Wu, C. F. (2014) On the choice of nugget in kriging modeling for deterministic computer experiments. *Journal of Computational and Graphical Statistics*, **23**, 151–168.

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

Pino-Cortés, E., Díaz-Robles, L., Cubillos, F., Cereceda-Balic, F., Santander, R., Fu, J., Carrasco, S., and Acosta, J. (2021) The black carbon dispersion in the Southern Hemisphere and its transport and fate to Antarctica, an Anthropocene evidence for climate change policies. *Science of The Total Environment*, **778**, 146242, https://doi.org/10.1016/j.scitotenv.2021.146242

Plant, R.E. (2019). *Spatial Data Analysis in Ecology and Agriculture Using R*. CRC Press Taylor & Francis Group: Boca Raton.

Pottchoff, R., Woodbury, M. and Manton K. (1992). "Equivalent sample size" and "equivalent degrees of freedom" refinements for inference using survey weights under superpopulation models. *Journal of the American Statistical Association* **87** 383–389.

R Core Team(2021). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. https://www.R-project.org/.

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

Rasmussen, C. E., and William, C. K. I. (2006), *Gaussian Processes for Machine Learning*, The MIT Press, Massachusetts.

Reimherr, M., Meng, X.-L., and Nicolae, D. L. (2021) Prior sample size extensions for assessing prior impact and prior-likelihood discordance *Journal of the Royal Statistical Society: Series B*, **83**(3), 413–437.

Renner, I. W., Warton, D. I., and Hui, F. K. (2021) What is the effective sample size of a spatial point process? *Australian & New Zealand Journal of Statistics*, **63**(1), 144–158.

Ribeiro, P. J., Diggle, P. J. (2020) geoR: Analysis of Geostatistical Data. R package version 1.8-1. https://CRAN.R-project.org/package=geoR

Rockafellar, R. T. (1970) *Convex Analysis*. Princeton: University Press

Rousseeuw, P. J., Croux, C., (1993). Alternatives of the median absolute deviation. *Journal of American Statistics Association*, **88**, 1273–1283.

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

Shabenberger, O., and Gotway, C. A. (2005) *Statistical Methods for Spatial Data Analysis*. Boca Raton, FL: Chapman & Hall/CRC.

Schoenberg (1938). Metric Spaces and Positive Definite Functions *American Mathematical Society*, **44**, 522–536.

Shapiro, A., and Botha, J.D. (1991) Variogram fitting with a general class of conditionally non-negative definite functions. *Computational Statistics and Data Analysis*, **11**, 87–96.

Simpson, D., Lindgren, F., and Rue, H. (2012) Think continuous: Markovian Gaussian models in spatial statistics *Spatial Statistics*, **1**, 16–29.

Stein, M. L. (2002) The Screening effect in kriging *The Annals of Statistics*, **30** (1), 298–323.

Stein, M. L. (2011) When does the screening effect hold? *The Annals of Statistics*, **39** (6), 2795–2819.

Stein, M. L. (2012) *Interpolation of Spatial Data: Some Theory for Kriging*, Springer Science & Business Media

Tajbakhsh, S.D, Aybat, N.S., and Del Castillo, E. (2020) On the Theoretical Guarantees for Parameter Estimation of Gaussian Random Field Models: A Sparse Precision Matrix Approach *Journal of Machine Learning Research*, **21**, 1–41.

Thiébaux, H. J, and Zwiers, F. W. (1984) Interpretation and estimation of effective sample size. *Journal of Climate and Applied Meteorology*, **23**, 800–811.

Tobler, W. R (1970) A computer movie simulanting urban growth in the Detroit region. *Economic Geography*, **46**, 234–240.

Tobler, W. R (2004) On the First Law of Geography: A Reply. *Annals of the Association of American Geographers*, **94**(2), 304–310.

Tucci, G. and Wang, K. (2019) New methods for handling singular sample covariance matrices. *Journal of Climate and Applied Meteorology*, **65**, 770–786.

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

Vallejos, R., Osorio, F., and Bevilacqua, M. (2020) *Spatial Relationships Between Two Georeferenced Variables: With Applications in R*. Springer.

Vallejos, R., and Acosta, J. (2021) The effective sample size for multivariate spatial processes with an application to soil contamination. *Resource Modeling* **34**, e12322

Varin, C., Reid, N. and Firth, D. (2011) An Overview of Composite Likelihood Methods. *Statistica Sinica*, **21**, 5–42.

Wakernagel, H. (2003). *Multivariate Geostatistics: An Introduction with Applications*. Springer: New York

Warnes, J. J., and Ripley, B. D. (1987) Problems with Likelihood Estimation of Covariance Functions of Spatial Gaussian Processes *Biometrika*, **74**, 640–642.

Won, J.-H. Lim, J., Kim, S.-J., and Rajaratnam, B. (2013) The effect of the nugget on Gaussian process emulators of computer models. *Journal of the Royal Statistical Society: Series B*, **75**, 427–450.

Yue, S., and Wang, C. (2004) The Mann-Kendall Test Modified by Effective Sample Size to Detect Trend in Serially Correlated Hydrological Series *Water Resources Management*, **18**, 201–218.

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