Assessing the Effective Sample Size for Large Spatial Datasets: A Block Likelihood Approach

Jonathan Acosta[a], Alfredo Alegría[1b], Felipe Osorio[b], Ronny Vallejos[b]

[a] *Departamento de Estadística, Pontificia Universidad Católica de Chile, Santiago, Chile.*

[b] *Departamento de Matemática, Universidad Técnica Federico Santa María, Valparaíso, Chile.*

Abstract

The development of new techniques for sample size reduction has attracted growing interest in recent decades. Recent findings allow us to quantify the amount of duplicated information within a sample of spatial data through the so-called effective sample size (ESS), whose definition arises from the Fisher information that is associated with maximum likelihood estimation. However, in all circumstances where the sample size is very large, maximum likelihood estimation and ESS evaluation are challenging from a computational viewpoint. An alternative definition of the ESS, in terms of the Godambe information from a block likelihood estimation approach, is presented. Several theoretical properties satisfied by this quantity are investigated. Our proposal is evaluated in some parametric correlation structures, including the intraclass, AR(1), Matérn, and simultaneous autoregressive models. Simulation experiments show that our proposal provides accurate approximations of the full likelihood-based ESS while maintaining a moderate computational cost. A large dataset is analyzed to quantify the effectiveness and limitations of the proposed framework in practice.

Keywords: Covariance Function, Fisher Information, Godambe Information, Matérn Model

1. Introduction

The effective sample size (ESS) has been developed to quantify the number of independent and identically distributed observations within a sample of size n (Cressie, 1993). Currently, with the rapid proliferation and acquisition of large datasets, the extraction of the relevant features and information within such datasets is particularly important, especially when a reduction of information occurs due to autocorrelation that is present in a set of georeferenced observations (Griffith, 2005). These considerations are predominantly useful in the analysis of remotely sensed satellite data because in such a context, the observations generally present high levels of spatial association, so the amount of duplicated information can be considerable (Griffith, 2015).

The literature on the ESS is substantial. While Vallejos and Osorio (2014) provided a novel definition for spatial random fields with constant means, Acosta and Vallejos (2018) extended this quantity for a general spatial regression model, including the asymptotic distribution of the

[1] Corresponding author. Email: alfredo.alegria@usm.cl

maximum likelihood estimator in the context of increasing domain sampling schemes. At the same time, the ESS has been studied for linear models with replicates (Faes et al., 2009). In a Bayesian context, Berger et al. (2014) addressed the ESS from a model selection perspective. Chatterjee and Diaconis (2018) and Elvira et al. (2018) studied the ESS in specific contexts of importance sampling. ESS applications can be found, for instance, in Solow and Polanky (1994) and Li et al. (2016), while Lenth (2001) provided practical guidelines for ESS determination.

The ESS introduced by Vallejos and Osorio (2014) is based on the Fisher information associated with maximum likelihood (ML) estimation under the Gaussian assumption. The objective functions of ML estimation and the ESS depend on the inverse of the correlation matrix, making their implementations challenging from a computational perspective in all circumstances where the sample size is very large. Modern approaches require a compromise between statistical and computational performances. For a thorough review of the current methods for analyzing large spatial datasets, the reader is referred to Heaton et al. (2019) and the references therein.

To circumvent the so-called "big n" problem, we propose an alternative sample size reduction approach, which is fully based on block likelihood inference (Caragea and Smith, 2007; Varin et al., 2011). First, parameter estimation is carried out by means of the block likelihood estimation method; this accomplishes a trade-off between statistical and computational efficiency, where the inverses of small correlation matrices are involved. Second, a new notion of the effective sample size (named ESS_B) comes from the Godambe information arising from this estimation framework. In particular, we focus on the "small blocks" method in Caragea and Smith (2007), which performs better than other similar competitors (Caragea and Smith, 2007; Varin et al., 2011; Stein, 2013). At the same time, the "small blocks" version generates a particularly amenable expression for ESS_B. To illustrate the use of ESS_B, some parametric correlation structures are revisited, including the intraclass, AR(1), and Matérn models that were previously analyzed by Vallejos and Osorio (2014), as well as the simultaneous autoregressive model (see, e.g., Griffith, 2015). Our proposal is also compared to the full likelihood-based ESS in terms of both statistical and computational efficiency through Monte Carlo simulation experiments. The proposed methodology is applied on a large dataset consisting of 21 million observations from a Harvard forest database.

The article is organized as follows. Section 2 briefly reviews the definition, some examples, and the main theoretical attributes of the traditional ESS. Section 3 introduces ESS_B and its properties. Section 4 discusses computational aspects that permit an efficient implementation of ESS_B for regularly-spaced locations. In Section 5, we calculate ESS_B for some parametric families of correlation functions. In Section 6, the discrepancies between the ESS and ESS_B estimates obtained through simulation experiments are assessed. The computational performances of these approaches are also explored. Section 7 presents a real data application. Finally, Section 8 is a discussion of the main findings, and this includes problems to study in future research. For a neater exposition, the proofs of the main results are given in Appendix A, and additional numerical studies that complement the main findings of the manuscript are contained in Appendix B.

2. Background

In this section, an approach that allows us to quantify the amount of duplicated information within a sample of spatial data due to the effect of spatial autocorrelation is described. This approach, proposed by Vallejos and Osorio (2014), is based on Fisher information about the mean.

Consider a spatial random field, $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$, and let $X(\boldsymbol{s}_1), \ldots, X(\boldsymbol{s}_n)$ be a realization at n spatial locations. For simplicity, we use the notations $X_i = X(\boldsymbol{s}_i)$ and $\boldsymbol{X} = (X_1, \ldots, X_n)^\top$, with $^\top$ standing for the transpose operator. Suppose that the random field has a constant mean $\mu \in \mathbb{R}$, such that $\mathbb{E}(\boldsymbol{X}) = \mu \mathbf{1}_n$, where $\mathbf{1}_n$ is an $n \times 1$ vector of ones, and a constant variance $\sigma^2 > 0$. Let $\boldsymbol{R}(\boldsymbol{\theta})$ denote the $n \times n$ correlation matrix of $\boldsymbol{X}$, where $\boldsymbol{\theta}$ is a vector of unknown parameters. Assume that $\boldsymbol{X}$ follows a multivariate Gaussian distribution. To discard redundant observations while preserving the essential information of a sample, Vallejos and Osorio (2014) proposed the definition of an effective sample size (ESS), a quantity that depends on the sample size, the covariance structure of the random field and the dimensionality of the spatial domain.

Definition 2.1. *Let $\boldsymbol{R}(\boldsymbol{\theta})$ be a non-singular correlation matrix. The ESS is defined as*

$$\text{ESS} = \text{ESS}(n, \boldsymbol{R}(\boldsymbol{\theta}), d) = \mathbf{1}_n^\top \boldsymbol{R}(\boldsymbol{\theta})^{-1} \mathbf{1}_n. \tag{2.1}$$

When the correlation matrix is singular, Definition 2.1 can be slightly modified. For such a case, an alternative definition (Vallejos and Osorio, 2014) is considered:

$$\text{ESS} = \mathbf{1}_n^\top \boldsymbol{R}(\boldsymbol{\theta})^{+} \mathbf{1}_n,$$

where $\boldsymbol{R}(\boldsymbol{\theta})^{+}$ denotes the uniquely determined Moore-Penrose pseudoinverse of $\boldsymbol{R}(\boldsymbol{\theta})$ (see, e.g., Magnus and Neudecker, 1988).

This definition of the ESS arose as a normalized version of the Fisher information about μ. More precisely, it can be written as

$$\text{ESS} = -\sigma^2 \mathbb{E}\left(\frac{\partial^2 \ell(\mu, \sigma^2, \boldsymbol{\theta} | \boldsymbol{X})}{\partial \mu^2} \right),$$

where

$$\ell(\mu, \sigma^2, \boldsymbol{\theta} | \boldsymbol{X}) = -\frac{1}{2} \log |\sigma^2 \boldsymbol{R}(\boldsymbol{\theta})| - \frac{1}{2\sigma^2} (\boldsymbol{X} - \mu \mathbf{1}_n)^\top \boldsymbol{R}(\boldsymbol{\theta})^{-1} (\boldsymbol{X} - \mu \mathbf{1}_n)$$

is the log-likelihood function (up to an additive constant) of $\boldsymbol{X}$.

The ESS satisfies several appealing properties. For example, under perfect positive correlation, $\boldsymbol{R}(\boldsymbol{\theta}) = \mathbf{1}_n \mathbf{1}_n^\top$ and $\boldsymbol{R}(\boldsymbol{\theta})^{+} = n^{-2} \mathbf{1}_n \mathbf{1}_n^\top$; thus,

$$\text{ESS} = \frac{\mathbf{1}_n^\top \mathbf{1}_n \mathbf{1}_n^\top \mathbf{1}_n}{n^2} = 1.$$

On the other hand, when all observations are independent, one has $\boldsymbol{R}(\boldsymbol{\theta}) = \boldsymbol{I}_n$, where $\boldsymbol{I}_n$ is the $n \times n$ identity matrix, and so

$$\text{ESS} = \mathbf{1}_n^\top \boldsymbol{I}_n \mathbf{1}_n = n.$$

Another interesting and intuitive property proved by Vallejos and Osorio (2014) is that the ESS is increasing with n. The following examples show that in some special situations, the ESS has closed-form expressions (Vallejos and Osorio, 2014).

Example 2.1. *For the intraclass correlation structure,*

$$\boldsymbol{R}(\rho) = (1 - \rho) \boldsymbol{I}_n + \rho \mathbf{1}_n \mathbf{1}_n^\top, \tag{2.2}$$

where $-1/(n-1) < \rho < 1$, *one has*

$$\text{ESS} = \frac{n}{1+(n-1)\rho}. \tag{2.3}$$

Example 2.2. *Another interesting case occurs when the correlation structure of an AR(1) process is considered. In this case, the* (i,j)*th entry of the correlation matrix* $\boldsymbol{R}(\rho)$ *is given by*

$$R_{ij}(\rho) = \rho^{|i-j|}, \tag{2.4}$$

where $0 < \rho < 1$. *The ESS for this particular structure is*

$$\text{ESS} = \frac{2+(n-2)(1-\rho)}{1+\rho}. \tag{2.5}$$

In Examples 2.1 and 2.2, as $\rho \to 0$, we recover the limiting case of independence (ESS $= n$). Similarly, $\rho \to 1$ corresponds to a perfect positive correlation (ESS $= 1$). For arbitrary correlation matrices with no explicit inverses, (2.1) must be evaluated numerically.

The ESS can be estimated through the plug-in estimate $\widehat{\text{ESS}} = \mathbf{1}_n^\top \boldsymbol{R}(\widehat{\boldsymbol{\theta}}_{\text{ML}})^{-1}\mathbf{1}_n$, where $\widehat{\boldsymbol{\theta}}_{\text{ML}}$ is the maximum likelihood estimate of $\boldsymbol{\theta}$. Since matrix inversion has a computational cost proportional to the cube of the sample size (Aho and Hopcroft, 1974), maximum likelihood estimation and the evaluation of $\widehat{\text{ESS}}$ become impractical for large datasets. This motivates the search for an alternative definition of the ESS.

3. Effective Sample Size Based on Godambe Information: Definition and Properties

An alternative method is proposed to carry out sample size reduction for spatial data. A new notion of the effective sample size is defined for Gaussian random fields in terms of the Godambe information about the mean, which comes from block likelihood inference. We shall use the notation ESS_B for our proposal. We expect ESS_B to be a reasonable approximation of the traditional ESS.

The block likelihood estimation framework (Caragea and Smith, 2007) is an estimation method within the class of composite-likelihood methods (Lindsay, 1988; Varin et al., 2011), and it consists of splitting the given data into m smaller pieces of information. Hence, a partition $\Delta = \{b_1, \ldots, b_m\}$ of the set $\{1, \ldots, n\}$ is considered, and the ith block of data is defined as $\boldsymbol{X}_{b_i} = (X_j : j \in b_i)$ for $i = 1, \ldots, m$. The cardinality of b_i (i.e., the length of $\boldsymbol{X}_{b_i}$) is denoted by $|b_i|$, and of course, we have that $|b_1| + \cdots + |b_m| = n$. The objective function of the block likelihood method is given by

$$\mathcal{L}_B(\mu, \sigma^2, \boldsymbol{\theta}|\boldsymbol{X}) = \prod_{i=1}^{m} \mathcal{L}_{b_i}(\mu, \sigma^2, \boldsymbol{\theta}|\boldsymbol{X}_{b_i}), \tag{3.1}$$

where $\mathcal{L}_{b_i}$ is the likelihood of the ith block. When the full sample in a single block is considered, this method is reduced to full likelihood. On the other hand, as the number of blocks increases, this approach can be seen as a misspecified likelihood, where misspecification comes from the assumption of independence among blocks. At the same time, an increase in the number of blocks implies a reduction in the computational burden since we need to compute the inverses of smaller

correlation matrices. Thus, the block likelihood method provides a trade-off between statistical and computational efficiency. The corresponding block log-likelihood is given by

$$\ell_B(\mu, \sigma^2, \boldsymbol{\theta}|\boldsymbol{X}) = \log \mathcal{L}_B(\mu, \sigma^2, \boldsymbol{\theta}|\boldsymbol{X}) = \sum_{i=1}^{m} \ell_{b_i}(\mu, \sigma^2, \boldsymbol{\theta}|\boldsymbol{X}_{b_i}), \tag{3.2}$$

where $\ell_{b_i}(\cdot) = \log \mathcal{L}_{b_i}(\cdot)$.

Since (3.2) is written as the sum of (sub)log-likelihoods, it is an unbiased estimation equation. Assuming adequate regularity conditions (see Guyon, 1995 for details), the block likelihood estimator is consistent and asymptotically Gaussian under increasing domain asymptotics, that is, when the spatial domain expands as the number of locations increases. In such a case, the Fisher information matrix must be substituted by the Godambe information matrix, which is defined as

$$\mathcal{G}(\boldsymbol{\xi}) = H(\boldsymbol{\xi}) J(\boldsymbol{\xi})^{-1} H(\boldsymbol{\xi}), \tag{3.3}$$

where the vector of parameters is written as $\boldsymbol{\xi} = (\mu, \sigma^2, \boldsymbol{\theta})^\top$, and $H(\boldsymbol{\xi}) = \mathbb{E}\left\{-\nabla^2_{\boldsymbol{\xi}} \ell_B(\boldsymbol{\xi})\right\}$ and $J(\boldsymbol{\xi}) = \mathbb{E}\left\{\nabla_{\boldsymbol{\xi}} \ell_B(\boldsymbol{\xi}) \nabla_{\boldsymbol{\xi}} \ell_B(\boldsymbol{\xi})^\top\right\}$ are the sensitivity and variability matrices, respectively. The asymptotic variance of the block likelihood estimator is given by the inverse of the Godambe information matrix. In particular, the Godambe information about μ, after being rescaled by σ^2, is given by

$$\sigma^2 \, \mathbb{E}\left(-\frac{\partial^2 \ell_B(\mu, \sigma^2, \boldsymbol{\theta})}{\partial \mu^2}\right)^2 \mathbb{E}\left(\left\{\frac{\partial \ell_B(\mu, \sigma^2, \boldsymbol{\theta})}{\partial \mu}\right\}^2\right)^{-1}. \tag{3.4}$$

Recall that under the Gaussian assumption, the block log-likelihood (3.2) is constructed as the sum of (sub) log-likelihoods in the following form:

$$\ell_{b_i}(\mu, \sigma^2, \boldsymbol{\theta}|\boldsymbol{X}_{b_i}) = -\frac{1}{2} \log |\sigma^2 \boldsymbol{R}_{b_i}(\boldsymbol{\theta})| - \frac{1}{2\sigma^2} \left(\boldsymbol{X}_{b_i} - \mu \mathbf{1}_{|b_i|}\right)^\top \boldsymbol{R}_{b_i}(\boldsymbol{\theta})^{-1} \left(\boldsymbol{X}_{b_i} - \mu \mathbf{1}_{|b_i|}\right), \tag{3.5}$$

where $\boldsymbol{R}_{b_i}(\boldsymbol{\theta})$ is the $|b_i| \times |b_i|$ correlation matrix of $\boldsymbol{X}_{b_i}$. Thus, we have the expressions

$$\frac{\partial \ell_{\mathrm{B}}(\mu, \sigma^2, \boldsymbol{\theta})}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^{m} \mathbf{1}^\top_{|b_i|} \boldsymbol{R}_{b_i}(\boldsymbol{\theta})^{-1} \left(\boldsymbol{X}_{b_i} - \mu \mathbf{1}_{|b_i|}\right),$$

and

$$\frac{\partial^2 \ell_B(\mu, \sigma^2, \boldsymbol{\theta})}{\partial \mu^2} = -\frac{1}{\sigma^2} \sum_{i=1}^{m} \mathbf{1}^\top_{|b_i|} \boldsymbol{R}_{b_i}(\boldsymbol{\theta})^{-1} \mathbf{1}_{|b_i|}.$$

The direct calculation of (3.4) provides the following definition of ESS_B.

Definition 3.1. *The effective sample size based on a block likelihood approach, denoted by* ESS_B*, for a nonsingular correlation matrix* $\boldsymbol{R}(\boldsymbol{\theta})$ *is given by*

$$\mathrm{ESS}_B = \mathrm{ESS}_B(n, \boldsymbol{R}(\boldsymbol{\theta}), d, \Delta) = \frac{\left(\sum_{i=1}^{m} \eta_{ii}\right)^2}{\sum_{j=1}^{m} \sum_{i=1}^{m} \eta_{ij}}, \tag{3.6}$$

where

$$\eta_{ij} = \mathbf{1}_{|b_i|}^{\top} \boldsymbol{R}_{b_i}^{-1}(\boldsymbol{\theta}) \boldsymbol{R}_{b_i b_j}(\boldsymbol{\theta}) \boldsymbol{R}_{b_j}^{-1}(\boldsymbol{\theta}) \mathbf{1}_{|b_j|}, \tag{3.7}$$

and

$$\boldsymbol{R}_{b_i b_j}(\boldsymbol{\theta}) = \sigma^{-2} \mathbb{E}\left\{ \left(\boldsymbol{X}_{b_i} - \mu \mathbf{1}_{|b_i|} \right) \left(\boldsymbol{X}_{b_j} - \mu \mathbf{1}_{|b_j|} \right)^{\top} \right\}$$

is the $|b_i| \times |b_j|$ *cross-correlation matrix between blocks* i *and* j*, where in particular,* $\boldsymbol{R}_{b_i b_i}(\boldsymbol{\theta})$ *is simply* $\boldsymbol{R}_{b_i}(\boldsymbol{\theta})$*.*

Definition 3.1 can be generalized to the case with singular correlation matrices by considering their pseudoinverses in a similar fashion to that in the definition provided by Vallejos and Osorio (2014). Certainly, if the ith block has a singular correlation matrix, we must replace $\boldsymbol{R}_{b_i}(\boldsymbol{\theta})^{-1}$ by $\boldsymbol{R}_{b_i}(\boldsymbol{\theta})^{+}$. Note that ESS_B depends strongly on the choice of the block partition Δ.

ESS_B preserves some features of the traditional ESS. The following proposition summarizes interesting properties established in this paper for ESS_B.

Proposition 3.1. *Let* ESS_B *be defined according to (3.6). Thus,*

(1) Under independence, $\mathrm{ESS}_B = n$*.*

(2) Under perfect positive correlation, $\mathrm{ESS}_B = 1$*.*

(3) It is always true that $1 \leq \mathrm{ESS}_B \leq \mathrm{ESS}$*.*

(4) If $m = 1$*, then* $\mathrm{ESS}_B = \mathrm{ESS}$*.*

(5) If $m = n$*, then* $\mathrm{ESS}_B = n^2 \left(\mathbf{1}_n^{\top} \boldsymbol{R}(\boldsymbol{\theta}) \mathbf{1}_n \right)^{-1}$*.*

Points (1) and (2) in Proposition 3.1 are elementary properties satisfied by the ESS and preserved by ESS_B. Point (3) implies that ESS_B always underestimates the value of the ESS when both measures are evaluated with the same parameter $\boldsymbol{\theta}$. The examples considered in the next section show that this discrepancy is often slight. In practice, the ESS and ESS_B are evaluated at $\widehat{\boldsymbol{\theta}}_{\mathrm{ML}}$ and $\widehat{\boldsymbol{\theta}}_{\mathrm{BL}}$ (the block likelihood estimate), respectively, obtaining the respective estimates $\widehat{\mathrm{ESS}}$ and $\widehat{\mathrm{ESS}}_B$, which do not necessarily satisfy the inequality in point (3). In points (4) and (5), the limit cases for ESS_B are described in terms of the number of blocks m. When the entire sample is taken in a single block, we recover the ESS, whereas for blocks of size one, we obtain the so-called effective geographic sample size (Griffith, 2005; Acosta et al., 2018), denoted throughout by ESS_G. For a neater exposition, a proof of Proposition 3.1 is available in Appendix A.

If we incorporate a new observation in the original sample, the information cannot decrease, so we expect ESS_B to be nondecreasing with n. The examples studied in the next section suggest that this is a reasonable conjecture. However, a general proof of this property is still elusive.

Let us recall that under regularity conditions (Guyon, 1995), the block likelihood estimator of $\boldsymbol{\theta} \in \mathbb{R}^p$ is asymptotically Gaussian with variance characterized by the Godambe information matrix. Defining

$$g(\boldsymbol{\theta}) = \mathrm{ESS}_B(n, \boldsymbol{R}(\boldsymbol{\theta}), d, \Delta),$$

the following proposition characterizes the asymptotic distribution of $g(\widehat{\boldsymbol{\theta}}_{\mathrm{BL}})$ under an increasing domain sampling scheme.

Proposition 3.2. *Suppose that as $n \to \infty$, the sequence $\mathcal{G}_{\boldsymbol{\theta}}^{1/2}(\widehat{\boldsymbol{\theta}}_{BL} - \boldsymbol{\theta})$ converges in distribution to the p-variate standard Gaussian distribution. Then, as $n \to \infty$, the sequence*

$$(\nabla g(\boldsymbol{\theta})^\top \mathcal{G}_{\boldsymbol{\theta}}^{-1} \nabla g(\boldsymbol{\theta}))^{-1/2}(g(\widehat{\boldsymbol{\theta}}_{BL}) - g(\boldsymbol{\theta}))$$

converges in distribution to the (univariate) standard Gaussian distribution.

The proof of this fact is omitted because it is a consequence of the delta method (see Acosta et al., 2018 for details).

4. Computational Aspects for Regular Grids

Regular grids are common in satellite data and image modeling. In the computation of ESS_B, a regularly gridded spatial design permits the reuse of some calculations. We now discuss some computational tools that will be used in the subsequent sections.

For a spatial random field defined on a rectangular grid of $\mathbb{Z}^2$, assume that all blocks are regular lattices with the same size. Thus, we can write $|b_i| = |b|$ for every $i = 1, \ldots, m$. Let $\boldsymbol{H}_{b_i b_j}$ be the distance matrix between blocks i and j, that is, the (k, r)th entry of $\boldsymbol{H}_{b_i b_j}$ corresponds to the distance between the kth site in the ith block and the rth site in the jth block. Observe that $\boldsymbol{H}_{b_i b_i} = \boldsymbol{H}_b$ for all $i = 1, \ldots, m$ and for some matrix $\boldsymbol{H}_b$. As a result, the correlation matrix of the ith block can be written as $\boldsymbol{R}_{b_i}(\boldsymbol{\theta}) = \boldsymbol{R}_b(\boldsymbol{\theta})$, regardless of the correlation structure used. In particular, we have that $\eta_{11} = \cdots = \eta_{mm}$. Let $\boldsymbol{v} = \boldsymbol{R}_b^{-1}(\boldsymbol{\theta})\mathbf{1}_{|b|}$; then, $\eta_{ii} = \mathbf{1}_{|b|}^\top \boldsymbol{v}$ and $\eta_{ij} = \boldsymbol{v}^\top \boldsymbol{R}_{b_i b_j}(\boldsymbol{\theta})\boldsymbol{v}$. According to Equation (3.6), ESS_B is completely determined by the trace of $\boldsymbol{\eta}$ and by the sum of its elements. Thus, a strategy to accelerate the computing time of ESS_B is to take advantage of the structure of $\boldsymbol{\eta}$, which under this setting contains repeated elements.

As an illustration, an example is provided in which a regular grid of size $n = 30 \times 24$ is partitioned into $m = 9$ blocks of the same size (see Figure 4.1). $\boldsymbol{\eta}$ is built explicitly by taking the arrangement between blocks into account:

$$\boldsymbol{\eta} = \begin{pmatrix}
\eta_{11} & \eta_{12} & \eta_{13} & \eta_{14} & \eta_{15} & \eta_{16} & \eta_{17} & \eta_{18} & \eta_{19} \\
 & \eta_{11} & \eta_{12} & \eta_{24} & \eta_{14} & \eta_{15} & \eta_{27} & \eta_{17} & \eta_{18} \\
 & & \eta_{11} & \eta_{34} & \eta_{24} & \eta_{14} & \eta_{37} & \eta_{27} & \eta_{17} \\
 & & & \eta_{11} & \eta_{12} & \eta_{13} & \eta_{14} & \eta_{15} & \eta_{16} \\
 & & & & \eta_{11} & \eta_{12} & \eta_{24} & \eta_{14} & \eta_{15} \\
 & & & & & \eta_{11} & \eta_{34} & \eta_{24} & \eta_{14} \\
 & & & & & & \eta_{11} & \eta_{12} & \eta_{13} \\
 & & & & & & & \eta_{11} & \eta_{12} \\
 & & & & & & & & \eta_{11}
\end{pmatrix}.$$

The terms below the main diagonal are omitted since $\boldsymbol{\eta}$ is symmetric. The general idea consists of summing across the diagonals of $\boldsymbol{\eta}$. Each diagonal has at most two different values, and the number of elements in the ith diagonal is $m-i$ for $i = 0, \ldots, m-1$, where $i = 0$ represents the main diagonal and so on. Hence, each η_{1i} is multiplied by $m-i$, and the extra elements are subsequently removed by considering the number of jumps in the batch (see the decomposition into glued blocks

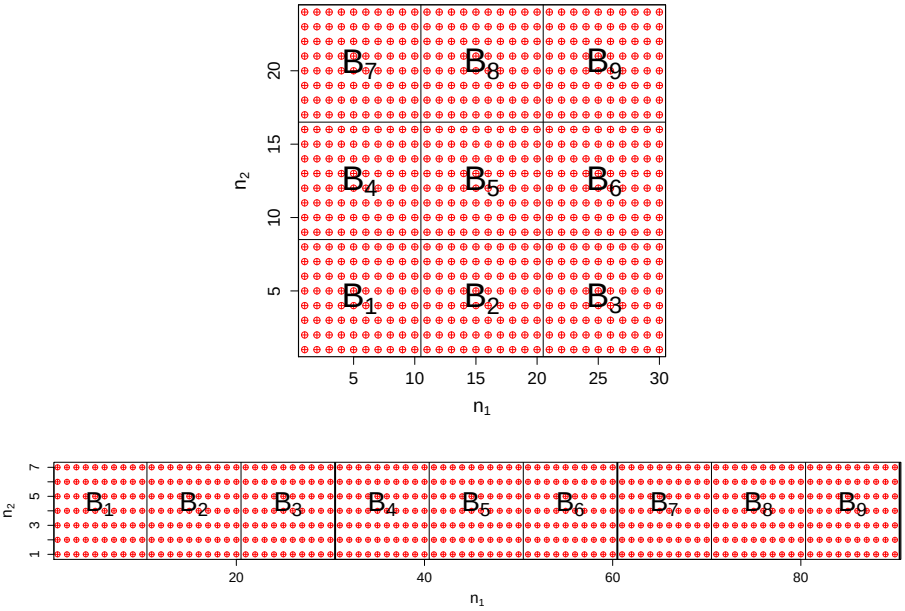

Figure 4.1: Grid of size 30×24 split into $m = 9$ identical blocks (top) and its decomposition into glued blocks (bottom).

in Figure 4.1). Following this scheme, we find that

$$\begin{aligned}\sum_{i=1}^{9}\sum_{j=1}^{9}\eta_{ij} &= 9\eta_{11} + 2\left[8\eta_{12} + 1 \times 2(\eta_{34} - \eta_{12})\right] + 2\left[7\eta_{13} + 2 \times 2(\eta_{35} - \eta_{13})\right] + \\ &\quad 2\left[6\eta_{14} + 0 \times 2(\eta_{36} - \eta_{14})\right] + 2\left[5\eta_{15} + 1 \times 1(\eta_{37} - \eta_{15})\right] + 2\left[4\eta_{16} + 2 \times 1(\eta_{38} - \eta_{16})\right] + \\ &\quad 2\left[3\eta_{17} + 0 \times 0(\eta_{39} - \eta_{17})\right] + 2\left[2\eta_{18} + 1 \times 0(\eta_{3,10} - \eta_{18})\right] + 2\left[1\eta_{19} + 2 \times 0(\eta_{3,11} - \eta_{19})\right].\end{aligned}$$

Note that $\eta_{3,10}$ and $\eta_{3,11}$ do not exist; however, these values have been written in the previous formula to illustrate the generic expression. In the final equation, these values are multiplied by zero.

In general, we consider a rectangular grid of size $n = n_1 \times n_2$ such that n_k is divided into m_k segments of equal length for $k = 1, 2$. Hence, the grid is split into $m = m_1 \times m_2$ equal blocks. Each block has size $|b_i| = n/m$. Again, by summing across the diagonals of $\boldsymbol{\eta}$, we obtain

$$\begin{aligned}\sum_{i=1}^{m}\eta_{ii} &= m\eta_{11}, \\ \sum_{i=1}^{m}\sum_{j=1}^{m}\eta_{ij} &= m\eta_{11} + 2\sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\eta_{ij} \\ &= m\eta_{11} + 2\sum_{i=1}^{m-1}\left\{\eta_{1,1+i}(m-i) + (i \mod m_1)\left(m_2 - \left\lceil\frac{i}{m_1}\right\rceil\right)(\eta_{m_1,m_1+i} - \eta_{1,1+i})\right\},\end{aligned}$$

where mod is the modulus function and $\lceil\cdot\rceil$ is the ceiling function.

The direct computation of the elements of $\boldsymbol{\eta}$ has an order of computation of m^2, while the computation across the diagonals has order m, so the computational time is considerably reduced in practice.

5. Examples

In this section, the ESS and ESS_B are compared for different correlation structures. We focus on the intraclass and AR(1) correlation matrices, which are described in the previous sections, as well as on the Matérn and simultaneous autoregressive correlation models, which are widely used in the spatial statistics literature.

5.1. Example 1: Intraclass Correlation

Consider the intraclass correlation matrix $\boldsymbol{R}(\rho)$ introduced in Equation (2.2). Each block has an intraclass correlation matrix of the form

$$\boldsymbol{R}_{b_i}(\rho) = (1-\rho)\boldsymbol{I}_{|b_i|} + \rho \mathbf{1}_{|b_i|}\mathbf{1}_{|b_i|}^\top,$$

and the cross-correlation matrix between distinct blocks is $\boldsymbol{R}_{b_i b_j}(\rho) = \rho \mathbf{1}_{|b_i|}\mathbf{1}_{|b_j|}^\top$. Using Equation (2.3), one has

$$\eta_{ii} = \frac{|b_i|}{1+(|b_i|-1)\rho}.$$

Additionally, a straightforward calculation shows that

$$\eta_{ij} = \rho\eta_{ii}\eta_{jj} = \frac{\rho|b_i||b_j|}{\{1+(|b_i|-1)\rho\}\{1+(|b_j|-1)\rho\}}.$$

In particular, when all blocks have the same size, that is, $|b_i| = n/m$, we find that $\mathrm{ESS_B} = \mathrm{ESS}$. Precisely, notice that

$$\eta_{ii} = \frac{n/m}{1+(n/m-1)\rho} \quad \text{and} \quad \eta_{ij} = \rho\left(\frac{n/m}{1+(n/m-1)\rho}\right)^2.$$

Then,

$$\begin{aligned}
\mathrm{ESS}_B &= \frac{m^2\left(\frac{n/m}{1+(n/m-1)\rho}\right)^2}{m\left(\frac{n/m}{1+(n/m-1)\rho}\right) + m(m-1)\rho\left(\frac{n/m}{1+(n/m-1)\rho}\right)^2} \\
&= \frac{m\left(\frac{n/m}{1+(n/m-1)\rho}\right)}{1+(m-1)\rho\left(\frac{n/m}{1+(n/m-1)\rho}\right)} \\
&= \frac{\frac{n}{1+(n/m-1)\rho}}{\frac{1+(n-1)\rho}{1+(n/m-1)\rho}} \\
&= \frac{n}{1+(n-1)\rho} \\
&= \mathrm{ESS}\,.
\end{aligned}$$

It is straightforward to show that the mapping $t \mapsto t/(1+(t-1)\rho)$ has a positive first-order derivative on the positive real line for any $\rho < 1$. Thus, ESS_B and ESS are increasing with n.

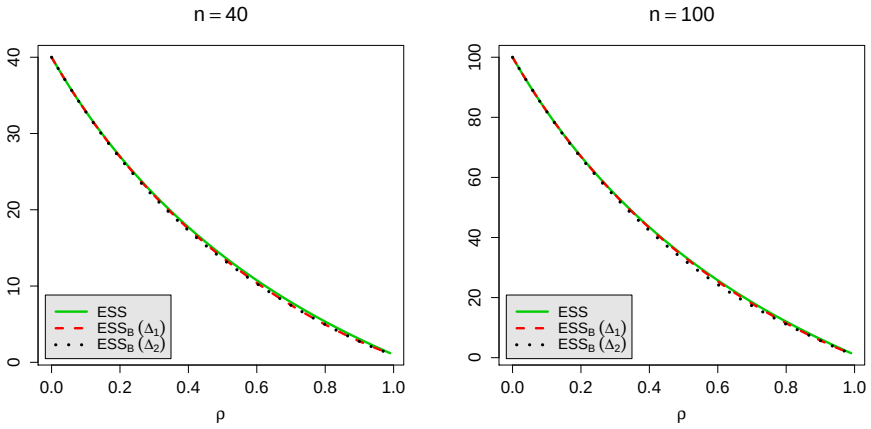

Figure 5.1: ESS_B and the ESS versus $0 < \rho < 1$ in the AR(1) model for $n = 40, 100$. For ESS_B, two different block partitions are considered: Δ_1 (two blocks of size $n/2$) and Δ_2 (ten blocks of size $n/10$).

5.2. Example 2: AR(1) Correlation

For the AR(1) correlation structure in Equation (2.4), the observations are split into m blocks of size n/m, i.e., $b_1 = \{1, \ldots, n/m\}, b_2 = \{n/m + 1, \ldots, 2n/m\}, \ldots, b_m = \{(m-1)n/m + 1, \ldots, n\}$. According to Equation (2.5), we have that

$$\eta_{ii} = \frac{2 + (n/m - 2)(1 - \rho)}{1 + \rho}.$$

To obtain η_{ij}, note that the (k, r)th entry of the matrix $\boldsymbol{R}_{b_i b_j}(\rho)$ for $i \neq j$ is given by

$$[\boldsymbol{R}_{b_i b_j}(\rho)]_{k,r} = \rho^{|n/m(i-j)+k-r|},$$

where $k, r = 1, \ldots, n/m$. In Figure 5.1, plots are shown for ESS_B and the ESS versus ρ for the AR(1) structure and for $0 < \rho < 1$, $n = 40$ and $n = 100$. For ESS_B, two block configurations are considered: Δ_1 (two blocks of size $n/2$) and Δ_2 (ten blocks of size $n/10$). From Figure 5.1, it is observed that in all cases, the curves are very similar, regardless of the number of blocks used. When $n = 40$, the maximum discrepancies between the ESS and $\text{ESS}_B(\Delta_1)$ and between the ESS and $\text{ESS}_B(\Delta_2)$ are 0.39 and 0.45, respectively. However, when $n = 100$, these values are 0.54 and 1.40, respectively. The sample size reduction percentages produced by the ESS and ESS_B relative to n are essentially the same (the differences between these percentages are approximately 1%). Thus, ESS_B provides a good approximation of the ESS under the AR(1) model.

Figure 5.2 shows ESS_B versus $n = 16k$ with $k = 1, \ldots, 10$, for different values of ρ. For each value of n, ESS_B is implemented with two blocks of size $n/2$. As expected, an increasing trend can be observed. Although it seems that ESS_B grows linearly, the increments are not constant. These experiments are repeated with additional block configurations, where they obtain quite similar results; thus, they are not reported here.

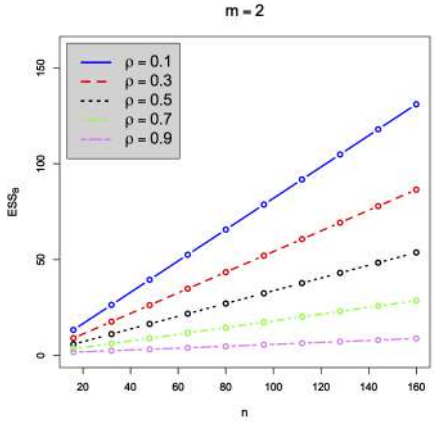

Figure 5.2: ESS_B versus $n = 16k$ with $k = 1, \ldots, 10$ for the AR(1) model with $\rho = 0.1, 0.3, 0.5, 0.7, 0.9$. For each value of n, ESS_B is considered with two blocks of size $n/2$.

5.3. Example 3: Matérn Correlation on a Plane

We consider random fields $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^2\}$ with a Matérn-type covariance function (see, e.g., Stein, 2012), i.e.,

$$\text{cov}\{Z(\boldsymbol{s}_1), Z(\boldsymbol{s}_2)\} = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{h}{\phi}\right)^{\nu} K_\nu \left(\frac{h}{\phi}\right), \tag{5.1}$$

where $h = \|\boldsymbol{s}_1 - \boldsymbol{s}_2\|$, K_ν is the modified Bessel function of the second kind and Γ is the gamma function (Abramowitz and Stegun, 1972). Here, $\phi > 0$ is a parameter that controls the (practical) range of the random field, that is, the distance h^* such that (5.1) is less than or equal to 0.05 for all $h \geq h^*$. The parameter $\nu > 0$ regulates the differentiability (in a mean square sense) of the random field. Specifically, the random field is k-times mean square differentiable if and only if $\nu > k$ (see Stein, 2012 for details). We focus on the following special cases:

- Matérn model with $\nu = 1/2$, which corresponds to an exponential structure:

$$h \mapsto \exp\left(-\frac{h}{\phi}\right). \tag{5.2}$$

  A random field with this covariance function is mean square continuous but nondifferentiable. As a result, the exponential model (5.2) is associated with random fields with rough sample paths.

- Matérn model with $\nu = 3/2$, which corresponds to the product between an exponential structure and a polynomial of degree one:

$$h \mapsto \left(1 + \frac{h}{\phi}\right) \exp\left(-\frac{h}{\phi}\right). \tag{5.3}$$

  This model generates once mean square differentiable random fields. In comparison with the exponential model, (5.3) produces random fields with smoother sample paths.

To study the discrepancy between ESS_B and ESS, the covariance structures defined in Equations (5.2) and (5.3) are considered for certain fixed sets of parameters. In all cases, a grid of the form

$\{1,\ldots,n_1\}\times\{1,\ldots,n_2\}$ is considered, where n_1 and n_2 vary from 8 to 64. ϕ is chosen in such a way that both models have the same range. Three different values for the range are considered: 3, 9 and 16. For simplicity, in the computation of ESS_B, the subdivision of the domain considers the same number of points in each block, and the points within each block are selected as close together as possible using the Euclidean distance measure (Hartigan and Wong, 1979).

The values of ESS_B are depicted in Figures 5.3 and 5.4, including the traditional ESS and ESS_G, which are obtained from the limit partitions. It is observed that the results are in agreement with the inequalities stated in Proposition 3.1. We also observe that ESS_B does not necessarily belong to the interval $[\mathrm{ESS}_G, \mathrm{ESS}]$. In percentage terms, the difference between the sample size reductions produced by the ESS and ESS_B with respect to n never exceeds 2% for every combination of block partition, sample size, and range.

Figure 5.5 reports ESS_B versus $n = 4k \times 2k$ with $k = 1,\ldots,10$ for the exponential correlation structure (5.2) and for different values of the range: 3, 9, and 16. In this example, ESS_B is implemented with two blocks of size $n/2$. An increasing trend is apparent. Similar studies are carried out for additional parametric families of covariance functions, such as the spherical model. Since similar results are obtained, they are omitted for the sake of brevity.

5.4. Example 4: Simultaneous Autoregressive Models on a Plane

Let $\boldsymbol{X}$ be a vector with entries $X_i = X(\boldsymbol{s}_i)$ for $\boldsymbol{s}_i \in \mathbb{R}^2$ and $i = 1,\ldots,n$, as described in Section 2. Assume that $\boldsymbol{X}$ has a simultaneous autoregressive (SAR) structure with zero mean, that is,

$$\boldsymbol{X} \;=\; \boldsymbol{B}\boldsymbol{X} + \boldsymbol{\varepsilon}, \tag{5.4}$$

where $\boldsymbol{\varepsilon}$ is a Gaussian random vector with zero mean and covariance matrix $\sigma^2\boldsymbol{I}_n$. Here, $\boldsymbol{B} = \rho\boldsymbol{W}$, with $\boldsymbol{W}$ standing for a row-standardized contiguity matrix and $0 < \rho < 1$ (Cressie, 1993). In other words, the (i,j)th entry of $\boldsymbol{W}$ is different from zero if and only if sites i and j are neighbors; otherwise, it is identically equal to 0. Consequently, the matrix $\boldsymbol{W}$ characterizes the association between the spatially indexed random variables, as it induces a dependency structure between neighbors, whereas ρ is a parameter that controls the degree of spatial autocorrelation. Recent interesting research about this model can be found in Ver Hoef et al. (2018), where several appealing attributes of SAR models are reviewed together with their importance in areas as diverse as disease mapping, econometrics and image analysis, as well as in Griffith (2015), where efficient algorithms and approximation techniques are proposed.

The definition of ESS_B can be directly used for this parametric model by noticing that the covariance matrix of $\boldsymbol{X}$ is given by $\sigma^2[(\boldsymbol{I}_n - \rho\boldsymbol{W})^\top(\boldsymbol{I}_n - \rho\boldsymbol{W})]^{-1}$. Of course, this strategy is not recommended for extremely large datasets. Hence, we now describe some procedures that allow us to obtain ESS_B efficiently. Let $\boldsymbol{W}_{b_ib_j}$ be the row-standardized contiguity matrix between blocks b_i and b_j (note that if blocks b_i and b_j are not adjacent, then $\boldsymbol{W}_{b_ib_j}$ is the zero matrix). We define $\boldsymbol{V}_{ij} = (\boldsymbol{I}_{|b_i|+|b_j|} - \rho\widetilde{\boldsymbol{W}}_{b_ib_j})^\top(\boldsymbol{I}_{|b_i|+|b_j|} - \rho\widetilde{\boldsymbol{W}}_{b_ib_j})$, such that the covariance matrix of $(\boldsymbol{X}_{b_i}, \boldsymbol{X}_{b_j})^\top$ is $\sigma^2\boldsymbol{V}_{ij}^{-1}$, where

$$\widetilde{\boldsymbol{W}}_{b_ib_j} = \begin{pmatrix} \boldsymbol{W}_{b_ib_i} & \boldsymbol{W}_{b_ib_j} \\ \boldsymbol{W}_{b_jb_i} & \boldsymbol{W}_{b_jb_j} \end{pmatrix}.$$

In addition, consider the partitions

$$\boldsymbol{V}_{ij} = \begin{pmatrix} \boldsymbol{A}_{ij}^{11} & \boldsymbol{A}_{ij}^{12} \\ \boldsymbol{A}_{ij}^{21} & \boldsymbol{A}_{ij}^{22} \end{pmatrix} \quad \text{and} \quad \boldsymbol{V}_{ij}^{-1} = \begin{pmatrix} \boldsymbol{G}_{ij}^{11} & \boldsymbol{G}_{ij}^{12} \\ \boldsymbol{G}_{ij}^{21} & \boldsymbol{G}_{ij}^{22} \end{pmatrix},$$

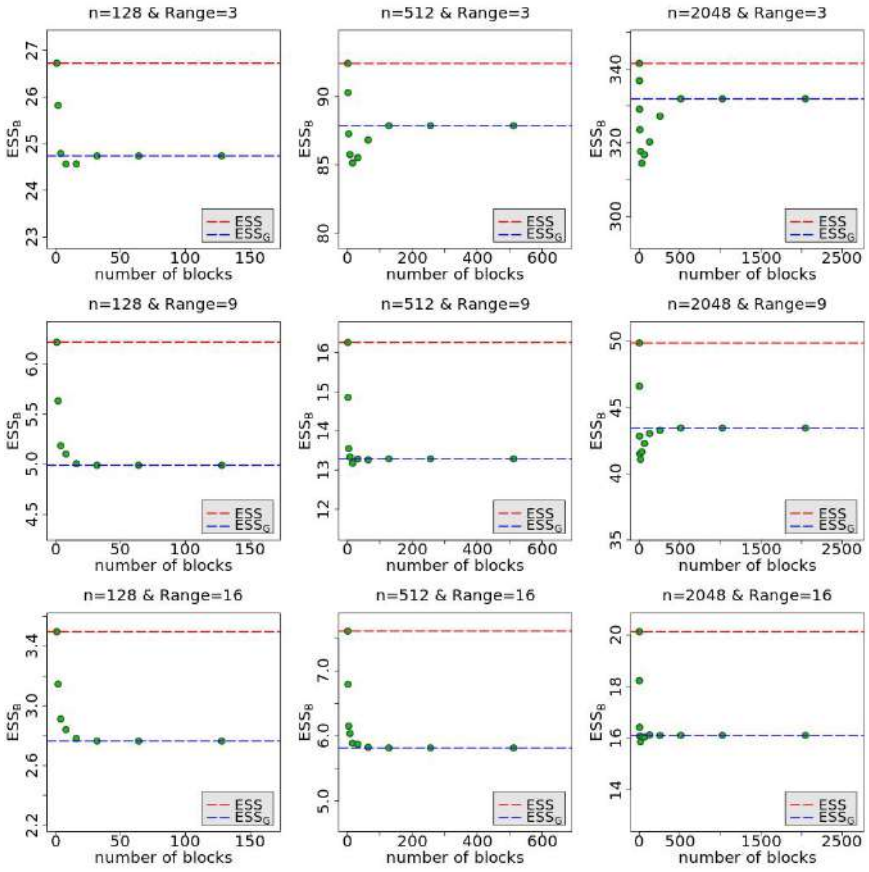

Figure 5.3: ESS_B (green circles) versus the number of blocks for the covariance model (5.2) and for different sample sizes ($16 \times 8 = 128$, $32 \times 16 = 512$, and $64 \times 32 = 2048$) and ranges (3, 9, and 16). The red and blue dashed lines represent the ESS and ESS_G, respectively.

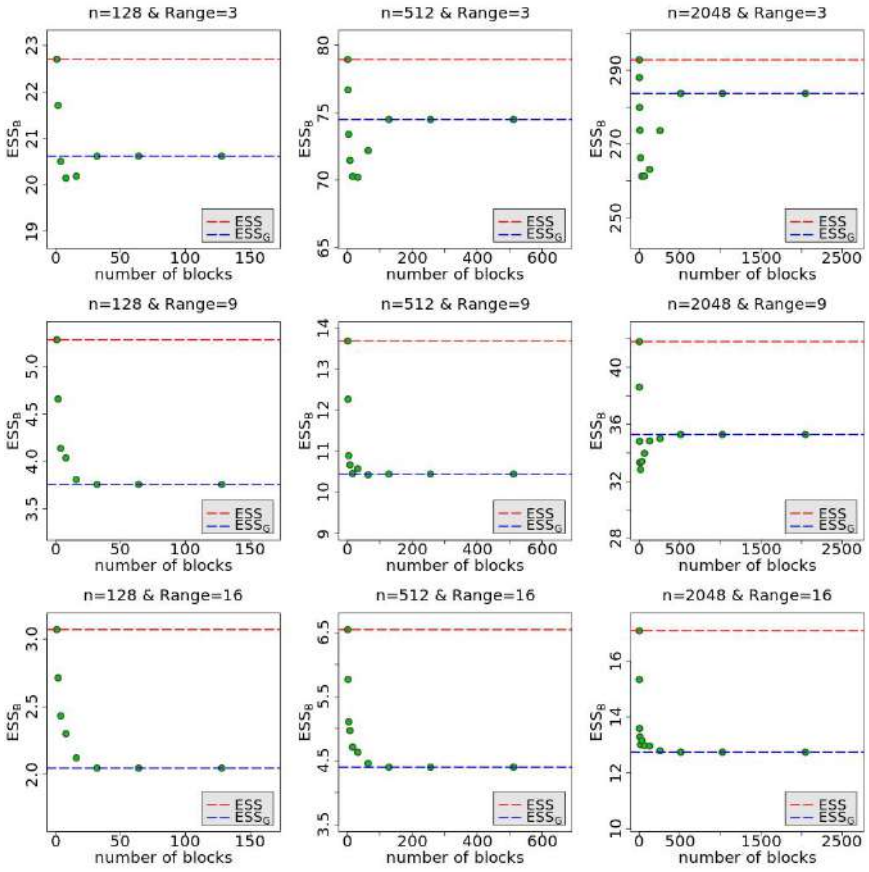

Figure 5.4: ESS_B (green circles) versus the number of blocks for the covariance model (5.3) and for different sample sizes ($16 \times 8 = 128$, $32 \times 16 = 512$, and $64 \times 32 = 2048$) and ranges (3, 9, and 16). The red and blue dashed lines represent the ESS and ESS_G, respectively.

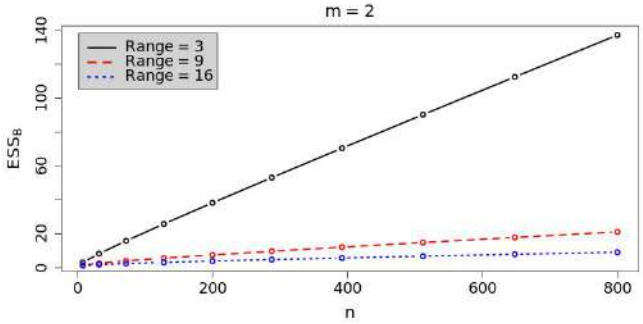

Figure 5.5: ESS_B versus $n = 4k \times 2k$, with $k = 1, \ldots, 10$, for the exponential model (5.2) with range $= 3,\ 9,\ 16$. For each value of n, we have considered ESS_B with two blocks of size $n/2$.

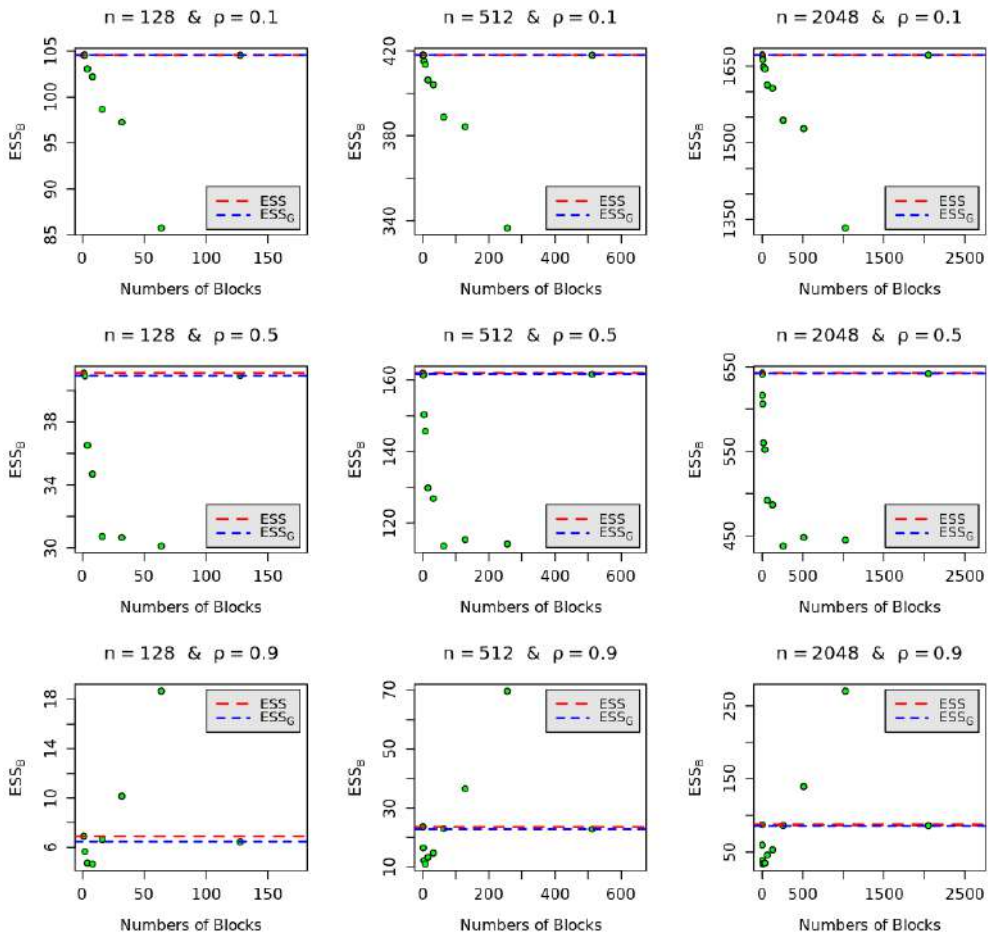

Figure 5.6: ESS_B (green circles) versus the number of blocks for the SAR model, for different sample sizes ($16 \times 8 = 128$, $32 \times 16 = 512$, and $64 \times 32 = 2048$) and different values of ρ (0.1, 0.5, and 0.9). The red and blue dashed lines represent the ESS and ESS_G, respectively.

and the diagonal matrix $\boldsymbol{C}_{b_i b_j} = \text{diag}\left\{\boldsymbol{V}_{ij}^{-1}\right\}^{1/2} = (\boldsymbol{C}_{(i)}, \boldsymbol{C}_{(j)})$. Thus, $\boldsymbol{R}_{b_i}^{-1} = \boldsymbol{C}_{(i)}\boldsymbol{A}_{ij}^{11}\boldsymbol{C}_{(i)}$, $\boldsymbol{R}_{b_j}^{-1} = \boldsymbol{C}_{(j)}\boldsymbol{A}_{ij}^{22}\boldsymbol{C}_{(j)}$, and $\boldsymbol{R}_{b_i b_j} = \boldsymbol{C}_{(i)}^{-1}\boldsymbol{G}_{ij}^{12}\boldsymbol{C}_{(j)}^{-1}$. Therefore

$$\eta_{ij} = \mathbf{1}_{|b_i|}^{\top}\boldsymbol{C}_{(i)}\boldsymbol{A}_{ij}^{11}\boldsymbol{G}_{ij}^{12}\boldsymbol{A}_{ij}^{22}\boldsymbol{C}_{(j)}\mathbf{1}_{|b_j|}, \tag{5.5}$$

$$\eta_{ii} = \mathbf{1}_{|b_i|}^{\top}\boldsymbol{C}_{(i)}\boldsymbol{A}_{ij}^{11}\boldsymbol{C}_{(i)}\mathbf{1}_{|b_i|}. \tag{5.6}$$

If blocks b_i and b_j are not adjacent, then $\eta_{ij} = 0$. Moreover, if the $m = m_1 \times m_2$ blocks are identical, as discussed in Section 4, then, ESS_B simplifies to

$$\text{ESS}_B = \frac{(m\eta_{11})^2}{m\eta_{11} + 2[(m_1 - 1)m_2 + (m_2 - 1)m_1]\eta_{12}}. \tag{5.7}$$

In Figure 5.6, the behavior of ESS_B is explored in terms of the number of blocks for different combinations of sample sizes and values of ρ. The ESS and ESS_G, which are the limit cases of our proposal, are similar in each scenario. The sample size reduction produced by ESS_B is similar to that produced by the ESS and ESS_G (the worst discrepancies, relative to n, are approximately 10%, but in most cases, they are much less than this percentage).

6. Simulation Study

6.1. Maximum Likelihood Versus Block Likelihood Estimation

The goal of this section is to compare the relative performances of the block likelihood (BL) and maximum likelihood (ML) estimation methods. We focus on the estimation of the range parameter of the covariance function because it is the aspect that most influences the assessment of the effective sample size. All computations described below were performed using a computer equipped with a 2.7 GHz processor and 8 GB of RAM.

In pursuance of running the experiments, lattices in $\mathbb{R}^2$ of sizes $16 \times 8 = 128$, $32 \times 16 = 512$, and $48 \times 32 = 1152$ were considered. These lattices were chosen to keep the computation of the full likelihood feasible. The Matérn-type correlation models detailed in Equations (5.2) and (5.3) were used with different ranges (3, 9, and 16). For each combination of correlation model, range, and lattice, 1000 independent random fields with zero mean and unit variance were simulated, and then the corresponding ML and BL estimates of the range were obtained. The BL method was implemented with three different block configurations: Δ_1 (two identical blocks), Δ_2 (four identical blocks), and Δ_3 (eight identical blocks), in a similar fashion to the examples illustrated in Section 5.3.

Figure 6.1 displays the boxplots of the range estimates for models (5.2) and (5.3). In each case, the estimations obtained by both methods appear unbiased, and it is also observed that an increase in the number of blocks leads to an increase in the variability of the BL estimates. When $n = 128$, many atypical BL estimates are observed for configuration Δ_3. This is not surprising because in this case, each block only contains 16 observations. Each estimation method shows less pronounced variability as the sample size increases. Table 6.1 summarizes the relative root mean squared errors (rRMSEs) of the ML estimates with respect to the BL estimates,

Table 6.1: Relative root mean squared errors (rRMSEs) between the ML and BL estimates based on 1000 independent realizations. We considered models (5.2) and (5.3), different samples sizes ($16 \times 8 = 128$, $32 \times 16 = 512$, and $48 \times 24 = 1152$) and different ranges (3, 9, and 16). The BL method was implemented with three different block configurations: Δ_1 (two identical blocks), Δ_2 (four identical blocks), and Δ_3 (eight identical blocks).

		Range = 3			Range = 9			Range = 16		
		Δ_1	Δ_2	Δ_3	Δ_1	Δ_2	Δ_3	Δ_1	Δ_2	Δ_3
Model (5.2)	$n = 128$	0.68	0.44	0.22	0.68	0.46	0.24	0.68	0.46	0.24
	$n = 512$	0.71	0.48	0.33	0.72	0.51	0.35	0.72	0.51	0.36
	$n = 1152$	0.70	0.49	0.34	0.70	0.50	0.35	0.70	0.50	0.35
Model (5.3)	$n = 128$	0.71	0.46	0.28	0.70	0.47	0.28	0.70	0.46	0.28
	$n = 512$	0.70	0.49	0.34	0.72	0.51	0.35	0.72	0.51	0.35
	$n = 1152$	0.70	0.49	0.35	0.70	0.50	0.36	0.70	0.51	0.36

i.e., rRMSE = $\text{RMSE}(\widehat{\phi}_{\text{ML}})/\text{RMSE}(\widehat{\phi}_{\text{BL}})$. For a given partition, the rRMSE does not vary remarkably across all combinations of sample sizes and ranges. The BL estimates based on Δ_1 always outperform the other BL alternatives.

Examining the behaviors of the different models, it is observed that for the Matérn model with $\nu = 1.5$, all methods provide more accurate estimates, which is consistent with previous findings related to composite likelihood inference (Bevilacqua and Gaetan, 2015). Thus, the estimations seem to be more precise for smoother random fields.

We now turn to a comparison of the computational time required for evaluating the objective functions associated with each estimation method. To accomplish this, lattices of sizes $n = 16k \times 8k$ for $k = 1, \ldots, 16$ were considered. Figure 6.2 reports the evaluation times (in seconds) for the ML and BL methods in terms of the sample size. The computational benefits are clear for the BL method. The computational burden of the ML rapidly increased for a few thousands observations.

For regularly spaced data, a convenient implementation of BL requires the consideration of all blocks with identical configurations. In such a case, the correlation matrix is the same for each block; hence, the only quantity that remains to be evaluated in each iteration of the maximization process is the inverse of one correlation matrix. As a result, the order of computation is $(n/m)^3$.

6.2. $\widehat{\text{ESS}}$ *Versus* $\widehat{\text{ESS}}_B$

We now investigate how the variabilities of the ML and BL estimates affect the variabilities of $\widehat{\text{ESS}} = \text{ESS}(\widehat{\phi}_{\text{ML}})$ and $\widehat{\text{ESS}}_B = \text{ESS}_B(\widehat{\phi}_{\text{BL}})$. To measure the absolute difference between the sample size reduction percentages generated by The $\widehat{\text{ESS}}$ and $\widehat{\text{ESS}}_B$ relative to n, the quantity

$$\frac{1}{n}\left|\widehat{\text{ESS}} - \widehat{\text{ESS}}_B\right|, \tag{6.1}$$

is analyzed. We evaluate (6.1) for the different scenarios described in the previous subsection.

Figure 6.3 displays the results obtained for both correlation models. The largest discrepancies occur in those cases where the range is 3; however, the median difference never exceeds 10%. The variability of (6.1) is considerable for the third block configuration when $n = 128$ and $n = 512$. Again, this pattern is expected because for small sample sizes, the third block partition generates

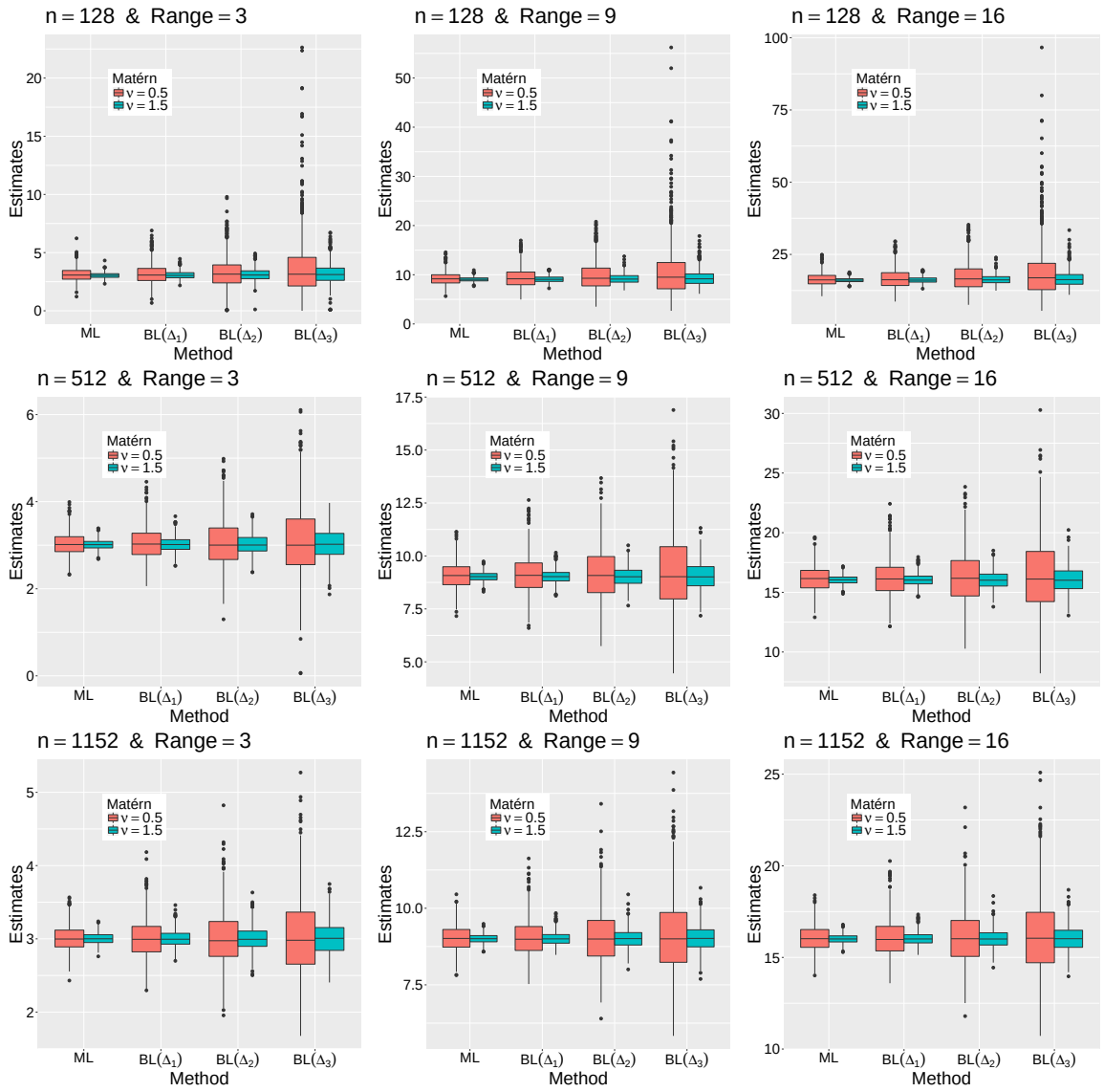

Figure 6.1: ML and BL estimates of the range based on 1000 independent realizations. We considered models (5.2) and (5.3) and different sample sizes ($16 \times 8 = 128$, $32 \times 16 = 512$, and $48 \times 24 = 1152$) and ranges (3, 9, and 16). The BL method was implemented with three different block configurations: Δ_1 (two identical blocks), Δ_2 (four identical blocks), and Δ_3 (eight identical blocks).

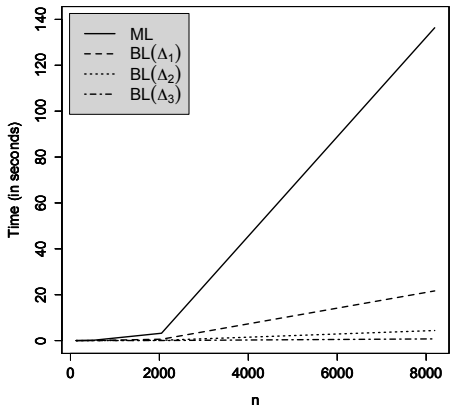

Figure 6.2: Time (in seconds) versus sample size when evaluating the objective functions of the ML and BL methods. The BL method was implemented with three different block configurations: Δ_1 (two identical blocks), Δ_2 (four identical blocks), and Δ_3 (eight identical blocks).

quite small blocks. In the scenarios where the range is 9 or 16, $\widehat{\text{ESS}}_B$ exhibits robust behavior, even for the third block configuration. As the sample size increases, the sample size reduction tends to be the same for all estimation methods. For example, when $n = 1152$, the maximum discrepancy is approximately 1.5% when the range is 9 and approximately 0.6% when the range is 16. The variability of (6.1) is less prominent for model (5.3).

7. Data Analysis

We illustrate the use of the block likelihood approach developed in Section 3 on a forest dataset, which consists of a three-band reference image of size 5616×3744 pixels and thus represents a dataset with a large sample size ($n = 21026304$). The image is shown in Figure 7.1, and it was taken above a section of forest at the Harvard Forest, Petersham, MA, USA. The image belongs to one of the comprehensive databases that are part of a long study carried out in Harvard Forest. The image and code for handling it in `R` are available from `https://github.com/JAcostaS/Code-and-Example-Codismap.git`. To apply our methodology, the reference image has been transformed to a grayscale image.

Computational Cost of the ESS. : For this image, it is infeasible to obtain the traditional ESS because it is not possible to obtain the pixel distance matrix. When trying to calculate the distance matrix, the software R indicates that it "cannot locate a size vector 1646971.1 GB" because such a large amount of memory is not available. In other words, a RAM size of 1608.37 TB is needed to be able to perform the calculation of the distance matrix, which is a necessary step for obtaining the correlation matrix. In addition, even more RAM is needed for parameter estimation. Even if such an amount of RAM were available, processing would be too slow because the inverse of the correlation matrix needs to be evaluated during the estimation process as well as in the ESS calculation.

The ESS_B introduced in this paper is computed for the abovementioned large dataset. First, for 324 blocks of size 312×208, the local empirical variograms are estimated (see Figure 7.2). The

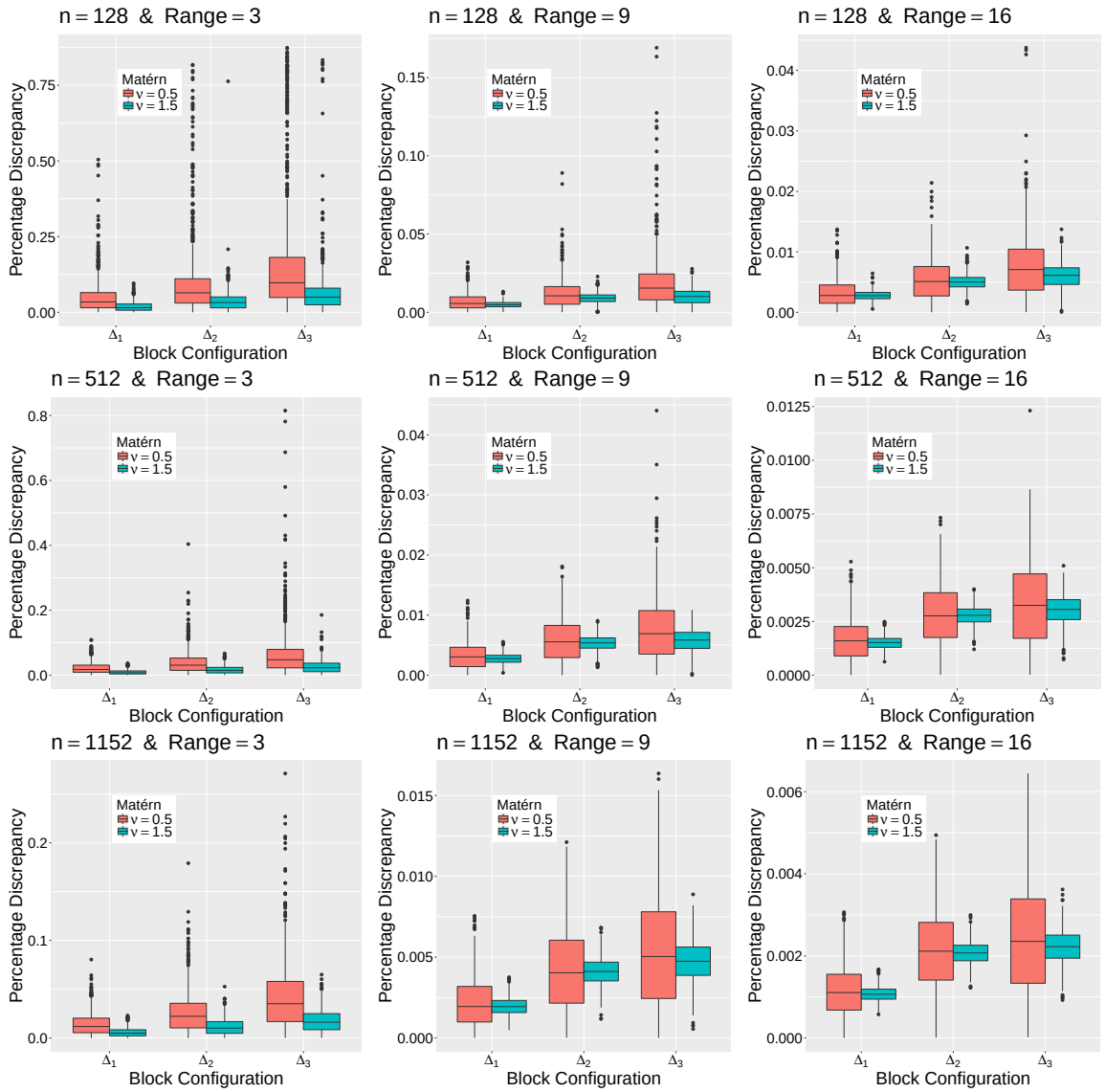

Figure 6.3: Absolute difference between the sample size reduction percentages generated by the $\widehat{\text{ESS}}$ and $\widehat{\text{ESS}}_B$ relative to n. We consider models (5.2) and (5.3), different sample sizes ($16 \times 8 = 128$, $32 \times 16 = 512$, and $48 \times 24 = 1152$) and different ranges (3, 9, and 16). The BL method is implemented with three different block configurations: Δ_1 (two identical blocks), Δ_2 (four identical blocks), and Δ_3 (eight identical blocks).

Figure 7.1: Reference image of size 5616×3744 pixels taken above a section of forest at the Harvard Forest, Petersham, MA, USA.

blue violet line represents the average across the 324 empirical variograms, and this gives us an idea of what the global empirical variogram looks like. Then, Matérn covariance models are fitted to this dataset by means of the block-likelihood estimation method. σ^2 and ϕ are estimated for different fixed values of ν. The red, green, blue and sky blue lines in Figure 7.2 correspond to the BL fits for $\nu = 0.5$, $\nu = 1.0$, $\nu = 1.5$ and $\nu = 2.0$, respectively. We consider three configurations in the block-likelihood estimation process: Δ_1: 54756 blocks of size 24×16, Δ_2: 10816 blocks of size 54×36, and Δ_3: 1521 blocks of size 144×96 (the results are displayed in Table 7.1). The involved distances never exceed 27.46, 63.51 and 171.68 for configurations Δ_1, Δ_2 and Δ_3, respectively. It is observed from Figure 7.2 that the practical range is approximately 50, which means that choosing blocks whose diagonals are greater than 50 ensures reasonable variogram estimates. Table 7.1 reports the respective ESS_B estimates for the Matérn models mentioned above. When $\nu = 0.5$, the estimates of σ^2 are very unsatisfactory, as shown in Figure 7.2. For the cases in which $\nu = 1.0$, $\nu = 1.5$ and $\nu = 2.0$, the estimations are highly stable across different configurations. Note that $\widehat{\text{ESS}}_B$ varies drastically among the covariance models.

We turn to a comparison between the full image and some subsamples. The estimated ESS_B for $\nu = 1.0$ and $\nu = 1.5$ are used to choose the sizes of these subsamples. We expect the subsamples to capture the essential statistical information of the original image. The sizes of the subsamples are selected in such a way that the proportion between rows and columns of 3:2 is preserved. Specifically, the sample sizes are $27 \times 18 = 486$, $33 \times 22 = 726$, $120 \times 80 = 9600$ and $129 \times 86 = 11094$. For each subsample, two types of sampling schemes are considered: random and regular. It is worth mentioning that the sampling process is performed on the entire image, i.e., there is no specific treatment within each block. In particular, in the random sampling scheme, a subsample of pixels is selected from the entire image through uniform sampling along the rows and columns (e.g., to obtain a subsample of size 486, we sample 27 columns and 18 rows out of the 5616 columns and 3744 rows of the entire image, respectively). Table 7.2 contains descriptive statistics for the original and subsample images. The mean and standard deviation are successfully reproduced regardless of

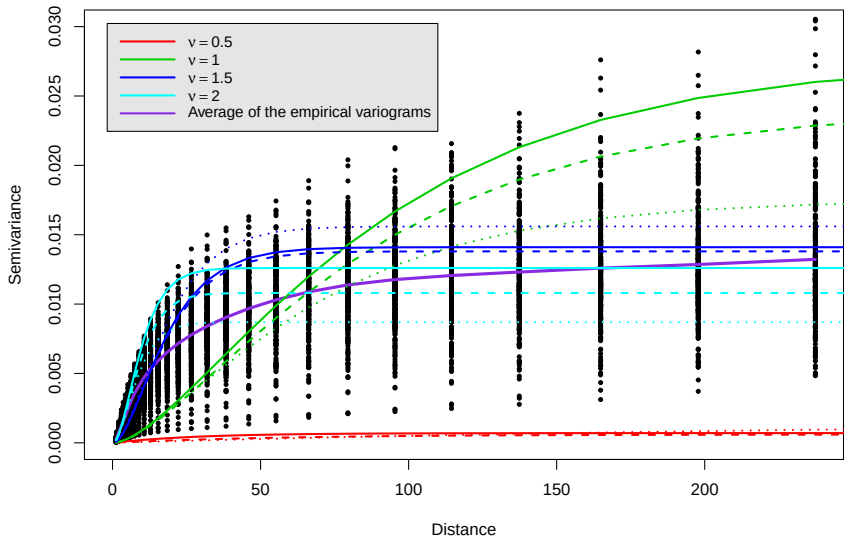

Figure 7.2: Estimated empirical variograms for the 324 selected blocks (black points). The blue violet line represents the average across the 324 estimations, while the red, green, blue, and sky blue lines represent the BL fits for $\nu = 0.5$, $\nu = 1.0$, $\nu = 1.5$ and $\nu = 2.0$, respectively. The different types of lines represent different block configurations: plane for Δ_1, segment for Δ_2 and dotted line for Δ_3.

Table 7.1: BL and ESS_B estimates for different block configurations.

Model	Block Size	Number of Blocks	$\widehat{\sigma}^2$	$\widehat{\phi}$	Estimated Range	$\widehat{\text{ESS}}_\text{B}$
$\nu = 0.5$	24×16	54756	0.0007	28.868	86.48	4051
	54×36	10816	0.0006	58.010	173.78	1018
	144×96	1521	0.0016	265.491	795.34	111
$\nu = 1.0$	24×16	54756	0.0275	60.678	242.62	474
	54×36	10816	0.0240	58.384	233.45	510
	144×96	1521	0.0175	46.601	186.33	765
$\nu = 1.5$	24×16	54756	0.0141	9.959	47.21	11206
	54×36	10816	0.0138	9.961	47.22	10522
	144×96	1521	0.0156	9.944	47.13	9315
$\nu = 2.0$	24×16	54756	0.0126	4.436	23.82	40538
	54×36	10816	0.0108	4.262	22.88	38503
	144×96	1521	0.0087	3.890	20.88	46238

the sampling scheme used, and these statistics improve as the sample size increases. In addition, it is observed that under a regular sampling scheme, the percentiles are also accurately estimated. Although the minimum and maximum values are reasonably approximated, they present a more significant imprecision.

Table 7.2: Summary statistics for the original data and the subsamples. The subsamples are obtained using random and regular sampling schemes.

	Original	Random Sampling				Regular Sampling			
Statistics	Image	486	726	9600	11094	486	726	9600	11094
Min.	0.0039	0.0458	0.0431	0.0157	0.0275	0.0431	0.0314	0.0118	0.0118
25%.	0.1281	0.1281	0.1350	0.1242	0.1294	0.1232	0.1333	0.1281	0.1281
50%.	0.1987	0.1954	0.1974	0.1908	0.1987	0.1908	0.2098	0.1987	0.1987
75%.	0.3098	0.3196	0.3134	0.3059	0.3072	0.2967	0.3134	0.3111	0.3098
Max.	0.8379	0.5386	0.5464	0.6719	0.6771	0.5320	0.5686	0.6261	0.6562
Mean	0.2245	0.2257	0.2278	0.2196	0.2243	0.2188	0.2279	0.2242	0.2243
s.d.	0.1150	0.1204	0.1147	0.1152	0.1141	0.1147	0.1126	0.1150	0.1150

Figure 7.3 shows the marginal empirical distribution functions for the original image (green line) and for the subsamples (red dashed lines for the random sampling scheme and black dotted lines for the regular scheme). The behaviors of both sampling schemes are very similar and impossible to distinguish for large sample sizes. Table 7.3 displays some quality measures, such as the bias, standard deviation and maximum difference between the original and subsample images. We observe substantial improvements in these measures when the sample size increases from 726 to 9600. To reinforce our results, the last column of Table 7.3 shows the p-values of the two-sample Kolmogorov-Smirnov test.

Table 7.3: Summary statistics for the differences between the empirical distributions of the original data and the subsamples. The p-values correspond to the two-sample Kolmogorov-Smirnov test.

Sampling	Size	Mean	Standard Deviation	Maximum Difference	p-value
Random	486	0.0012	0.0096	0.0319	0.3711
	726	0.0033	0.0081	0.0379	0.3871
	9600	-0.0003	0.0023	0.0079	0.2452
	11094	-0.0049	0.0073	0.0004	0.1861
Regular	486	0.0034	0.0094	0.0371	0.6605
	726	-0.0057	0.0089	0.0095	0.5078
	9600	-0.0003	0.0013	0.0023	0.3505
	11094	0.0058	0.0029	0.0093	0.7158

We also employ the SAR model to perform additional analysis regarding this dataset. For a clearer exposition, these results are available in Appendix B.

8. Discussion

This paper introduced a new way to address the computation of the effective sample size. The methodology is based on block likelihood inference. We showed that ESS_B preserves some relevant

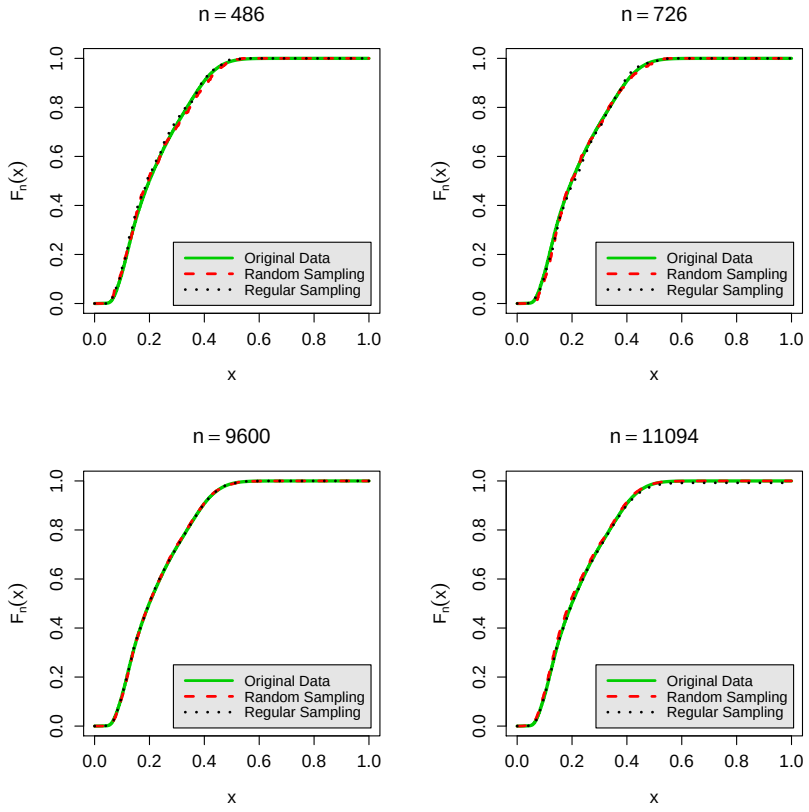

Figure 7.3: Empirical distributions of the original data and the subsamples: (top left) subsample of size 486, (top right) subsample of size 726, (bottom left) subsample of size 9600, and (bottom right) subsample of size 11094.

attributes of the traditional ESS. The approach, equipped with powerful computational machinery, is appropriate for large spatial datasets and reduces to the original ESS when the number of blocks is equal to one. The use of our findings has been illustrated with a real forest dataset. Once ESS_B was determined, we showed that subsamples of such a size preserve the main properties of the original image, thereby illustrating the effectiveness of our proposal. Not only is the effective sample size a relevant feature when planning a sampling design, but how to choose such locations in an optimal way is also a crucial problem that deserves special attention. Several alternatives related to this task have been discussed in past studies, including random, systematic (regular), and stratified schemes, as well as hybrid strategies (see, e.g., Griffith, 2008 and Li et al., 2016).

Although Monte Carlo simulations and applications were developed and illustrated for spatial datasets defined on regular grids in this paper, our methodology may be applied for sample size reduction with irregularly spaced spatial data. In addition, we focused on spatial random fields with a constant mean, so the extension of this approach to general spatial regression models is a natural research topic that we expect to tackle in further research, along the lines of Acosta and Vallejos (2018). Another promising research direction is the formulation of alternative sample size reduction techniques based on different estimating equations (see, e.g., Bevilacqua and Gaetan, 2015; Sun et al., 2018; Bachoc et al., 2019; Litvinenko et al., 2019). Here, the derivation of an

amenable expression for the Godambe information could be challenging. Generalizations to non-Gaussian random fields could be useful for analyzing data in the presence of asymmetry or heavy tails (Xu and Genton, 2015, 2017).

Acknowledgements

Alfredo Alegría acknowledges the funding of the National Agency for Research and Development of Chile, through grant ANID/FONDECYT/INICIACIÓN/No. 11190686. Felipe Osorio was partially supported by UTFSM through grant PI_LI_19_11. Ronny Vallejos was partially supported by the AC3E, UTFSM, under grant FB-0008, and by UTFSM under grant PI_L_18_20. Ronny Vallejos and Felipe Osorio also acknowledge financial support from CONICYT through the MATH-AMSUD program, grant 20-MATH-03.

Appendices

A. Proof of Proposition 3.1

We now proceed with a point-to-point proof of Proposition 3.1.

(1) Suppose that all the observations are independent. It follows that $\boldsymbol{R}_{b_i}(\boldsymbol{\theta}) = \boldsymbol{I}_{|b_i|}$ for all $i = 1, \ldots, m$ and that $\boldsymbol{R}_{b_i b_j}(\boldsymbol{\theta})$ is identically equal to the matrix of zeros of dimension $|b_i| \times |b_j|$ for all $i \neq j$. As a result, $\eta_{ii} = |b_i|$ and $\eta_{ij} = 0$; hence, $\mathrm{ESS}_B = |b_1| + \cdots + |b_m| = n$.

(2) Suppose now that all the observations are perfectly correlated. Then, $\boldsymbol{R}_{b_i b_j}(\boldsymbol{\theta}) = \mathbf{1}_{|b_i|}\mathbf{1}_{|b_j|}^\top$ for all $i, j = 1, \ldots, m$. Furthermore, the pseudoinverse of $\boldsymbol{R}_{b_i}(\boldsymbol{\theta})$ is given by $\boldsymbol{R}_{b_i}^+(\boldsymbol{\theta}) = \mathbf{1}_{|b_i|}\mathbf{1}_{|b_i|}^\top/|b_i|^2$. Thus, it is straightforward to see that $\eta_{ii} = \eta_{ij} = 1$ for all $i, j = 1, \ldots, m$. Consequently, $\mathrm{ESS}_B = 1$.

(3) The inequality $\mathrm{ESS}_B \leq \mathrm{ESS}$ is true because the Fisher information is always greater than the Godambe information (Godambi and Kale, 1991, pp. 3-20). The goal now is to prove that $\mathrm{ESS}_B \geq 1$. Note that ESS_B can be written as

$$\mathrm{ESS}_B = \frac{\mathrm{tr}(\boldsymbol{\eta})^2}{\mathbf{1}_m^\top \boldsymbol{\eta} \mathbf{1}_m},$$

where $\boldsymbol{\eta}$ is an $m \times m$ matrix with elements η_{ij} as in (3.7). Additionally, note that $\boldsymbol{\eta}$ is nonnegative definite since it has a representation of the form

$$\boldsymbol{\eta} = \mathbb{E}\left\{\boldsymbol{L}\boldsymbol{L}^\top\right\},$$

where $\boldsymbol{L} = (\partial \ell_{b_1}/\partial\mu, \ldots, \partial \ell_{b_m}/\partial\mu)^\top \in \mathbb{R}^m$. Observe that $|\eta_{ij}| \leq \sqrt{\eta_{ii}\eta_{jj}}$ for all $i, j = 1, \ldots, m$, so we have that

$$\mathbf{1}_m^\top \boldsymbol{\eta} \mathbf{1}_m = \sum_{i,j=1}^m \eta_{ij} \leq \sum_{i,j=1}^m \sqrt{\eta_{ii}}\sqrt{\eta_{jj}} = \left(\sum_{i=1}^m \sqrt{\eta_{ii}}\right)^2.$$

Using Jensen's inequality in the last expression, one obtains that

$$\mathbf{1}_m^\top \boldsymbol{\eta} \mathbf{1}_m \leq m \sum_{i=1}^m \eta_{ii} = m\,\mathrm{tr}(\boldsymbol{\eta}).$$

Finally, we recall that η_{ii} is the traditional ESS of the ith block; thus, $\eta_{ii} \geq 1$ (see Vallejos and Osorio, 2014), and of course, $\text{tr}(\boldsymbol{\eta}) \geq m$. Thus,

$$\mathbf{1}_m^\top \boldsymbol{\eta} \mathbf{1}_m \leq \text{tr}(\boldsymbol{\eta})^2,$$

and the result follows.

(4) When $m = 1$, we only have a single block of length $|b_1| = n$. Thus,

$$\text{ESS}_B = \eta_{11} = \mathbf{1}_n^\top \boldsymbol{R}^{-1}(\boldsymbol{\theta}) \mathbf{1}_n = \text{ESS}\,.$$

(5) When $m = n$, the length of each block is $|b_i| = 1$ for $i = 1, \ldots, m$. The matrices involved in Definition 3.1 are given by $\boldsymbol{R}_{b_i}(\boldsymbol{\theta}) = 1$ and $\boldsymbol{R}_{b_i b_j}(\boldsymbol{\theta}) = r_{ij}$, where r_{ij} denotes the (i, j)th element of $\boldsymbol{R}(\boldsymbol{\theta})$. Therefore, $\eta_{ij} = r_{ij}$ and $\eta_{ii} = 1$. Thus,

$$\text{ESS}_B = \frac{n^2}{\sum_{j=1}^n \sum_{i=1}^n r_{ij}} = \frac{n^2}{\mathbf{1}_n^\top \boldsymbol{R}(\boldsymbol{\theta}) \mathbf{1}_n}.$$

B. Numerical Experiments with the SAR Model

B.1. ML versus BL in the SAR Model

In addition to the experiments conducted in Section 6.1 related to the Matérn family, we also compare the statistical performances of the ML and BL estimates when an SAR model, as in Equation (5.4), is assumed. This study allows us to have a more complete idea about the performance of the BL estimation method.

We focus on the estimation of ρ, which fully characterizes the different notions of the effective sample size. Here, $\sigma^2 = 1$ is fixed. A lattice in $\mathbb{R}^2$ of size $32 \times 16 = 512$ and three different values of ρ (0.1, 0.5, and 0.9) are considered. For each case, we simulate 1000 independent realizations, and then the ML and BL estimates of ρ are obtained. The BL method is implemented with three different block configurations: Δ_1 (two identical blocks), Δ_2 (four identical blocks), and Δ_3 (eight identical blocks). Figure B.1 shows the corresponding boxplots. We perceive that both methods have similar levels of variability when an SAR intrinsic structure is present. This is different from previous experiments performed with the Matérn model, where there was a decrease in the quality of the estimates as the number of blocks increased.

This brief simulation example indicates that the BL estimation approach is particularly compatible with the SAR model. The results reported here are not surprising, as we are aware that even small blocks contain key information about neighboring sites (the only information lost occurs at the border that separates neighboring blocks), so the BL method is extremely competitive for estimating ρ.

B.2. SAR Analysis of the Forest Dataset

We look at the ESS_B estimates for the forest dataset analyzed in Section 7 when the observations are assumed to have an SAR structure. Table B.1 shows the parameter estimates obtained through the BL method with different block configurations, together with the corresponding ESS_B estimates. Note that the ESS_G estimates are also reported, and these are based on the approximation given in Equation (3) of Griffith (2005).

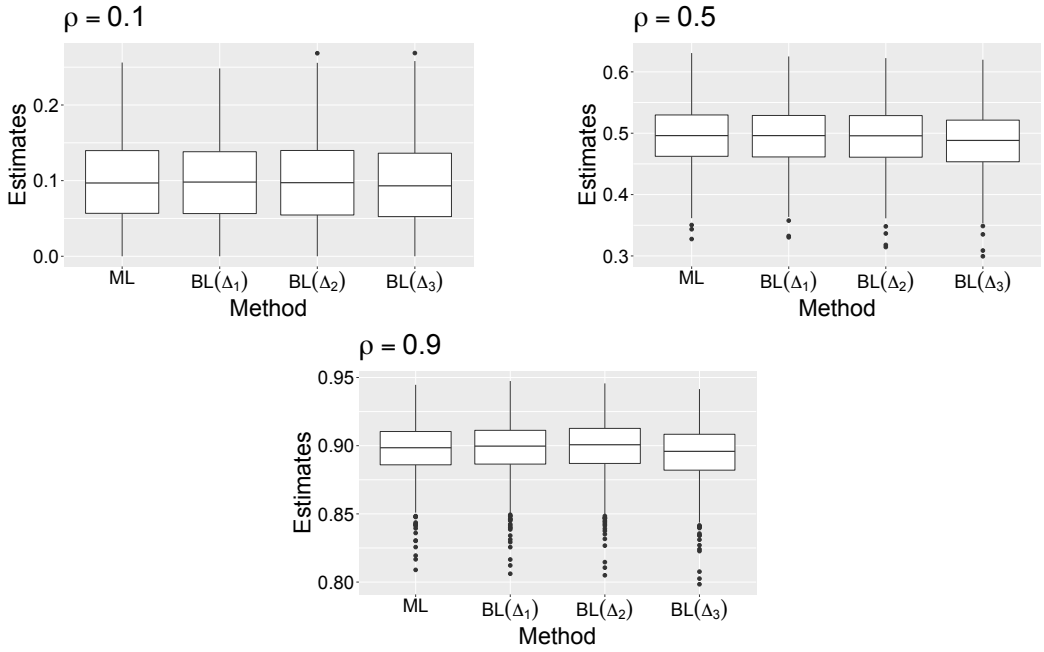

Figure B.1: ML and BL estimates of ρ based on 1000 independent realizations for the SAR model. The sample size used is $32 \times 16 = 512$. We consider three scenarios for ρ (0.1, 0.5, and 0.9). The BL method is implemented with three different block configurations: Δ_1 (two identical blocks), Δ_2 (four identical blocks), and Δ_3 (eight identical blocks).

We observe that the estimates of ρ are close to 1, indicating a high spatial association, which is consistent with similar studies reported by Griffith (2005) in the context of remotely sensed data. The obtained results tell us that the reduction of the sample size can be considerable when we work with this type of image. In this specific case, less than 1% of the pixels of the original image contain the essential statistical information about this dataset.

Table B.1: BL and ESS_B estimates for different block configurations under the SAR model. The ESS_G estimates are also reported.

Block Size	Number of Blocks	$\widehat{\sigma}^2$	$\widehat{\rho}$	$\widehat{\text{ESS}}_G$	$\widehat{\text{ESS}}_B$
24×16	54756	0.00016	0.94782	401107	72370
54×36	10816	0.00009	0.99686	23642	4968
144×96	1521	0.00007	0.99890	8761	1000

References

Abramowitz, M., Stegun, I.A., 1972. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Dover Publications.

Acosta, J., Vallejos, R., 2018. Effective sample size for spatial regression processes. Electronic Journal of Statistics 12, 3147–3180.

Acosta, J., Vallejos, R., Griffth, D., 2018. On the effective geographic sample size. Journal of Statistical Computation and Simulation 88, 1958–1975.

Aho, A.V., Hopcroft, J.E., 1974. The design and analysis of computer algorithms. Pearson Education India.

Bachoc, F., Bevilacqua, M., Velandia, D., 2019. Composite likelihood estimation for a Gaussian process under fixed domain asymptotics. Journal of Multivariate Analysis 174, 104534.

Berger, J., Bayarri, M.J., Pericchi, L.R., 2014. The effective sample size. Econometric Reviews 33, 197–217.

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

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

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

Cressie, N.A.C., 1993. Statistics for Spatial Data. Wiley, New York.

Elvira, V., Martino, L., Robert, C.P., 2018. Rethinking the Effective Sample Size. arXiv:1809.04129 [stat.CO] .

Faes, C., Molenberghs, G., Aerts, M., Verbeke, G., Kenward, M., 2009. The effective sample size and an alternative small-sample degrees-of-freedom method. The American Statistician 88, 389–399.

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

Griffith, D., 2005. Effective geographic sample size in the presence of spatial autorrelation. Annals of the Association of American Geographers 95, 740–760.

Griffith, D.A., 2008. Geographic sampling of urban soils for contaminant mapping: how many samples and from where. Environmental geochemistry and health 30, 495–509.

Griffith, D.A., 2015. Approximation of gaussian spatial autoregressive models for massive regular square tessellation data. International Journal of Geographical Information Science 29, 2143–2173.

Guyon, X., 1995. Random Fields on a Network. Springer, New York.

Hartigan, J.A., Wong, M.A., 1979. Algorithm AS 136: A k-means clustering algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics) 28, 100–108.

Heaton, M.J., Datta, A., Finley, A.O., Furrer, R., Guinness, J., Guhaniyogi, R., Gerber, F., Gramacy, R.B., Hammerling, D., Katzfuss, M., Lindgren, F., Nychka, D.W., Sun, F., Zammit-Mangion, A., 2019. A Case Study Competition Among Methods for Analyzing Large Spatial Data. Journal of Agricultural, Biological and Environmental Statistics 24, 398–425.

Lenth, R.V., 2001. Some practical guidelines for effective sample size determination. The American Statistician 55, 187–193.

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

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

Litvinenko, A., Sun, Y., Genton, M.G., Keyes, D.E., 2019. Likelihood approximation with hierarchical matrices for large spatial datasets. Computational Statistics & Data Analysis 137, 115–132.

Magnus, J., Neudecker, H., 1988. Matrix differential calculus with applications in statistics and econometrics. Wiley Series in Probability and Statistics - Applied Probability and Statistics Section, Wiley.

Solow, A.R., Polanky, S., 1994. Measuring biological diversity. Environmental and Ecological Statistics 1, 95–107.

Stein, M.L., 2012. Interpolation of spatial data: some theory for kriging. Springer Science & Business Media.

Stein, M.L., 2013. Statistical properties of covariance tapers. Journal of Computational and Graphical Statistics 22, 866–885.

Sun, Y., Chang, X., Guan, Y., 2018. Flexible and efficient estimating equations for variogram estimation. Computational Statistics & Data Analysis 122, 45–58.

Vallejos, R., Osorio, F., 2014. Effective sample size of spatial process models. Spatial Statistics 9, 66–92.

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

Ver Hoef, J.M., Hanks, E.M., Hooten, M.B., 2018. On the relationship between conditional (car) and simultaneous (sar) autoregressive models. Spatial statistics 25, 68–85.

Xu, G., Genton, M.G., 2015. Efficient maximum approximated likelihood inference for Tukey's g-and-h distribution. Computational Statistics & Data Analysis 91, 78–91.

Xu, G., Genton, M.G., 2017. Tukey g-and-h random fields. Journal of the American Statistical Association 112, 1236–1249.