Effective Sample Size of Spatial Process Models

Ronny Vallejos*, Felipe Osorio

Departamento de Matemática, Universidad Técnica Federico Santa María, Valparaíso, Chile.

Abstract

This paper focuses on the reduction of sample sizes due to the effect of autocorrelation for the most common models used in spatial statistics. This work is an extension of a simple illustration highlighted in several books for an autoregressive-type correlation structure. The paper briefly reviews existing proposals to quantify the effective sample size and proposes a new definition that is a function of the correlation structure, sample size, and dimension of the space where the coordinates are defined. It describes the properties of and explicit expression for the effective sample size for processes with patterned correlation matrices, including elliptical contoured distributions. The estimation of the effective sample size is achieved using restricted maximum likelihood. Additionally, the paper describes the monotonicity of the effective sample size when two random points are uniformly distributed on the unit sphere and includes several Monte Carlo simulations to explore monotonic features of the effective sample size and to compare its behavior with respect to other proposals. Finally, this paper analyzes two real datasets, and the discussion includes topics that should be addressed in further research.

Key words: Spatial process, Effective sample size, Elliptically contoured distributions, REML estimator

1. Introduction

Spatial analysis has developed considerably in recent decades. Particularly, problems such as determining sample sizes and how and where to sample have been studied in many different contexts. In spatial statistics, it is well known that as the spatial autocorrelation latent in georeferenced data increases, the amount of duplicated information in these data also increases. This property has many implications for the subsequent analysis of spatial data. A similar problem has been discussed by Box (1954a, 1954b) for approximating the distribution of a quadratic form in the normal random vector by a chi-square distribution that has the same first two moments. Clifford et al. (1989) used this approach to suggest effective degrees of freedom in a modified t-test designed to assess the association between two spatial processes; a detailed discussion can be found in Dutilleul (1993). An extension,

*Corresponding author

Email address: ronny.vallejos@usm.cl (Ronny Vallejos)

[1]Address for correspondence: Departamento de Matemática, Universidad Técnica Federico Santa María, Avenida España 1680, Casilla 110-V, Valparaíso, Chile.

in the context of multiple correlation, involving one spatial process and several others is described in Dutilleul et al. (2008). A study on the effective number of spatial degrees of freedom of a time-varying field was conducted by Bretherton et al. (1999). The effective sample size of two glaciers determined by analyzing the spatial correlation between point mass balance measurements is discussed in Cogley (1999). The relative information content of the mean for various hydrologic series is addressed in Matalas and Langbein (1962). The method required to determine the number of independent observations in an autocorrelated time series is noted by Bayley and Hammersley (1946).

The effect of spatial correlation on statistical inference, more specifically the problem of how many uncorrelated samples provide the same precision as correlated observations, is mentioned and illustrated in classical spatial statistics books such as Cressie (1993), Haining (1990), and Schabenberger and Gotway (2005). Griffith (2005) developed a new method to determine the effective sample size for normally distributed georeferenced data for a single mean and also provided extensions for multiple sample means. Griffith's proposal is based on a regression model for which the expected value of the estimated variance of the response variable is calculated. Using this expression and the variance inflation factor, an effective sample size formula is obtained as a function of the covariance structure of the model. Other model-based alternatives and extensions to two processes are also provided. Griffith (2008) later used this method with soil samples from Syracuse, NY. Another approach based on the integral range for which the estimation process is philosophically different (i.e., it is not based on the likelihood) from that mentioned above is described in Lantuejoul (1991).

This paper addresses the following problem: if we have n data points located on a general grid in an r-dimensional space, what is the effective sample size (ESS) associated with these points? If the observations are independent and if a regional mean is being estimated, then, given a suitable definition, the answer is n. Intuitively, when perfect positive spatial autocorrelation prevails, ESS should be equal to 1; with dependence, less than n. Getis and Ord (2000) studied this type of reduction of information in the context of multiple testing of local indices of spatial autocorrelation. Note that the general approach to addressing this question does not depend on the data values; however, it does depend on the spatial locations of the points in the range of the spatial process and on the spatial dimension. We suggest a definition of the spatial effective sample size based on an alternative way of calculating the reduction of information due to the existing spatial association in the data. Our definition can be explored analytically given certain assumptions. We explore certain patterned correlation matrices that commonly arise in spatial statistics, study the effective sample size for a single normal process and extended it to CAR and SAR processes. Additionally, we consider a single mean process with errors that have an elliptically contoured distribution. We present theoretical results and examples to illustrate the features of our proposed method.

Estimation of the effective sample size is addressed via the restricted maximum likelihood estimator for the normal and elliptical cases. We discuss both how sampling design affects effective sample size and how to select a sample after sample size has been reduced to account for autocorrelation, highlighting that effective sample size is intimately related to the way in which data are collected. Additionally, we carried out numerical experiments and Monte Carlo simulations to measure the performance of the estimators of effective

sample size and determined the effect of the selected variogram model. We use two examples with real data to illustrate the practical scope of our proposal. The first dataset consists of georeferenced samples from a contaminated area in Utah, USA, and the second dataset consists of forest variables based on a study of Pinus radiata plantations in southern Chile. In both cases, we calculate the effective sample size to explore the reduction of the sample size.

This paper is organized as follows. In Section 2, a motivation is given in the context of a time series previously discussed by Cressie (1993) and an extension to a spatial process is considered to illustrate the effect of dimension. Section 3 develops the effective sample size for normal, CAR, SAR, and elliptical models, including several illustrative examples. Section 4 describes the REML estimation approach for normal and elliptical spatial processes. Through a numerical example, the effect of sampling design on effective sample size is briefly addressed in Section 5. An analysis supports the fact that when two points are randomly distributed on a unit sphere in an r-dimensional space, the effective sample size increases, as described in Section 6. Two simulation studies explore the performance of our proposed method in Section 7. Effective sample size is calculated for two real datasets in Section 8. Additionally, conventional sampling schemes to obtain samples after determining effective sample size are outlined in Section 9. Finally, Section 10 is a discussion which includes problems to study in future research. The derivations for the results presented in this paper are relegated to the Appendix.

2. Motivation

Cressie (1993), p. 14-15 illustrated the effect of spatial correlation on the variance of the sample mean using an AR(1) correlation structure with spatial data. As a result, the new sample size (effective sample size) could be interpreted as an equivalent number of independent observations. Precisely, consider $Y(t) = \rho Y(t-1) + \epsilon(t)$, where $t = 1, 2, \ldots, n$, ϵ_t is white noise with variance σ^2, and $|\rho| < 1$. Then, if $1 \leq t, t+h \leq n$, $\text{cor}[Y(t), Y(t+h)] = \rho^h$ and the effective sample size is given by

$$\text{ESS} = \frac{n}{[1 + 2\rho/(1-\rho)(1-1/n) - 2(\rho/(1-\rho)^2)(1-\rho^{n-1})/n]}. \tag{1}$$

Formula (1) is associated with the variance of the sample mean $\text{var}(\overline{Y}) = \sigma^2/\text{ESS}$, which is equal to σ^2/n under independence. The AR(1) process is defined over an equispaced one dimensional set. A similar formula can be obtained for a two dimensional index set. In fact, for an index set of the form $D = \{(i,j) \in \mathbb{Z}^2 : 0 \leq i \leq m, 0 \leq j \leq n\}$, let us consider a random field $\{Y(i,j) : (i,j) \in D\}$ with $\text{var}[Y(i,j)] = \sigma^2$ such that for $(i,j), (k,l) \in D$,

$$\text{cor}[Y(i,j), Y(k,l)] = \rho^{\sqrt{(i-k)^2+(j-l)^2}}.$$

Then, defining

$$\begin{aligned}
S(\rho,m,n) &= \sum_{\substack{(i,j),(k,l)\in D\\ i>k,j>l}} \text{cor}[Y(i,j),Y(k,l)] \\
&= \sum_{\substack{(i,j),(k,l)\in D\\ i>k,j>l}} \rho^{\sqrt{(i-k)^2+(j-l)^2}} \\
&= (m+1)\frac{\rho}{1-\rho}\left(n-\rho\frac{1-\rho^n}{1-\rho}\right)+(n+1)\frac{\rho}{1-\rho}\left(m-\rho\frac{1-\rho^m}{1-\rho}\right) \\
&+ 2\sum_{i=1}^{n}\sum_{j=1}^{m}(m+1-i)(n+1-j)\rho^{\sqrt{i^2+j^2}},
\end{aligned}$$

one has

$$\begin{aligned}
\text{var}[\overline{Y}] &= \frac{1}{(m+1)^2(n+1)^2}\text{var}\left[\sum_{(i,k)\in D} Y(i,j)\right] \\
&= \frac{\sigma^2}{(m+1)^2(n+1)^2}\left[(m+1)(n+1)+2S(\rho,m,n)\right] \\
&= \frac{\sigma^2}{(m+1)(n+1)}\left[1+\frac{2S(\rho,m,n)}{(m+1)(n+1)}\right].
\end{aligned}$$

Therefore, the effective sample size is given by

$$\text{ESS} = \frac{(m+1)(n+1)}{\left[1+\frac{2S(\rho,m,n)}{(m+1)(n+1)}\right]}. \tag{2}$$

We recall that (2) is only valid if the locations are arranged on a rectangular grid D. Here, we define a less restrictive setting to address data points located on a general grid on the plane.

3. Effective Sample Size

In this section, we provide expressions to quantify the effective sample size for a wide class of spatial models. We start with a single mean spatial regression model under normality. We then extend the notion of the effective sample size to the case of two or more variables. Because we developed this concept for CAR and SAR processes, this section also considers a similar idea for elliptical distributions.

3.1. The Case of a Single Mean

Consider the random field $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in D \subset \mathbb{R}^r\}$. We suppose that $Y(\cdot)$ has been observed at each of n distinct locations in D. Denote the n-vector of values $Y(\cdot)$ observed at the data locations $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_n$ as $\boldsymbol{Y} = [Y(\boldsymbol{s}_1),\, Y(\boldsymbol{s}_2), \ldots, Y(\boldsymbol{s}_n)]^\top$. Assume that $\mathbb{E}[\boldsymbol{Y}] = \mu\mathbf{1}$, where $\mu \in \mathbb{R}$ and $\mathbf{1}$ is an $n \times 1$ vector of ones. Let $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ denote the

$n \times n$ covariance matrix of $\boldsymbol{Y}$ where $\boldsymbol{\theta} \in \Theta$ is a k-vector of unknown parameters and $\Theta \subset \mathbb{R}^k$. For multivariate distributions with finite second moments, effective sample size can be characterized by the correlation matrix $\boldsymbol{R}(\boldsymbol{\theta}) = (\sigma_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}}) = \boldsymbol{C}^{-1}\boldsymbol{\Sigma}(\boldsymbol{\theta})\boldsymbol{C}^{-1}$, where $\boldsymbol{C} = \text{diag}(\sigma_{11}^{1/2}, \sigma_{22}^{1/2}, \cdots, \sigma_{nn}^{1/2})$, whose (i,j)th element is $\rho(\boldsymbol{s}_i, \boldsymbol{s}_j, \boldsymbol{\theta})$. We emphasize the dependence of the spatial correlation matrix on the data locations. For the set $\mathcal{C} = \{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_n\}$ the goal is to suggest a function

$$\text{ESS} = \text{ESS}(n, \boldsymbol{R}(\boldsymbol{\theta}, \mathcal{C}), r)$$

that satisfies the following desirable properties: i) ESS $= n$, when there is no spatial correlation; ii) ESS $= 1$, for a perfect correlation between all possible pairs of variables; iii) ESS is increasing in n; and iv) ESS decreases as the correlation increases. For simplicity, the notation for the covariance and correlation matrices will be $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ and $\boldsymbol{R}(\boldsymbol{\theta})$. The dependency of effective sample size on $\mathcal{C}$ will be discussed in Section 5. Because in many different cases, effective sample size lies in the interval $[1, n]$, it can be observed as a reduction of the information due to the spatial association present in the data. There are many possible reductions of $\boldsymbol{R}(\boldsymbol{\theta})$ to a single number and many appropriate but arbitrary transformations of that number to the interval $[1, n]$. For the case in which $\boldsymbol{C} = \boldsymbol{I}$, a reduction is provided by the Kullback-Leibler distance from $\mathcal{N}(\mu\mathbf{1}, \boldsymbol{R}(\boldsymbol{\theta}))$ to $\mathcal{N}(\mu\mathbf{1}, \boldsymbol{I})$. Straightforward calculations indicate that $\text{KL} = \frac{1}{2}\left(\log|\boldsymbol{R}(\boldsymbol{\theta})| + \text{tr}(\boldsymbol{R}(\boldsymbol{\theta})^{-1} - \boldsymbol{I})\right)$. For an isotropic spatial process with spatial variance σ^2 and an exponential correlation function $\rho(\boldsymbol{s}_i - \boldsymbol{s}_j) = \exp(-\phi||\boldsymbol{s}_i - \boldsymbol{s}_j||)$, $\phi > 0$, KL needs to be inversely scaled to $[1, n]$ and decrease in ϕ. Another way to avoid making an arbitrary choice of transformation is to use the relative efficiency of $\overline{Y}$, the sample mean, to estimate the constant mean μ under the process and compare it with μ estimated under independence. Scaling by n readily indicates this quantity to be

$$n^2(\mathbf{1}^\top \boldsymbol{R}(\boldsymbol{\theta})\ \mathbf{1})^{-1}. \tag{3}$$

If $\phi = 0$, $n^2(\mathbf{1}^\top \boldsymbol{R}(\boldsymbol{\theta})\ \mathbf{1})^{-1} = 1$, and as ϕ increases to ∞, (3) increases to n. This equation is attractive in that it assumes no distributional model for the process. The existence of $\text{var}(\overline{Y})$ is implied by the assumption of an isotropic correlation function. A negative feature of this process, however, is that for a fixed ϕ, the effective sample size need not increase in n. As an alternative to the previous suggestions regarding effective sample size, Griffith (2005) suggested a measure of the size of a geographic sample based on a model with a constant mean given by

$$\boldsymbol{Y} = \mu\mathbf{1} + \boldsymbol{e} = \mu\mathbf{1} + \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1/2}\boldsymbol{e}^*,$$

where $\boldsymbol{Y} = [Y(\boldsymbol{s}_1), Y(\boldsymbol{s}_2), \ldots, Y(\boldsymbol{s}_n)]^\top$, $\boldsymbol{e} = [e(\boldsymbol{s}_1), e(\boldsymbol{s}_2), \ldots, e(\boldsymbol{s}_n)]^\top$, and $\boldsymbol{e}^* = [e^*(\boldsymbol{s}_1), e^*(\boldsymbol{s}_2), \ldots, e^*(\boldsymbol{s}_n)]^\top$, respectively, denote $n \times 1$ vectors of spatially autocorrelated and non-autocorrelated errors such that $\text{var}(\boldsymbol{e}) = \sigma^2_{\boldsymbol{e}^*}\boldsymbol{\Sigma}(\boldsymbol{\theta})$ and $\text{var}(\boldsymbol{e}^*) = \sigma^2_{\boldsymbol{e}^*}\boldsymbol{I}_n$. This measure is

$$n^* = \text{tr}(\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1})n/(\mathbf{1}^\top\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\ \mathbf{1}), \tag{4}$$

where tr denotes the trace operator. Alternatively, assume that $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^r\}$ is a random field with mean μ, variance σ^2 and correlation ρ. Suppose that a realization of the

process is observed on the domain $V \subset \mathbb{R}^r$. Consider the following definition of ergodicity

$$\lim_{V \to \infty} \text{var}[Y(V)] = 0,$$

where the meaning of $V \to \infty$ is defined in Lantuejoul (1991) and

$$\text{var}[Y(V)] = \frac{\sigma^2}{|V|^2} \int_V \int_V \rho(\boldsymbol{x} - \boldsymbol{y}) \, \mathrm{d}\boldsymbol{x} \, \mathrm{d}\boldsymbol{y}.$$

Lantuejoul (1991) defined the integral range, denoted by A, as follows:

$$A = \lim_{V \to \infty} |V| \frac{\text{var}[Y(V)]}{\sigma^2}$$

If $A < \infty$, for large V

$$\text{var}[Y(V)] \approx \frac{\sigma^2 A}{|V|}.$$

Then, if $A \neq 0$ we can find an integer N such that

$$N \approx \frac{|V|}{A}.$$

In an alternative reduction of $\boldsymbol{R}(\boldsymbol{\theta})$, one can compare the reciprocal of the variance of the BLUE unbiased estimator of μ under $\boldsymbol{R}(\boldsymbol{\theta})$, which is readily shown to be $\mathbf{1}^\top \boldsymbol{R}(\boldsymbol{\theta})^{-1} \mathbf{1}$. As ϕ increases to ∞, this quantity increases to n. Again, no distributional model for the process is assumed. However, this expression arises as the Fisher information about μ under normality. In fact, for $\boldsymbol{Y} \sim \mathcal{N}(\mu \mathbf{1}, \boldsymbol{R}(\boldsymbol{\theta}))$, $\mathcal{I}(\mu) = \mathbf{1}^\top \boldsymbol{R}(\boldsymbol{\theta})^{-1} \mathbf{1}$, yielding the following definition.

Definition 1. Let $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in D \subset \mathbb{R}^r\}$ be a random field such that for $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_n \in D$, the vector $\boldsymbol{Y} = [Y(\boldsymbol{s}_1), Y(\boldsymbol{s}_2), \ldots, Y(\boldsymbol{s}_n)]^\top \sim \mathcal{N}(\mu \mathbf{1}, \boldsymbol{R}(\boldsymbol{\theta}))$, where $\boldsymbol{R}(\boldsymbol{\theta})$ is a non-singular correlation matrix. The quantity

$$\text{ESS} = \text{ESS}(n, \boldsymbol{R}(\boldsymbol{\theta}), r) = \mathbf{1}^\top \boldsymbol{R}(\boldsymbol{\theta})^{-1} \, \mathbf{1} \tag{5}$$

is called the effective sample size of $\boldsymbol{Y}$.

Notice from Definition 1 that if the n observations are independent and $\boldsymbol{R}(\boldsymbol{\theta}) = \boldsymbol{I}$, then ESS $= n$. If perfect positive spatial correlation prevails, then $\boldsymbol{R}(\boldsymbol{\theta}) = \mathbf{1}\mathbf{1}^\top$ is a singular matrix. Nonetheless, Definition 1 can be extended by considering a pseudoinverse of $\boldsymbol{R}(\boldsymbol{\theta})$. For instance, the Moore-Penrose pseudoinverse $\boldsymbol{R}(\boldsymbol{\theta})^+$of $\boldsymbol{R}(\boldsymbol{\theta})$ yields

$$\text{ESS} = \mathbf{1}^\top \boldsymbol{R}(\boldsymbol{\theta})^+ \mathbf{1} = \frac{\mathbf{1}^\top \mathbf{1} \mathbf{1}^\top \mathbf{1}}{n^2} = 1.$$

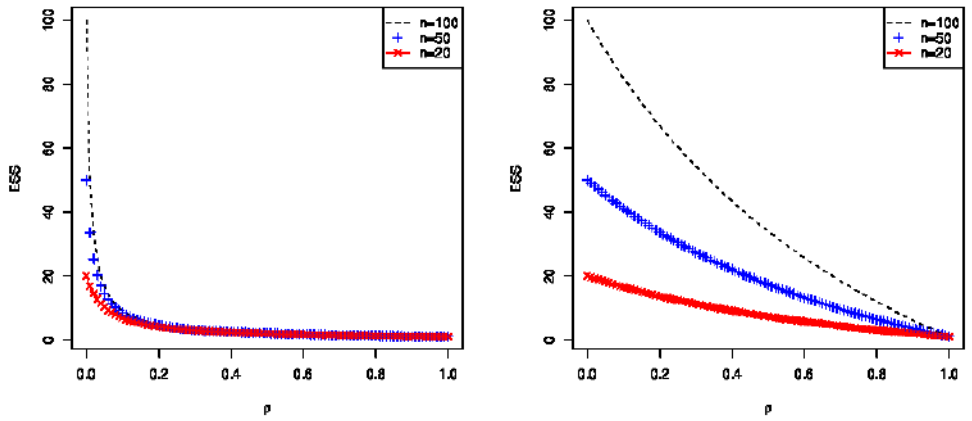
Note that the Moore-Penrose pseudoinverse of $\boldsymbol{R}(\boldsymbol{\theta})$ is unique (Magnus and Neudecker, 1999, p. 37), although the effective sample size is not affected by choosing a non-unique pseudoinverse, $\boldsymbol{R}(\boldsymbol{\theta})^-$ because for any matrix X, $X R(\boldsymbol{\theta})^- X^\top$ is invariant to $R(\boldsymbol{\theta})^-$. Particularly, these results can be used for the treatment of any positive semidefinite correlation

matrix. Here, the correlation matrix $\boldsymbol{R}(\theta)$ will be defined by a spatial covariance function as is customary in spatial statistics. Examples of these matrices will be discussed below. Haining (1990, p.163) noted that spatial dependency implies a loss of information in the estimation of the mean. One way to quantify that loss is through (5). Moreover, the asymptotic variance of the generalized least squares estimator of μ is 1/ESS.

Example 1. Let us consider $\boldsymbol{Y} \sim \mathcal{N}(\mu\mathbf{1}, \boldsymbol{R}(\rho))$, with the intra-class correlation structure $\boldsymbol{R}(\rho) = (1-\rho)\boldsymbol{I} + \rho\boldsymbol{J}$, $\boldsymbol{J} = \mathbf{1}\mathbf{1}^\top$, and $-1/(n-1) < \rho < 1$. Then

$$\mathrm{ESS}_{\mathrm{IC}} = n/(1+(n-1)\rho).$$

In Figure 1 (a) we observe that the reduction that takes place in this case is quite severe. For example, for $n = 100$ and $\rho = 0.1$, ESS $= 9.17$, and for $n = 100$ and $\rho = 0.5$, ESS $=$ 1.98. Generally, such noticeable reductions in sample size are not expected. However, the intra-class correlation does not take into account the spatial association between the observations. With richer correlation structures, the effective sample size is better at reducing the information from $\boldsymbol{R}(\rho)$ (the reduction is not so severe). It should be stressed that when a negative autocorrelation is present some counterintuitive phenomena can occur. For example, there are specific correlation structures for which the effective sample size is greater than n. This inconvenience has been noted by Richardson (1990). Additional examples and effects of negative spatial autocorrelation have been indicated by Anselin (1988); Griffith et al. (2003) among others.



(a) (b)

Figure 1: (a) ESS for the intra-class correlation; (b) ESS for the a Toeplitz correlation.

Example 2. Consider the vector $\boldsymbol{Y} \sim \mathcal{N}(\mu\mathbf{1}, \boldsymbol{R}(\rho_1, \rho_2, \ldots, \rho_{n-1}))$, where for $|\rho_i| < 1$, the

correlation structure is defined by the Jacobi matrix as follows:

$$\boldsymbol{R}(\rho_1, \rho_2, \ldots, \rho_{n-1}) = \begin{pmatrix} 1 & \rho_1 & \rho_1\rho_2 & \rho_1\rho_2\rho_3 & \cdots & \prod_{i=1}^{n-1} \rho_i \\ \rho_1 & 1 & \rho_2 & \rho_2\rho_3 & \cdots & \prod_{i=2}^{n-1} \rho_i \\ \rho_1\rho_2 & \rho_2 & 1 & \rho_3 & \ddots & \vdots \\ \rho_1\rho_2\rho_3 & \rho_2\rho_3 & \rho_3 & 1 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \rho_{n-1} \\ \prod_{i=1}^{n-1} \rho_i & \prod_{i=2}^{n-1} \rho_i & \cdots & \cdots & \rho_{n-1} & 1 \end{pmatrix} \tag{6}$$

The inverse of $\boldsymbol{R}(\rho_1, \rho_2, \ldots, \rho_{n-1})$ in (6) is given in the appendix. Consequently, the ESS has a closed form. Particularly, if $\rho_1 = \rho_2 = \ldots = \rho_{n-1} = \rho$, the correlation matrix corresponds to a AR(1) process, and the ESS is given by (see Appendix)

$$\mathrm{ESS}_{\mathrm{AR}} = (2 + (n-2)(1-\rho))/(1+\rho).$$

Furthermore, straightforward calculations show that for $0 < \rho < 1$ and $n > 2$,

$$\mathrm{ESS}_{\mathrm{IC}} < \mathrm{ESS}_{\mathrm{AR}}.$$

Therefore, the reduction in $\boldsymbol{R}(\rho)$ under the AR(1) structure is not as severe as in the intra-class correlation case. Based on Figure 1 (a) and (b), we see that the effective sample size is decreasing in ρ in both cases.

Example 3. Let us consider

$$Y(t) = \mu + W(t), \tag{7}$$

where $t \in [0, L]$ and $W(t)$ is a mean zero stationary Gaussian process with a covariance function of the form

$$k(s-t) = \sigma^2 \exp(-\phi|s-t|),$$

where σ is the standard deviation of $W(t)$ and $\phi > 0$ is a parameter that determines the correlation function ρ. For instance, consider $\rho(s-t) = \exp(-|s-t|)$, $L = 1$ and $\phi = 1$, corresponding to a Ornstein-Uhlenbeck process. Xia et al. (2006) studied the problem of estimating the mean μ in the following two cases: when the parameters σ and ϕ are known and when the scale and location are assumed unknown and the correlation function is assumed known. They also found an explicit expression for the effective sample size. In fact, when the Ornstein-Uhlenbeck process is sampled at n locations, say $\Delta, 2\Delta, \ldots, n\Delta$, the effective sample size allows the following closed form expression:

$$\mathrm{ESS} = \mathbf{1}^\top \boldsymbol{R}(\rho)^{-1}\ \mathbf{1} = 1 - (n-1)\frac{1-\rho}{1+\rho},$$

where $\rho = \exp(-\Delta)$.

Example 4. Consider $\boldsymbol{Y}_j = [Y_j(\boldsymbol{s}_1), Y_j(\boldsymbol{s}_2), \ldots, Y_j(\boldsymbol{s}_n)]^\top \sim \mathcal{N}(\mu\mathbf{1}, \boldsymbol{R}(\boldsymbol{\theta}))$, for $j = 1, 2, \ldots, n$, such that $\boldsymbol{Y}_j$ and $\boldsymbol{Y}_k$ are independent for $i \neq k$. Let us define $\boldsymbol{Z} = [\boldsymbol{Y}_1^\top, \boldsymbol{Y}_2^\top, \ldots, \boldsymbol{Y}_n^\top]^\top$. Then, $\boldsymbol{Z} \sim \mathcal{N}(\mu\mathbf{1}_{n^2}, \boldsymbol{I} \otimes \boldsymbol{R}(\boldsymbol{\theta}))$, and

$$\mathrm{ESS}_{\boldsymbol{Z}} = n\ \mathrm{ESS}_{\boldsymbol{Y}}.$$

The information provided by independent replicates of the vector $\boldsymbol{Y}$ is captured by the effective sample size $\text{ESS}_{\boldsymbol{Z}}$. Generally, new information is also accounted for by the effective sample size when new observations are correlated with existing data. This feature can be observed as a monotonic property below.

Proposition 1. Assume that $\boldsymbol{Y} \sim \mathcal{N}(\mu\mathbf{1}, \boldsymbol{R}(\boldsymbol{\theta}))$. Then, for r fixed

i) ESS is increasing in n.

ii) $\text{ESS} \geq 1$.

Proof. See Appendix □

3.2. The Case of Two Means

A spatial scientist may be simultaneously interested in more than one variable. This motivates the study of extensions of equation (5) to the case of two attribute variables measured at the same locations in D. Let $\{X(\boldsymbol{s}) : \boldsymbol{s} \in D \subset \mathbb{R}^r\}$ and $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in D \subset \mathbb{R}^r\}$ be two random fields representing the variables of interest. The joint treatment of two correlated spatial variables must consider the following two sources of redundant information (Griffith, 2005): correlation between the two variables and spatial association within each variable. We suggest a weighted version of equation (5) to avoid imposing a correlation structure between $X(\cdot)$ and $Y(\cdot)$. We assume that the information of variables $X(\cdot)$ and $Y(\cdot)$ is available when both processes are non-spatially correlated and when neither of them depend on spatial location. We will use X and Y to differentiate these variables from $X(\cdot)$ and $Y(\cdot)$. A weighted average of X and Y is $wX + (1-w)Y$, for $0 \leq w \leq 1$. Let us call the respective variances σ_X^2 and σ_Y^2, and the correlation between X and Y, ρ_{XY}. Then

$$\text{var}[wX + (1-w)Y] = w^2\sigma_X^2 + (1-w)^2\sigma_Y^2 + 2w(1-w)\sigma_X^2\sigma_Y^2\rho_{XY}.$$

Following Griffith (2005), we used this variance to construct a weighted version of the effective sample size for two means. Indeed, if $\boldsymbol{X} \sim \mathcal{N}(\mu_X\mathbf{1}, \boldsymbol{R}_X(\boldsymbol{\theta}))$ and $\boldsymbol{Y} \sim \mathcal{N}(\mu_Y\mathbf{1}, \boldsymbol{R}_Y(\boldsymbol{\theta}))$, then

$$\text{ESS}_w = \frac{w^2\sigma_X^2\mathbf{1}^\top\boldsymbol{R}_X(\boldsymbol{\theta})^{-1}\,\mathbf{1} + (1-w)^2\sigma_Y^2\mathbf{1}^\top\boldsymbol{R}_Y(\boldsymbol{\theta})^{-1}\,\mathbf{1} + 2w(1-w)\rho_{XY}\sigma_X\sigma_Y\mathbf{1}^\top\boldsymbol{R}_X(\boldsymbol{\theta})^{-1/2}\boldsymbol{R}_Y(\boldsymbol{\theta})^{-1/2}\,\mathbf{1}}{w^2\sigma_X^2 + (1-w)^2\sigma_Y^2 + 2w(1-w)\rho_{XY}\sigma_X\sigma_Y}. \tag{8}$$

If $\boldsymbol{R}_X(\boldsymbol{\theta}) = \boldsymbol{R}_Y(\boldsymbol{\theta}) = \boldsymbol{I}$, then this expression is reduced to n. If $w = 0$, $w = 1$ or $\boldsymbol{R}_X(\boldsymbol{\theta}) = \boldsymbol{R}_Y(\boldsymbol{\theta})$, then this expression is reduced to (5). Therefore, the effective sample size (8) is a weighted average of the individual effective sample sizes. If $\boldsymbol{R}_X(\boldsymbol{\theta})$ and $\boldsymbol{R}_Y(\boldsymbol{\theta})$ approach the case of perfect positive spatial autocorrelation for both variables, ESS_w approaches one. Here, the weight w is not a parameter to be estimated; w is chosen in advance by considering the importance of each sample with respect to the other. Defining the 2×2 matrices

$$\begin{aligned}
\boldsymbol{A}_d &= \text{diag}(w, 1-w),\\
\boldsymbol{\Phi}_d &= \text{diag}(\sigma_X, \sigma_Y),\\
\mathcal{R} &= \begin{pmatrix} 1 & \rho_{XY} \\ \rho_{XY} & 1 \end{pmatrix},
\end{aligned}$$

and

$$\boldsymbol{V}_d = \text{diag}(\boldsymbol{R}_X(\boldsymbol{\theta})^{-1/2}, \boldsymbol{R}_Y(\boldsymbol{\theta})^{-1/2}),$$

we have

$$\begin{aligned}
&\mathbf{1}^\top \boldsymbol{A}_d \boldsymbol{\Phi}_d \mathcal{R} \boldsymbol{\Phi}_d \boldsymbol{A}_d \mathbf{1} = w^2\sigma_X^2 + (1-w)^2\sigma_Y^2 + 2w(1-w)\sigma_X^2\sigma_Y^2\rho_{XY}, \\
&\mathbf{1}^\top \left((\boldsymbol{A}_d\boldsymbol{\Phi}_d \otimes \boldsymbol{I})\boldsymbol{V}_d^{1/2}\right)^\top (\mathcal{R}\otimes \boldsymbol{I}) \left((\boldsymbol{A}_d\boldsymbol{\Phi}_d \otimes \boldsymbol{I})\boldsymbol{V}_d^{1/2}\right)\mathbf{1} = w^2\sigma_x^2 \mathbf{1}^\top \boldsymbol{R}_X(\boldsymbol{\theta})^{-1}\mathbf{1} \\
&+ (1-w)^2\sigma_Y^2 \mathbf{1}^\top \boldsymbol{R}_Y(\boldsymbol{\theta})^{-1}\mathbf{1} + 2w(1-w)\rho_{XY}\sigma_X\sigma_Y \mathbf{1}^\top \boldsymbol{R}_X(\boldsymbol{\theta})^{-1/2}\boldsymbol{R}_Y(\boldsymbol{\theta})^{-1/2}\mathbf{1},
\end{aligned}$$

where $\otimes$ denotes the Kronecker product. Thereafter, the weighted effective sample size can be written in matrix form using

$$\text{ESS}_w = \frac{\mathbf{1}^\top \left((\boldsymbol{A}_d\boldsymbol{\Phi}_d \otimes \boldsymbol{I})\boldsymbol{V}_d^{1/2}\right)^\top (\mathcal{R}\otimes \boldsymbol{I}) \left((\boldsymbol{A}_d\boldsymbol{\Phi}_d \otimes \boldsymbol{I})\boldsymbol{V}_d^{1/2}\right)\mathbf{1}}{\mathbf{1}^\top \boldsymbol{A}_d \boldsymbol{\Phi}_d \mathcal{R} \boldsymbol{\Phi}_d \boldsymbol{A}_d \mathbf{1}}. \tag{9}$$

Equation (9) can easily be generalized to a multivariate situation involving p variables. In such a case, $\boldsymbol{A}_d$, $\boldsymbol{\Phi}_d$, and $\mathcal{R}$ are $p \times p$ matrices, $\boldsymbol{I}$ is an $n \times n$ matrix and $\boldsymbol{V}_d$ is an $np \times np$ block diagonal matrix.

3.3. CAR and SAR Processes

Consider a CAR model of the following form:

$$Y(\boldsymbol{s}_i) \mid Y(\boldsymbol{s}_j), j \neq i \sim \mathcal{N}(\mu + \rho \sum_j b_{ij} Y(\boldsymbol{s}_j), \tau_i^2), i = 1, 2, \cdots, n. \tag{10}$$

where ρ determines the direction and magnitude of the spatial neighborhood effect, b_{ij} are weights that determine the relative influence of location j on location i, and τ_i^2 is the conditional variance. If n is finite, we form the matrices $\boldsymbol{B} = (b_{ij})$ and $\boldsymbol{D} = diag(\tau_1^2, \tau_2^2, \cdots, \tau_n^2)$, and according to the factorization theorem,

$$\boldsymbol{Y} \sim \mathcal{N}(\mu\mathbf{1}, (\boldsymbol{I} - \rho\boldsymbol{B})^{-1}\boldsymbol{D}).$$

We assume that the parameter ρ satisfies the necessary conditions for a positive definite matrix (Banerjee et al., 2004, p. 70-82). A common way to construct $\boldsymbol{B}$ is to use a defined neighborhood matrix $\boldsymbol{W}$ that indicates whether the areal units associated with the measurements $Y(\boldsymbol{s}_1)$, $Y(\boldsymbol{s}_2)$, $\cdots, Y(\boldsymbol{s}_n)$ are neighbors. For example, if $b_{ij} = w_{ij}/w_{i+}$ and $\tau_i^2 = \tau^2/w_{i+}$, then

$$\boldsymbol{Y} \sim \mathcal{N}(\mu\mathbf{1}, \tau^2(\boldsymbol{D}_w - \rho\boldsymbol{W})^{-1}), \tag{11}$$

where $\boldsymbol{D}_w = \text{diag}(w_{i+})$. For a CAR process, as described in equation (11), $\boldsymbol{R}_{\text{CAR}}^{-1} = \boldsymbol{C}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}$ for a suitable diagonal matrix $\boldsymbol{C}$.

Proposition 2. For a CAR model with $\boldsymbol{\Sigma} = \tau^2(\boldsymbol{D}_w - \rho\boldsymbol{W})^{-1}$ where $\sigma_i = \boldsymbol{\Sigma}_{ii}^{1/2}$ and $\boldsymbol{C} = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n)$

$$\text{ESS}_{\text{CAR}} = \mathbf{1}^\top \boldsymbol{R}_{\text{CAR}}^{-1}\mathbf{1} = \frac{1}{\tau^2}\left[\sum_i \sigma_i^2 w_{i+} - \rho \sum_i \sum_j \sigma_i\sigma_j w_{ij}\right], \tag{12}$$

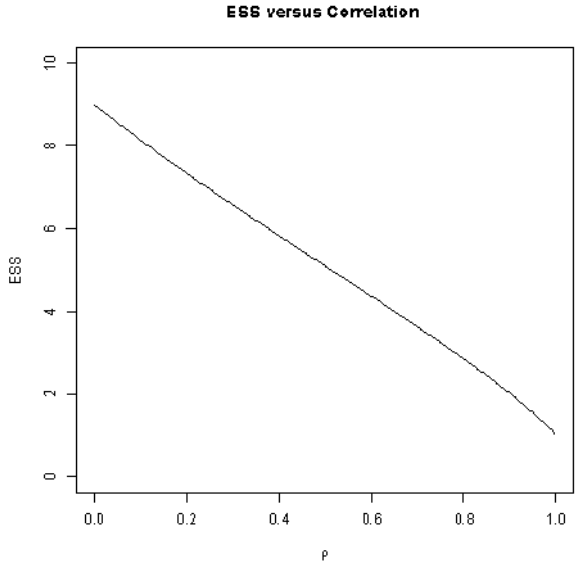


Figure 2: ESS for a CAR process defined on a 9×9 rectangular grid, when $\tau = 1$ and $\mu = 0$.

where $w_{i+} = \sum_j w_{ij}$. Moreover if $\sigma_i^2 = \tau^2$, then

$$\mathrm{ESS}_{\mathrm{CAR}} = (1-\rho)\sum_i w_{i+}.$$

Example 5. Let $\boldsymbol{Y}$ be a CAR process as in (11), defined on a 3×3 rectangular lattice. Assume that each site located at the edges of the grid has two neighbors and the central site has four neighbors as described by the matrix $\boldsymbol{W}$ defined as

$$\boldsymbol{W} = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}.$$

The effective sample size (12) as a function of ρ is shown in Figure 2 for $\tau = 1$ and $\mu = 0$. This allows us to find the effective sample size when an estimate of ρ is available. We observe, for example, that for $\rho = 0.5$ the sample size decreases from 9 to 5.

Now, let us consider a SAR process of the form

$$\begin{aligned} \boldsymbol{Y} &= \boldsymbol{X} + \boldsymbol{e} \\ \boldsymbol{e} &= \boldsymbol{B}\boldsymbol{e} + \boldsymbol{v} \end{aligned}$$

where $\boldsymbol{B}$ is a matrix of spatial dependency, $\mathbb{E}[\boldsymbol{v}] = \mathbf{0}$, and $\boldsymbol{\Sigma}_{\boldsymbol{v}} = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2)$. Then, $\boldsymbol{\Sigma} = \text{var}[\boldsymbol{Y}] = (\boldsymbol{I} - \boldsymbol{B})^{-1}\boldsymbol{\Sigma}_{\boldsymbol{v}}(\boldsymbol{I} - \boldsymbol{B}^\top)^{-1}$ and as before $\boldsymbol{R}_{\text{SAR}}^{-1} = \boldsymbol{C}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}$ for a suitable diagonal matrix $\boldsymbol{C}$. We can state the following result.

Proposition 3. For a SAR process with $\boldsymbol{B} = \rho\boldsymbol{W}$ where $\boldsymbol{W}$ is any contiguity matrix, $\boldsymbol{\Sigma}_{\boldsymbol{v}} = \sigma^2\boldsymbol{I}$, $\sigma_i = \boldsymbol{\Sigma}_{ii}^{1/2}$ and $\boldsymbol{C} = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n)$, the effective sample size is given by

$$\text{ESS}_{\text{SAR}} = \mathbf{1}^\top \boldsymbol{R}_{\text{SAR}}^{-1}\mathbf{1} = \frac{1}{\sigma^2}\left[\sum_i \sigma_i^2 - 2\rho\sum_i\sum_j \sigma_i\sigma_j w_{ij} + \rho^2\sum_i\sum_j\sum_k \sigma_i\sigma_j w_{ki}w_{kj}\right]. \tag{13}$$

Proof. See Appendix □

The definition of effective sample size can be generalized if we consider the Fisher information quantity for other non-normal multivariate distributions, which we will develop in the next section for a multivariate elliptical distribution.

3.4. Elliptical Distributions

Definition 2. Let $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in D \subset \mathbb{R}^r\}$ be a random field and the locations $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_n \in D$. Suppose that the random vector $\boldsymbol{Y} = [Y(\boldsymbol{s}_1), Y(\boldsymbol{s}_2), \ldots, Y(\boldsymbol{s}_n)]^\top$ has a density of the form

$$f(\boldsymbol{Y}) = |\boldsymbol{\Sigma}(\boldsymbol{\theta})|^{-1/2} g((\boldsymbol{Y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}(\boldsymbol{Y} - \boldsymbol{\mu})),$$

where $\boldsymbol{\mu} \in \mathbb{R}^n$ is a location parameter, $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is an $n \times n$ positive definite matrix, and $g : \mathbb{R} \to [0, \infty)$ such that $\int_0^\infty u^{n/2-1} g(u)\, \mathrm{d}u < \infty$. Therefore, the vector $\boldsymbol{Y}$ has an elliptically contoured distribution and is denoted as $\boldsymbol{Y} \sim EC_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}(\boldsymbol{\theta}), g)$.

Definition 3. Assume that $\boldsymbol{Y} \sim EC_n(\mu\mathbf{1}, \boldsymbol{R}(\boldsymbol{\theta}), g)$. The effective sample size associated with $\boldsymbol{Y}$ is defined as

$$\text{ESS}_e = 4\,\mathbb{E}\left[\|\boldsymbol{Z}\|^2 W_g^2(\|\boldsymbol{Z}\|^2)\right]\frac{1}{n}\mathbf{1}^\top \boldsymbol{R}(\boldsymbol{\theta})^{-1}\mathbf{1} = \kappa \cdot \text{ESS}, \tag{14}$$

where $W_g(u) = g'(u)/g(u)$ and $\boldsymbol{Z} = \boldsymbol{R}(\boldsymbol{\theta})^{-1/2}(\boldsymbol{Y} - \mu\mathbf{1})$.

The definition of ESS_e stems from the form of the expected information matrix for the case of elliptically contoured models (Lange et al., 1989). Particularly, if $\boldsymbol{Y}$ follows a multivariate t-distribution, say $\boldsymbol{Y} \sim t_n(\mu\mathbf{1}, \boldsymbol{R}(\boldsymbol{\theta}), \nu), \nu > 2$, then

$$\text{ESS}_t = \frac{\nu + n}{\nu + n + 2}\text{ESS}. \tag{15}$$

From (15) we see that $\text{ESS}_t \leq \text{ESS}$ and from Proposition 1 for $\nu > 2$ and $n \geq 1$, $\text{ESS}_t \geq 3/2$.

Example 6. The power exponential distribution, $PE_n(\boldsymbol{\mu}, \boldsymbol{R}(\boldsymbol{\theta}), \lambda)$ with $\lambda > 0$, is an extensively studied distribution because it presents both behaviors, light- ($\lambda > 1$) and heavy-tailed ($\lambda < 1$) distributions, while also including the normal case ($\lambda = 1$). Theoretical developments for this distribution can be found in Gómez et al. (1998). An

application in the context of repeated measure data may be found in Lindsey (1999). If $\boldsymbol{Y} \sim PE_n(\mu\mathbf{1}, \boldsymbol{R}(\boldsymbol{\theta}), \lambda)$, $\lambda > 0$, then using results given in Osorio et al. (2007) we obtain

$$\mathrm{ESS}_{pe} = \frac{\lambda^2 \Gamma\left(\frac{n-2}{2\lambda} + 2\right)}{2^{1/\lambda}\Gamma(\frac{n}{2\lambda})}\mathrm{ESS}. \tag{16}$$

The constant κ for the power exponential distribution has been plotted as a function of n and λ to explore those values for which the effective sample size is greater or smaller than the effective sample size for the Gaussian case, ESS. In Figure 3 we observe that κ increases as long as λ and n increase. Therefore, there are values of λ and n for which $\mathrm{ESS}_{pe} < \mathrm{ESS}$ and there are also values of λ and n for which $\mathrm{ESS}_{pe} > \mathrm{ESS}$. For the light-tailed case ($\lambda > 1$), the behavior of κ is similar (not shown here), yielding the same conclusions as for ($\lambda < 1$).

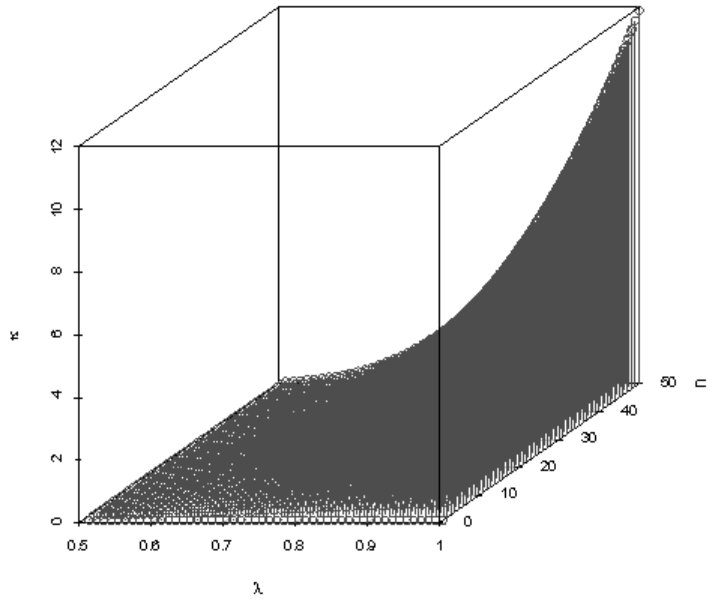


Figure 3: $\kappa = \lambda^2\Gamma\left(\frac{n-2}{2\lambda} + 2\right)/(2^{1/\lambda}\Gamma(\frac{n}{2\lambda}))$ versus λ and n for the exponential power distribution. The grid was generated for $\lambda \in [0.5, 1]$ and $n = 1, 2, \ldots, 50$.

4. Estimation

We now discuss the estimation of the effective sample size for the previously introduced models. The estimation of the effective sample size can be handled by maximum likelihood for the models used in examples 1-3 because there are few parameters. We note that

in spatial statistics, there is no replication of the random vector $\boldsymbol{Y}$. Therefore, efficient estimation requires defining a small number of parameters in the covariance structure $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ with respect to n. Commonly used covariance structures in spatial statistics involve few parameters, which are related to the spatial range, sill, nugget effect, and smoothness of the process $Y(\cdot)$. Let $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in D \subset \mathbb{R}^r\}$ be a random field. Let $\{\epsilon(\boldsymbol{s}) : \boldsymbol{s} \in D \subset \mathbb{R}^r\}$ be a zero mean Gaussian random field such that $\text{cov}(\epsilon(\boldsymbol{t}), \epsilon(\boldsymbol{t}+\boldsymbol{s})) = \sigma(\boldsymbol{s}, \boldsymbol{\theta}), \boldsymbol{s}, \boldsymbol{t} \in D, \boldsymbol{\theta} \in \Theta$. Suppose that for the locations $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_n \in D$, $\boldsymbol{Y} := [Y(\boldsymbol{s}_1), Y(\boldsymbol{s}_2), \ldots, Y(\boldsymbol{s}_n)]^\top$ and $\boldsymbol{\epsilon} := [\epsilon(\boldsymbol{s}_1), \epsilon(\boldsymbol{s}_2), \ldots, \epsilon(\boldsymbol{s}_n)]^\top$. Consider a model of the form

$$\boldsymbol{Y} = \mu \mathbf{1} + \boldsymbol{\epsilon}, \tag{17}$$

where $\mu \in \mathbb{R}$, and $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$. Our goal here is to provide an estimate for the unknown parameter $\boldsymbol{\theta}$ so that the effective sample size can be estimated by substituting this value in the correlation matrix. We consider that the mean of the process is also unknown; that is, the estimation should consider both of the following unknown parameters: $\boldsymbol{\theta}$ and μ. One way to approach this problem is by using the REML estimator of $\boldsymbol{\theta}$, which is called $\widehat{\boldsymbol{\theta}}_{REML}$ and was first introduced by Patterson and Thompson (1971). The REML estimator of $\boldsymbol{\theta}$ is based on performing maximum likelihood estimations for $\boldsymbol{K}^\top \boldsymbol{Y}$, where $\boldsymbol{K}$ is an $n \times (n-1)$ matrix chosen so that $\mathbb{E}[\boldsymbol{K}^\top \boldsymbol{Y}] = \mathbf{0}$ and $\text{rank}[\boldsymbol{K}] = n-1$. Then, minus twice the log likelihood of $\boldsymbol{K}^\top \boldsymbol{Y}$ we have

$$\ell_{\text{REML}}(\boldsymbol{\theta}) = \log|\boldsymbol{K}^\top \boldsymbol{\Sigma}(\boldsymbol{\theta}) \boldsymbol{K}| + (n-1)\log(2\pi) + \boldsymbol{Y}^\top \boldsymbol{K}(\boldsymbol{K}^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})\boldsymbol{K})^{-1} \boldsymbol{K}^\top \boldsymbol{Y}.$$

Harville (1974, 1977) established sufficient conditions so that ℓ_{REML} does not depend on $\boldsymbol{K}$. Then

$$\ell_{\text{REML}}(\boldsymbol{\theta}) = \log|\boldsymbol{\Sigma}(\boldsymbol{\theta})| + \log(\mathbf{1}^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\mathbf{1}) + \boldsymbol{r}^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{r} + (n-1)\log(2\pi),$$

where

$$\boldsymbol{r} = \boldsymbol{Y} - \widehat{\mu}\,\mathbf{1}, \qquad \widehat{\mu} = (\mathbf{1}^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\mathbf{1})^{-1}\mathbf{1}^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{Y}. \tag{18}$$

Accordingly, an estimator of the effective sample size is

$$\widehat{\text{ESS}} = \mathbf{1}^\top \boldsymbol{R}(\widehat{\boldsymbol{\theta}}_{\text{REML}})^{-1}\mathbf{1}. \tag{19}$$

The restricted maximum likelihood estimation can be extended in a very simple way to the context of elliptically distributed errors. In this case, the REML estimator has the same form as in the normal case. Indeed, assuming the model $\boldsymbol{Y} \sim EC_n(\mu\mathbf{1}, \boldsymbol{\Sigma}(\boldsymbol{\theta}), g)$, one has

$$\boldsymbol{Z} = \boldsymbol{K}^\top \boldsymbol{Y} \sim EC_{n-1}(\mathbf{0}, \boldsymbol{K}^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})\boldsymbol{K}, g),$$

with density function

$$f(\boldsymbol{Z}; \boldsymbol{\theta}) = |\boldsymbol{K}^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})\boldsymbol{K}|^{-1/2} g(\boldsymbol{Y}^\top \boldsymbol{K}(\boldsymbol{K}^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})\boldsymbol{K})^{-1}\boldsymbol{K}^\top \boldsymbol{Y}). \tag{20}$$

Let $\widetilde{\boldsymbol{\theta}}_{\text{REML}}$ be the REML based on equation (20). Then, following the results given by Anderson et al. (1986), we have

$$\widetilde{\boldsymbol{\theta}}_{\text{REML}} = \frac{n-1}{u_g}\,\widehat{\boldsymbol{\theta}}_{\text{REML}},$$

where u_g is the maximum of the function $h(u) = u^{(n-1)/2}g(u)$. Therefore, minus twice the log-likelihood function associated with the REML estimation for elliptically contoured distribution errors we have

$$\ell_{\mathrm{REML}}(\boldsymbol{\theta}) = \log|\boldsymbol{\Sigma}(\boldsymbol{\theta})| + \log(\mathbf{1}^\top\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\mathbf{1}) - 2\log g(\boldsymbol{r}^\top\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{r}),$$

where $\boldsymbol{r}$ and $\widehat{\mu}$ are given in (18). When g is a continuously differentiable function, u_g can be obtained as a solution to the equation

$$\frac{n-1}{u_g} + W_g(u) = 0.$$

When g is the density generating function associated with the normal or Student's-t distribution, $u_g = n-1$, and the power exponential distribution $u_g = (\frac{n-1}{\lambda})^{1/\lambda}$. The estimation of parameters associated with the density generating function g has been suggested in the literature. For example, the degrees of freedom of Student's-t distribution or the shape parameter λ of the exponential power distribution have been used. However, some authors (Breusch et al., 1997; Fernández and Steel, 1999) have noted drawbacks of the maximum likelihood estimation of these parameters. Without replication, as in the spatial regression case given in (17), the kurtosis parameter selection cannot be conducted using procedures based on the log-likelihood function because it is not bounded (Zellner, 1976). The use of parameter selection via cross-validation (Rao and Wu, 2001) may overcome this inconvenience. Cressie (1993) and Cressie and Lahiri (1996) studied the asymptotic properties of the REML estimator for regression models with covariance parameters that follow a Gaussian linear model, including linear spatial regression. Particularly, the following result Corollary 3.1 Cressie and Lahiri (1996) was established to obtain a more general model than (17).

Proposition 4. Define $\boldsymbol{\mathcal{J}}_n(\boldsymbol{\theta}) = (\mathbb{E}f_{n}(\boldsymbol{\theta}))^{1/2}$, where $f_{n}(\boldsymbol{\theta}) = (\partial^2\ell_{\mathrm{REML}}(\boldsymbol{\theta})/\partial\theta_i\partial\theta_j)$, $1 \le i,j \le k$, and $\boldsymbol{B}_n(\boldsymbol{\theta}) = \mathrm{diag}(\|\boldsymbol{\Pi}(\boldsymbol{\theta})\boldsymbol{\Sigma}_1(\boldsymbol{\theta})\|, \ldots, \|\boldsymbol{\Pi}(\boldsymbol{\theta})\boldsymbol{\Sigma}_k(\boldsymbol{\theta})\|)$, where $\boldsymbol{\Pi}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} - \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}(\mathbf{1}(\mathbf{1}^\top\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\mathbf{1})^{-1}\mathbf{1}^\top\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}$, and $\boldsymbol{\Sigma}_i(\boldsymbol{\theta}) = \partial\boldsymbol{\Sigma}(\boldsymbol{\theta})/\partial\theta_i$. Suppose that there exists a nonsingular matrix $\tilde{\boldsymbol{W}}(\boldsymbol{\theta})$, continuous in $\boldsymbol{\theta}$, such that

$$\boldsymbol{\mathcal{J}}_n(\boldsymbol{\theta})\boldsymbol{B}_n(\boldsymbol{\theta}) \xrightarrow{u} \tilde{\boldsymbol{W}}(\boldsymbol{\theta}).$$

Then, under the conditions of Theorem 3.1 in Cressie and Lahiri (1996),

$$\boldsymbol{\mathcal{J}}_n(\widehat{\boldsymbol{\theta}}_n)(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{I}_k),$$

where $\widehat{\boldsymbol{\theta}}_n = \widehat{\boldsymbol{\theta}}_{\mathrm{REML}}$, the REML estimator based on the first n observations.

The following result provides the asymptotic normality of the effective sample size.

Proposition 5. Suppose that the assumptions of Proposition 4 hold, and assume that $g(\boldsymbol{\theta}) = \mathbf{1}^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{1}$ has continuous first partial derivatives. Then

$$[\nabla^\top g(\boldsymbol{\theta})\boldsymbol{\mathcal{J}}_n(\widehat{\boldsymbol{\theta}}_n)\boldsymbol{\mathcal{J}}_n(\widehat{\boldsymbol{\theta}}_n)^\top\nabla g(\boldsymbol{\theta})]^{-1/2}(g(\widehat{\boldsymbol{\theta}}_n) - g(\boldsymbol{\theta})) \xrightarrow{d} \mathcal{N}(0,1).$$

Proof. This result is an immediate consequence of the delta method and Proposition 4. □

5. Dependence of ESS on the Spatial Sites

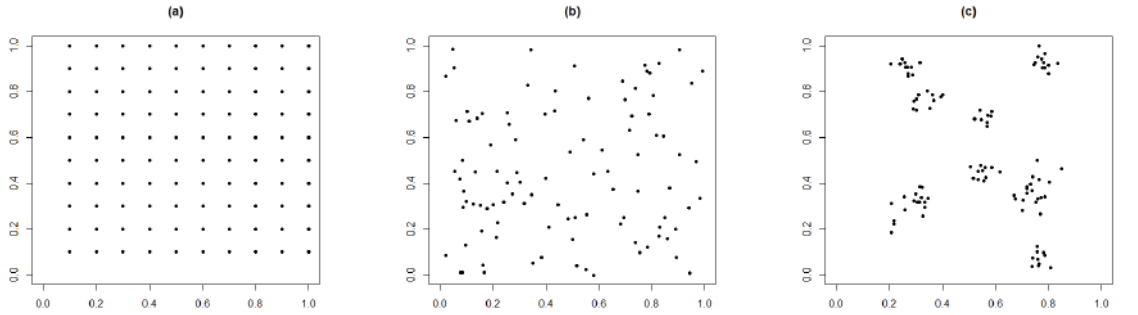
In Section 3, we emphasized the dependence of the effective sample size on sampling design. This dependence exists through the correlation structure of the process. In other words, the ESS depends on the configuration of the sites that belong to $\mathcal{C}$ through $\boldsymbol{R}(\boldsymbol{\theta})$. Therefore, it is natural to have different ESS values for datasets that are arranged differently in space. Additionally, ESS depends on the parametric model that is used for the spatial covariance structure. Hence, the first property described above is related to $\mathcal{C}$, and the second is related to the selection of the covariance structure.

To elucidate the dependence of ESS on $\mathcal{C}$, we performed a numerical experiment for the following three configurations: a regular grid, a random pattern, and an aggregate pattern. For each case, 100 observations were generated from a zero-mean Gaussian random field with the following correlation structures:

$$\text{Spherical:} \quad \rho(h) = \begin{cases} 1 - 1.5\frac{h}{\phi} + 0.5\left(\frac{h}{\phi}\right)^3, & h < \phi, \\ 0, & \text{otherwise,} \end{cases}$$

$$\text{Exponential:} \quad \rho(h) = \exp\left(-\frac{h}{\phi}\right),$$

with $\phi = 1$. Then, the effective sample size (5) and its estimation were computed as is shown in Figure 4.



$\widehat{\text{ESS}}_S = 9.063,\ \widehat{\text{ESS}}_E = 7.954$	$\widehat{\text{ESS}}_S = 1.001,\ \widehat{\text{ESS}}_E = 1.213$	$\widehat{\text{ESS}}_S = 3.897,\ \widehat{\text{ESS}}_E = 5.610$
$\text{ESS}_S = 4.401,\ \text{ESS}_E = 4.504$	$\text{ESS}_S = 1.965,\ \text{ESS}_E = 1.994$	$\text{ESS}_S = 5.904,\ \text{ESS}_E = 6.304$

Figure 4: ESS_S and ESS_E denote the effective sample size, respectively, for the spherical and exponential correlation structures. Similarly for the estimations $\widehat{\text{ESS}}_S$ and $\widehat{\text{ESS}}_E$.

For this particular experiment, effective sample size is greatly affected by sampling design. Indeed, whereas the ESS is overestimated for a regular design, it is underestimated for a design with clusters. Moreover, in Figure 4 (b) we see that estimators from the random design are closer to the true values than those from other designs. The same patterns hold for an extensive simulation study (not shown here) that was carried out to explore the sensitivity of ESS to sampling design for several spatial correlation structures.

The problem of finding an optimal spatial sampling design has been widely considered in the literature. In spatial statistics, at least three different approaches have been studied

to tackle this problem. The first approach considers the optimal sampling design for predicting the spatial process at some unobserved location (Su and Cambanis, 1993; Ritter, 1996). The second approach studies the optimal sampling design using an empirical variogram to estimate the covariance structure (Russo, 1984; Bogaert and Russo, 1999). The third approach determines the optimal design for covariance parameter estimation (Lark, 2002). Improvements to existing methods have been suggested by Zhu and Stein (2005). These authors have suggested the following three design criteria for parameter estimation: locally optimal designs, minimax designs using relative efficiency, and a simulated annealing algorithm.

In any experiment aiming to determine effective sample size, an appropriate sampling design should be planned in advance to provide sufficiently good ESS estimates (using optimality criteria). We have described the REML estimation for the variogram parameters, therefore, the locally optimal design developed by Zhu and Stein (2005), based on the inverse of the Fisher information matrix of $\boldsymbol{\theta}$ to approximate the covariance matrix of $\widehat{\boldsymbol{\theta}}_{\text{REML}}$, could be used in practice.

6. Random Locations

In this section, we explore the notion of the effective sample size when two locations defined over a compact set on a Euclidean space are randomly distributed. This problem has been studied repeatedly in the context of geometrical probabilities. We start by considering a compact subset D of the r-dimensional Euclidean space $\mathbb{R}^r$. Following Burgstaller and Pillichshammer (2009), we define

$$\begin{aligned} a(D) &= \mathbb{E}[||\boldsymbol{x} - \boldsymbol{y}||], \\ d(D) &= \max\{||\boldsymbol{x} - \boldsymbol{y}|| : \boldsymbol{x}, \boldsymbol{y} \in D\}, \end{aligned}$$

where $||\cdot||$ denotes the Euclidean distance and $\boldsymbol{x} = (x_1, x_2, \ldots, x_r), \boldsymbol{y} = (y_1, y_2, \ldots, y_r) \in D$. Assume that we choose $\boldsymbol{x}$ and $\boldsymbol{y}$ uniformly and independently from D. The problem of finding $a(D)$ as a function of the dimension r was already stated in the literature (Alagar, 1976; Dumbar, 1997). For example, for all compact subsets of $\mathbb{R}$, $a(D) = d(D)/3$. If $D \subseteq \mathbb{R}^r$ with diameter $d(D)$, then

$$a(D) = \frac{r}{2r+1}\beta_r d(D), \tag{21}$$

where

$$\beta_r = \begin{cases} \frac{2^{3r+1}((r/2)!)^2 r!}{(r+1)(2r)!\pi}, & \text{for even } r, \\ \\ \frac{2^{r+1}(r)!^3}{(r+1)(((r-1)/2)!)^2(2r)!}, & \text{for odd } r. \end{cases}$$

Moreover, for specific geometric figures, the distance $d(D)$ has been calculated explicitly. In fact, if $D \subseteq \mathbb{R}^2$ is a rectangle of sides $a \geq b$ or if D is a cube in $\mathbb{R}^r$, then $a(D)$ can also be written as a function of $d(D)$ (Burgstaller and Pillichshammer, 2009). These authors also provided bounds for the average distance between two points that are uniformly and independently chosen from a compact subset of the r-dimensional Euclidean space.

Philip (2007) determined the probability distribution for the distance between two random points in a box, including the expected distance, which is known as Robbins' constant. Similar calculations for the expected value of the distance between two randomly defined points on circles and rectangles in $\mathbb{R}^2$, in the context of random networks, can be found in Moltchanov (2012). Now, let $\boldsymbol{x}$ and $\boldsymbol{y}$ be two independent random vectors uniformly distributed on the set $B := B(0,1) = \{\boldsymbol{z} \in \mathbb{R}^r : |\boldsymbol{z}| \leq 1\}$. That is,

$$f(\boldsymbol{x}) = \begin{cases} \frac{\Gamma(r/2+1)}{\pi^{r/2}}, & \text{if } \boldsymbol{x} \in B(0,1) \\ 0, & \text{otherwise} \end{cases}$$

A formula for the expected value of the square of the distance between $\boldsymbol{x}$ and $\boldsymbol{y}$, $\mathbb{E}[||\boldsymbol{x}-\boldsymbol{y}||^2]$, is (see Appendix)

$$\mathbb{E}[||\boldsymbol{x}-\boldsymbol{y}||^2] = \frac{2r}{r+2}. \tag{22}$$

The expressions (21) and (22) for $a(D)$ and $\mathbb{E}[||\boldsymbol{x}-\boldsymbol{y}||^2$, respectively, are increasing functions of the dimension r. As a result, the distributions of the expressions described in (21) and (22) stochastically increase in dimension. Hence, for a fixed $\boldsymbol{\theta}$, $\rho(\boldsymbol{s}_i, \boldsymbol{s}_j, \boldsymbol{\theta})$ stochastically decreases in dimension. On average, the correlation grows weaker as the dimension grows larger, and the effective sample size increases in r.

7. Simulations

To illustrate how our method uncovers the effective sample size, we considered several numerical experiments; two of them are reported here. The first was designed to explore the behavior of the ESS for different parametric covariance models. The second experiment compares our proposal with Griffith's proposal and the integral range.

7.1. Simulation Study 1

We used Monte Carlo simulation to address the following questions: 1. How does the effective sample size change with different covariance structures (models)? 2. How does the nugget effect impact the estimation of the effective sample size? 3. What is the performance of the estimated effective sample size in terms of bias and variance when variogram model parameters vary? One hundred points were fixed in the region $[0,1] \times [0,1]$, and random samples were generated from a zero-mean Gaussian distribution with covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = C(||\boldsymbol{s}_i - \boldsymbol{s}_j||, \boldsymbol{\theta})$ such that

$$C(h, \boldsymbol{\theta}) = \begin{cases} \tau^2 + \sigma^2, & h = 0, \\ \sigma^2 \rho(h), & h > 0, \end{cases}$$

and the elements of the correlation matrix are

$$\boldsymbol{R}(\boldsymbol{\theta})_{ij} = \begin{cases} 1, & i = j, \\ \frac{\sigma^2 \rho(h)}{\sigma^2 + \tau^2}, & i \neq j. \end{cases}$$

The correlation functions considered in the study are listed in Table 1.

Model	Correlation Function
Spherical	$\rho(h) = \begin{cases} 1 - 1.5\frac{h}{\phi} + 0.5\left(\frac{h}{\phi}\right)^3, & h < \phi \\ 0, & \text{otherwise.} \end{cases}$
Exponential	$\rho(h) = \exp\left(-\frac{h}{\phi}\right)$
Matérn	$\rho(h) = \frac{1}{2^{\kappa-1}\Gamma(\kappa)}\left(\frac{h}{\phi}\right) K_\kappa\left(\frac{h}{\phi}\right)$
Gaussian	$\rho(h) = \exp\left[-\left(\frac{h}{\phi}\right)^2\right]$

Table 1: Correlation models used in the study. In the Matern model $\Gamma(\kappa)$ is the gamma function and K_κ is the modified Bessel function of the second kind. The exponential model is a particular case of the Matérn function for $\kappa = 1/2$ and the Gaussian model corresponds to the Matérn function for $\kappa \to \infty$.

For all models, simulations were generated using $\sigma^2 = 1$, $\tau^2 \in \{0.01, 0.1, 1, 10\}$, and $\phi \in \{0.1, 0.5, 1\}$. For the Matérn model $\kappa = \{0.5, 1, 1.5, 2.5\}$, all κ values were estimated for starting values belonging to the set $\{0.5, 1, 2\}$. Our results are based on 500 replicates [2]. In each run, the REML estimator of $\boldsymbol{\theta} = (\sigma^2, \tau^2, \phi, \kappa)^\top$ and the estimation of the effective sample size given in (19) were calculated. Based on these 500 replicates, the standard deviation of the ESS values were also recorded. The results are shown in Table 2.

Table 2 shows the model parameters, theoretical value of the effective sample size and its estimates and standard deviations. We observe an increase in the ESS when the nugget effect increases. This agrees with the fact that a spatial sequence associated with a pure nugget effect model is a white noise process for which ESS $= n$, as observed in several entries in Table 2 for $\tau = 10$. In most cases we considered, the standard deviation associated with the estimation of the ESS also increases when the nugget effect increases; the exception being $\tau = 10$, in which the maximum sample size is attained. Additionally, for each covariance model, the ESS decreases as the range increases, as expected. Although there are no clear monotonic properties when the smoothing parameter κ increases, we found that the largest bias and standard deviation for the estimates occurred for the Matérn model with $\kappa \in \{1, 1.5, 2.5\}$. This behavior may be explained by the fact that in the Matérn class, κ was also estimated via REML, increasing the complexity of the overall process because the REML estimator usually involves a highly nonlinear system of equations. Additionally, in Table 2 we observe that the standard deviation of the estimates decreases as ϕ increases only for the exponential model.

7.2. Simulation Study 2

In this section, we carried out another Monte Carlo simulation to compare the performance of the three measures of effective sample size reviewed in Section 3; Griffith's proposal (n^*), our proposal (the ESS), and effective sample size based on the integral range

[2] We performed Monte Carlo experiments increasing the number of simulation replicates to 1,000. No pattern changes were observed.

(N). We used a sample size of 360 to compare these three methods because the estimation of N requires us to split the region into small windows (Lantuejoul, 2010, p. 32-34). Consequently, the chosen sample size was larger to ensure that each window had enough observations. The random samples were generated from a zero-mean Gaussian distribution with three different covariance functions (exponential, spherical and Gaussian), which had the same functional form as in the first simulation study. For all models, simulations were generated for $\sigma^2 = 1$, $\tau^2 \in \{0.01, 0.1, 1, 10\}$ and $\phi \in \{0.1, 0.5, 1\}$. Estimation of the covariance parameters was addressed via an REML estimation for n^* and the ESS, and an estimate of N was obtained using a method suggested in (Lantuejoul, 2010, p. 32-34). Our results are based on 500 replicates for which the three estimates of effective sample size were calculated. Standard deviations of the ESS values were obtained from these 500 replicates.

There is no theoretical value for the estimate of N; therefore, it is not possible to compare all three coefficients in terms of bias. Table 3 shows that in some cases, $\widehat{n}^*$ and $\widehat{\text{ESS}}$ overestimate the true values. This is also observed in the first simulation study and is associated with the sampling design, which in this case is closer to a cluster pattern because the observed area ($[0,1] \times [0,1]$) was fixed for both studies, but the sample size was larger in the second study. Clearly, the estimates obtained for n^* and the ESS are comparable in terms of bias and precision, but neither is comparable with those obtained for N, which in most cases are much greater than n. Although the nugget effect has a clear impact on the ESS, in all cases $\widehat{\text{ESS}}$ reaches the sample size n, which is also observed for $\widehat{n}^*$. Moreover, both estimators increase with τ. The correlations $\rho_1 = \text{cor}(\widehat{\text{ESS}}, \widehat{n}^*)$ and $\rho_2 = \text{cor}(\widehat{\text{ESS}}, \widehat{N})$ are displayed in the last two columns of Table 3. Additionally, we explored the linear association between $\widehat{\text{ESS}}$ and $\widehat{n}^*$. In all cases, $\rho_1 = 0.99$ and $0 \leq \rho_2 \leq 0.44$, which highlights both the strong linear association between $\widehat{\text{ESS}}$ and $\widehat{n}^*$ and the poor linear relationship between $\widehat{\text{ESS}}$ and $\widehat{N}$. For the spherical and exponential cases, $\widehat{n}^*$ has smaller variance than $\widehat{\text{ESS}}$, while the variance of $\widehat{N}$ is much larger than the other two estimators. Overall, the estimates of n^* and the ESS provide comparable results but differ completely from the approach based on the integral range. This may be due to the fact that the estimation procedure for N is not based on the statistical inference of a normal model in contrast to the other two estimation methods.

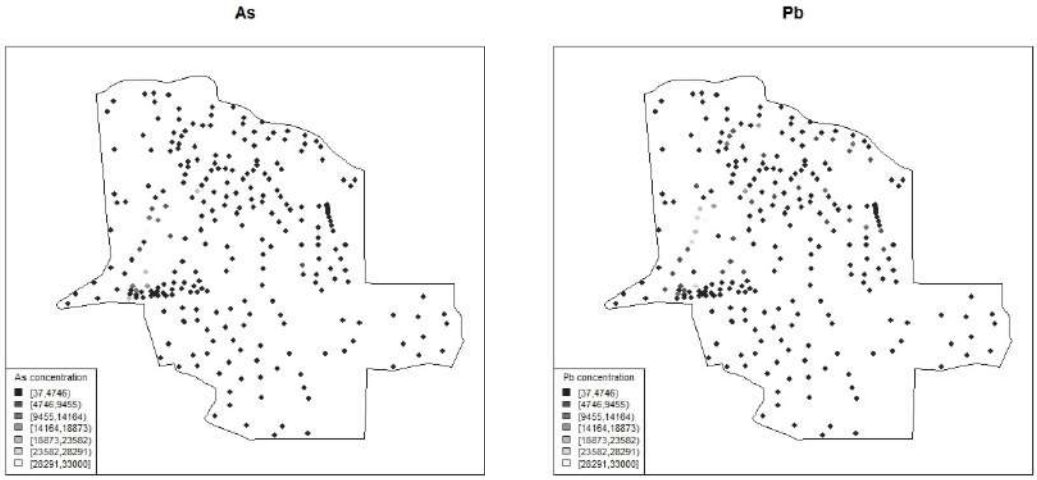
8. Real Data Examples

In this section, we discuss two examples with real data to illustrate practical applications of this work. The first example uses the well-known Murray smelter site dataset in which the variables of interest are arsenic (As) and lead (Pb). The second example pertains to a forest inventory where several variables concerning the trees and terrain have been measured. These two examples illustrate the calculation of the effective sample size.

8.1. The Murray Smelter Site Dataset

The dataset consists of soil samples collected in and around the vacant, industrially contaminated, Murray smelter site (Utah, USA). This area was polluted by airborne emissions and the disposal of waste slag from the smelting process. A total of 253 locations

were included in the study, and soil samples were taken from each location. Each georeferenced sample point is a pool composite of four closely adjacent soil samples in which the concentration of the heavy metals arsenic (As) and lead (Pb) was determined. A detailed description of this dataset can be found in Griffith and Paelinck (2011), and a plot of the locations appears in Figure 5 (a). For each location, the As and Pb attributes are shown in Figure 5 (b) and (c).



(a) (b)

Figure 5: Locations of 253 geocoded aggregated surface soil samples collected in a 0.5 square mile area of Murray, Utah, and their concentrations of As and Pb measured. Of these 173 were collected in a facility superfund site, and 80 were collected in two of its adjacent residential neighborhood located along the western and southern borders of the smelter site. (a) As measurements; (b) Pb measurements.

Recently, a study of the spatial association between As and Pb for the Murray dataset was conducted by Vallejos et al. (2013). They found certain directions for which the spatial association can be quantified through the codispersion coefficient.

Nine semivariogram models were fit to As and Pb, both of which had previously been transformed by the Box-Cox function to achieve symmetry. The estimates of the parameters of the variograms, mean square error (MSE), and effective sample sizes for both variables are reported in Table 4. The effective sample sizes here range from 15.60 to 50.09 for As and 14.89 to 101.6 for Pb. Although the effective sample sizes are of the same order of magnitude as those provided by Griffith (2005), they tend to be noticeably smaller. The best fit (using the MSE) for both variables was achieved with the spherical model with effective sample sizes of 41.20 and 87.65, respectively. For the Matérn models, we used the following initial values for the smoothing parameter: $\kappa = 0.5.$, $\kappa = 1$, and $\kappa = 2$. The effective sample sizes for the last two Matérn models do not differ, whereas the first Matérn model represents the worst fit in both cases; it has very different MSE values than the other models. Table 4 shows that in both cases the worst fit has the smallest nugget effect value, which produces a small effective sample size for As and Pb. The values are consistent with what would be obtained with a time series for which $\rho = 0.52180$ for

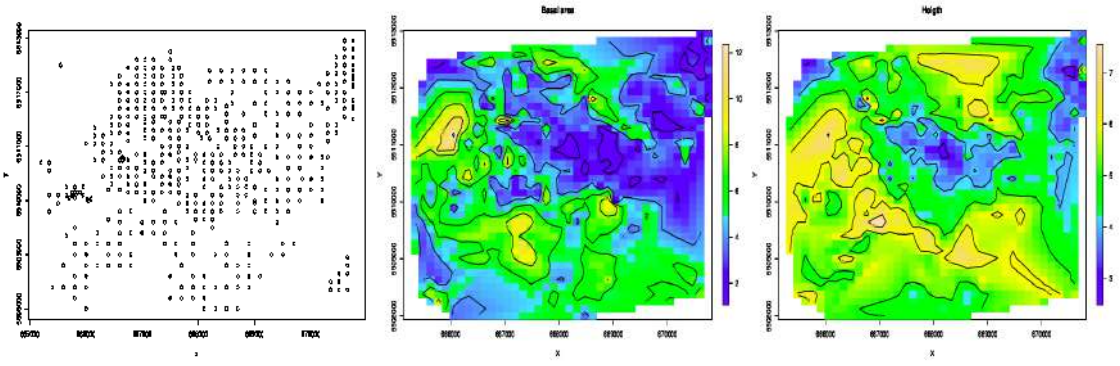
As and $\rho = 0.49363$ for Pb, as follows:

$$\text{As} : \text{ESS}_{\text{AR}} = 78.02,$$
$$\text{Pb} : \text{ESS}_{\text{AR}} = 86.43.$$

In the As case, the effective sample size, which was computed by fitting a variogram model, is smaller than the sample size produced by assuming an autoregressive correlation structure.

8.2. The Pinus radiata Dataset

Pinus radiata is one of the most widely planted species in Chile and is planted in a wide array of soil types and regional climates. Two important measures of plantation development are dominant tree height and basal area; research shows that both these measures are correlated with the regional climate and local growing conditions (Snowdon, 2001). The study site is located in the *Escuadrón* sector, south of Concepción, in the southern portion of Chile (36° 54' S, 73°54' O) and has an area of 1244.43 hectares. Aside from mature stands, we were also interested in areas that contain young (i.e., four years old) stands of Pinus radiata. These areas have an average density of 1600 trees per hectare. The basal area and dominant tree height in the year of the plantation's establishment (1993, 1994, 1995, and 1996) were used to represent stand attributes. These three variables were obtained from 200 m^2 circular sample plots and point-plant sample plots. For the latter, four quadrants are established around the sample point; the four closest trees in each quadrant (16 trees in total) are then selected and measured. The samples were located systematically using a mean distance of 150 meters between samples. The total number of plots available for this study was 468 (Figure 6a). Figure 6 shows a simple bilinear interpolation and the corresponding contours for the two variables.



(a) (b) (c)

Figure 6: (a) Locations; (b) Bilinear interpolation of the three basal area; (c) Bilinear interpolation of the three height.

The spatial correlation between the variables that correspond to the Pinus radiata dataset has been addressed by using a Nadaraya-Watson version of the codispersion coefficient in Cuevas et al. (2013). The spatial association between pairs of variables was visualized through a graphical tool called a codispersion map.

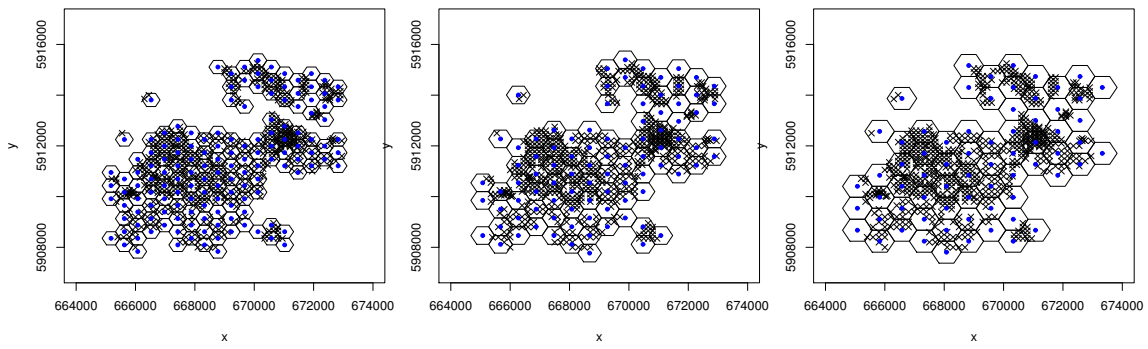
As before, several variogram models were fit to the Pinus radiata dataset, which contains each tree's basal area and height. We also transformed these variables using the Box-Cox function to achieve symmetry. Summaries of the parameter estimates, MSE, and effective sample size can be found in Table 5. To provide a measure of the uncertainty of the effective sample size for these data, we implemented the block bootstrap introduced by Sherman (1996) for statistics calculated from a spatial lattice. The spatial locations shown in Figure 6(a) were divided into three, four, and six non-overlapping blocks to ensure that there would be enough points to estimate the variance of the effective sample size in each block. The results are displayed in columns 8-10 of Table 5. The effective sample sizes in this case range from 8.64 to 58.33 for the basal area and 7.60 to 28.76 for height. The exponential and spherical covariance functions achieved the best fit for both variables using the MSE. The block bootstrap estimations for the variance of the effective sample size displayed in Table 5 show large variability that depends on the block size, making interpretation difficult and comparison of this technique with others problematic.

9. Sampling Schemes

In Section 5, we noted that sampling design affects sample size calculations. Here, we briefly discuss how to select the samples once the effective sample size has been determined. A recent guide with several methods and discussions can be found in Müller (2010). It is desirable for the selected observations to preserve the original correlation structure. This can be tested by measuring the difference between estimates of the covariance parameters before and after selecting the final sample. Another relevant case is when a pilot study with enough information to achieve a high-quality estimation of the variogram is available. Because there are many different configurations for the original sampling region, a stratified sampling design seems to be a suitable method for the analysis of spatial data (Gelfand et al., 2010). One way to cover the whole region is by using tessellation, as conducted by Griffith (2005) (see also Overton and Stehman, 1993) with a method involving hexagonal tessellation. The radius of the desired hexagon can be approximated as a function of the area of the landscape to be sampled. Then, a starting point is generated, which is the origin of a sequence of hexagon centroids. Consequently, the hexagons are generated with a standard Thiessen polygon algorithm. For the Pinus radiata dataset, the hexagonal tessellations, centroids, and sample locations are shown in Figure 7.

10. Conclusions and Final Remarks

We have proposed using the Fisher information quantity about the mean in a single-mean linear spatial regression model to assess sample size reduction due to the effect of spatial autocorrelation. Additionally, the sample size reduction (i.e., the reduction of the effective sample size) has been extended to the most common process models in spatial statistics. These include the CAR and SAR models and the single-mean linear process with several different patterned correlation matrices. Furthermore, we discussed the dependent case of an elliptically distributed process. A brief review with alternative measures to calculate the effective sample size and examples supplemented our discussion.



(a) (b) (c)

Figure 7: (a) Hexagonal tesselation with radius 300 m; (b) Hexagonal tesselation with radius 400 m ; (c) Hexagonal tesselation with radius 500 m. The locations are denoted by crosses (x) and the centroids are denoted by solid circles (•), for the forestry dataset.

The estimation of the parameters in the correlation structure was handled via REML, which in the elliptical case coincides with the normal distribution when the degrees of freedom are known. In this case, the computation of the effective sample size only differs from the normal case by a constant. The estimation of the degrees of freedom in the elliptical case and the selection of spatial models when a single replicate is available are open problems that deserve further research.

The developments for random locations discussed in Section 6 and the simulation studies described in Section 7 support our proposal and highlight the superior performance of the ESS with respect to other methods. Furthermore, our results enable extending these ideas to more general contexts. Of particular interest is the extension of (5) for processes of the form $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{R}(\boldsymbol{\theta}))$. However, there are some restrictions on the number of independent variables that need to be taken into account. For example, $\dim(\boldsymbol{\beta}) + \dim(\boldsymbol{\theta}) << n$; therefore, no replicates are needed in the estimation process. The use of an information matrix to extend the notion of the effective sample size is an open problem for future research.

There is a relationship between the ESS and the variance of an estimate of the spatial mean described by the quantity

$$M(A) = \frac{\int_A Y(\boldsymbol{s}) d\boldsymbol{s}}{\int_A d\boldsymbol{s}},$$

where $A \subset D$. The usual best linear unbiased (BLU) estimator is $\tilde{\mu} = \boldsymbol{\lambda}_\mu \boldsymbol{Y}(\boldsymbol{s})$, where $\boldsymbol{\lambda}_\mu = (\mathbf{1}^\top \boldsymbol{R}\, \mathbf{1})^{-1} \mathbf{1}^\top \boldsymbol{R}^{-1}$ represents the vector of optimal weights. In this kriging context, Barnes (1988) and de Gruijter and ter Braak (1990) have derived the variance of $\tilde{\mu}$, which is

$$\text{var}(\tilde{\mu}) = (\mathbf{1}^\top \boldsymbol{R}\, \mathbf{1})^{-1}.$$

An apparent paradox arises here because if the ESS increases, the kriging variance decreases. Therefore, instead of reducing the sample size, we would try to increase the

sample size to obtain a smaller kriging variance. Nonetheless, we do not describe how to reduce kriging variance; instead, we only introduce the ESS to quantify the equivalent number of independent observations that occur due to the spatial correlation that is present in the dataset. This phenomenon occurs when the main goal is to estimate an overall mean μ in a single-mean spatial linear model, such as the one described in Definition 1.

In Section 8.2, a block bootstrap approach was used to calculate the uncertainty measures of the estimates. Although Sherman's approach is valid for statistics calculated from a rectangular lattice, no definite proposals exist in a general spatial setting. Further research should be developed to obtain confidence regions for the effective sample size.

The use of line transects to observe animal or plant species is very common in several areas of study (Buckland et al., 1992), and the questions addressed in this paper can be reformulated in this context. If we have n observations that have been taken over a transect line, it is important to determine the effective sample size for this dataset when the goal is to estimate an overall mean over that transect line. To resolve this issue, it is necessary to define a process over the transect that includes an appropriate correlation structure for the observations. If information is available for two variables defined in a region that includes the transect line of interest, the codispersion coefficient (Rukhin and Vallejos, 2008) can be used to quantify the spatial association in the particular direction of the transect line.

11. Acknowledgements

Ronny Vallejos was partially supported by Fondecyt grant 1120048, Chile and and from UTFSM grant 12.12.05. Felipe Osorio was partially supported by CONICYT grant 791100007 and from UTFSM grant 12.12.05. We are indebted to Alan Gelfand for helpful discussions and for initially sugesting this topic of research to Ronny Vallejos while the latter was a graduate student at the University of Conncecticut. We also thank Manuel Galea, Oscar Bustos, Renato Assunção, Moreno Bevilacqua and Consuelo Moreno for helpful discussions.

Appendix

Matrices of Example 2

The inverse of $\boldsymbol{R}(\rho_1, \rho_2, \ldots, \rho_{n-1})$ in (6) is the tri-diagonal matrix given by

$$\boldsymbol{R}(\rho_1, \rho_2, \ldots, \rho_{n-1})^{-1} = \begin{pmatrix} \frac{1}{1-\rho_1^2} & -\frac{\rho_1}{1-\rho_1^2} & 0 & \cdots & 0 \\ -\frac{\rho_1}{1-\rho_1^2} & \frac{1-\rho_1^2\rho_2^2}{(1-\rho_1^2)(1-\rho_2^2)} & -\frac{\rho_2}{(1-\rho_2^2)} & \cdots & 0 \\ 0 & -\frac{\rho_2}{(1-\rho_2^2)} & \ddots & \ddots & 0 \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \frac{1-\rho_{n-2}^2\rho_{n-1}^2}{(1-\rho_{n-2}^2)(1-\rho_{n-1}^2)} & -\frac{\rho_{n-1}}{(1-\rho_{n-1}^2)} \\ 0 & \cdots & 0 & -\frac{\rho_{n-1}}{(1-\rho_{n-1}^2)} & \frac{1}{(1-\rho_{n-1}^2)} \end{pmatrix}.$$

If $\rho_1 = \rho_2 = \ldots = \rho_{n-1} = \rho$ the correlation matrix of an AR(1) process is

$$\boldsymbol{R}_n(\rho) = \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{pmatrix}.$$

The inverse is given by

$$\boldsymbol{R}_n^{-1}(\rho) = \begin{cases} 1/(1-\rho^2), & \text{if } i = j = 1, n, \\ (1+\rho^2)/(1-\rho^2), & \text{if } i = j = 2, \cdots, n-1, \\ -\rho/(1-\rho^2), & \text{if } |j-i| = 1, \\ 0 & \text{otherwise.} \end{cases}$$

and the effective sample size is

$$\text{ESS}_{\text{AR}} = (2 + (n-2)(1-\rho))/(1+\rho).$$

Proof of Proposition 1

We will use a subscript on $\boldsymbol{R}(\boldsymbol{\theta})$, $\mathbf{1}$, and ESS to emphasize the dependency of these quantities on the sample size n.

To prove i) it is enough to show that $\text{ESS}_{n+1} - \text{ESS}_n \geq 0$, for all $n \in \mathbb{N}$. First, we define the $(n+1) \times (n+1)$ correlation matrix $R_{n+1}(\boldsymbol{\theta})$ as follows

$$\boldsymbol{R}_{n+1}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{R}_n(\boldsymbol{\theta}) & \boldsymbol{\gamma} \\ \boldsymbol{\gamma}^\top & 1 \end{pmatrix},$$

where $\boldsymbol{\gamma}^\top = (\gamma_1, \gamma_2, \cdots, \gamma_n), 0 \leq \gamma_i \leq 1,$. Since $\boldsymbol{R}_{n+1}(\boldsymbol{\theta})$ is positive definite, the Schur complement $(1 - \boldsymbol{\gamma}^\top \boldsymbol{R}_n(\boldsymbol{\theta})^{-1}\boldsymbol{\gamma})$ of $\boldsymbol{R}_n(\boldsymbol{\theta})$ is positive definite (Harville 1997, p. 244). Thus $(1 - \boldsymbol{\gamma}^\top \boldsymbol{R}_n(\boldsymbol{\theta})^{-1}\boldsymbol{\gamma}) > 0$. Then

$$\begin{aligned} \text{ESS}_{n+1} &= \mathbf{1}_{n+1}^\top \boldsymbol{R}_{n+1}(\boldsymbol{\theta})^{-1} \mathbf{1}_{n+1} = \mathbf{1}_{n+1}^\top \begin{pmatrix} \boldsymbol{R}_n(\boldsymbol{\theta}) & \boldsymbol{\gamma} \\ \boldsymbol{\gamma}^\top & 1 \end{pmatrix}^{-1} \mathbf{1}_{n+1} \\ &= \text{ESS}_n + \frac{(\mathbf{1}_n^\top \boldsymbol{R}_n(\boldsymbol{\theta})^{-1}\boldsymbol{\gamma})^2 - 2\mathbf{1}_n^\top \boldsymbol{R}_n(\boldsymbol{\theta})^{-1}\boldsymbol{\gamma} + 1}{1 - \boldsymbol{\gamma}^\top \boldsymbol{R}_n(\boldsymbol{\theta})^{-1}\boldsymbol{\gamma}}, \end{aligned}$$

where $\mathbf{1}_{n+1}^\top = (\mathbf{1}_n \ 1)^\top$. Since the function $f(x) = x^2 - 2x + 1 = (x-1)^2 \geq 0$, for all x, we have that $\text{ESS}_{n+1} - \text{ESS}_n \geq 0$, for all $n \in \mathbb{N}$. □

To prove ii) we use the Cauchy-Schwartz inequality for matrices. Let us denote $\boldsymbol{R}_n(\boldsymbol{\theta}) = (r_{ij})$. Then $n^2 \leq (\mathbf{1}_n^\top \boldsymbol{R}_n(\boldsymbol{\theta})\mathbf{1}_n)(\mathbf{1}_n^\top \boldsymbol{R}_n(\boldsymbol{\theta})^{-1}\mathbf{1}_n)$. Since $0 \leq r_{ij} \leq 1$, one has $\mathbf{1}_n^\top \boldsymbol{R}_n \mathbf{1}_n = n + \sum_{j=1}^n \sum_{i \neq j} r_{ij} \leq n + n(n-1)$. Hence $\mathbf{1}_n^\top \boldsymbol{R}_n(\boldsymbol{\theta})\mathbf{1}_n \leq n^2$, and this implies that $\mathbf{1}_n^\top \boldsymbol{R}_n(\boldsymbol{\theta})^{-1}\mathbf{1}_n \ \geq 1$. □

Proof of Proposition 3

Equation (13) can be derived from the following facts:

$$\begin{aligned}
\boldsymbol{R}_{\text{SAR}}^{-1} &= \boldsymbol{C}\boldsymbol{\Sigma}_{\text{SAR}}^{-1}\boldsymbol{C} = \frac{1}{\sigma^2}\boldsymbol{C}(I-\rho\boldsymbol{W}^\top)(I-\rho\boldsymbol{W})\boldsymbol{C} \\
&= \frac{1}{\sigma^2}\boldsymbol{C}(I-\rho\boldsymbol{W}-\rho\boldsymbol{W}^\top+\rho^2\boldsymbol{W}^\top\boldsymbol{W})\boldsymbol{C}, \\
\rho\mathbf{1}^\top\boldsymbol{C}\boldsymbol{W}\boldsymbol{C}\mathbf{1} &= \rho\mathbf{1}^\top\boldsymbol{C}\boldsymbol{W}^\top\boldsymbol{C}\mathbf{1} = \rho\sum_i\sum_j\sigma_i\sigma_j w_{ij}, \\
\rho^2\mathbf{1}^\top\boldsymbol{C}\boldsymbol{W}^\top\boldsymbol{W}\boldsymbol{C}\mathbf{1} &= \rho^2\sum_i\sum_j\sum_k\sigma_i\sigma_j w_{ki}w_{kj}.
\end{aligned}$$

□

Derivation of Equation (16)

The required expectations can be obtained by direct integration. Note that the density of the random variable $U = \|\boldsymbol{Z}\|^2$ is given by (Fang et al., 1990, p. 36)

$$h(u) = \frac{\pi^{n/2}}{\Gamma(n/2)}\,u^{n/2-1}g(u) = \frac{\lambda}{\Gamma(\frac{n}{2\lambda})2^{n/(2\lambda)}}u^{n/2-1}\exp(-\tfrac{1}{2}u^\lambda), \quad u>0.$$

Moreover, for $\boldsymbol{Y}\sim PE_n(\boldsymbol{\mu},\boldsymbol{\Sigma},\lambda)$ we have $W_g(u) = -\frac{1}{2}\lambda u^{\lambda-1}$, with $\lambda\neq 1$. Thus, we need to compute

$$\mathbb{E}(W_g^2(U)U) = \frac{\lambda^2}{4}\mathbb{E}(U^{2\lambda-1}).$$

Using that $\int_0^\infty h(u)\,\mathrm{d}u = 1$, one gets

$$\begin{aligned}
\mathbb{E}(U^{2\lambda-1}) &= \frac{\lambda}{\Gamma(\frac{n}{2\lambda})2^{n/(2\lambda)}}\int_0^\infty u^{(4\lambda+n-2)/2}\exp(-\tfrac{1}{2}u^\lambda)\,\mathrm{d}u \\
&= \frac{\Gamma(\frac{4\lambda+n-2}{2\lambda})}{\Gamma(\frac{n}{2\lambda})}\,2^{2-1/\lambda},
\end{aligned}$$

from where the result can be obtained.

Derivation of Equation (22)

Since $\boldsymbol{x}$ and $\boldsymbol{y}$ are independent, we have

$$\mathbb{E}[||\boldsymbol{x}-\boldsymbol{y}||^2] = \mathbb{E}[||\boldsymbol{x}||^2] + \mathbb{E}[||\boldsymbol{y}||^2].$$

By symmetry, we compute $\mathbb{E}[||\boldsymbol{x}||^2]$, which is

$$\mathbb{E}[||\boldsymbol{x}||^2] = C_r^{-1}\int_B\int_B||\boldsymbol{x}||\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y} = C_r^{-1}\int_B||\boldsymbol{x}||\,\mathrm{d}\boldsymbol{x},$$

where $C_r^{-2} = \frac{\pi^{r/2}}{\Gamma(r/2+1)}$. To obtain $\int_B||\boldsymbol{x}||\,\mathrm{d}\boldsymbol{x}$, we use the formula for the volume of an r-dimensional sphere with radius ϱ which is given by

$$V_r(\varrho) = C_r\varrho^r.$$

Using radial symmetry one obtains

$$\int_B ||\boldsymbol{x}|| \, \mathrm{d}\boldsymbol{x} = \int_0^1 \rho 2 \, \mathrm{d}V_r = C_r \int_0^1 r\rho^2 \rho^{r-1} \, \mathrm{d}\rho = rC_r \int_0^1 \rho^{r+1} \, \mathrm{d}\rho = \frac{rC_r}{r+2}.$$

This implies that

$$\mathbb{E}[||\boldsymbol{x} - \boldsymbol{y}||^2] = \frac{2r}{r+2}.$$

References

Alagar, V. S., 1976. The distribution of the distance between random points. Journal of Applied Probability 13, 558-566.

Anderson, T.W., Fang, K.T., Hsu, H., 1986. Maximum-likelihood estimates and likelihood-ratio criteria for multivariate elliptically contoured distributions. The Canadian Journal of Statistics 14, 55-59.

Anselin, L., 1988. Spatial econometrics: Methods and models. Dordrecht, the Netherlands: Martinus Nijhoff.

Banerjee, S., Carlin, B., Gelfand, A. E., 2004. Hierarchical Modeling and Analysis for Spatial Data. Chapman Hall/CRC, Boca Raton.

Barnes, R. J., 1988. Bounding the require sample size for geologic site characterization. Matematical Geology 20, 477-489.

Bayley, G. V., Hammersley J. M., 1946. The "Effective" Number of Independent Observations in an Autocorrelated Time Series. Journal of the Royal Statistical Society B 8, 184-197.

Bogaert, P., Russo, D., 1999. Optimal spatial sampling design for the estimation of the variogram based on a least squares approach. Water Resources research 35, 1275-1289.

Breusch, T.S., Robertson, J.C., Welsh, A.H., 1997. The emperor's new clothes: A critique of the multivariate t regression model. Statistica Neerlandica 51, 269-286.

Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., 1992. Distance Sampling: Estimating Abundance of Biological Populations. Chapman & Hall, London.

Burgstaller B., Pillichshammer, F., 2009. The average distance between two points. Bulletin of the Australian Mathematical Society 80, 353-359.

Clifford, P., Richardson, S., 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. Wiley, New York.

Cressie, N., Lahiri, S. N., 1993. Asymptotic distribution of REML estimators. Journal of Multivariate Analysis 45, 217-233.

Cressie, N., Lahiri, S. N., 1996. Asymptotics for REML estimation of spatial covariance parameters. Journal of Statistical planning and inference 50, 327-341.

Cuevas, F., Porcu, E., Vallejos, R., 2013. Study of spatial relationships between two sets of variables: A nonparametric approach. Journal of Nonparametric Statistics 25, 695-714.

de Gruijter, J. J., ter Braak, C. J. F., 1990. Model-free estimation from spatial samples: A reappraisal of classical sampling theory. Mathematical Geology 22, 407-415.

Dumbar, S. R., 1997. The average distance between points in geometric figures. College Math. J. 28, 187-197.

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

Duttilleul, P. Pelletier, B. 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

Fang, K.T., Kotz, S., Ng, K.W., 1990. Symmetric Multivariate and Related Distributions. Chapman & Hall, London.

Fernández, C., Steel, M.F.J., 1999. Multivariate Student-t regression models: Pitfalls and inference. Biometrika 86, 153-177.

Gelfand, A., Digle, P., Fuentes, M., Guttorp, P., 2010. Handbook of Spatial Statistics. Chapman & Hall/CRC, Boca Raton.

Getis, A., Ord, J., 2000. Seemingly independent tests: Addressing the problem of multiple simultaneous and dependent tests. Paper presented at the 39th Annual Meeting of the Western Regional Science Association, Kanuai, HI, 28 February.

Gómez, E., Gómez-Villegas, M.A., Marín, J.M., 1998. A multivariate generalization of the power exponential family of distributions. Commun.Statist. Theory Methods 27, 589-600.

Graybill, F., 2001. Matrices with Applications in Statistics. Duxbury Classic series.

Griffith, D., Wong, D., and Whitfield, T. T., 2003. Exploring relationships between the global and regional measures of spatial autocorrelation. Journal of Regional Science 43, 683-710.

Griffith, D., 2005. Effective geographic sample size in the presence of spatial autocorrelation. Annals of the Association of American Geographers 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., Paelinck, J. H. P., 2011. Non-Stardard Spatial Statistics. Springer, New York.

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

Harville, D., 1974. Bayesian inference for variance components using only error contrats. Biometrika 61, 383-385.

Harville, D., 1977. Maximum likelihood approaches to variance component estimation and to related problems. Journal of American Statistics Association 72, 320-340.

Harville, D., 1997. Matrix algebra from a statistician's perspective. Springer, New York.

Lange, K. L., Little, R. J. A., Taylor, M. G., 1989. Robust statistical modeling using the t distribution. Journal of the American Statistical Association 84, 881-895.

Lantuejoul, C., 1991. Ergodicity and Integral Range. Journal of Microscopy 161, 387-403.

Lantuejoul, C., 2010. Geostatistical Simulation Models and Algorithms. Springer, Berlin.

Lark, R. M., 2002. Optimized spatial sampling of soil for estimation of the variogram by maximum likelihood. Geoderma 105, 49-80.

Lindsey, J. K., 1999. Multivariate elliptically contoured distributions for repeated measurements. Biometrics 55, 1277-1280.

Magnus, J. R., Neudecker, H., 1999. Matrix Differential Calculus with Applications in Statistics and Econometrics. Wiley & Sons Ltd., West Sussex.

Matalas, N. C., Langbein, W. B., 1962. Information content of the mean. Journal of Geophysical Research 67, 3441-3448.

Moltchanov, D., 2012. Distance distributions in random networks. Ad Hoc Networks 10, 1146-1166.

Müller, W., 2010. Collecting Spatial Data. Springer, Berlin.

Osorio, F., Paula, G. A., Galea, M., 2007. Assessment of local influence in elliptical linear models with longitudinal structure. Computational Statistics & Data Analysis 51, 4354-4368.

Overton, W., Stehman, S., 1993. Properties of designs for sampling continuous spatial resources from a triangular grid. Communications in Statistics: Theory and Methods 22, 2641-2660.

Patterson, H. D., Thompson, R., 1971. Recovery of interblock information when block sizes are unequal. Biometrika 58, 545-554.

Philip, J., 2007. The probability distribution of the distance between two random points in a box. TRITA MAT 7(10).

Rao, C. R., Wu, Y., 2001. On model selection, in: P. Lahiri (Ed.), Model Selection, Lecture Notes-Monograph Series, Institute of Mathematical Statistics, Beachwood, 38, 1-57.

Richardson, S., 1990. Some remarks on the testing of association between spatial processes. In Spatial statistics: Past, present, and future, ed. D. Griffith, 277-309. Ann Arbor, MI: Institute of Mathematical Geography.

Ritter, K., 1996. Asymptotic optimality of regular sequence designs. The annals of Statistics 24, 2081-2096.

Rukhin, A., Vallejos, R., 2008. Codispersion Coefficient for Spatial and Temporal Series. Statistics & Probability Letters 78, 1290-1300.

Russo, D., 1984. Design of an optimal sampliing network for estimating the variogram. Soil Science Society of America Journal 48, 708-716.

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

Sherman, M., 1996. Variance Estimation for Statistics Computed From Spatial Lattice Data. Journal of the Royal Statistical Society 58, 509-523.

Snowdon, P., 2001. Short-term predictions of growth of Pinus radiata with models incorporating indices of annual climatic variation. Forest Ecology and Management 152, 1-11.

Su, Y., Cambanis, S., 1993. Sampling designs for estimation of a random process. Stochastic Processes and their Applications 46, 47-89.

Vallejos, R., Osorio, F., Cuevas, F., 2013. `SpatialPack` - An R package for computing the association between two spatial or temporal processes. Submitted.

Xia, G., Hjort, N. L., Gelfand, A. E., 2006. Information growth and asymptotic Inference under stochastic process models with structures dependence. Personal communication.

Zellner, A., 1976. Bayesian and non-bayesian analysis of the regression model with multivariate Student-t error terms. Journal of the American Statistical Association 71, 400-405.

Zhu, Z., Stein, M., 2005. Spatial sampling design for parameter estimation of the covariance function 134, 583-603.

Model	τ^2	ϕ	κ	ESS	$\widehat{\text{ESS}}$	sd($\widehat{\text{ESS}}$)	Model	τ^2	ϕ	κ	ESS	$\widehat{\text{ESS}}$	sd($\widehat{\text{ESS}}$)
(1)	0,01	0.1	0.5	19.85	22.10	13.89	(4)	0,01	0.1	1	12.23	50.70	33.23
	0,1	0.1	0.5	21.08	24.28	14.48		0,1	0.1	1	12.96	48.63	32.93
	1	0.1	0.5	31.58	42.18	28.96		1	0.1	1	20.23	55.30	35.00
	10	0.1	0.5	69.96	100	0.00		10	0.1	1	55.72	91.02	21.12
	0,01	0.5	0.5	3.18	5.43	4.98		0,01	0.5	1	2.79	19.18	24.21
	0,1	0,5	0.5	3.40	7.03	6.92		0,1	0,5	1	2.37	23.26	26.84
	1	0,5	0.5	5.67	17.21	21.62		1	0,5	1	3.88	48.19	39.74
	10	0,5	0.5	22.84	100	0.00		10	0,5	1	16.29	20.79	22.00
	0,01	1	0.5	1.97	4.41	4.51		0,01	1	1	1.53	11.00	15.65
	0,1	1	0.5	2.11	5.37	5.87		0,1	1	1	1.60	17.91	24.80
	1	1	0.5	3.57	25.58	33.12		1	1	1	2.65	62.98	40.84
	10	1	0.5	15.77	100	0.00		10	1	1	12.18	95.92	15.16
(2)	0,01	0.1	–	68.57	68.93	19.69	(5)	0,01	0.1	1.5	9.34	43.96	34.11
	0,1	0.1	–	69.96	71.65	18.47		0,1	0.1	1.5	9.79	46.21	34.93
	1	0.1	–	79.30	70.52	30.67		1	0.1	1.5	15.44	56.97	34.25
	10	0.1	–	94.93	100	0.00		10	0.1	1.5	47.24	86.14	26.27
	0,01	0.5	–	10.20	11.01	5.97		0,01	0.5	1.5	1.99	12.37	16.36
	0,1	0,5	–	10.80	11.84	6.20		0,1	0,5	1.5	2.03	17.48	23.25
	1	0,5	–	16.94	23.52	21.15		1	0,5	1.5	3.26	44.51	40.19
	10	0,5	–	49.88	100	0.00		10	0,5	1.5	13.89	92.91	19.22
	0,01	1	–	19.85	6.04	4.24		0,01	1	1.5	1.42	4.13	6.66
	0,1	1	–	21.08	7.29	6.03		0,1	1	1.5	1.45	10.68	19.50
	1	1	–	31.58	15.52	18.72		1	1	1.5	2.35	68.40	41.40
	10	1	–	69.96	100	0.00		10	1	1.5	11.13	31.28	37.19
(3)	0,01	0.1	∞	32.28	32.80	12.11	(6)	0,01	0.1	2.5	6.85	32.9	31.03
	0,1	0.1	∞	33.19	34.25	12.09		0,1	0.1	2.5	7.02	44.76	33.55
	1	0.1	∞	44.85	50.20	26.23		1	0.1	2.5	11.07	51.43	35.11
	10	0.1	∞	79.84	100	0.00		10	0.1	2.5	37.47	24.52	27.61
	0,01	0.5	∞	4.45	5.46	1.97		0,01	0.5	2.5	1.77	5.07	5.15
	0,1	0,5	∞	4.21	5.09	2.10		0,1	0,5	2.5	1.78	8.38	15.10
	1	0,5	∞	6.45	12.81	17.90		1	0,5	2.5	2.72	41.51	41.70
	10	0,5	∞	23.72	100	0.00		10	0,5	2.5	11.99	35.24	39.70
	0,01	1	∞	2.52	2.77	0.83		0,01	1	2.5	1.38	3.67	8.33
	0,1	1	∞	2.49	4.28	7.49		0,1	1	2.5	1.32	17.25	28.04
	1	1	∞	3.55	25.85	33.33		1	1	2.5	2.13	67.08	41.81
	10	1	∞	13.70	100	0.00		10	1	2.5	10.50	93.13	19.63

Table 2: ESS $= \mathbf{1}^\top R(\boldsymbol{\theta})^{-1}\mathbf{1}$ and $\widehat{\text{ESS}} = \mathbf{1}^\top R(\widehat{\boldsymbol{\theta}}_{REML})^{-1}\mathbf{1}$. The values reported under the $\widehat{\text{ESS}}$ columns are the simulation averages obtained using 500 runs. sd($\widehat{\text{ESS}}$) corresponds to the simulation standard deviation. Model (1): Exponential; Model (2): Spherical; Model (3): Gaussian; Model (4), (5) and (6): Matérn.

Spherical Model											
τ^2	ϕ	ESS	$\widehat{\text{ESS}}$	sd($\widehat{\text{ESS}}$)	n^*	$\widehat{n}^*$	sd($\widehat{n}^*$)	$\widehat{N}$	sd($\widehat{N}$)	ρ_1	ρ_2
0.01	0.10	139.05	154.56	40.64	123.45	141.36	37.73	2256.80	9235.18	0.99	0.07
0.10	0.10	145.17	157.64	35.82	130.40	145.87	33.22	1206.58	2661.64	0.99	0.07
1.00	0.10	192.31	218.42	57.17	182.95	211.09	57.04	1311.35	4164.02	0.99	0.03
10.00	0.10	307.24	360.00	0.00	306.12	360.01	0.00	381.94	915.89	–	–
0.01	0.50	11.35	12.97	5.32	8.67	10.18	4.83	124.82	341.91	0.99	0.21
0.10	0.50	12.17	13.54	5.72	9.39	10.80	5.24	102.14	238.64	0.99	026
1.00	0.50	20.16	24.48	14.89	16.63	21.03	14.38	204.69	1996.54	0.99	0.00
10.00	0.50	81.42	360.02	0.00	75.82	360.05	0.00	114.72	424.87	–	–
0.01	1.00	4.76	7.33	5.10	3.03	5.44	4.41	269.75	2672.58	0.99	0.09
0.10	1.00	5.06	8.65	7.07	3.39	6.60	6.34	111.28	478.42	0.99	0.32
1.00	1.00	8.48	27.87	60.53	5.98	25.11	60.20	154.65	1214.33	0.99	0.03
10.00	1.00	36.68	360.08	0.00	30.33	360.03	0.00	102.02	523.53	–	–
Exponential Model											
τ^2	ϕ	ESS	$\widehat{\text{ESS}}$	sd($\widehat{\text{ESS}}$)	n^*	$\widehat{n}^*$	sd($\widehat{n}^*$)	$\widehat{N}$	sd($\widehat{N}$)	ρ_1	ρ_2
0.01	0.10	23.96	27.82	13.58	20.67	24.59	12.77	310.81	819.74	0.99	0.19
0.10	0.10	25.76	30.16	16.54	22.45	26,86	15.63	260.63	676,88	0.99	0.18
1.00	0.10	42.61	47.87	32.13	38.73	44,14	31.57	289.65	744,65	0.99	0.28
10.00	0.10	147.05	360.09	0.00	143.48	360,09	0.00	321.24	3568,27	–	–
0.01	0.50	3.47	5.90	5.11	2.60	4.75	4.36	145.32	438.89	0.99	0.36
0.10	0.50	3.69	7.52	7.40	2.88	6.28	6.57	181.36	810.53	0.99	0.20
1.00	0.50	6.26	15.54	20.96	5.06	13.77	19.88	814.38	14317.26	0.99	0.01
10.00	0.50	28.70	360.08	0.00	26.17	360.03	0.00	428.92	4974.51	–	–
0.01	1.00	2.059	4.69	4.92	1.75	3.84	4.18	401.31	3015.71	0.99	0.45
0.10	1.00	2.20	5.86	6.82	1.80	4.87	6.00	230.04	977.68	0.99	0.38
1.00	1.00	3.84	18.66	39.79	3.33	17.11	39.00	297.32	2678.93	0.99	0.10
10.00	1.00	18.40	360.02	0.00	17.28	360.09	0.00	129.57	363.65	–	–
Gaussianl Model											
τ^2	ϕ	ESS	$\widehat{\text{ESS}}$	sd($\widehat{\text{ESS}}$)	n^*	$\widehat{n}^*$	sd($\widehat{n}^*$)	$\widehat{N}$	sd($\widehat{N}$)	ρ_1	ρ_2
0.01	0.10	41.17	43.71	30.24	34.82	38.33	30.38	394.38	810.08	0.99	0.30
0.10	0.10	42.98	45.74	17.70	37.61	40.62	17.39	485.76	1701.27	0.99	0.07
1.00	0.10	67.64	93.19	82.22	63.08	88.93	83.33	476.89	2124.72	0.99	0.12
10.00	0.10	196.01	360.05	0.00	193.80	360.08	0.00	309.27	1851.47	–	–
0.01	0.50	4.87	5.57	1.43	2.55	3.19	1.12	96.26	340.16	0.98	0.23
0.10	0.50	4.75	5.84	2.55	2.72	3.71	2.21	84.63	338.29	0.99	0.21
1.00	0.50	7.28	11.35	12.26	4.94	8.77	11.51	664.50	13002.22	0.99	0.15
10.00	0.50	31.39	360.02	0.00	25.60	360.01	0.00	83.18	330.72	–	–
0.01	1.00	2.76	3.30	1.04	1.49	1.89	0.75	261.13	1324.11	0.98	0.22
0.10	1.00	2.63	4.00	3.44	1.56	2.60	2.90	142.53	458.66	0.99	0.37
1.00	1.00	4.28	20.65	46.75	2.76	18.53	46.12	364.21	2196.59	0.99	0.21
10.00	1.00	16.52	360.04	0.00	14.33	360.01	0.00	136.38	645.65	–	–

Table 3: Study of the behavior of $\widehat{\text{ESS}}$, $\widehat{n}^*$, and $\widehat{N}$. In the last two columns $\rho_1 = \text{cor}(\widehat{\text{ESS}}, \widehat{n}^*)$ and $\rho_2 = \text{cor}(\widehat{\text{ESS}}, \widehat{N})$.

Arsenic (As)						
Model	κ	σ^2	ϕ	τ^2	MSE	$\widehat{\text{ESS}}$
Spherical	–	1.10	1000.61	1.27	0.03	41.24
Exponential	0.5	1.88	286.74	0.90	0.07	44.01
Gaussian	∞	1.29	442.76	1.55	0.05	49.47
Cubic	–	1.36	1104.26	1.55	0.04	47.19
Circular	–	1.64	788.54	1.28	0.04	50.09
Cauchy	–	1.81	143.37	1.11	0.07	15.69
Matern ($\kappa = 0.5$)	0.16	3.08	884.13	0.04	0.63	17.01
Matern ($\kappa = 1$)	0.35	2.12	429.12	0.77	0.09	30.84
Matern ($\kappa = 2$)	0.35	2.12	429.04	0.77	0.09	30.84
Lead (Pb)						
Model	κ	σ^2	ϕ	τ^2	MSE	$\widehat{\text{ESS}}$
Spherical	–	1.28	551.87	0.79	0.03	87.65
Exponential	0.5	1.67	215.07	0.45	0.03	58.42
Gaussian	∞	1.02	286.74	1.04	0.03	83.03
Cubic	–	1.02	673.94	1.03	0.03	87.08
Circular	–	1.26	435.93	0.77	0.03	101.62
Cauchy	–	1.53	143.37	0.79	0.04	14.89
Matern ($\kappa = 0.5$)	0.18	2.55	884.13	0.00	0.43	15.75
Matern ($\kappa = 1$)	0.29	2.02	312.04	0.14	0.09	45.90
Matern ($\kappa = 2$)	0.29	2.02	312.17	0.14	0.09	45.88

Table 4: $\widehat{\text{ESS}}$ for As and Pb for different covariance models. MSE $= \frac{\sum_h (\gamma(h) - \widehat{\gamma}(h))^2}{M}$, where M is the number of points where the empirical variogram was computed. The first Matérn model was fitted using $\kappa = 0.5$ as a starting point. The other two Matérn models were fitted using respectively $\kappa = 1$ and $\kappa = 2$ as starting points.

Basal Area									
Model	κ	σ^2	ϕ	τ^2	MSE	$\widehat{\text{ESS}}$	sd(3)	sd(4)	sd(6)
Spherical	–	2.58	1803.87	1.94	0.08	42.31	13.17	4.52	5.60
Exponential	0.5	2.91	1115.82	1.88	0.08	17.09	2.86	2.53	5.47
Gaussian	∞	2.15	691.22	2.20	0.09	58.33	4.31	12.18	5.24
Cubic	–	3.22	2479.60	2.31	0.96	29.80	11.98	5.83	6.41
Circular	–	3.39	2247.08	2.01	0.66	24.41	10.48	5.14	8.10
Cauchy	0.5	3.15	565.10	2.16	0.12	8.64	0.94	2.13	3.21
Matern ($\kappa = 0.5$)	0.64	3.69	1239.80	2.02	0.69	14.13	2.55	1.45	20.70
Matern ($\kappa = 1$)	0.72	3.31	991.83	2.05	0.50	19.62	14.65	4.56	20.29
Matern ($\kappa = 2$)	1.27	2.57	495.89	2.13	0.27	56.47	24.76	7.64	20.94
Height									
Model	κ	σ^2	ϕ	τ^2	MSE	$\widehat{\text{ESS}}$	sd(3)	sd(4)	sd(6)
Spherical	–	0.82	2252.61	0.18	0.08	22.83	5.21	5.52	6.37
Exponential	0.5	0.64	619.49	0.12	0.05	28.76	8.08	2.06	2.03
Cubic	–	0.82	2626,87	0,28	0,12	22,56	3,96	3,71	4,25
Circular	–	0.75	1751,06	0,18	0,06	28,68	3,10	9,27	7,65
Cauchy	0.5	0.77	495,59	0,24	0,07	7,60	1,95	3,21	0,92
Matern ($\kappa = 0.5$)	0.70	0.83	867.29	0.20	0.09	18.62	3.08	7.37	1.79
Matern ($\kappa = 1$)	0.72	0.82	841.22	0.20	0.08	19.49	5.40	8.85	3.50
Matern ($\kappa = 2$)	0.72	0.82	841.20	0.20	0.08	19.49	17.59	7.49	17.95

Table 5: $\widehat{\text{ESS}}$ for basal area and height for different covariance models. MSE = $\frac{\sum_h (\gamma(h) - \widehat{\gamma}(h))^2}{M}$, where M is the number of points where the empirical variogram was computed (by default $M = 13$). The first Matérn model was fitted using $\kappa = 0.5$ as a starting point. The other two Matérn models were fitted using respectively $\kappa = 1$ and $\kappa = 2$ as starting points. In columns 8-10, sd(i) stands for the Sherman block bootstrap variance of the effective sample size computed using i blocks.