Effective sample size for spatial regression models

Jonathan Acosta

Insituto de Estadística, Pontificia Universidad Católica de Valparaíso
Valparaíso, Chile
e-mail: jonathan.acosta@pucv.cl

and

Ronny Vallejos

Departamento de Matemática, Universidad Técnica Federico Santa María
Valparaíso, Chile
e-mail: ronny.vallejos@usm.cl

Abstract: We propose a new definition of effective sample size. Although the recent works of Griffith (2005, 2008) and Vallejos and Osorio (2014) provide a theoretical framework to address the reduction of information in a spatial sample due to spatial autocorrelation, the asymptotic properties of the estimations have not been studied in those studies or in previously ones. In addition, the concept of effective sample size has been developed primarily for spatial regression processes with a constant mean. This paper introduces a new definition of effective sample size for general spatial regression models that is coherent with previous definitions. The asymptotic normality of the maximum likelihood estimation is obtained under an increasing domain framework. In particular, the conditions for which the limiting distribution holds are established for the Matérn covariance family. Illustrative examples accompany the discussion of the limiting results, including some cases where the asymptotic variance has a closed form. The asymptotic normality leads to an approximate hypothesis testing that establishes whether there is redundant information in the sample. Simulation results support the theoretical findings and provide information about the behavior of the power of the suggested test. A real dataset in which a transect sampling scheme has been used is analyzed to estimate the effective sample size when a spatial linear regression model is assumed.

Contents

1. Introduction

This paper introduces a new definition of effective sample size for spatial regression processes. By definition (Cressie, 1993, p. 16), the effective sample size (ESS), which in this paper will be denoted as n^*, is the number of equivalent independent observations associated with a spatial sample of size n. The correct computation of the ESS is important due to the possible existing duplicated information in the data, and it has many implications for the subsequent analysis and inference of spatial models (Clifford et al., 1989; Dutilleul, 1993; Dutilleul et al., 2008). A similar problem has been reported by Box (1954a,b) for approximating the distribution of a quadratic form. This approach was later used by Clifford et al. (1989) to assess the spatial association between two spatial processes. Directly related to the content of this paper is the work of Griffith (2005), who defined the effective sample size for georeferenced data for simple and multiple means and applied the methodology to soil samples in Syracuse, NY (Griffith, 2008). Alternatively, Faes et al. (2009) proposed another way of computing the ESS based on the Fisher information quantity, which is applicable to linear models with replicates. In a Bayesian context, model selection procedures often depend explicitly on the sample size of the experiment. Extensions of the Bayesian information criteria (BIC) for non-iid vectors require a definition of effective sample size that also applies in such cases (Berger et al., 2014). Although discoveries about ESS are still disperse in the literature, standard books on spatial statistics have mentioned the problem of how many uncorrelated samples provide the same precision as correlated observations (Cressie, 1993; Haining, 1990; Schabenberger and Gotway, 2005; Griffith and Paelinck, 2011; Dale and Fortin, 2009). Some applications involving the computation of the ESS can be found in Thiébaux and Zwiers (1984); de Gruijter and ter Braak (1990); Cogley (1999). Practical guidelines for ESS determination can be found in Lenth (2001).

Any given formula to quantify the reduction of information from a spatial sample should depend on the covariance matrix. Consequently, the estimation

of the ESS and posterior asymptotics rely on the estimation of the covariance matrix, which is a subject that has been extensively studied in the literature; however, in the context of spatial statistics, there are no formal treatments regarding the estimation and limiting properties of the ESS. Vallejos and Osorio (2014) proposed a definition of ESS and developed several examples to show the behavior of such a definition. The restricted maximum likelihood (REML) estimation of the ESS was addressed as an application of the conditions stated in Cressie and Lahiri (1993) for the estimation of the variance components of a spatial linear process. If the random field $\left\{Y(\boldsymbol{s}) : \boldsymbol{s} \in D \subset \mathbb{R}^d\right\}$ has been observed on the set of n spatial sites $D_n = \{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_n\} \subset D$, then a model of the form

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

where $\boldsymbol{Y} = (Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n))^\top$, $\boldsymbol{\epsilon} = (\epsilon(\boldsymbol{s}_1), \ldots, \epsilon(\boldsymbol{s}_n))^\top \sim \mathcal{N}(\mathbf{0}, \boldsymbol{R}(\boldsymbol{\theta}))$, $\boldsymbol{\theta} \in \mathbb{R}^q$, $\mathbf{1} = (1, \ldots, 1)^\top$, and $\mu \in \mathbb{R}$, was assumed to define the ESS as

$$n^* = \mathbf{1}^\top \boldsymbol{R}(\boldsymbol{\theta})^{-1}\mathbf{1}, \tag{1.1}$$

which is the Fisher information quantity with respect to μ.

Here, we consider a regression of a response $\boldsymbol{Y}$ given the predictors $\boldsymbol{x}_j = (x_j(\boldsymbol{s}_1), x_j(\boldsymbol{s}_2), \ldots, x_j(\boldsymbol{s}_n))^\top$, $j = 1, \ldots, p$, such that $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p)$, and

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{1.2}$$

where $\boldsymbol{\beta}$ is a vector parameter of size p. The information matrix of $\boldsymbol{\beta}$ in (1.2) is $\boldsymbol{\mathcal{I}}(\boldsymbol{\beta}) = \boldsymbol{X}^\top \boldsymbol{R}(\boldsymbol{\theta})^{-1}\boldsymbol{X}$. To obtain a formula for the effective sample size of $\boldsymbol{Y}$, several issues should be addressed in advance, for example, how to reduce the Fisher information matrix about $\boldsymbol{\beta}$ to a single number, lying on the interval $[1, n]$, which is a desirable property among others established by Vallejos and Osorio (2014). One way to generalize (1.1) is to consider a weighted version of the ESS of the form

$$n^* = \sum_{i=1}^{p} v_i \boldsymbol{x}_i^\top \boldsymbol{R}(\boldsymbol{\theta})^{-1}\boldsymbol{x}_i, \tag{1.3}$$

where v_i are suitable weights to be determined. Combining Equations (1.1) and (1.3), we can build a definition of the ESS of $\boldsymbol{Y}$ through

$$n^* = \frac{n \cdot \operatorname{tr}(\boldsymbol{X}^\top \boldsymbol{R}(\boldsymbol{\theta})^{-1}\boldsymbol{X})}{\operatorname{tr}(\boldsymbol{X}^\top \boldsymbol{X})}, \tag{1.4}$$

where $\operatorname{tr}(\cdot)$ denotes the trace operator. This definition is consistent with that given in Vallejos and Osorio (2014) in the sense that for $\boldsymbol{X} = \mathbf{1}$, we recover (1.1), but for $\boldsymbol{X} = (\mathbf{1}, \boldsymbol{x}_2)$ and $\widetilde{\boldsymbol{X}} = (\mathbf{1}, \boldsymbol{x}_3)$, with $\boldsymbol{x}_3 = k\boldsymbol{x}_2$ and $k \in \mathbb{R} \setminus \{0\}$, the effective sample size (1.4) for $\boldsymbol{X}$ and $\widetilde{\boldsymbol{X}}$ respectively are

$$\begin{aligned}
n_1^* &= n \left(\frac{\mathbf{1}^\top \boldsymbol{R}^{-1}\mathbf{1} + \boldsymbol{x}_2^\top \boldsymbol{R}^{-1}\boldsymbol{x}_2}{n + \boldsymbol{x}_2^\top \boldsymbol{x}_2} \right), \\
n_2^* &= n \left(\frac{\mathbf{1}^\top \boldsymbol{R}^{-1}\mathbf{1} + k^2 \boldsymbol{x}_2^\top \boldsymbol{R}^{-1}\boldsymbol{x}_2}{n + k^2 \boldsymbol{x}_2^\top \boldsymbol{x}_2} \right),
\end{aligned}$$

and it is affected by the scales of the variables because $n_1^* = n_2^*$ if and only if $k^2 = 1$. In addition, the simple intraclass and Toeplitz correlation structures respectively described by $\boldsymbol{R}(\rho) = (1-\rho)\mathbf{I} + \rho\boldsymbol{J}$ with $\boldsymbol{J} = \mathbf{1}\mathbf{1}^\top$, $-1/(n-1) < \rho < 1$, and $\boldsymbol{R}(\rho)_{ij} = \rho^{|i-j|}$, if $i \neq j$, or $\boldsymbol{R}(\rho)_{ij} = 1$, otherwise, for $|\rho| < 1$, (correlation function of an AR(1) process) show a nondecreasing pattern as a function of ρ, when $\boldsymbol{x}_2$ is a random sequence in the $[0, 1]$, as shown in Figure 1. The behavior of the effective sample size is similar for other choices of $\boldsymbol{x}_2$. This is a considerable difference from the constant mean case, in which the patterns of both correlation structures decrease for increasing values of $\rho \in [0, 1]$.

On the basis of model (1.2), we first provide a new definition of effective sample size as a function of the correlation structure of a process that does not depend on the scale of the variables. For most correlations used in spatial statistics, the ESS decreases when the correlation increases. We then present several examples and preliminary results that support the given definition. In particular, the problem of computing the ESS for variables that are normal after applying a logarithm transformation is addressed via Taylor approximations of the elements of the correlation structure. We next turn to studying the limiting distribution of the maximum likelihood (ML) estimator of $\boldsymbol{\theta}$ under an increasing domain framework. The conditions given by Mardia and Marshall (1984) to obtain the asymptotic distribution of the ML estimator of $\boldsymbol{\theta}$ are established, and the limiting distribution of the ESS is obtained using the delta method. In particular, these conditions are particularized for a Gaussian random field with the Matérn covariance, as well as for CAR and directional CAR processes. The asymptotic variance in each case is derived. An approximate hypothesis testing for the null hypothesis $H_0 : n^* = n$ is developed. Although under a suitable definition of ESS, if $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \sigma^2\boldsymbol{I}$, $n^* = n$, this type of test does not exactly correspond to the test of the form $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \sigma^2\boldsymbol{I}$, and cannot be compared with the existing proposals, mainly due to the lack of replicates in the spatial case. However, for the alternative hypothesis of the form $H_1 : n^* < n$, the procedure admits testing for information reduction due to the presence of autocorrelation. Simulation studies and a real data example support the theoretical findings and provide a closer insight into the estimation of the effective sample size. The broader impacts resulting from this research lie in its contributions to understanding the reduction of information due to the effect of autocorrelation in a spatial statistics context.

The remainder of this paper is organized as follows. In Section 2, we present the definition of effective sample size for spatial regression models accompanied by several illustrative examples for models commonly used in spatial statistics. More examples to illustrate how the ESS looks like for SAR, multivariate CAR, and partitioned normal models are provided in Appendix A. In Section 3, we present three propositions that characterize the ESS. In Section 4, we develop the asymptotics for the ML estimation of the ESS for an increasing domain. In Section 5, the hypothesis testing for $H_1 : n^* < n$ is presented with illustrative examples. The performance of the asymptotic methods is ex-
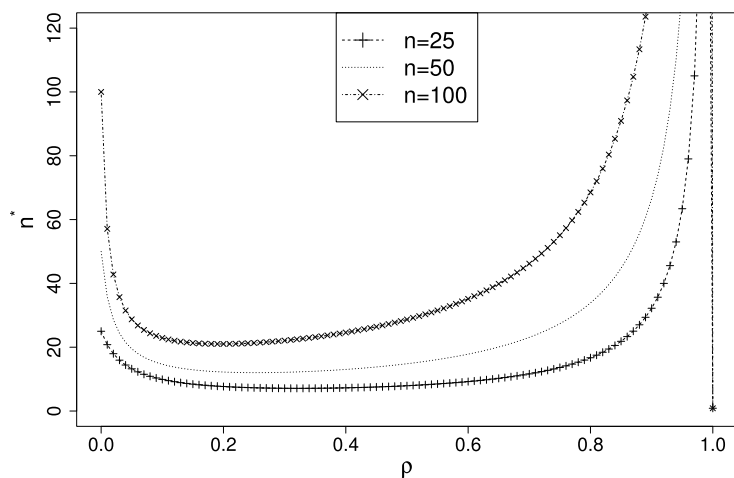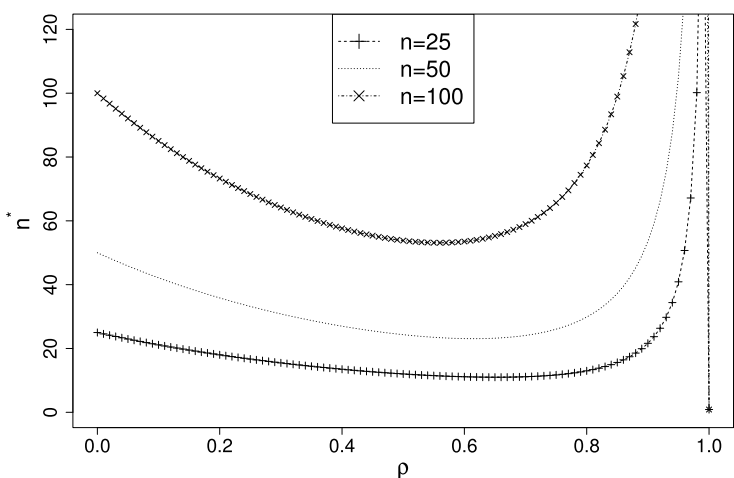
FIG 1. *(a) ESS for the intraclass correlation; (b) ESS for the Toeplitz correlation.*

plored via two Monte Carlo simulation studies in Section 6. The estimation of the effective sample size for a macroalgae dataset collected off the central coast of Chile is discussed in Section 7. We conclude the paper with a discussion in Section 8. We relegate all technical proofs to Appendix B. The programs that were used to analyze the data and the dataset can be obtained from http://spatialpack.mat.utfsm.cl.

2. The effective sample size

To avoid the scale problem mentioned in Section 1, we consider a spatial regression model such as (1.2) with $\boldsymbol{X} = (\boldsymbol{1}, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p)$ such that $\boldsymbol{x}_j^\top \boldsymbol{x}_j = n$ for all $j = 2, \ldots, p$, where $\boldsymbol{x}_j = (x_j(\boldsymbol{s}_1), x_j(\boldsymbol{s}_2), \ldots, x_j(\boldsymbol{s}_n))^\top$. Then, the effective sample size of $\boldsymbol{Y}$ is defined as

$$n^* = \frac{\mathrm{tr}(\boldsymbol{X}^\top \boldsymbol{R}(\boldsymbol{\theta})^{-1} \boldsymbol{X})}{p} = \frac{1}{p} \sum_{j=1}^{p} \boldsymbol{x}_j^\top \boldsymbol{R}(\boldsymbol{\theta})^{-1} \boldsymbol{x}_j. \tag{2.1}$$

This definition is consistent with that given in Vallejos and Osorio (2014) because if $\boldsymbol{X} = \boldsymbol{1}$ in (2.1), $n^* = \boldsymbol{1}^\top \boldsymbol{R}(\boldsymbol{\theta})^{-1} \boldsymbol{1}$. If $\boldsymbol{R}(\boldsymbol{\theta}) = \boldsymbol{I}$, then $n^* = n$.

Now, we explore how the ESS appears for different correlation structures.

Example 1. Assume that $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{R}(\rho))$, where $\boldsymbol{R}(\rho) = (1 - \rho)\boldsymbol{I} + \rho\boldsymbol{J}$ with $\boldsymbol{J} = \boldsymbol{1}\boldsymbol{1}^\top$ and $-1/(n-1) < \rho < 1$. Suppose that the columns of $\boldsymbol{X}$ have Euclidean norm equal to $\sqrt{n}$. Then,

$$n_{\mathrm{IC}}^* = \frac{1}{p} \left(\frac{n}{1 + (n-1)\rho} + \frac{1}{1-\rho} \sum_{j=2}^{p} \left[n - \frac{n^2 \overline{\boldsymbol{x}}_j^2 \rho}{1 + (n-1)\rho} \right] \right). \tag{2.2}$$

In particular, if $\boldsymbol{X} = (\boldsymbol{1}, \boldsymbol{x}_2)$ with $\boldsymbol{x}_2 = (x_1, \ldots, x_n)^\top$ and $x_i = i$, then $\|\boldsymbol{x}_2\|^2 = n(n+1)(2n+1)/6$, $\overline{\boldsymbol{x}}_2^2 = n^2(n+1)^2/4$, and (2.2) becomes

$$n_{\mathrm{IC}}^* = \frac{n}{2} \left(\frac{1}{1 + (n-1)\rho} + \frac{(n^2 - 5n - 2)\rho + 2(2n+1)}{2(1-\rho)(1 + (n-1)\rho)(2n+1)} \right).$$

Let $a = \sum_{j=2}^{p}(1 - \overline{\boldsymbol{x}}_j^2)$. Because $\boldsymbol{x}_j^\top \boldsymbol{x}_j/n - \overline{\boldsymbol{x}}_j^2 \geq 0$ and $\boldsymbol{x}_j^\top \boldsymbol{x}_j = n$ for all $j = 2, \ldots, p$, we have that $0 \leq 1 - \overline{\boldsymbol{x}}_j^2 \leq 1$, therefore $0 \leq a \leq p - 1 < p$. Expanding (2.2) in term of a we obtain,

$$n_{\mathrm{IC}}^* = \frac{1}{p} \left(\frac{n(p-a)}{1 + (n-1)\rho} + \frac{na}{1-\rho} \right). \tag{2.3}$$

Clearly, n_{IC}^* in (2.3) is a convex function, if and only if, $-1/(n-1) < \rho < 1$. Moreover, n_{IC}^* is the sum of two convex functions defined in the interval $(-1/(n-1), 1)$, one strictly increasing function tending to infinity when $\rho \to 1^-$, and another strictly decreasing function tending to infinity when $\rho \to -1/(n-1)^+$. Thus, there exist a unique critical point that can be obtained by solving the equation

$$\frac{dn_{\mathrm{IC}}^*}{d\rho} = 0$$

in the interval $(-1/(n-1), 1)$, which is a global minimum of n_{IC}^* given by

$$\rho_0 = 1 - \frac{n}{n - 1 + \sqrt{\dfrac{(p-a)(n-1)}{a}}}.$$

Example 2. Suppose that $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{R}(\rho))$, where $\boldsymbol{R}(\rho)$ is the AR(1) correlation structure, and that the columns of $\boldsymbol{X}$ have an Euclidean norm equal to $\sqrt{n}$. Then,

$$n^*_{\mathrm{AR}} = \frac{1}{p}\left(\frac{n(1-\rho)+2\rho}{(1+\rho)} + \frac{1}{1-\rho^2}\sum_{j=2}^{p}\left[n + \rho^2 \sum_{i=2}^{n-1} x_{i,j}^2 - 2\rho \sum_{i=1}^{n-1} x_{i,j}x_{i+1,j}\right]\right).$$
(2.4)

Because $\boldsymbol{R}^{-1}(\rho) = \left(\boldsymbol{I} - \rho\boldsymbol{B} + \rho^2\boldsymbol{A}\right)/(1-\rho^2)$, where $\boldsymbol{A} = \mathrm{diag}\{0, 1, \ldots, 1, 0\}$ and $\boldsymbol{B} = (b_{ij})$ with $b_{ij} = 1$ if $|i-j| = 1$ and 0 otherwise, a matrix version of Equation (2.4) is

$$n^*_{\mathrm{AR}} = \frac{1}{p}\left(\frac{n(1-\rho)+2\rho}{(1+\rho)} + \frac{1}{1-\rho^2}\sum_{j=2}^{p}\left[n + \rho^2 \boldsymbol{x}_j^\top \boldsymbol{A}\boldsymbol{x}_j - \rho\boldsymbol{x}_j^\top \boldsymbol{B}\boldsymbol{x}_j\right]\right). \quad (2.5)$$

Let $a = \sum_{j=2}^{p} \boldsymbol{x}_j^\top \boldsymbol{A}\boldsymbol{x}_j$, $b = \sum_{j=2}^{p} \boldsymbol{x}_j^\top \boldsymbol{B}\boldsymbol{x}_j$ and $c = n(p-1)$. Clearly $a > 0$, $c > 0$ and $b \in \mathbb{R}$. Then

$$
\begin{aligned}
n^*_{\mathrm{AR}} &= \frac{1}{p}\left(\frac{(2-n)\rho+n}{(1+\rho)} + \frac{a\rho^2 - b\rho + c}{1-\rho^2}\right), \\
&= \frac{1}{p}\left((2-n) + \frac{2(n-1)}{(1+\rho)} + \frac{a\rho^2 - b\rho + c}{1-\rho^2}\right), \\
&= \frac{1}{2p}\left(2(2-n-a) + \frac{4(n-1)+a+b+c}{1+\rho} + \frac{a-b+c}{1-\rho}\right). \quad (2.6)
\end{aligned}
$$

In order to prove that n^*_{AR} in (2.6) is convex, note that by using the triangular and Young's inequality we have that $a+b \geq a-|b|$, $a-b \geq a-|b|$ and $a-|b| \geq -c$. Indeed, the latter inequality is coming because

$$
\begin{aligned}
a - |b| &= \sum_{j=2}^{p}\left[\sum_{i=2}^{n-1} x_{i,j}^2 - 2\left|\sum_{i=1}^{n-1} x_{i,j}x_{i+1,j}\right|\right], \\
&\geq \sum_{j=2}^{p}\left[\sum_{i=2}^{n-1} x_{i,j}^2 - 2\sum_{i=1}^{n-1} |x_{i,j}x_{i+1,j}|\right], \\
&\geq \sum_{j=2}^{p}\left[\sum_{i=2}^{n-1} x_{i,j}^2 - \sum_{i=1}^{n-1}\left(x_{i,j}^2 + x_{i+1,j}^2\right)\right], \\
&= -\sum_{j=2}^{p}\sum_{i=1}^{n} x_{i,j}^2, \\
&= -n(p-1), \\
&= -c.
\end{aligned}
$$

Then $a+b+c \geq 0$ and $a-b+c \geq 0$. Therefore, the function $(4(n-1)+a+b+c)/(1+\rho)$ is convex if and only if $\rho > -1$, and the function $(a-b+c)/(1-\rho)$ is convex if and only if $\rho < 1$. In consequence, n^*_{AR} is a convex function if only

if $-1 < \rho < 1$. Moreover, n_{AR}^* corresponds to the sum of two convex functions defined in the interval $(-1, 1)$, similarity to case developed in Example 1, there exist only one critical point of n_{AR}^* in the interval $(-1, 1)$ corresponding to a global minimum, that can be obtained by solving the equation

$$\frac{dn_{\mathrm{AR}}^*}{d\rho} = 0. \tag{2.7}$$

A simple calculation shows that

$$\frac{dn_{\mathrm{AR}}^*}{d\rho} = \frac{1}{2p}\left(-\frac{4(n-1) + a + b + c}{(1+\rho)^2} + \frac{a - b + c}{(1-\rho)^2}\right).$$

If $b \neq -2(n-1)$, then the only solution of (2.7) in the interval $(-1, 1)$ is

$$\rho_0 = \frac{2(n-1) + (a+c) - \sqrt{(a+c)^2 + 4(n-1)(a+c) - 4b(n-1) - b^2}}{2(n-1) + b}.$$

If $b = -2(n-1)$, the only solution of (2.7) in the interval $(-1, 1)$ is $\rho_0 = 0$, and thus $n_{\mathrm{AR}}^*(0) = n$. This means that the lowest possible value for n_{AR}^* is n. However, this case is valid for very specific covariates. In fact, consider $\boldsymbol{X} = (\boldsymbol{1}, \boldsymbol{x}_2)$, and $\boldsymbol{x}_2 = (x_1, \ldots, x_n)^\top$. If $x_i \cdot x_{i+1} = -1$, then $b = -2(n-1)$. Note that $x_i \cdot x_{i+1} = -1$ implies that $x_{i+1} = x_{i-1}$, which is a difference equation, whose general solution is $x_i = C_1 + C_2(-1)^i$, where $C_1 = (x_1^2 - 1)/2x_1$, and $C_2 = -(x_1^2 + 1)/2x_1$, because $x_2 = -1/x_1$. Now, $\sum_{i=1}^n x_i^2 = n$ implies that $|x_1| = 1$, for each $n \in \mathbb{N}$ (when n is odd there is another solution, namely $|x_1| = \sqrt{(n-1)/(n+1)}$). Therefore, we conclude that $x_i = (-1)^i$ or $x_i = (-1)^{i+1}$. A model for which the effective sample size is greater than n is described in Equation (3.2).

A convex pattern of functions n_{IC}^* and n_{AR}^* can be observed in Figure 1, where $\boldsymbol{Y} \sim \mathcal{N}((\boldsymbol{1}, \boldsymbol{x})\boldsymbol{\beta}, \boldsymbol{R}(\rho))$, $\boldsymbol{R}(\rho)$ is the intraclass (Figure 1(a)) or the Toeplitz AR(1) correlation structure (Figure 1(b)), $n = 25, 50, 100$, $\rho \in [0, 1]$, and the covariate was simulated as a uniform random variable in the interval $[0, 1]$.

The ESS for some covariance structures is explored next.

Example 3. Consider a CAR model of the form $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = \tau^2(\boldsymbol{D}_w - \rho\boldsymbol{W})^{-1}$, where the columns of $\boldsymbol{X}$ have norm $\sqrt{n}$, $\boldsymbol{W} = (w_{ij})$ is a contiguity matrix, and $\boldsymbol{D}_w = \mathrm{diag}(w_{i+})$ with $w_{i+} = \sum_j w_{ij}$. Let $\boldsymbol{R}_{\mathrm{CAR}}^{-1} = \boldsymbol{C}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}$, where $\boldsymbol{C} = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n)$ and $\sigma_i = \Sigma_{ii}^{1/2}$. Then,

$$n_{\mathrm{CAR}}^* = \frac{1}{\tau^2 p}\left(\sum_{k=1}^p \sum_{i=1}^n x_{ik}^2 \sigma_i^2 w_{i+} - \rho \sum_{k=1}^p \sum_{i=1}^n \sum_{j=1}^n x_{ik} x_{jk} \sigma_i \sigma_j w_{ij}\right). \tag{2.8}$$

Example 4. Let $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in D \subset \mathbb{R}^d\}$ be a random field observed on D_n. Suppose that $\boldsymbol{Y} = (Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n))^\top \sim \mathcal{N}(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$, where $\boldsymbol{\Sigma}(\boldsymbol{\theta})_{ij} = C(\|\boldsymbol{s}_i - \boldsymbol{s}_j\|, \boldsymbol{\theta})$, with $\boldsymbol{\theta} = (\sigma^2, \tau^2, \phi)$ such that

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

TABLE 1
Correlation functions used in Example 1.

Model	Correlation function
Spherical	$\rho(h, \phi) = \begin{cases} 1 - 1.5\dfrac{h}{\phi} + 0.5\left(\dfrac{h}{\phi}\right)^3 & ; \quad h < \phi \\ 0 & ; \quad \text{otherwise} \end{cases}$
Exponential	$\rho(h, \phi) = \exp\left(-\dfrac{h}{\phi}\right)$
Matérn, $\kappa = 1.5$	$\rho(h, \phi) = \left(1 + \dfrac{h}{\phi}\right)\exp\left(-\dfrac{h}{\phi}\right)$
Matérn, $\kappa = 2.5$	$\rho(h, \phi) = \left[1 + \dfrac{h}{\phi} + \dfrac{1}{3}\left(\dfrac{h}{\phi}\right)^2\right]\exp\left(-\dfrac{h}{\phi}\right)$
Gaussian	$\rho(h, \phi) = \exp\left[-\left(\dfrac{h}{\phi}\right)^2\right]$

and $\rho(h, \phi)$ is a correlation function (for instance, $\rho(h, \phi)$ can be any of the correlation functions listed in Table 1). Then, the elements of $\boldsymbol{R}(\boldsymbol{\theta})$ are

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

For illustrative purposes, we plotted the ESS versus ϕ for $n = 100$, where the sites $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_{100}$ were generated from uniform random variables in the region $[0, 1] \times [0, 1]$, while $\sigma^2 = 1$ and $\tau^2 = 0.1$. Figure 2 shows that the ESS decays as a function of $\phi \in (0, 1)$, as expected.
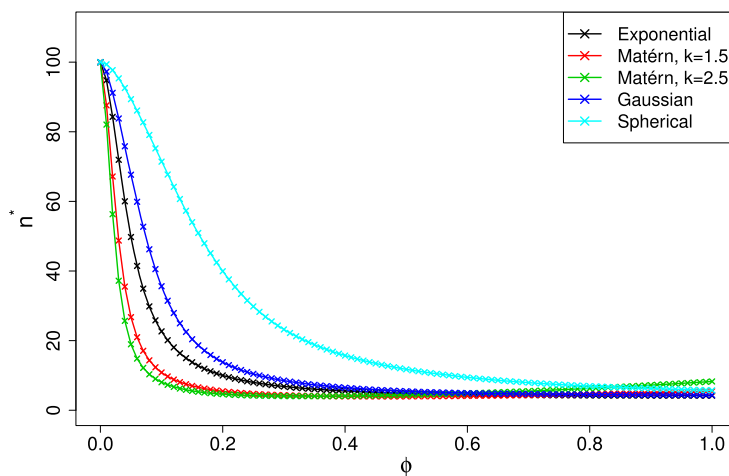

FIG 2. *ESS for the correlation functions shown in Table 1.*

The ESS for other related processes is provided in Appendix A.

We want to emphasize that the examples described above are strongly dependent on the knowledge we have about the covariance parameters. This has

been addressed in Vallejos and Osorio (2014). As an illustration, consider the intraclass correlation model of the form

$$Y(\boldsymbol{s}) = \mu + Z_0 + Z(\boldsymbol{s}), \tag{2.9}$$

where Z_0 is a random variable with variance ρ, $Z(\boldsymbol{s})$ is a sequence of non-correlated (pure nugget) variables with variance equal to $(1 - \rho)$, and Z_0 and $Z(\boldsymbol{s})$ are independent. To get any information about μ it is necessary to have some knowledge about the variance of Z_0. Moreover, the estimation of μ is dispairing when model (2.9) has an additional variance parameter σ^2.

3. Preliminary results

From definition (2.1), the following propositions for the ESS can be obtained.

Proposition 1. *Let $\boldsymbol{R}(\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \mathbb{R}^q$ be a correlation matrix, and $\boldsymbol{X}$ be a matrix of order $n \times p$ as in (2.1), $n \geq 1$. Then, $n^* > 1$.*

There is no upper bound for the ESS in the general case, although computational experiments provide empirical evidence in favor of the upper bound n. The next result provides the maximum number of covariates allowed in a spatial regression process for the intraclass correlation structure, and it also provides an upper bound for the ESS.

Proposition 2. *Suppose that $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{X\beta}, \boldsymbol{R}(\rho))$, where $\boldsymbol{R}(\rho)$ is the intraclass correlation matrix and the covariates are the eigenvectors of $\boldsymbol{R}(\rho)$. Then,*

$$n^*_{\mathrm{IC}} \leq n \iff p \leq \frac{n}{1 + (n-1)\rho}.$$

As an example, assume that $n = 10$ and $\rho = 0.6$. Then, $n/(1+(n-1)\rho) \approx 1.56$. Now, $\rho = 1 \implies p = 1$; thus, under perfect correlation, we can include only one covariate in the model. Moreover, if $\rho = 0.25$, $p \leq 4n/(n+3) \to 4$ as $n \to \infty$.

An immediate consequence of Proposition 2 is the equivalence

$$n^*_{\mathrm{IC}} > n \iff \rho > \frac{1}{p} - \frac{p-1}{p(n-1)}, \tag{3.1}$$

which provides a range for the correlation for a fixed number of covariates to include in a regression model. For instance, if $p = 2$, $\rho \leq 1/2 - 1/(2(n-1)) \to 1/2$ when $n \to \infty$. This helps to explain the behavior shown in Figure 1(a).

It should be stressed that there are examples where the effective sample size is greater than n. Let us consider the model

$$\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{X\beta}, \boldsymbol{R}(\rho)), \tag{3.2}$$

where $\boldsymbol{R}(\rho)$ is the AR(1) correlation structure, $\boldsymbol{X} = (\boldsymbol{1}, \boldsymbol{x}_2)$ with covariate $\boldsymbol{x}_2 = (x_1, \ldots, x_n)^\top$, and $x_i = (-1)^i$. Then the constants used in Example 2 are

$a = n-2$, $b = -2(n-1)$, and $c = n$. Therefore, $a+b+c = 0$, $a-b+c = 4(n-1)$, and by Equation (2.6)

$$
\begin{aligned}
n_{\mathrm{AR}}^* &= \frac{1}{4}\left(4(2-n) + \frac{4(n-1)}{1+\rho} + \frac{4(n-1)}{1-\rho}\right), \\
&= \frac{(n-2)\rho^2 + n}{1-\rho^2}.
\end{aligned}
\tag{3.3}
$$

Because in this case $b = -2(n-1)$, from Example 2, we have that $n_{\mathrm{AR}}^* \geq n$, for all $\rho \in (-1,1)$; and $n_{\mathrm{AR}}^* = n$, if and only if, $\rho = 0$.

For models with constant mean there are specific correlation structures (with negative autocorrelation) for which the effective sample size is greater than n (Richardson, 1990). For models with non-constant mean, the example describe in (3.2) alert us that the properties of the effective sample size also depend on the structure of the covariates.

In practice, the normality assumption could be not satisfied. One way to address this problem is a transformation approach. Box and Cox (1964) proposed the transformation

$$
Z_{\boldsymbol{\delta}}(\boldsymbol{s}_i) = \begin{cases} \dfrac{(Y(\boldsymbol{s}_i) + \delta_2)^{\delta_1} - 1}{\delta_1} & ; \quad \delta_1 \neq 0 \\ \ln(Y(\boldsymbol{s}_i) + \delta_2) & ; \quad \delta_1 = 0 \end{cases}
$$

where $Y(\cdot)$ is the original variable and $\boldsymbol{\delta} = (\delta_1, \delta_2)^\top$ is an unknown parameter vector to be estimated to achieve normality of $Z(\cdot)$. Given the vector of spatial observations $(Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_n))^\top$, $\boldsymbol{\delta}$ can be estimated by maximizing the likelihood function

$$
L(\boldsymbol{\delta}) = -\frac{n}{2}\ln\left(\frac{1}{n}\sum_{i=1}^n (Z_{\boldsymbol{\delta}}(\boldsymbol{s}_i) - \bar{Z}(\boldsymbol{s}_i)^2\right) + (\delta_1 - 1)\sum_{i=1}^n \ln(Y(\boldsymbol{s}_i) + \delta_2).
$$

Alternatively, an estimation for $\boldsymbol{\delta}$ can be obtained by finding the optimal value that maximizes the correlation between $\Phi^{-1}((i - 0.5)/n)$ and $Z_{(i)}$, where Φ^{-1} is the inverse of the cumulative distribution function of $Z(\boldsymbol{s}_i)$ and $Z_{(i)}$ is the order statistic associated with $Z(\boldsymbol{s}_i)$, for $i = 1, \ldots, n$ (Kutner et al., 2004).

Proposition 3. *Suppose that $\boldsymbol{Z} = (Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_n))^\top \sim \mathcal{N}(\boldsymbol{\mu_Z}, \boldsymbol{\Sigma_Z})$ and*

$$
Z(\boldsymbol{s}_i) = \ln(Y(\boldsymbol{s}_i) + \delta_2), \quad i = 1, \ldots, n,
$$

where $\boldsymbol{Y} = (Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n))^\top$ is the original spatial sample. Let $\boldsymbol{R_Z}$ and $\boldsymbol{R_Y}$ be the correlation matrices of $\boldsymbol{Z}$ and $\boldsymbol{Y}$, respectively. If $(\boldsymbol{\mu_Z})_i = \mu_i$, $(\boldsymbol{\Sigma_Z})_{ij} = \sigma_{ij}$ and $\sigma_{ii} = \tilde{\sigma}^2$, $i, j = 1, \ldots, n$, then $(\boldsymbol{R_Y})_{ii} = (\boldsymbol{R_Z})_{ii} = 1$ and

$$
(\boldsymbol{R_Y})_{ij} = c \cdot (\boldsymbol{R_Z})_{ij} \cdot k_{ij} + o\left(|\tilde{\sigma}^2|^m\right), \quad \textit{for all } i \neq j,
$$

where $c = \dfrac{\tilde{\sigma}^2}{\exp(\tilde{\sigma}^2) - 1}$, $k_{ij} = 1 + \dfrac{\sigma_{ij}}{2} + \cdots + \dfrac{\sigma_{ij}^{m-1}}{m!}$, for all $i \neq j$, and $k_{ii} = 1$. Furthermore, as $m \to \infty$, $(\boldsymbol{R_Y})_{ij} = \dfrac{c}{\tilde{\sigma}^2}(e^{\sigma_{ij}} - 1)$.

Observe that $\boldsymbol{R}_Y = \boldsymbol{K}_m \odot \boldsymbol{R}_Z$, where $\odot$ denotes the Hadamard product and $\boldsymbol{K}_m$ is the matrix with elements k_{ij}. In particular, $\boldsymbol{K}_1 = \boldsymbol{1}\boldsymbol{1}^\top$. Thus, for $m = 1$,

$$n^*(\boldsymbol{Y}) = \frac{n}{p}\,\mathrm{tr}\left(\boldsymbol{X}^\top \boldsymbol{R}_Y^{-1}\boldsymbol{X}\right) = \frac{n}{p}\,\mathrm{tr}\left(\boldsymbol{X}^\top \boldsymbol{R}_Z^{-1}\boldsymbol{X}\right) = n^*(\boldsymbol{Z}).$$

4. Estimation and asymptotics under increasing domain

Let $\left\{Y(\boldsymbol{s}) : \boldsymbol{s} \in D \subset \mathbb{R}^d\right\}$ be a Gaussian random field such that $Y(\cdot)$ is observed on $D_n \subset D$. It is assumed that D_n is a non-random set satisfying $\|\boldsymbol{s} - \boldsymbol{t}\| \geq \gamma > 0$ for all $\boldsymbol{s}, \boldsymbol{t} \in D_n$. This ensures that the sampling set is increasing as n increases. Denote $\boldsymbol{Y} = (Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n))^\top$ and assume that $\mathbb{E}[\boldsymbol{Y}] = \boldsymbol{X}\boldsymbol{\beta}$, $\mathrm{cov}(Y(\boldsymbol{t}), Y(\boldsymbol{s})) = \sigma(\boldsymbol{t}, \boldsymbol{s}; \boldsymbol{\theta})$, $\boldsymbol{X}$ is $n \times p$ with $\mathrm{rank}(\boldsymbol{X}) = p$, $\boldsymbol{\beta} \in \mathbb{R}^p$, $\boldsymbol{\theta} \in \Theta$, where Θ is an open set of $\in \mathbb{R}^q$. Let $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ be the covariance matrix of $\boldsymbol{Y}$ such that the ij-th element of $\boldsymbol{\Sigma}$ is $\sigma_{ij} = \sigma(\boldsymbol{s}_i, \boldsymbol{s}_j; \boldsymbol{\theta})$. The estimation of $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ can be made by ML estimation, maximizing

$$L = L(\boldsymbol{\beta}, \boldsymbol{\theta}) = \mathrm{Conts} - \frac{1}{2}\ln|\boldsymbol{\Sigma}| - \frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}). \qquad (4.1)$$

Let $\boldsymbol{L}_n^{(1)} = \nabla L = (\boldsymbol{L}_{\boldsymbol{\beta}}^\top, \boldsymbol{L}_{\boldsymbol{\theta}}^\top)^\top$ and

$$\boldsymbol{L}_n^{(2)} = \begin{pmatrix} \boldsymbol{L}_{\boldsymbol{\beta}\boldsymbol{\beta}} & \boldsymbol{L}_{\boldsymbol{\beta}\boldsymbol{\theta}} \\ \boldsymbol{L}_{\boldsymbol{\theta}\boldsymbol{\beta}} & \boldsymbol{L}_{\boldsymbol{\theta}\boldsymbol{\theta}} \end{pmatrix}$$

be the gradient vector and Hessian matrix, respectively, obtained from (4.1). Let $\boldsymbol{B}_n = -\mathbb{E}[\boldsymbol{L}_n^{(2)}]$ be the Fisher information matrix with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$. Then, $\boldsymbol{B}_n = \mathrm{diag}(\boldsymbol{B}_{\boldsymbol{\beta}}, \boldsymbol{B}_{\boldsymbol{\theta}})$, where $\boldsymbol{B}_{\boldsymbol{\beta}} = -\mathbb{E}[\boldsymbol{L}_{\boldsymbol{\beta}\boldsymbol{\beta}}]$ and $\boldsymbol{B}_{\boldsymbol{\theta}} = -\mathbb{E}[\boldsymbol{L}_{\boldsymbol{\theta}\boldsymbol{\theta}}]$.

For a twice differentiable covariance function $\sigma(\cdot, \cdot; \boldsymbol{\theta})$ on Θ with continuous second derivatives, Mardia and Marshall (1984) provided sufficient conditions on $\boldsymbol{\Sigma}$ and $\boldsymbol{X}$ such that the limiting distribution of $(\boldsymbol{\beta}^\top, \boldsymbol{\theta}^\top)^\top$ is normal as is stated in the following result.

Theorem 1. *Let $\lambda_1 \leq \cdots \leq \lambda_n$ be the eigenvalues of $\boldsymbol{\Sigma}$, and let those of $\boldsymbol{\Sigma}_i = \dfrac{\partial \boldsymbol{\Sigma}}{\partial \theta_i}$ and $\boldsymbol{\Sigma}_{ij} = \dfrac{\partial^2 \boldsymbol{\Sigma}}{\partial \theta_i \partial \theta_j}$ be λ_k^i and λ_k^{ij}, $k = 1, \ldots, n$, such that $|\lambda_1^i| \leq \cdots \leq |\lambda_n^i|$ and $|\lambda_1^{ij}| \leq \cdots \leq |\lambda_n^{ij}|$ for $i, j = 1, \cdots, q$. Suppose that as $n \to \infty$*

> *(i) $\lim \lambda_n = C < \infty$, $\lim|\lambda_n^i| = C_i < \infty$ and $\lim|\lambda_n^{ij}| = C_{ij} < \infty$ for all $i, j = 1, \ldots, q$.*
> *(ii) $\|\boldsymbol{\Sigma}_i\|^{-2} = \mathcal{O}(n^{-\frac{1}{2}-\delta})$ for some $\delta > 0$, for $i = 1, \ldots, q$.*
> *(iii) For all $i, j = 1, \ldots, q$, $a_{ij} = \lim\left[t_{ij}/(t_{ii}t_{jj})^{\frac{1}{2}}\right]$ exists,*
>   *where $t_{ij} = \mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j\right)$ and $\boldsymbol{A} = (a_{ij})$ is nonsingular.*
> *(iv) $\lim(\boldsymbol{X}^\top\boldsymbol{X})^{-1} = 0$.*

Then, $(\widehat{\boldsymbol{\beta}}^\top, \widehat{\boldsymbol{\theta}}^\top)^\top \xrightarrow{\mathcal{L}} \mathcal{N}\left((\boldsymbol{\beta}^\top, \boldsymbol{\theta}^\top)^\top, \boldsymbol{B}_n^{-1}\right)$ as $n \to \infty$, in an increasing domain sense.

For the application of Theorem 1, the matrix $\boldsymbol{X}$ can be chosen in such a way that condition (iv) is satisfied. When the process $Y(\cdot)$ has a stationary covariance function on $\mathbb{R}^d$ with $\sigma(\boldsymbol{s}_i, \boldsymbol{s}_j; \boldsymbol{\theta}) = \sigma(\boldsymbol{s}_i - \boldsymbol{s}_j; \boldsymbol{\theta})$ and D_n represents a regular but not necessarily rectangular grid with a fixed distance between any pair of locations, conditions (i) and (ii) are satisfied if $\sigma_{\boldsymbol{k}}$, $\sigma_{\boldsymbol{k},i}$ and $\sigma_{\boldsymbol{k},ij}$ are absolutely summable over $\mathbb{Z}^d$ for all $i, j = 1, \ldots, p$, where $\sigma_{\boldsymbol{k}}$, $\boldsymbol{k} \in \mathbb{Z}^d$, is the covariance for lag $\boldsymbol{k} = (k_1, \ldots, k_d)$ of the lattice, $\sigma_{\boldsymbol{k},i} = \partial \sigma_{\boldsymbol{k}} / \partial \theta_i$, and $\sigma_{\boldsymbol{k},ij} = \partial^2 \sigma_{\boldsymbol{k}} / \partial \theta_i \partial \theta_j$ (Mardia and Marshall, 1984). These conditions are established next for a particular parametrization of the Matérn covariance function.

Theorem 2. *Let $\left\{ Y(\boldsymbol{s}) : \boldsymbol{s} \in D \subset \mathbb{R}^d \right\}$ be a covariance stationary Gaussian process sampled on a regular lattice of fixed spacing, with the Matérn covariance structure $\sigma_m(h; \sigma^2, \phi) = \sigma^2 \exp\left(-\dfrac{h}{\phi} \right) P_m\left(\dfrac{h}{\phi} \right)$, where $\nu = m + \frac{1}{2}$, $m \in \mathbb{N}_0$, $h = \|\boldsymbol{s} - \boldsymbol{t}\|$ with $\boldsymbol{s}, \boldsymbol{t} \in D_n$ and*

$$P_m(x) = \sum_{i=0}^{m} a_i x^i, \quad \text{with} \ \ a_i = \frac{2^i}{i!} \binom{m}{i} \Big/ \binom{2m}{i}.$$

Let $\widehat{\boldsymbol{\theta}}$ be the ML estimator of $\boldsymbol{\theta} = (\sigma^2, \phi)^\top$. Then, $\widehat{\boldsymbol{\theta}} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\boldsymbol{\theta}, \boldsymbol{B}_{\boldsymbol{\theta}}^{-1} \right)$, as $n \to \infty$, where

$$\boldsymbol{B}_{\boldsymbol{\theta}} = \frac{1}{2} \begin{pmatrix} \dfrac{n}{\sigma^4} & \dfrac{\operatorname{tr}(\boldsymbol{V})}{\sigma^2 \phi^2} \\ \dfrac{\operatorname{tr}(\boldsymbol{V})}{\sigma^2 \phi^2} & \dfrac{\operatorname{tr}(\boldsymbol{V}^2)}{\phi^4} \end{pmatrix},$$

$\boldsymbol{V} = (\boldsymbol{Q} \odot \boldsymbol{P})^{-1/2} \left(\boldsymbol{Q} \odot \boldsymbol{H} \odot \left[\boldsymbol{P} - \dot{\boldsymbol{P}} \right] \right) (\boldsymbol{Q} \odot \boldsymbol{P})^{-1/2}$, $\boldsymbol{H} = (h_{ij})$ *with* $h_{ij} = \|\boldsymbol{s}_i - \boldsymbol{s}_j\|$, $\boldsymbol{Q} = (q_{ij})$, $\boldsymbol{P} = (p_{ij})$, $\dot{\boldsymbol{P}} = (p'_{ij})$, $q_{ij} = \exp(-h_{ij}/\phi)$, $p_{ij} = P_m(h_{ij}/\phi)$, *and* $p'_{ij} = P'_m(h_{ij}/\phi)$, *for all* $i, j = 1, \ldots, n$.

Next, we establish the asymptotic normality of the ML estimate for the effective sample size (2.1).

Theorem 3. *Let $\boldsymbol{X}$ be a normalized design matrix as in Equation (2.1), and let $\widehat{\boldsymbol{\theta}}$ be the ML estimator of $\boldsymbol{\theta}$ as in Theorem 1. If $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} \xrightarrow{\mathcal{L}} \mathcal{N}_q(\boldsymbol{0}, \boldsymbol{B}_{\boldsymbol{\theta}}^{-1})$, as $n \to \infty$, and $g(\boldsymbol{\theta}) = \operatorname{tr}\left(\boldsymbol{X}^\top \boldsymbol{R}^{-1}(\boldsymbol{\theta}) \boldsymbol{X} \right)/p$, then*

$$g(\widehat{\boldsymbol{\theta}}) - g(\boldsymbol{\theta}) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \nabla^\top g(\boldsymbol{\theta}) \boldsymbol{B}_{\boldsymbol{\theta}}^{-1} \nabla g(\boldsymbol{\theta}) \right), \text{ as } n \to \infty.$$

Let $\left\{ Y(\boldsymbol{s}) : \boldsymbol{s} \in D \subset \mathbb{R}^d \right\}$ be a random field observed on n sites on a space such that $\mathbb{E}[\boldsymbol{Y}] = \boldsymbol{X} \boldsymbol{\beta}$. Assume that there are two valid covariance functions denoted as $\sigma_1(\cdot, \cdot, \boldsymbol{\theta}_1)$, $\boldsymbol{\theta}_1 \in \mathbb{R}^{q_1}$ and $\sigma_2(\cdot, \cdot, \boldsymbol{\theta}_2)$, $\boldsymbol{\theta}_2 \in \mathbb{R}^{q_2}$. Then, we can state the following result.

Theorem 4. *Assume that $\sigma_1(\cdot, \cdot, \boldsymbol{\theta}_1)$ and $\sigma_2(\cdot, \cdot, \boldsymbol{\theta}_2)$ satisfy conditions $(i) - (iii)$ of Theorem 1. Then, the covariance function*

$$\sigma(\boldsymbol{s}_i, \boldsymbol{s}_j, \boldsymbol{\theta}) = \sigma_1(\boldsymbol{s}_i, \boldsymbol{s}_j, \boldsymbol{\theta}_1) + \sigma_2(\boldsymbol{s}_i, \boldsymbol{s}_j, \boldsymbol{\theta}_2), \ i, j = 1, \ldots, n \ , \qquad (4.2)$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$, satisfies conditions (i) and (ii) of Theorem 1.

It is not immediate to establish condition (iii) of Theorem 1. It is sufficient to prove that $\lim_{n \to \infty} |\boldsymbol{A}_n| \neq 0$, where

$$\boldsymbol{A}_n = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{12}^\top & \boldsymbol{A}_{22} \end{pmatrix},$$

$\boldsymbol{A}_{ll}$ is obtained from $\boldsymbol{\theta}_l$, $l = 1, 2$, while $\boldsymbol{A}_{12}$ is obtained from $\mathrm{tr}(\boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_i \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_j)$ for $i = 1, \ldots, q_1$ and $j = q_1 + 1, \ldots, q_1 + q_2$. By hypothesis $\lim_{n \to \infty} |\boldsymbol{A}_{ll}| = a_l \neq 0$, $l = 1, 2$. It follows that

$$\begin{aligned} \lim_{n \to \infty} |\boldsymbol{A}_n| &= a_1 \cdot a_2 \cdot \left| \boldsymbol{I}_{q_2} - \boldsymbol{A}_{22}^{-1/2} \boldsymbol{A}_{12}^\top \boldsymbol{A}_{11}^{-1} \boldsymbol{A}_{12} \boldsymbol{A}_{22}^{-1/2} \right| \\ &= a_1 \cdot a_2 \cdot \left| \boldsymbol{I}_{q_1} - \boldsymbol{A}_{11}^{-1/2} \boldsymbol{A}_{12} \boldsymbol{A}_{22}^{-1} \boldsymbol{A}_{12}^\top \boldsymbol{A}_{11}^{-1/2} \right|. \end{aligned}$$

Thus, $\lim_{n \to \infty} |\boldsymbol{A}_n| \neq 0$ if and only if $\left| \boldsymbol{I}_{q_2} - \boldsymbol{A}_{22}^{-1/2} \boldsymbol{A}_{12}^\top \boldsymbol{A}_{11}^{-1} \boldsymbol{A}_{12} \boldsymbol{A}_{22}^{-1/2} \right| \neq 0$ or $\left| \boldsymbol{I}_{q_1} - \boldsymbol{A}_{11}^{-1/2} \boldsymbol{A}_{12} \boldsymbol{A}_{22}^{-1} \boldsymbol{A}_{12}^\top \boldsymbol{A}_{11}^{-1/2} \right| \neq 0$. However, it is not straightforward to establish any of these conditions for a general case.

5. Hypothesis testing

As a consequence of the limiting distribution of the ESS established in Theorem 3, an approximate hypothesis testing for the ESS is constructed. Consider the null hypothesis

$$\mathrm{H}_0 : g(\boldsymbol{\theta}) = n_0,$$

versus one of the following three alternative hypotheses $\mathrm{H}_1 : g(\boldsymbol{\theta}) \neq n_0$, $\mathrm{H}_1 : g(\boldsymbol{\theta}) < n_0$ or $\mathrm{H}_1 : g(\boldsymbol{\theta}) > n_0$, where $1 < n_0 < n$. When $n_0 = n$, this hypothesis testing problem is relevant because it leads to the case when $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}$. For the Matérn covariance function, the parameter ϕ controls the range of spacial association such that if $\phi = 0$, then $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}$. Thus, the test about the parameter ϕ, $\mathrm{H}_0 : \phi = 0$ versus $\mathrm{H}_1 : \phi > 0$ implies the test about the effective sample size $\mathrm{H}_0 : g(\boldsymbol{\theta}) = n$ versus $\mathrm{H}_1 : g(\boldsymbol{\theta}) < n$. For a given size α, the critical regions $\mathcal{C}_\alpha$ for the three alternative hypotheses are respectively $\mathcal{C}_\alpha = \left\{ g(\widehat{\boldsymbol{\theta}}) : \ |g(\widehat{\boldsymbol{\theta}}) - n_0| > z_{\alpha/2} \sqrt{\sigma_g^2} \right\}, \mathcal{C}_\alpha = \left\{ g(\widehat{\boldsymbol{\theta}}) : \ g(\widehat{\boldsymbol{\theta}}) - n_0 < z_{1-\alpha} \sqrt{\sigma_g^2} \right\},$ $\mathcal{C}_\alpha = \left\{ g(\widehat{\boldsymbol{\theta}}) : \ g(\widehat{\boldsymbol{\theta}}) - n_0 > z_\alpha \sqrt{\sigma_g^2} \right\},$ where z_α denotes the upper quantile of the standard normal distribution and $\sigma_g^2 = \mathrm{var}(g(\boldsymbol{\theta}))$.

The following two examples illustrate the computation of σ_g^2 for a Gaussian random field with the exponential covariance function and for the CAR and DCAR processes.

Example 5. Let $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{Z}^2\}$ be a stationary Gaussian process having an exponential covariance function. Let $n = NM$ be the sample size and $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{Q}(\phi)$ be the covariance matrix of $\boldsymbol{Y} = (Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n))^\top$, where the ij-th element of $\boldsymbol{Q}$ is $q_{ij} = \exp(-\|\boldsymbol{s}_i - \boldsymbol{s}_j\|/\phi)$. Let $\boldsymbol{X}$ be an $n \times p$ normalized matrix and $\boldsymbol{\theta} = (\sigma^2, \phi)$. By Theorem 2 ($m = 0$), the asymptotic normality of $\widehat{\boldsymbol{\theta}}$ is granted, $g(\phi) = \operatorname{tr}(\boldsymbol{X}^\top \boldsymbol{Q}(\phi)^{-1} \boldsymbol{X})/p$, and

$$\sigma_g^2 = [g'(\phi)]^2 \operatorname{var}[\widehat{\phi}],$$

where $g'(\phi) = -\dfrac{1}{p\phi^2} \operatorname{tr}\left\{\boldsymbol{X}^\top \boldsymbol{Q}^{-1}(\boldsymbol{Q} \odot \boldsymbol{H})\boldsymbol{Q}^{-1}\boldsymbol{X}\right\}$ and $\operatorname{var}[\widehat{\phi}]$ is the element located in position 22 of the inverse of

$$\boldsymbol{B_\theta} = \frac{1}{2} \begin{pmatrix} \dfrac{n}{\sigma^4} & \dfrac{1}{\sigma^2\phi^2} \operatorname{tr}\left(\boldsymbol{Q}^{-1}(\boldsymbol{Q} \odot \boldsymbol{H})\right) \\ \dfrac{1}{\sigma^2\phi^2} \operatorname{tr}\left(\boldsymbol{Q}^{-1}(\boldsymbol{Q} \odot \boldsymbol{H})\right) & \dfrac{1}{\phi^4} \operatorname{tr}\left(\left(\boldsymbol{Q}^{-1}(\boldsymbol{Q} \odot \boldsymbol{H})\right)^2\right) \end{pmatrix}.$$

Because $|\boldsymbol{B_\theta}| = \dfrac{2n^2}{\sigma^4\phi^4} S^2$, where S^2 is given by

$$S^2 = \frac{1}{n} \operatorname{tr}\left(\left(\boldsymbol{Q}^{-1}(\boldsymbol{Q} \odot \boldsymbol{H})\right)^2\right) - \left(\frac{1}{n} \operatorname{tr}\left(\boldsymbol{Q}^{-1}(\boldsymbol{Q} \odot \boldsymbol{H})\right)\right)^2,$$

we have that $\operatorname{var}[\widehat{\phi}] = \dfrac{\phi^4}{2nS^2}$. Consequently,

$$\sigma_g^2 = \frac{\operatorname{tr}^2\left(\boldsymbol{X}^\top \boldsymbol{Q}^{-1}(\boldsymbol{Q} \odot \boldsymbol{H})\boldsymbol{Q}^{-1}\boldsymbol{X}\right)}{2np^2S^2}.$$

When $n_0 = n$, it is meaningful to compare the tests

$$\mathrm{H}_0 : \ g(\phi) = n \quad \text{versus} \quad \mathrm{H}_1 : \ g(\phi) < n, \tag{5.1}$$

$$\mathrm{H}_0 : \ \phi = 0 \quad \text{versus} \quad \mathrm{H}_1 : \ \phi > 0, \tag{5.2}$$

where the critical region for (5.2) is $\mathcal{C}_\alpha = \left\{\widehat{\phi} : \ \widehat{\phi} < z_\alpha^{-1}\sqrt{nS^2}\right\}$. Moreover, under H_0, the test (5.2) implies that $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}$. Thus, the test (5.2) is an alternative to the hypothesis testing problem $H_0 : \boldsymbol{\Sigma} = \sigma^2 \mathbf{I}$ versus $H_1 : \boldsymbol{\Sigma} \neq \sigma^2 \mathbf{I}$, commonly studied when there are replicates (Ledoit and Wolf, 2002).

Example 6. Let $\boldsymbol{Y} \sim \mathcal{N}_n(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$ be a DCAR process with

$$\boldsymbol{\Sigma} = \tau^2 \left(\mathbf{I} - \delta_1 \widetilde{\boldsymbol{W}}^{(1)} - \delta_2 \widetilde{\boldsymbol{W}}^{(2)}\right)^{-1} \boldsymbol{D}^{-1},$$

where $\boldsymbol{D}$ is as in Section A.3 and $\widetilde{\boldsymbol{W}}^{(1)}$ and $\widetilde{\boldsymbol{W}}^{(2)}$ are the normalized directional contiguity matrices satisfying $\boldsymbol{W} = \widetilde{\boldsymbol{W}}^{(1)} + \widetilde{\boldsymbol{W}}^{(2)}$. Kyung and Ghosh

(2010) proved that the conditions of Theorem 1 for this model are satisfied for an increasing domain asymptotic scheme. For $\boldsymbol{\theta} = (\tau^2, \delta_1, \delta_2)$, the Fisher information matrix is

$$B_{\boldsymbol{\theta}} = \frac{1}{2\tau^4} \begin{pmatrix} n & \operatorname{tr}(\boldsymbol{G}_1) & \operatorname{tr}(\boldsymbol{G}_2) \\ \operatorname{tr}(\boldsymbol{G}_1) & \operatorname{tr}(\boldsymbol{G}_1^2) & \operatorname{tr}(\boldsymbol{G}_1\boldsymbol{G}_2) \\ \operatorname{tr}(\boldsymbol{G}_2) & \operatorname{tr}(\boldsymbol{G}_1\boldsymbol{G}_2) & \operatorname{tr}(\boldsymbol{G}_2^2) \end{pmatrix}, \tag{5.3}$$

with $\boldsymbol{G}_k = \widetilde{\boldsymbol{W}}^{(k)} \boldsymbol{D} \boldsymbol{\Sigma}$, $k = 1, 2$. Moreover, $\operatorname{var}\left[\widehat{\tau}^2, \widehat{\delta}_1, \widehat{\delta}_2\right] = \boldsymbol{B}_{\boldsymbol{\theta}}^{-1}$, where

$$B_{\boldsymbol{\theta}}^{-1} = \frac{2\tau^4(\boldsymbol{M}_1 - \boldsymbol{M}_2)}{a},$$

with $a = n \operatorname{tr}(\boldsymbol{G}_1^2) \operatorname{tr}(\boldsymbol{G}_2^2) + 2 \operatorname{tr}(\boldsymbol{G}_1) \operatorname{tr}(\boldsymbol{G}_2) \operatorname{tr}(\boldsymbol{G}_1\boldsymbol{G}_2) - n \operatorname{tr}^2(\boldsymbol{G}_1\boldsymbol{G}_2) - \operatorname{tr}(\boldsymbol{G}_1^2) \operatorname{tr}^2(\boldsymbol{G}_2) - \operatorname{tr}(\boldsymbol{G}_2^2) \operatorname{tr}^2(\boldsymbol{G}_1)$ and

$$M_1 = \begin{pmatrix} \operatorname{tr}(\boldsymbol{G}_1^2) \operatorname{tr}(\boldsymbol{G}_2^2) & \operatorname{tr}(\boldsymbol{G}_2) \operatorname{tr}(\boldsymbol{G}_1\boldsymbol{G}_2) & \operatorname{tr}(\boldsymbol{G}_1) \operatorname{tr}(\boldsymbol{G}_1\boldsymbol{G}_2) \\ \operatorname{tr}(\boldsymbol{G}_2) \operatorname{tr}(\boldsymbol{G}_1\boldsymbol{G}_2) & n \operatorname{tr}(\boldsymbol{G}_2^2) & \operatorname{tr}(\boldsymbol{G}_1) \operatorname{tr}(\boldsymbol{G}_2) \\ \operatorname{tr}(\boldsymbol{G}_1) \operatorname{tr}(\boldsymbol{G}_1\boldsymbol{G}_2) & \operatorname{tr}(\boldsymbol{G}_1) \operatorname{tr}(\boldsymbol{G}_2) & n \operatorname{tr}(\boldsymbol{G}_1^2) \end{pmatrix},$$

$$M_2 = \begin{pmatrix} \operatorname{tr}^2(\boldsymbol{G}_1\boldsymbol{G}_2) & \operatorname{tr}(\boldsymbol{G}_1) \operatorname{tr}(\boldsymbol{G}_2^2) & \operatorname{tr}(\boldsymbol{G}_2) \operatorname{tr}(\boldsymbol{G}_1^2) \\ \operatorname{tr}(\boldsymbol{G}_1) \operatorname{tr}(\boldsymbol{G}_2^2) & \operatorname{tr}^2(\boldsymbol{G}_2) & n \operatorname{tr}(\boldsymbol{G}_1\boldsymbol{G}_2) \\ \operatorname{tr}(\boldsymbol{G}_1^2) \operatorname{tr}(\boldsymbol{G}_2) & n \operatorname{tr}(\boldsymbol{G}_1\boldsymbol{G}_2) & \operatorname{tr}^2(\boldsymbol{G}_1) \end{pmatrix}.$$

Equation (5.3) implies that $n^* = g(\tau^2, \delta_1, \delta_2) = \dfrac{1}{\tau^2 p} \operatorname{tr}\left(\boldsymbol{X}^\top \boldsymbol{C} \boldsymbol{A}(\delta_1, \delta_2) \boldsymbol{D} \boldsymbol{C} \boldsymbol{X}\right)$, where $\boldsymbol{C} = \operatorname{diag}\left(\Sigma_{ii}^{1/2}\right)$, $i = 1, \ldots, n$ and $\boldsymbol{A}(\delta_1, \delta_2) = \boldsymbol{I} - \delta_1 \widetilde{\boldsymbol{W}}^{(1)} - \delta_2 \widetilde{\boldsymbol{W}}^{(2)}$. Defining $g_{\tau^2} = \dfrac{\partial g}{\partial \tau^2}$, $g_{\delta_i} = \dfrac{\partial g}{\partial \delta_i}$, $i = 1, 2$,

$$g_{\tau^2} = -\frac{1}{\tau^4 p} \operatorname{tr}\left(\boldsymbol{X}^\top \boldsymbol{C} \boldsymbol{A}(\delta_1, \delta_2) \boldsymbol{D} \boldsymbol{C} \boldsymbol{X}\right) + \frac{2}{\tau^2 p} \operatorname{tr}\left(\boldsymbol{X}^\top \boldsymbol{C}_{\tau^2} \boldsymbol{A}(\delta_1, \delta_2) \boldsymbol{D} \boldsymbol{C} \boldsymbol{X}\right), \tag{5.4}$$

$$g_{\delta_1} = -\frac{1}{\tau^2 p} \operatorname{tr}\left(\boldsymbol{X}^\top \boldsymbol{C} \widetilde{\boldsymbol{W}}^{(1)} \boldsymbol{D} \boldsymbol{C} \boldsymbol{X}\right) + \frac{2}{\tau^2 p} \operatorname{tr}\left(\boldsymbol{X}^\top \boldsymbol{C}_{\delta_1} \boldsymbol{A}(\delta_1, \delta_2) \boldsymbol{D} \boldsymbol{C} \boldsymbol{X}\right), \tag{5.5}$$

$$g_{\delta_2} = -\frac{1}{\tau^2 p} \operatorname{tr}\left(\boldsymbol{X}^\top \boldsymbol{C} \widetilde{\boldsymbol{W}}^{(2)} \boldsymbol{D} \boldsymbol{C} \boldsymbol{X}\right) + \frac{2}{\tau^2 p} \operatorname{tr}\left(\boldsymbol{X}^\top \boldsymbol{C}_{\delta_2} \boldsymbol{A}(\delta_1, \delta_2) \boldsymbol{D} \boldsymbol{C} \boldsymbol{X}\right), \tag{5.6}$$

where $\boldsymbol{C}_{\tau^2} = \operatorname{diag}\left(\partial(\boldsymbol{\Sigma}_{ii}^{1/2})/\partial\tau^2\right)$, $\boldsymbol{C}_{\delta_k} = \operatorname{diag}\left(\partial(\boldsymbol{\Sigma}_{ii}^{1/2})/\partial\delta_k\right)$, $i = 1, \ldots, n$, $k = 1, 2$, $\boldsymbol{\Sigma}_{ii} = \tau^2 \boldsymbol{A}^{ii}(\delta_1, \delta_2)/w_{i+}$ and $\boldsymbol{A}^{ii}$ is as before, we have that

$$\sigma_g^2 = \frac{2\tau^4 b}{a}, \tag{5.7}$$

where

$$b = g_{\tau^2}^2 \left[\mathrm{tr}\left(\boldsymbol{G}_1^2\right) \mathrm{tr}\left(\boldsymbol{G}_2^2\right) - \mathrm{tr}^2\left(\boldsymbol{G}_1\boldsymbol{G}_2\right) \right] + g_{\delta_1}^2 \left[n\,\mathrm{tr}\left(\boldsymbol{G}_2^2\right) - \mathrm{tr}^2\left(\boldsymbol{G}_2\right) \right]$$
$$+ g_{\delta_2}^2 \left[n\,\mathrm{tr}\left(\boldsymbol{G}_1^2\right) - \mathrm{tr}^2\left(\boldsymbol{G}_1\right) \right] - 2g_{\tau^2}g_{\delta_1} \left[Tr\left(\boldsymbol{G}_2\right) \mathrm{tr}\left(\boldsymbol{G}_1\boldsymbol{G}_2\right) - \mathrm{tr}\left(\boldsymbol{G}_1\right) \mathrm{tr}\left(\boldsymbol{G}_2^2\right) \right]$$
$$- 2g_{\tau^2}g_{\delta_2} \left[\mathrm{tr}\left(\boldsymbol{G}_1\right) \mathrm{tr}\left(\boldsymbol{G}_1\boldsymbol{G}_2\right) - \mathrm{tr}\left(\boldsymbol{G}_2\right) \mathrm{tr}\left(\boldsymbol{G}_1^2\right) \right]$$
$$- 2g_{\delta_1}g_{\delta_2} \left[\mathrm{tr}\left(\boldsymbol{G}_1\right) \mathrm{tr}\left(\boldsymbol{G}_2\right) - n\,\mathrm{tr}\left(\boldsymbol{G}_1\boldsymbol{G}_2\right) \right],$$

and g_{τ^2}, g_{δ_1} and g_{δ_2} are as in Equations (5.4), (5.5), and (5.6), respectively.

If $\delta_1 = \delta_2 = \rho$, then

$$\sigma_g^2 = \frac{\tau^4 \left(g_{\tau^2}^2 \,\mathrm{tr}\left(\left(\widetilde{\boldsymbol{W}}\boldsymbol{D}\boldsymbol{\Sigma}\right)^2 \right) - 2g_{\tau^2}g_\rho\,\mathrm{tr}\left(\widetilde{\boldsymbol{W}}\boldsymbol{D}\boldsymbol{\Sigma}\right) + ng_\rho^2 \right)}{n\,\mathrm{tr}\left(\left(\widetilde{\boldsymbol{W}}\boldsymbol{D}\boldsymbol{\Sigma}\right)^2 \right) - \mathrm{tr}^2\left(\widetilde{\boldsymbol{W}}\boldsymbol{D}\boldsymbol{\Sigma}\right)},$$

where

$$g_{\tau^2} = \frac{\partial g}{\partial \tau^2} = -\frac{1}{\tau^4 p}\,\mathrm{tr}\left(\boldsymbol{X}^\top \boldsymbol{C}\boldsymbol{A}(\rho)\boldsymbol{D}\boldsymbol{C}\boldsymbol{X}\right) + \frac{2}{\tau^2 p}\,\mathrm{tr}\left(\boldsymbol{X}^\top \boldsymbol{C}_{\tau^2}\boldsymbol{A}(\rho)\boldsymbol{D}\boldsymbol{C}\boldsymbol{X}\right),$$

$$g_\rho = \frac{\partial g}{\partial \rho} = -\frac{1}{\tau^2 p}\,\mathrm{tr}\left(\boldsymbol{X}^\top \boldsymbol{C}\widetilde{\boldsymbol{W}}\boldsymbol{D}\boldsymbol{C}\boldsymbol{X}\right) + \frac{2}{\tau^2 p}\,\mathrm{tr}\left(\boldsymbol{X}^\top \boldsymbol{C}_\rho\boldsymbol{A}(\rho)\boldsymbol{D}\boldsymbol{C}\boldsymbol{X}\right),$$

with $\boldsymbol{A}(\rho) = \boldsymbol{I} - \rho\widetilde{\boldsymbol{W}}$.

6. Numerical experiments

In this section, we conducted two Monte Carlo simulation studies. The first one is designed to observe the performance of the ML estimates of the ESS for different correlation structures and sample sizes. The second study is dedicated to exploring the power function of the test stated in Section 5 as a function of the sample size.

The first experiment involves the generation of 500 replicates from a Gaussian random field, sampled on a regular lattice of size $n = r \times r$, for $r = 8, 16, 32$, where the sites $\boldsymbol{s}_1, \dots, \boldsymbol{s}_n$ are of the form $\boldsymbol{s}_i = (x_i, y_i)$. Each replicate was generated from the model

$$Y(\boldsymbol{s}_i) = \beta_0 + \beta_1 x_i y_i + \varepsilon(\boldsymbol{s}_i),$$

where $\mathrm{cov}(\varepsilon(\boldsymbol{s}_i), \varepsilon(\boldsymbol{s}_j)) = \sigma^2 \rho(\|\boldsymbol{s}_i - \boldsymbol{s}_j\|, \phi)$ for four correlation functions $\rho(\cdot, \phi)$ given in Table 1, with $\beta_0 = 2, \beta_1 = 0.1, \sigma^2 = 2$, and $\phi = 1$. The spherical correlation was not included because it does not satisfy the conditions of Theorem 1. Table 2 displays the simulation estimates and mean square error (in parentheses) for the parameters β_0, β_1, σ^2 and ϕ. In addition, n^*, n^*/n and its estimates were included to make the estimates comparable when n varies. n^*/n represents the proportion of equivalent independent observations associated with a spatial sample of size n. There is empirical evidence to support the consistency of the estimates of n^* and n^*/n. In practice, the size of the grid

TABLE 2

Monte Carlo estimates based on 500 replicates. The true parameters are $\beta_0 = 2$, $\beta_1 = 0.1$, $\sigma^2 = 2$ and $\phi = 1$, and the numbers in brackets are the root mean square error of the simulation estimates.

Covariance	n	n^*	n^*/n	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\sigma}^2$	$\widehat{\phi}$	$\widehat{n^*}$	$\widehat{n^*}/n$
Exponential	8×8	16.06	0.251	2.0340 (0.5574)	0.0992 (0.0207)	1.7738 (0.4541)	0.8161 (0.3519)	24.31 (14.41)	0.3800 (0.2252)
	16×16	51.32	0.200	1.9906 (0.2943)	0.1000 (0.0030)	1.9501 (0.2390)	0.9581 (0.1748)	58.13 (17.59)	0.2271 (0.0687)
	32×32	180.78	0.177	2.0065 (0.1588)	0.1000 (0.0004)	1.9941 (0.1175)	0.9947 (0.0780)	184.90 (24.79)	0.1806 (0.0242)
Gaussian	8×8	24.65	0.385	2.0169 (0.4528)	0.0996 (0.0183)	1.8940 (0.4504)	0.9742 (0.1017)	26.32 (5.15)	0.4113 (0.0805)
	16×16	89.88	0.351	2.0211 (0.2449)	0.0999 (0.0025)	1.9810 (0.2086)	0.9980 (0.0389)	90.56 (6.32)	0.3537 (0.0247)
	32×32	342.55	0.335	2.0071 (0.1141)	0.1000 (0.0003)	1.9893 (0.1141)	0.9976 (0.0192)	344.46 (12.50)	0.3364 (0.0122)
Matérn $m = 1$	8×8	8.12	0.127	2.0012 (0.7590)	0.1003 (0.0248)	1.6425 (0.7557)	0.8966 (0.2064)	9.63 (2.78)	0.1505 (0.0435)
	16×16	21.90	0.086	1.9812 (0.4453)	0.1001 (0.0044)	1.8737 (0.4868)	0.9651 (0.1067)	23.50 (3.75)	0.0918 (0.0146)
	32×32	69.87	0.068	2.0032 (0.2437)	0.1000 (0.0007)	1.9677 (0.2214)	0.9907 (0.0509)	71.40 (6.29)	0.0697 (0.0061)
Matérn $m = 2$	8×8	6.73	0.105	1.9804 (0.8072)	0.1002 (0.0239)	1.6293 (0.8888)	0.9248 (0.1540)	7.2624 (0.9761)	0.1135 (0.0153)
	16×16	15.81	0.062	1.9971 (0.5473)	0.1000 (0.0048)	1.8181 (0.7972)	0.9633 (0.0964)	16.84 (2.92)	0.0658 (0.0114)
	32×32	46.35	0.045	2.0036 (0.3040)	0.0999 (0.0008)	2.0549 (1.4618)	0.9955 (0.0651)	46.96 (3.33)	0.0459 (0.0033)

could be considerably large. We explore this by performing the same simulation experiment for $r = 64, 128$, and 256 (not shown here), obtaining the same results with respect to the consistency of the estimates; however, the computational cost have increased considerably. In Figure 3, we show the average user time that the system takes to compute the ML estimate of the ESS based on 100 replicates. A breaking point in the slope of the curves for $n \geq 40 \times 40$ is observed regardless of the type of covariance function that is used. In the Gaussian case, there are 32 observations for which the user time was greater than three hours for the gird sizes 40×40 and 48×48, making it not comparable with the other covariance functions. In practice, the computation of the effective sample size can be performed on a regular computer for grid sizes of less than 32×32. Our simulation experiments were developed using a server HP ProLiant DL360 Gen9 E5-2630v3 2.4 GHz 8-core 1P 16 GB-R P440ar 500 W PS Base SAS.

The second simulation study is designed to explore the power of the hypothesis testing problem $H_0 : n^* = n$ versus $H_1 : n^* < n$. The true value of the n^* is denoted by n_0^*. Based on 500 replicates of a Gaussian random field, the empirical power function was computed for $r = 8, 16, 32$. For $r = 8$, $n_0^* = 38.67, 60.41, 20.69, 15.26$, for $r = 16$, $n_0^* = 142.02, 240.02, 67.82, 45.84$, and for $r = 32$, $n_0^* = 142.02, 240.02, 67.82, 45.84$, for the exponential, Gaussian and Matérn $(m = 1, 2)$ covariance functions, respectively. In contrast to a usual hy-
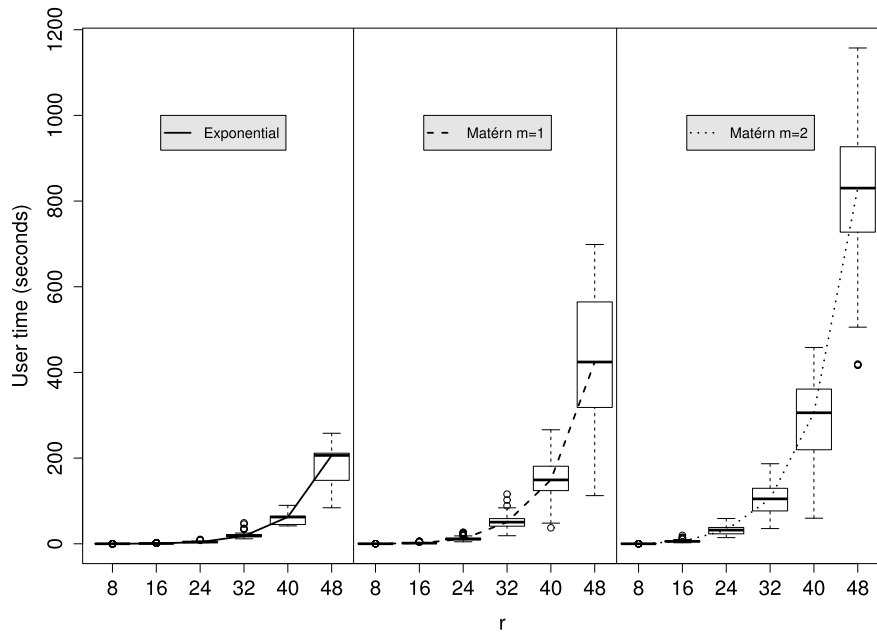
FIG 3. *User time in seconds for the exponential, Matérn (m = 1, m = 2), and Gaussian correlation functions based in each case on 100 replicates from a Gaussian random field.*

pothesis testing, the value of n_0^* is variable. The effect of this behavior is shown in Figure 4. For large values of n, the power function become small, regardless of the covariance structure that is used. In all cases, the Gaussian covariance function has the worst behavior in terms of power, whereas the Matérn model with $m = 2$ has the best.

7. Real data analysis

We analyzed the macroalgae dataset from Acosta et al. (2016), which consists of 427 observations related to the density of Lessonia trabeculata (scientific name of macroalgae in the study) per 20 m^2. This dataset was collected by the Fisheries Research Institute (IFOP in Spanish) in a protected area near Quintero, Chile. In Acosta et al. (2016), the effective sample size computations were obtained under the assumption of a constant mean Gaussian process using 26 perpendicular line transects to the coastline to study those species located at a depth of no more than 20 m because the experts are knowledgeable of the existence of species with different sizes. In addition to the density, the depth was measured for 144 georeferenced sites (see Figure 5). We examined the effective sample size for the subset of 144 sites for which both the density and depth are available. Acosta et al. (2016) found a high asymmetry (Fisher asymmetry
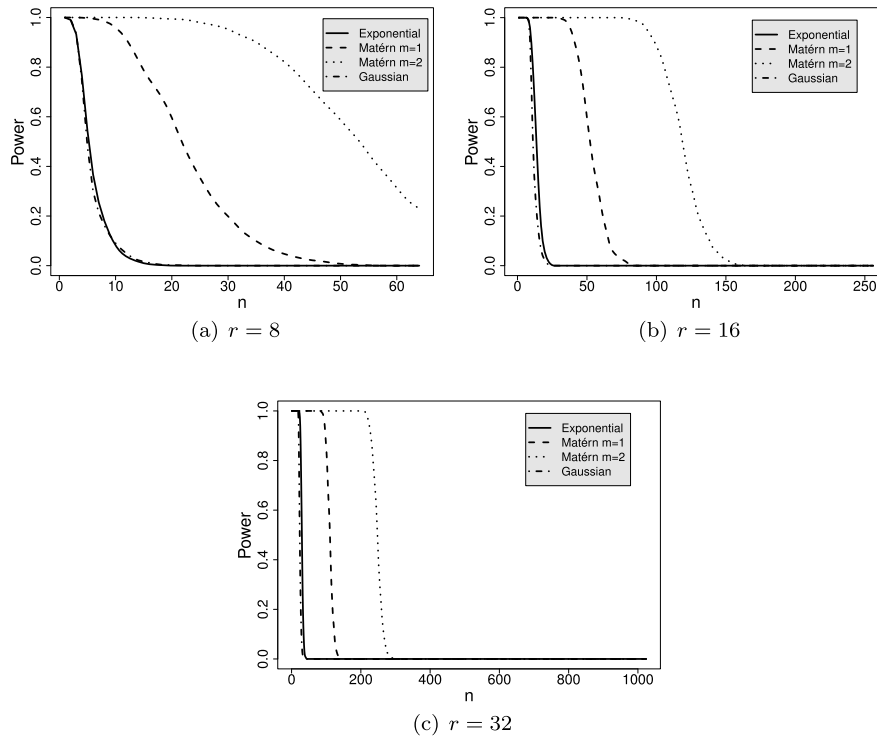
FIG 4. *Empirical power functions for the test $H_0 : n^* = n$ versus $H_1 : n^* < n$, for the exponential, Gaussian and Matérn $(m = 1, 2)$ covariance functions. $r = 8, 16, 32$.*

coefficient equal to 0.803); then, the logarithm transformation

$$Z(\boldsymbol{s}_i) = \ln(Y(\boldsymbol{s}_i) + \delta)$$

was applied to the original dataset, where δ is the value that maximizes the correlation (0.9878) between the percentiles of the normal distribution and the order statistics $Z_{(i)}$, obtaining $\widehat{\delta} = 0.588$ (Acosta et al., 2016, Appendix 2). The Fisher asymmetry coefficient for the 144 observations is -0.00176.

Let $Z(\cdot)$ be a random field and assume that $\boldsymbol{Z} = (Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_{144}))$ is a Gaussian random vector such that $\mathbb{E}[\boldsymbol{Z}] = \boldsymbol{X}\boldsymbol{\beta}$, where $\boldsymbol{X}$ is the design matrix and $\boldsymbol{\Sigma}_{ij} = \operatorname{cov}(Z(\boldsymbol{s}_i), Z(\boldsymbol{s}_j)) = \tau^2 + \sigma^2$ if $i = j$, and $\boldsymbol{\Sigma}_{ij} = \operatorname{cov}(Z(\boldsymbol{s}_i, \boldsymbol{s}_j)) = \tau^2 + \sigma^2 \rho(\|\boldsymbol{s}_i - \boldsymbol{s}_j\|, \phi)$ if $i \neq j$, where $\rho(\cdot)$ is one of the functions presented in Table 1. A preliminary model might be

$$\mathbb{E}[Z(\boldsymbol{s}_i)] = \beta_0 + \beta_1(x_i - \bar{x}) + \beta_2(y_i - \bar{y}) + \beta_3 d(\boldsymbol{s}_i),$$

where $\boldsymbol{s}_i = (x_i, y_i)$, $\bar{x} = \sum x_i/n$, $\bar{y} = \sum y_i/n$ and $d(\boldsymbol{s}_i)$ is the depth at the site $\boldsymbol{s}_i, i = 1, \ldots, 144$. For each of the covariance structures listed in Table 1, the ML estimates of the parameters were computed, obtaining that the overall F test led
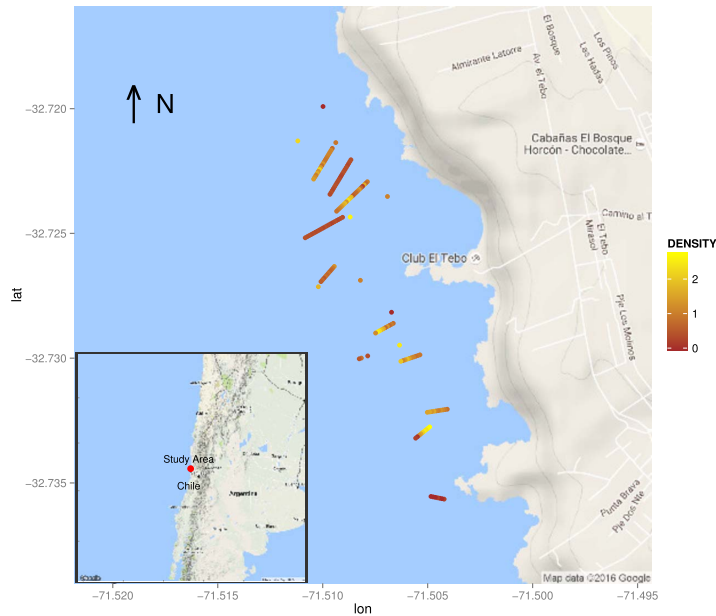
Fig 5. *Study area in the central coast of Chile, including 144 observations collected by using a transect lines scheme. This area was chosen because the species are located at a depth of no more than 20 meters and because the experts are knowledgeable of the existence of species with different sizes. The region to be sampled has an area from north to south that is approximately 2.12 km long. In addition, a free access area 1 km long was considered in the study.*

us to a significant model with a p-value in each case of less than 0.05. However, the t-tests on single regression coefficients yielded nonsignificant estimates for β_1 and β_2 and significant estimates for β_3, regardless of the covariance function used. Then, we consider the reduced trend model of the form

$$\mathbb{E}[Z(\boldsymbol{s}_i)] = \beta_0 + \beta_1 d(\boldsymbol{s}_i). \tag{7.1}$$

The overall F test, AIC coefficient, mean square error, n^* and the ML estimates for the parameters of model (7.1) are presented in Table 3. The Gaussian covariance structure minimized the AIC and the MSE. For this case, the effective sample size of the transformed variable is 27.5, which is equal to the estimate obtained by using the Taylor approximation of order 1. In Acosta et al. (2016), the obtained ESS value was 66.1 for the Gaussian covariance, but the sample size in that case was 427; therefore, the results are not comparable.

The modeling of the density of Lessonia trabeculata constitutes a valuable advance in the design of protocols in IFOP that can be applied in the future when they plan to collect new datasets in the same study area for this or other variables of interest, under similar conditions to those described in this paper, where the inclusion of covariates plays an important role when constructing the spatial regression model to be used.

TABLE 3
Parameter estimation and summary of principal statistics of the regression model (7.1).
The value in parentheses is the p-value of t-test.

Model	F	AIC	MSE	n_Z^*	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\sigma}^2$	$\widehat{\phi}$	$\widehat{\tau}^2$
Exponential	0.030	23.62	25.63	20.22	0.067 (0.719)	0.018 (0.102)	0.219	100.00	0.009
Matérn $\kappa = 1.5$	0.008	17.84	25.09	24.13	0.056 (0.755)	0.022 (0.061)	0.225	42.02	0.024
Matérn $\kappa = 2.5$	0.006	17.25	24.99	25.21	0.056 (0.751)	0.023 (0.054)	0.226	30.12	0.026
Gaussian	0.002	15.94	24.83	27.51	0.054 (0.751)	0.024 (0.037)	0.224	81.65	0.028

8. Discussion

This paper extended the known methodology for effective sample size computations for general spatial regression models. The approach, equipped with powerful computational machinery, is appropriate for large spatial datasets, and it provides formulas for a number of spatial processes. We conjecture that the n^* in (2.1) is increasing in n. This is suggested by a comprehensive simulation study that we performed. However, at this stage, we can only provide a proof with additional hypotheses.

The definition of ESS can easily be extended to more general models. For instance, for a nonlinear model of the form (Golub and Pereyra, 1973)

$$\boldsymbol{Y} = \boldsymbol{X}(\boldsymbol{\alpha})\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{8.1}$$

where $\boldsymbol{\varepsilon} \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{R}(\boldsymbol{\theta})\right)$, $\boldsymbol{\beta} \in \mathbb{R}^p$, $\boldsymbol{\alpha} \in \mathbb{R}^r$ is a conditionally linear parameter (Soo and Bates, 1992), $X(\cdot)$ is normalized by columns up to $\sqrt{n}$ design matrix and $\boldsymbol{R}(\cdot)$ is as before, it is straightforward to obtain

$$n^* = n \cdot \frac{\operatorname{tr}\left(\boldsymbol{X}^\top(\boldsymbol{\alpha})\boldsymbol{R}^{-1}(\boldsymbol{\theta})\boldsymbol{X}(\boldsymbol{\alpha})\right) + \sum_{i=1}^r \boldsymbol{\beta}^\top \boldsymbol{X}_i^\top(\boldsymbol{\alpha})\boldsymbol{R}^{-1}(\boldsymbol{\theta})\boldsymbol{X}_i(\boldsymbol{\alpha})\boldsymbol{\beta}}{np + \sum_{i=1}^r \boldsymbol{\beta}^\top \boldsymbol{X}_i^\top(\boldsymbol{\alpha})\boldsymbol{X}_i(\boldsymbol{\alpha})\boldsymbol{\beta}}.$$

where $\boldsymbol{X}_i(\boldsymbol{\alpha}) = \frac{\partial \boldsymbol{X}(\boldsymbol{\alpha})}{\partial \alpha_i}$, $i = 1, \ldots, r$. After simple algebra, we have that

$$n^* = \lambda n_{\boldsymbol{\beta}}^* + (1 - \lambda)n_{\boldsymbol{\alpha}}^*,$$

where $n_{\boldsymbol{\beta}}^* = \operatorname{tr}\left(\boldsymbol{X}^\top(\boldsymbol{\alpha})\boldsymbol{R}^{-1}(\boldsymbol{\theta})\boldsymbol{X}(\boldsymbol{\alpha})\right)/p$ and

$$n_{\boldsymbol{\alpha}}^* = n \cdot \sum_{i=1}^r \|\boldsymbol{X}_i(\boldsymbol{\alpha})\boldsymbol{\beta}\|_{\boldsymbol{R}^{-1}}^2 / \sum_{i=1}^r \|\boldsymbol{X}_i(\boldsymbol{\alpha})\boldsymbol{\beta}\|^2,$$

which correspond to the effective sample size associated with the linear and conditionally linear parameters respectively, and $\lambda = np/(np + \sum_{i=1}^r \|\boldsymbol{X}_i(\boldsymbol{\alpha})\boldsymbol{\beta}\|^2)$.

The asymptotic results provide support for the inference when using the ML estimates for processes defined on rectangular grids. The conditions of Theorem 1 are not easy to establish for a particular covariance function. The results

and examples developed here highlight the limitation of the technique. For a nonlinear model as in Equation (8.1), the conditions of Theorem 1 are not sufficient to guarantee the normality assumption of the ML estimate of $(\boldsymbol{\beta}^\top, \boldsymbol{\alpha}^\top, \boldsymbol{\theta}^\top)^\top$. A hypothesis ensuring that the Fisher information matrix converges to a nonsingular matrix is needed. Such a condition is assumed for the Fisher information matrix associated with the mean of a nonlinear model in Crujeiras and Van Keilegom (2010).

The hypothesis testing discussed in Section 5 provides a framework for testing whether there is redundant information in the spatial sample of size n due to autocorrelation. Although the test is simple, for more sophisticated models, the computation of the asymptotic variance is a challenging problem.

We make specific recommendations regarding the calculation of the effective sample size in practice.

(a) The computation of the ESS requires the estimation of the covariance function. To select a good isotropic model, the classical cross-validation techniques might be used to choose from among the competing models. However, Stein (1999) noted that making parametric estimates of semivariograms match the empirical one is a considerable flaw in classical geostatistics (Gelfand et al., 2010).

(b) At the first stage, we suggest addressing the hypothesis testing problem $H_0 : n^* = n$ $H_1 : n^* < n$, to elucidate whether there is a reduction of information. Then other values for n_0 (e.g., $n/2$ or $n/3$) can be used to quantify the percentage of reduction with respect to the original sample size. For a single mean spatial regression model, the ESS might be greater than n when there is negative autocorrelation (Richardson, 1990). Hence, the alternative hypothesis $H_1 : n^* > n$ is not very meaningful as a description of a conventional situation in spatial statistics.

Acknowledgments

Appendix A: ESS for related models

A.1. ESS for SAR models

Consider an SAR model of the form

$$
\begin{aligned}
\boldsymbol{Y} &= \boldsymbol{X\beta} + \boldsymbol{e}, \\
\boldsymbol{e} &= \boldsymbol{Be} + \boldsymbol{\nu},
\end{aligned}
$$

where $\boldsymbol{B}$ is a fixed matrix of spatial interactions, the columns of $\boldsymbol{X}$ have been normalized, $\mathbb{E}[\boldsymbol{\nu}] = \boldsymbol{0}$, and $\boldsymbol{\Sigma_\nu} = \mathrm{diag}(\tilde{\sigma}_1^2, \ldots, \tilde{\sigma}_n^2)$. Then, $\mathbb{V}[\boldsymbol{Y}] = \boldsymbol{\Sigma} = (\boldsymbol{I} - \boldsymbol{B})^{-1} \boldsymbol{\Sigma_\nu} (\boldsymbol{I} - \boldsymbol{B}^\top)^{-1}$, $\boldsymbol{R}_{\mathrm{SAR}}^{-1} = \boldsymbol{C} \boldsymbol{\Sigma}^{-1} \boldsymbol{C}$ with $\boldsymbol{C} = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n)$ and $\sigma_i = \boldsymbol{\Sigma}_{ii}^{1/2}$ is the ii-th element of $\boldsymbol{\Sigma}$. If $\boldsymbol{B} = \rho \boldsymbol{W}$, where $\boldsymbol{W}$ is any contiguity matrix and $\boldsymbol{\Sigma_\nu} = \sigma^2 \boldsymbol{I}$, the effective sample size is

$$
\begin{aligned}
n_{\mathrm{SAR}}^* &= \frac{1}{p} \mathrm{tr}(\boldsymbol{X}^\top \boldsymbol{R}_{\mathrm{SAR}}^{-1} \boldsymbol{X}) \\
&= \frac{1}{\sigma^2 p} \left(\mathrm{tr}(\boldsymbol{X}^\top \boldsymbol{C}^2 \boldsymbol{X}) - 2\rho \, \mathrm{tr}(\boldsymbol{X}^\top \boldsymbol{C} \boldsymbol{W} \boldsymbol{C} \boldsymbol{X}) + \rho^2 \, \mathrm{tr}(\boldsymbol{X}^\top \boldsymbol{C} \boldsymbol{W}^\top \boldsymbol{W} \boldsymbol{C} \boldsymbol{X}) \right).
\end{aligned}
$$

n_{SAR}^* is a quadratic function of ρ, where the coefficient of ρ^2 is positive; thus, n_{SAR}^* is decreasing when $\rho \leq \rho_0$, where ρ_0 is given by

$$
\rho_0 = \frac{\mathrm{tr}(\boldsymbol{X}^\top \boldsymbol{C} \boldsymbol{W} \boldsymbol{C} \boldsymbol{X})}{\mathrm{tr}(\boldsymbol{X}^\top \boldsymbol{C} \boldsymbol{W}^\top \boldsymbol{W} \boldsymbol{C} \boldsymbol{X})}.
$$

A.2. ESS for multivariate CAR models

Suppose that $\boldsymbol{\Phi}^\top = (\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2, \ldots, \boldsymbol{\Phi}_n)$, where each $\boldsymbol{\Phi}_i$ is a matrix of size $m \times 1$. A multivariate CAR (MCAR) process is described by Gelfand and Vounatsou (2003) as

$$
\boldsymbol{\Phi}_i | \boldsymbol{\Phi}_{j \neq i}, \boldsymbol{\Sigma}_i \sim \mathcal{N} \left(\sum_j \boldsymbol{B}_{ij} \boldsymbol{\Phi}_j, \boldsymbol{\Sigma}_i \right), \quad i = 1, \ldots, n,
$$

where each $\boldsymbol{B}_{ij}$ and $\boldsymbol{\Sigma}_i$ are of size $m \times m$. Brook's lemma provides the joint density of $\boldsymbol{\Phi}$ given by

$$
f(\boldsymbol{\phi} | \{\boldsymbol{\Sigma}_i\}) \propto \exp \left(-\frac{1}{2} \boldsymbol{\phi}^\top \boldsymbol{\Gamma}^{-1} (I - \tilde{\boldsymbol{B}}) \boldsymbol{\phi} \right),
$$

where $\boldsymbol{\Gamma}$ is a block diagonal matrix with blocks $\boldsymbol{\Sigma}_i$, and $\widetilde{\boldsymbol{B}}$ is $nm \times nm$ with (i, j)th block $\boldsymbol{B}_{ij}$. Given the contiguity matrix $\boldsymbol{W} = (w_{ij})$, define $\boldsymbol{B}_{ij} = b_{ij} \mathbf{I}_{m \times m}$, with $b_{ij} = w_{ij}/w_{i+}$, $w_{i+} = \sum_j w_{ij}$. Additionally, consider $\boldsymbol{\Sigma}_i = w_{i+}^{-1} \boldsymbol{\Sigma}$. Then, $\widetilde{\boldsymbol{B}} = \boldsymbol{B} \otimes \mathbf{I}$, $\boldsymbol{\Gamma} = \boldsymbol{D}_w^{-1} \otimes \boldsymbol{\Sigma}$, where $\boldsymbol{D}_w = \mathrm{diag}(w_{i+})$, and

$\boldsymbol{\Gamma}^{-1}(I - \widetilde{\boldsymbol{B}}) = (\boldsymbol{D}_w - \boldsymbol{W}) \otimes \boldsymbol{\Sigma}^{-1}$. Since $(\boldsymbol{D}_w - \boldsymbol{W})$ could be singular, we consider $\boldsymbol{\Sigma}_{\boldsymbol{\Phi}}^{-1} = (D_w - \rho \boldsymbol{W}) \otimes \boldsymbol{\Sigma}^{-1}$, where ρ is a parameter that can be considered as the autocorrelation that stabilizes the covariance matrix (Banerjee et al., 2004, Sections 3.3 and 7.4). This model is consequently called MCAR$(\rho, \boldsymbol{\Sigma})$.

Let $\boldsymbol{Y}^{\top} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n)$, where each $\boldsymbol{Y}_i$ is $m \times 1$ and $\boldsymbol{Y}$ is $n \times m$. Let $\boldsymbol{X}, \boldsymbol{\beta}$, and $\boldsymbol{\Phi}$ be of sizes $n \times p$, $p \times 1$, and $n \times m$, respectively, such that $\boldsymbol{\Phi} \sim MCAR(\rho, \boldsymbol{\Sigma})$. $\boldsymbol{Y} = \boldsymbol{1}_m^{\top} \otimes \boldsymbol{X} \boldsymbol{\beta} + \boldsymbol{\Phi}$ implies that $\boldsymbol{Y} \sim \mathcal{N}\left(\boldsymbol{1}_m^{\top} \otimes \boldsymbol{X} \boldsymbol{\beta}, (\boldsymbol{D}_w - \rho \boldsymbol{W})^{-1} \otimes \boldsymbol{\Sigma}\right)$, or equivalently in vectorial form, $\widetilde{\boldsymbol{Y}} = (\boldsymbol{1}_m \otimes \boldsymbol{X}) \boldsymbol{\beta} + \widetilde{\boldsymbol{\Phi}}$ implies that $\boldsymbol{Y} \sim \mathcal{N}\left((\boldsymbol{1}_m \otimes \boldsymbol{X}) \boldsymbol{\beta}, \boldsymbol{\Sigma}_{\widetilde{\boldsymbol{Y}}}\right)$, where $\widetilde{\boldsymbol{Y}} = \text{vec}(\boldsymbol{Y})$, $\widetilde{\boldsymbol{\Phi}} = \text{vec}(\boldsymbol{\Phi})$, and $\boldsymbol{\Sigma}_{\widetilde{\boldsymbol{Y}}} = \boldsymbol{\Sigma} \otimes (\boldsymbol{D}_w - \rho \boldsymbol{W})^{-1}$. If the columns of $\boldsymbol{X}$ have norm $\sqrt{n}$, we have that

$$\text{tr}((\boldsymbol{1}_m \otimes \boldsymbol{X})^{\top}(\boldsymbol{1}_m \otimes \boldsymbol{X})) = \text{tr}(\boldsymbol{1}_m^{\top} \boldsymbol{1}_m \otimes \boldsymbol{X}^{\top} \boldsymbol{X}) = m \, \text{tr}(\boldsymbol{X}^{\top} \boldsymbol{X}) = mpn .$$

Finally, setting $\boldsymbol{C} = \text{diag}\left((\boldsymbol{\Sigma}_{\widetilde{\boldsymbol{Y}}})_{ii}^{1/2}\right)$ and $\boldsymbol{R}^{-1} = \boldsymbol{C}\left(\boldsymbol{\Sigma}^{-1} \otimes (D_w - \rho \boldsymbol{W})\right) \boldsymbol{C}$, we obtain

$$n_{\text{MCAR}}^* = \frac{1}{mp} \text{tr}\left((\boldsymbol{1}_m^{\top} \otimes \boldsymbol{X}^{\top}) \boldsymbol{R}^{-1} (\boldsymbol{1}_m \otimes \boldsymbol{X})\right). \tag{A.1}$$

If $\boldsymbol{X} = \boldsymbol{1}_n$, $n_{\text{MCAR}}^* = \boldsymbol{1}_{nm}^{\top} \boldsymbol{C}\left(\boldsymbol{\Sigma}^{-1} \otimes (\boldsymbol{D}_w - \rho \boldsymbol{W})\right) \boldsymbol{C} \boldsymbol{1}_{mn}/m$. In addition, if $\boldsymbol{\Sigma}$ is diagonal, the CAR specification is recovered, i.e., $n_{MCAR}^* = n_{CAR}^*$.

A.3. ESS for DCAR models

The directional conditionally autoregressive processes (DCAR) were defined by Kyung and Ghosh (2010) as an extension of the CAR processes for which the correlation between the process observed in different sites can change with the direction. The contiguity matrix is $\boldsymbol{W} = \boldsymbol{W}^{(1)} + \boldsymbol{W}^{(2)}$, where $\boldsymbol{W}^{(1)} = \left(w_{ij}^{(1)}\right)$ and $\boldsymbol{W}^{(2)} = \left(w_{ij}^{(2)}\right)$ are directional weighted matrices satisfying $w_{ij} = w_{ij}^{(1)} + w_{ij}^{(2)}$, and $w_{ij}^{(1)} = 1$ if $j \in \mathcal{N}_{i1}^*$, $w_{ij}^{(2)} = 1$ if $j \in \mathcal{N}_{i2}^*$, where $\mathcal{N}_{i1}^*$ and $\mathcal{N}_{i2}^*$ are as in Kyung and Ghosh (2010). The vector $\boldsymbol{Y}$ follows a DCAR process if $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \tau^2 \left(\boldsymbol{I} - \delta_1 \boldsymbol{W}^{(1)} - \delta_2 \boldsymbol{W}^{(2)}\right)^{-1} \boldsymbol{D}^{-1}$, $\boldsymbol{D} = \text{diag}(w_{i+})$ and $w_{i+} = \sum_j w_{ij}$. If the columns of $\boldsymbol{X}$ have norm $\sqrt{n}$, $\boldsymbol{R}_{\text{DCAR}}^{-1} = \boldsymbol{C} \boldsymbol{\Sigma}^{-1} \boldsymbol{C}$ with $\boldsymbol{C} = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n)$, and $\sigma_i = \Sigma_{ii}^{1/2}$, then

$$
\begin{aligned}
n_{\text{DCAR}}^* &= \frac{1}{\tau^2 p} \left[\sum_{k=1}^{p} \sum_{i=1}^{n} x_{ik}^2 \sigma_i^2 w_{i+} - \delta_1 \sum_{k=1}^{p} \sum_{i=1}^{n} \sum_{j=1}^{n} x_{ik} x_{jk} \sigma_i \sigma_j w_{ij}^{(1)} w_{i+} \right. \\
&\quad \left. - \delta_2 \sum_{k=1}^{p} \sum_{i=1}^{n} \sum_{j=1}^{n} x_{ik} x_{jk} \sigma_i \sigma_j w_{ij}^{(2)} w_{i+} \right]
\end{aligned}
\tag{A.2}
$$

If $\delta_1 = \delta_2 = \rho$, then the ESS for the DCAR and CAR models coincide.

A.4. ESS for a partitioned normal model

We assume that $\boldsymbol{Y}_1$ is $n \times 1$ and $\boldsymbol{Y}_2$ is $m \times 1$ such that

$$\widetilde{\boldsymbol{Y}} = \begin{pmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \boldsymbol{X}_1\boldsymbol{\beta}_1 \\ \boldsymbol{X}_2\boldsymbol{\beta}_2 \end{pmatrix}, \boldsymbol{R} \right), \qquad \text{with} \quad \boldsymbol{R} = \begin{pmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} \\ \boldsymbol{R}_{12}^\top & \boldsymbol{R}_{22} \end{pmatrix},$$

where $\boldsymbol{X}_i$ is $n \times p_i$ and $\boldsymbol{\beta}_i$ is $p_i \times 1$ for $i = 1, 2$. Suppose that the columns of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ have been normalized by $\sqrt{n}$ and $\sqrt{m}$, respectively. We define

$$\widetilde{\boldsymbol{X}} = \begin{pmatrix} \boldsymbol{X}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{X}_2 \end{pmatrix}, \text{and} \quad \widetilde{\boldsymbol{\beta}} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}.$$

Then,

$$
\begin{aligned}
n^*_{\widetilde{Y}} &= (n+m)\frac{\operatorname{tr}\left(\widetilde{\boldsymbol{X}}^\top \boldsymbol{R}^{-1}\widetilde{\boldsymbol{X}}\right)}{\operatorname{tr}\left(\widetilde{\boldsymbol{X}}^\top \widetilde{\boldsymbol{X}}\right)} \\
&= \frac{(n+m)}{np_1 + mp_2}\left(\operatorname{tr}\left(\boldsymbol{X}_1^\top \boldsymbol{R}^{11}\boldsymbol{X}_1\right) + \operatorname{tr}\left(\boldsymbol{X}_2^\top \boldsymbol{R}^{22}\boldsymbol{X}_2\right)\right),
\end{aligned}
\qquad \text{(A.3)}
$$

where

$$
\begin{aligned}
\boldsymbol{R}^{-1} &= \begin{pmatrix} \boldsymbol{R}^{11} & \boldsymbol{R}^{12} \\ \boldsymbol{R}^{21} & \boldsymbol{R}^{22} \end{pmatrix}, \; \boldsymbol{R}^{11} = \left(\boldsymbol{R}_{11} - \boldsymbol{R}_{12}\boldsymbol{R}_{22}^{-1}\boldsymbol{R}_{12}^\top\right)^{-1}, \\
\boldsymbol{R}^{22} &= \boldsymbol{R}_{22}^{-1} + \boldsymbol{R}_{22}^{-1}\boldsymbol{R}_{12}^\top \boldsymbol{R}^{11}\boldsymbol{R}_{12}\boldsymbol{R}_{22}^{-1}, \; \boldsymbol{R}^{12} = -\boldsymbol{R}^{11}\boldsymbol{R}_{12}\boldsymbol{R}_{22}^{-1}, \\
\boldsymbol{R}^{21} &= \left(\boldsymbol{R}^{12}\right)^\top.
\end{aligned}
$$

Some specific cases are the following:

1. If $\boldsymbol{R}_{12} = 0$, then $\boldsymbol{R}^{11} = \boldsymbol{R}_{11}^{-1}$, $\boldsymbol{R}^{22} = \boldsymbol{R}_{22}^{-1}$, and
   $n^*_{\widetilde{Y}} = \frac{(n+m)}{np_1 + mp_2}\left(p_1 n^*_{Y_1} + p_2 n^*_{Y_2}\right)$, where $n^*_{Y_i} = \operatorname{tr}\left(\boldsymbol{X}_i^\top \boldsymbol{R}_{ii}^{-1}\boldsymbol{X}_i\right)/p_i$, $i = 1, 2$. In particular, if $\boldsymbol{X}_1 = \boldsymbol{X}_2 = \boldsymbol{1}$, we have that $n^*_{\widetilde{Y}} = \operatorname{tr}\left(\boldsymbol{1}^\top \boldsymbol{R}_{11}^{-1}\boldsymbol{1}\right) + \operatorname{tr}\left(\boldsymbol{1}^\top \boldsymbol{R}_{22}^{-1}\boldsymbol{1}\right) = n^*_{Y_1} + n^*_{Y_2}$.
2. If $p_1 = p_2 = p$, then $n^*_{\widetilde{Y}} = \frac{1}{p}\left(\operatorname{tr}\left(\boldsymbol{X}_1^\top \boldsymbol{R}^{11}\boldsymbol{X}_1\right) + \operatorname{tr}\left(\boldsymbol{X}_2^\top \boldsymbol{R}^{22}\boldsymbol{X}_2\right)\right)$.
3. If $n = m$, then $n^*_{\widetilde{Y}} = \frac{2}{p_1 + p_2}\left(\operatorname{tr}\left(\boldsymbol{X}_1^\top \boldsymbol{R}^{11}\boldsymbol{X}_1\right) + \operatorname{tr}\left(\boldsymbol{X}_2^\top \boldsymbol{R}^{22}\boldsymbol{X}_2\right)\right)$.

Appendix B: Proofs

Proof of Proposition 1. Since $\boldsymbol{R}(\boldsymbol{\theta})$ is a correlation matrix, let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be the eigenvalues of $\boldsymbol{R}(\boldsymbol{\theta})$; then, $\lambda_i > 0$, $i = 1, 2, \ldots, n$, and $\lambda_1 + \lambda_2 + \cdots + \lambda_n = \operatorname{tr}(\boldsymbol{R}(\boldsymbol{\theta})) = n$. Thus, the maximum eigenvalue of $\boldsymbol{R}(\boldsymbol{\theta})$, say $\lambda_{(n)}$, satisfies that $\lambda_{(n)} < n$ or $\lambda_{(n)}^{-1} > n^{-1}$. That is, the minimum eigenvalue of $\boldsymbol{R}^{-1}(\boldsymbol{\theta})$ is $\lambda_{(n)}^{-1}$.

Now, assume that $\boldsymbol{x}_j^\top \boldsymbol{x_j} = n$ for all j; then, $\boldsymbol{x}_j^\top \boldsymbol{R}^{-1}(\boldsymbol{\theta})\boldsymbol{x}_j \geq \lambda_{(n)}^{-1}\boldsymbol{x}_j^\top \boldsymbol{x}_j \geq n^{-1}n = 1, \; \forall\, j = 1, \ldots, p$. This implies that

$$n^* = \frac{1}{p}\operatorname{tr}(\boldsymbol{X}^\top \boldsymbol{R}^{-1}(\boldsymbol{\theta})\boldsymbol{X}) = \frac{1}{p}\sum_{j=1}^{p} \boldsymbol{x}_j^\top \boldsymbol{R}^{-1}(\boldsymbol{\theta})\boldsymbol{x}_j > \frac{1}{p}\sum_{j=1}^{p} 1 = 1. \qquad \square$$

Proof of Proposition 2. Consider the intraclass correlation matrix $\boldsymbol{R}(\rho)$ as in Example 1. The eigenvalues of $\boldsymbol{R}(\rho)$ are $\lambda_{\max} = 1 + (n-1)\rho$ with algebraic dimension 1 and $\lambda_{\min} = 1 - \rho$ with algebraic dimension $n - 1$. Denote the normalized eigenvalues of $\boldsymbol{R}(\rho)$ as v_λ such that the vector associated with $\lambda_{\max}$ is $v_{\max} = \boldsymbol{1}$. Assume that the covariates of the model are the eigenvectors associated with $\lambda_{\min}$. Then, $\boldsymbol{v}_\lambda^\top \boldsymbol{R}^m(\rho)\boldsymbol{v}_\lambda = \lambda^m \boldsymbol{v}_\lambda^\top \boldsymbol{v}_\lambda = n\lambda^m$ and

$$n_{\mathrm{IC}}^* = \frac{1}{p}\left(\boldsymbol{1}^\top \boldsymbol{R}^{-1}(\rho)\boldsymbol{1} + \sum_{j=2}^{p} \boldsymbol{v}_{\min_j}^\top \boldsymbol{R}^{-1}(\rho)\boldsymbol{v}_{\min_j}\right) = \frac{n}{p}\left(\frac{1}{\lambda_{\max}} + \frac{p-1}{\lambda_{\min}}\right).$$

Note that $\dfrac{n}{p}\left(\dfrac{1}{\lambda_{\max}} + \dfrac{p-1}{\lambda_{\min}}\right) \leq n \Leftrightarrow \dfrac{\lambda_{\min} + (p-1)\lambda_{\max}}{\lambda_{\max}\lambda_{\min}} \leq p$. Thus, since $\lambda_{\max} - \lambda_{\min} = n\rho$ and $1 - \lambda_{\min} = \rho$, we have that $p \leq \dfrac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max}(1 - \lambda_{\min})} = \dfrac{n}{1 + (n-1)\rho}$. $\qquad \square$

Proof of Proposition 3. Because $Z(\boldsymbol{s}_i)$ is normally distributed, $Y(\boldsymbol{s}_i)$ has a log-normal distribution with

$$\mathbb{E}[Y(\boldsymbol{s}_i)] = e^{\mu_i + 0.5\tilde{\sigma}^2} - \delta_2, \tag{B.1}$$

$$\operatorname{cov}(Y(\boldsymbol{s}_i), Y(\boldsymbol{s}_j)) = e^{\mu_i + \mu_j + \tilde{\sigma}^2}\left(e^{\sigma_{ij}} - 1\right), \tag{B.2}$$

for $i, j = 1, \ldots, n$. For $i = j$, one obtains that

$$\operatorname{cor}(Y(\boldsymbol{s}_i), Y(\boldsymbol{s}_i)) = \frac{e^{2\mu_i + \tilde{\sigma}^2}\left(e^{\tilde{\sigma}^2} - 1\right)}{e^{2\mu_i + \tilde{\sigma}^2}\left(e^{\tilde{\sigma}^2} - 1\right)} = 1,$$

while that for $i \neq j$, using a Taylor expansion for $e^{\sigma_{ij}}$, one obtains

$$\begin{aligned}
\operatorname{cor}(Y(\boldsymbol{s}_i), Y(\boldsymbol{s}_j)) &= \frac{e^{\mu_i + \mu_j + \tilde{\sigma}^2}\left(e^{\sigma_{ij}} - 1\right)}{e^{\mu_i + \mu_j + \tilde{\sigma}^2}\left(e^{\tilde{\sigma}^2} - 1\right)} \\
&= \frac{\sigma_{ij}}{e^{\tilde{\sigma}^2} - 1}\left(1 + \frac{\sigma_{ij}}{2} + \cdots + \frac{\sigma_{ij}^{m-1}}{m!}\right) + \frac{1}{e^{\tilde{\sigma}^2} - 1}o\left(|\tilde{\sigma}^2|^m\right) \\
&= \frac{\tilde{\sigma}^2}{e^{\tilde{\sigma}^2} - 1}\frac{\sigma_{ij}}{\tilde{\sigma}^2}k_{ij} + o\left(|\tilde{\sigma}^2|^m\right) \\
&= c \cdot (\boldsymbol{R_Z})_{ij} \cdot k_{ij} + o\left(|\tilde{\sigma}^2|^m\right),
\end{aligned}$$

where $\text{cor}(Z(\boldsymbol{s}_i), Z(\boldsymbol{s}_j)) = (\boldsymbol{R_Z})_{ij} = \dfrac{\sigma_{ij}}{\tilde{\sigma}^2}$, $k_{ij} = 1 + \dfrac{\sigma_{ij}}{2} + \cdots + \dfrac{\sigma_{ij}^{m-1}}{m!}$, $i \neq j$, $k_{ii} = 1$, and $c = \dfrac{\tilde{\sigma}^2}{\exp(\tilde{\sigma}^2) - 1}$. Furthermore, if $m \to \infty$, we have that

$$\text{cor}(Y(\boldsymbol{s}_i), Y(\boldsymbol{s}_j)) = \frac{e^{\sigma_{ij}} - 1}{e^{\tilde{\sigma}^2} - 1} = \frac{c}{\tilde{\sigma}^2}(e^{\sigma_{ij}} - 1). \qquad \square$$

Proof of Theorem 2.

Lemma 1. *Let $P_m(y) = \sum_{i=0}^{m} a_i y^i$, be a polynomial of degree $m \in \mathbb{N}$. Then, for $d < \infty$,*

$$\int_{\mathbb{R}^d} P_m(\|\boldsymbol{x}\|) \exp(-\|\boldsymbol{x}\|) d\boldsymbol{x} = \frac{2\pi^{d/2}}{\Gamma(d/2)} \sum_{i=0}^{m} a_i \Gamma(i + d) < \infty,$$

where $\Gamma(\cdot)$ is the gamma function.

Proof. Applying polar coordinates on $\mathbb{R}^d$, with $r = \|\boldsymbol{x}\|$, one has

$$\begin{aligned}
\int_{\mathbb{R}^d} P_m(\|\boldsymbol{x}\|) \exp(-\|\boldsymbol{x}\|) d\boldsymbol{x} &= c_d \int_0^{\infty} P_m(r) \exp(-r) r^{d-1} dr \\
&= c_d \int_0^{\infty} \sum_{i=0}^{m} a_i r^i \exp(-r) r^{d-1} dr \\
&= \frac{2\pi^{d/2}}{\Gamma(d/2)} \sum_{i=0}^{m} a_i \Gamma(d + i),
\end{aligned}$$

where $c_d = \dfrac{2\pi^{d/2}}{\Gamma(d/2)}$ represents the surface area of the unit sphere S^{d-1}. $\qquad \square$

Lemma 2. *For $d < \infty$ and $m \geq 0$ fixed,*

$$\int_{\mathbb{R}^d} \|\boldsymbol{x}\|^m \exp(-\|\boldsymbol{x}\|^2) d\boldsymbol{x} = \frac{\pi^{d/2} \Gamma\left(\frac{d+m}{2}\right)}{2\Gamma\left(\frac{d}{2}\right)} < \infty.$$

Proof. The result is obtained by emulating the proof of Lemma 1. $\qquad \square$

Proof of Theorem 2 (continued). We assume that $\boldsymbol{X}$ is suitable chosen such that $\lim(\boldsymbol{X}^{\top}\boldsymbol{X})^{-1} = 0$ holds, and hence, condition (iv) in Theorem 1 holds. The proof of conditions $(i) - (iii)$ is divided into two parts.

Case 1: $m < \infty$ fixed. To prove conditions (i) and (ii), we use the summability of the covariances. Precisely, define $\sigma_{\boldsymbol{k}} = \sigma_m(\|\boldsymbol{k}\|; \sigma^2, \phi)$. Then,

$$\sigma_{\boldsymbol{k}} = \sigma^2 P_m \left(\frac{\|\boldsymbol{k}\|}{\phi} \right) \exp \left(-\frac{\|\boldsymbol{k}\|}{\phi} \right),$$

$$\sigma_{\boldsymbol{k},\sigma^2} = P_m \left(\frac{\|\boldsymbol{k}\|}{\phi} \right) \exp \left(-\frac{\|\boldsymbol{k}\|}{\phi} \right),$$

$$\sigma_{\boldsymbol{k},\phi} = \sigma^2 \left(\frac{\|\boldsymbol{k}\|}{\phi^2} \right) \left[P_m \left(\frac{\|\boldsymbol{k}\|}{\phi} \right) - P'_m \left(\frac{\|\boldsymbol{k}\|}{\phi} \right) \right] \exp \left(-\frac{\|\boldsymbol{k}\|}{\phi} \right),$$

$$\sigma_{\boldsymbol{k},\sigma^2\sigma^2} = 0,$$

$$\sigma_{\boldsymbol{k},\sigma^2\phi} = \left(\frac{\|\boldsymbol{k}\|}{\phi^2} \right) \left[P_m \left(\frac{\|\boldsymbol{k}\|}{\phi} \right) - P'_m \left(\frac{\|\boldsymbol{k}\|}{\phi} \right) \right] \exp \left(-\frac{\|\boldsymbol{k}\|}{\phi} \right),$$

$$\sigma_{\boldsymbol{k},\phi\phi} = \sigma^2 \left\{ \left(\frac{\|\boldsymbol{k}\|}{\phi^2} \right)^2 \left[P_m \left(\frac{\|\boldsymbol{k}\|}{\phi} \right) - 2P'_m \left(\frac{\|\boldsymbol{k}\|}{\phi} \right) + P''_m \left(\frac{\|\boldsymbol{k}\|}{\phi} \right) \right] \right.$$
$$\left. - \left(\frac{2\|\boldsymbol{k}\|}{\phi^3} \right) \left[P_m \left(\frac{\|\boldsymbol{k}\|}{\phi} \right) - P'_m \left(\frac{\|\boldsymbol{k}\|}{\phi} \right) \right] \right\} \exp \left(-\frac{\|\boldsymbol{k}\|}{\phi} \right),$$

where $P'_m(x) = \sum_{i=0}^{m-1} (i+1)a_{i+1}x^i$ and $P''_m(x) = \sum_{i=0}^{m-2} (i+2)(i+1)a_{i+2}x^i$. Moreover,

$$\sum_{\boldsymbol{k}\in\mathbb{Z}^d} |\sigma_{\boldsymbol{k}}| \leq \int_{\mathbb{R}^d} \sigma^2 P_m \left(\frac{\|\boldsymbol{x}\|}{\phi} \right) \exp \left(-\frac{\|\boldsymbol{x}\|}{\phi} \right) d\boldsymbol{x}$$

$$= \sigma^2 \phi \int_{\mathbb{R}^d} P_m(\|\boldsymbol{z}\|) \exp(-\|\boldsymbol{z}\|) d\boldsymbol{z}$$

$$< \infty,$$

$$\sum_{\boldsymbol{k}\in\mathbb{Z}^d} |\sigma_{\boldsymbol{k},\sigma^2}| \leq \int_{\mathbb{R}^d} P_m \left(\frac{\|\boldsymbol{x}\|}{\phi} \right) \exp \left(-\frac{\|\boldsymbol{x}\|}{\phi} \right) d\boldsymbol{x}$$

$$= \phi \int_{\mathbb{R}^d} P_m(\|\boldsymbol{z}\|) \exp(-\|\boldsymbol{z}\|) d\boldsymbol{z}$$

$$< \infty,$$

$$\sum_{\boldsymbol{k}\in\mathbb{Z}^d} |\sigma_{\boldsymbol{k},\phi}| \leq \int_{\mathbb{R}^d} \frac{\sigma^2}{\phi} \left(\frac{\|\boldsymbol{x}\|}{\phi} \right) \left| P_m \left(\frac{\|\boldsymbol{x}\|}{\phi} \right) - P'_m \left(\frac{\|\boldsymbol{x}\|}{\phi} \right) \right| \exp \left(-\frac{\|\boldsymbol{x}\|}{\phi} \right) d\boldsymbol{x}$$

$$= \sigma^2 \int_{\mathbb{R}^d} \|\boldsymbol{z}\| \left| P_m(\|\boldsymbol{z}\|) - P'_m(\|\boldsymbol{z}\|) \right| \exp \left(-\|\boldsymbol{z}\| \right) d\boldsymbol{z}$$

$$\leq \sigma^2 \int_{\mathbb{R}^d} \|\boldsymbol{z}\| \left[P_m(\|\boldsymbol{z}\|) + P'_m(\|\boldsymbol{z}\|) \right] \exp(-\|\boldsymbol{z}\|) d\boldsymbol{z}$$

$$< \infty,$$

$$\sum_{\boldsymbol{k}\in\mathbb{Z}^d} |\sigma_{\boldsymbol{k},\sigma^2\sigma^2}| = 0,$$

$$\sum_{\boldsymbol{k}\in\mathbb{Z}^d} |\sigma_{\boldsymbol{k},\sigma^2\phi}| \leq \int_{\mathbb{R}^d} \frac{1}{\phi}\left(\frac{\|\boldsymbol{x}\|}{\phi}\right)\left|P_m\left(\frac{\|\boldsymbol{x}\|}{\phi}\right) - P_m'\left(\frac{\|\boldsymbol{x}\|}{\phi}\right)\right|\exp\left(-\frac{\|\boldsymbol{x}\|}{\phi}\right) d\boldsymbol{x}$$

$$= \int_{\mathbb{R}^d} \|\boldsymbol{z}\|\,|P_m(\|\boldsymbol{z}\|) - P_m'(\|\boldsymbol{z}\|)|\exp\left(-\|\boldsymbol{z}\|\right) d\boldsymbol{z}$$

$$\leq \int_{\mathbb{R}^d} \|\boldsymbol{z}\|\,[P_m(\|\boldsymbol{z}\|) + P_m'(\|\boldsymbol{z}\|)]\exp\left(-\|\boldsymbol{z}\|\right) d\boldsymbol{z}$$

$$< \infty,$$

$$\sum_{\boldsymbol{k}\in\mathbb{Z}^d} |\sigma_{\boldsymbol{k},\phi\phi}| \leq \int_{\mathbb{R}^d} \frac{\sigma^2}{\phi^2}\left|\left(\frac{\|\boldsymbol{x}\|}{\phi}\right)^2\left[P_m\left(\frac{\|\boldsymbol{x}\|}{\phi}\right) - 2P_m'\left(\frac{\|\boldsymbol{x}\|}{\phi}\right) + P_m''\left(\frac{\|\boldsymbol{x}\|}{\phi}\right)\right]\right.$$

$$\left. -2\left(\frac{\|\boldsymbol{x}\|}{\phi}\right)\left[P_m\left(\frac{\|\boldsymbol{x}\|}{\phi}\right) - P_m'\left(\frac{\|\boldsymbol{x}\|}{\phi}\right)\right]\right|\exp\left(-\frac{\|\boldsymbol{x}\|}{\phi}\right) d\boldsymbol{x}$$

$$= \frac{\sigma^2}{\phi}\int_{\mathbb{R}^d}\left|\|\boldsymbol{z}\|^2\left[P_m\left(\|\boldsymbol{z}\|\right) - 2P_m'\left(\|\boldsymbol{z}\|\right) + P_m''\left(\|\boldsymbol{z}\|\right)\right]\right.$$

$$-2\|\boldsymbol{z}\|\left[P_m\left(\|\boldsymbol{z}\|\right) - P_m'\left(\|\boldsymbol{z}\|\right)\right]|\exp\left(-\|\boldsymbol{z}\|\right) d\boldsymbol{z}$$

$$\leq \frac{\sigma^2}{\phi}\left\{\int_{\mathbb{R}^d}\|\boldsymbol{z}\|^2\left[P_m\left(\|\boldsymbol{z}\|\right) + 2P_m'\left(\|\boldsymbol{z}\|\right) + P_m''\left(\|\boldsymbol{z}\|\right)\right]\exp\left(-\|\boldsymbol{z}\|\right) d\boldsymbol{z}\right.$$

$$\left. +2\int_{\mathbb{R}^d}\|\boldsymbol{z}\|\left[P_m\left(\|\boldsymbol{z}\|\right) + P_m'\left(\|\boldsymbol{z}\|\right)\right]\exp\left(-\|\boldsymbol{z}\|\right) d\boldsymbol{z}\right\}$$

$$< \infty.$$

All of the above expressions have been obtained by using Lemma 1. Now, we establish condition (iii). The Fisher information matrix is given by

$$\boldsymbol{B_\theta} = \frac{1}{2}\begin{pmatrix} \mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{\sigma^2}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{\sigma^2}\right) & \mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{\sigma^2}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\phi\right) \\ \mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{\sigma^2}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\phi\right) & \mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\phi\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\phi\right) \end{pmatrix},$$

where $\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{\sigma^2} = \boldsymbol{I}/\sigma^2$ and $\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\phi = \frac{1}{\phi^2}\left(\boldsymbol{Q}\odot\boldsymbol{P}\right)^{-1}\left(\boldsymbol{Q}\odot\boldsymbol{H}\odot\left[\boldsymbol{P}-\boldsymbol{P}'\right]\right)$. Since $\boldsymbol{Q}\odot\boldsymbol{P}$ is symmetric and positive definite, it follows that

$$\mathrm{tr}\left(\left(\boldsymbol{Q}\odot\boldsymbol{P}\right)^{-1}\left(\boldsymbol{Q}\odot\boldsymbol{H}\odot\left[\boldsymbol{P}-\boldsymbol{P}'\right]\right)\right)$$
$$= \mathrm{tr}\left(\left(\boldsymbol{Q}\odot\boldsymbol{P}\right)^{-1/2}\left(\boldsymbol{Q}\odot\boldsymbol{H}\odot\left[\boldsymbol{P}-\boldsymbol{P}'\right]\right)\left(\boldsymbol{Q}\odot\boldsymbol{P}\right)^{-1/2}\right)$$
$$= \mathrm{tr}\left(\boldsymbol{V}\right),$$

and

$$\boldsymbol{B_\theta} = \frac{1}{2}\begin{pmatrix} \dfrac{n}{\sigma^4} & \dfrac{\mathrm{tr}\left(\boldsymbol{V}\right)}{\sigma^2\phi^2} \\ \dfrac{\mathrm{tr}\left(\boldsymbol{V}\right)}{\sigma^2\phi^2} & \dfrac{\mathrm{tr}\left(\boldsymbol{V}^2\right)}{\phi^4} \end{pmatrix}.$$

Now, $\boldsymbol{A} = \lim_{n \to \infty} \boldsymbol{A}_n$, where

$$
\boldsymbol{A}_n = \begin{pmatrix} 1 & \dfrac{\operatorname{tr}(\boldsymbol{V})}{\sqrt{n \operatorname{tr}(\boldsymbol{V}^2)}} \\ \dfrac{\operatorname{tr}(\boldsymbol{V})}{\sqrt{n \operatorname{tr}(\boldsymbol{V}^2)}} & 1 \end{pmatrix}. \tag{B.3}
$$

By algebra of limits, $\boldsymbol{A}$ is nonsingular if $\lim_{n \to \infty} \det(\boldsymbol{A}_n) \neq 0$. This is satisfied because

$$
|\boldsymbol{A}_n| = 1 - \frac{\operatorname{tr}^2(\boldsymbol{V})}{n \operatorname{tr}(\boldsymbol{V}^2)} = \frac{n \operatorname{tr}(\boldsymbol{V}^2) - \operatorname{tr}^2(\boldsymbol{V})}{n \operatorname{tr}(\boldsymbol{V}^2)} = \frac{\sum_{i=1}^n (\lambda_i - \bar{\lambda})^2}{\sum_{i=1}^n \lambda_i^2},
$$

where λ_i are the eigenvalues of $\boldsymbol{V}$ and $\bar{\lambda}$ is the arithmetic average of λs, which are different real numbers because $\boldsymbol{V}$ is symmetric and $\boldsymbol{V} \neq c\boldsymbol{I}$. Additionally, $|\lambda_i| < C$ for all i, which implies that $\bar{\lambda} < \infty$; thus, $\lim_{n \to \infty} |\boldsymbol{A}_n| = 1$.

Case 2: $m \to \infty$. In this case,

$$
\sigma_\infty(h; \sigma^2, \phi) = \sigma^2 \exp\left(-\left(\frac{h}{\phi}\right)^2\right),
$$

and

$$
\boldsymbol{B}_{\boldsymbol{\theta}} = \frac{1}{2} \begin{pmatrix} \dfrac{n}{\sigma^4} & \dfrac{2 \operatorname{tr}(\widetilde{\boldsymbol{V}})}{\sigma^2 \phi^3} \\ \dfrac{2 \operatorname{tr}(\widetilde{\boldsymbol{V}})}{\sigma^2 \phi^3} & \dfrac{4 \operatorname{tr}(\widetilde{\boldsymbol{V}}^2)}{\phi^6} \end{pmatrix},
$$

where $\widetilde{\boldsymbol{V}} = \boldsymbol{Q}^{-1/2}(\boldsymbol{Q} \odot \boldsymbol{H} \odot \boldsymbol{H})\boldsymbol{Q}^{-1/2}$, $\boldsymbol{H} = (h_{ij})$, $\boldsymbol{Q} = (q_{ij})$ with $h_{ij} = \|\boldsymbol{s}_i - \boldsymbol{s}_j\|$ and $q_{ij} = \exp(-(h_{ij}/\phi)^2)$, for all $i, j = 1, \ldots, n$.

Conditions (i) and (ii) are proved in a similar way because Lemma 2 implies that $\sigma_{\boldsymbol{k}}, \sigma_{\boldsymbol{k}, \sigma^2}, \sigma_{\boldsymbol{k}, \phi}, \sigma_{\boldsymbol{k}, \sigma^2 \sigma^2}, \sigma_{\boldsymbol{k}, \phi\phi}$ and $\sigma_{\boldsymbol{k}, \sigma^2 \phi}$ are absolutely summable, where

$$
\sigma_{\boldsymbol{k}, \sigma^2} = \exp\left(-\left(\frac{\|\boldsymbol{k}\|}{\phi}\right)^2\right),
$$

$$
\sigma_{\boldsymbol{k}, \phi} = \frac{2\sigma^2 \|\boldsymbol{k}\|^2}{\phi^3} \exp\left(-\left(\frac{\|\boldsymbol{k}\|}{\phi}\right)^2\right),
$$

$$
\sigma_{\boldsymbol{k}, \sigma^2 \sigma^2} = 0,
$$

$$
\sigma_{\boldsymbol{k}, \phi\phi} = \frac{2\sigma^2 \|\boldsymbol{k}\|^2 (2 - 3\phi^2)}{\phi^6} \exp\left(-\left(\frac{\|\boldsymbol{k}\|}{\phi}\right)^2\right),
$$

$$
\sigma_{\boldsymbol{k}, \sigma^2 \phi} = \frac{2\|\boldsymbol{k}\|^2}{\phi^3} \exp\left(-\left(\frac{\|\boldsymbol{k}\|}{\phi}\right)^2\right).
$$

Condition (iii) is granted by replacing $\boldsymbol{V}$ with $\widetilde{\boldsymbol{V}}$, which is also symmetric, $\widetilde{\boldsymbol{V}} \neq c\mathbf{I}$, and the eigenvalues of $\widetilde{\boldsymbol{V}}$ are all distinct real numbers and bounded. $\qquad \square$

Proof of Theorem 3. Because of the invariance property of the ML estimator of $\boldsymbol{\theta}$, $g(\widehat{\boldsymbol{\theta}})$ is the ML estimator of $g(\boldsymbol{\theta})$. Since $g(\boldsymbol{\theta})$ is a continuous and differentiable function for all $\boldsymbol{\theta} \in \Theta$, the result follows by direct application of the delta method to $g(\boldsymbol{\theta})$. $\square$

Proof of Theorem 4. Denote $\boldsymbol{\Sigma}^{(1)}$ and $\boldsymbol{\Sigma}^{(2)}$ as the covariance matrices obtained from $\sigma_1(\boldsymbol{s}_i, \boldsymbol{s}_j, \boldsymbol{\theta}_1)$ and $\sigma_2(\boldsymbol{s}_i, \boldsymbol{s}_j, \boldsymbol{\theta}_2)$, respectively, for the vector $\boldsymbol{Y}$. Then, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{(1)} + \boldsymbol{\Sigma}^{(2)}$. Denote by $\alpha_1 \leq \cdots \leq \alpha_n$, $|\alpha_1^i| \leq \cdots \leq |\alpha_n^i|$, and $|\alpha_1^{ij}| \leq \cdots \leq |\alpha_n^{ij}|$ the ordered eigenvalues of $\boldsymbol{\Sigma}^{(1)}$, $\boldsymbol{\Sigma}_i^{(1)}$, and $\boldsymbol{\Sigma}_{ij}^{(1)}$, respectively, for $i, j = 1, \ldots, q_1$, and similarly we denote $\gamma_1 \leq \cdots \leq \gamma_n$, $|\gamma_1^i| \leq \cdots \leq |\gamma_n^i|$, and $|\gamma_1^{ij}| \leq \cdots \leq |\gamma_n^{ij}|$ the ordered eigenvalues of $\boldsymbol{\Sigma}^{(2)}$, $\boldsymbol{\Sigma}_i^{(2)}$, and $\boldsymbol{\Sigma}_{ij}^{(2)}$, for $i, j = q_1 + 1, \ldots, q_1 + q_2$, and $\lambda_1 \leq \cdots \leq \lambda_n$, $|\lambda_1^i| \leq \cdots \leq |\lambda_n^i|$, and $|\lambda_1^{ij}| \leq \cdots \leq |\lambda_n^{ij}|$ the ordered eigenvalues of $\boldsymbol{\Sigma}$, $\boldsymbol{\Sigma}_i$, and $\boldsymbol{\Sigma}_{ij}$, for $i, j = 1, \ldots, q_1 + q_2$.

Proof of Condition (i). Condition (i) is granted for λ_n by applying the triangular inequality for the spectral norm denoted by $||\cdot||_e$. Then

$$
\begin{aligned}
\lim_{n\to\infty} \lambda_n &= \lim_{n\to\infty} \|\boldsymbol{\Sigma}\|_e \\
&= \lim_{n\to\infty} \|\boldsymbol{\Sigma}^{(1)} + \boldsymbol{\Sigma}^{(2)}\|_e \\
&\leq \lim_{n\to\infty} \|\boldsymbol{\Sigma}^{(1)}\|_e + \lim_{n\to\infty} \|\boldsymbol{\Sigma}^{(2)}\|_e \\
&= \lim_{n\to\infty} \alpha_n + \lim_{n\to\infty} \gamma_n \\
&< \infty,
\end{aligned}
$$

and similarly for λ_n^i and λ_n^{ii}.

Proof of Condition (ii). It is straightforward to see that $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_i^{(1)}$ for $i = 1, \ldots, q_1$ and $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_i^{(2)}$ for $i = q_1 + 1, \ldots, q_1 + q_2$. Since $\|\boldsymbol{\Sigma}_i^{(1)}\|^{-2} = \mathcal{O}(n^{-\frac{1}{2}-\delta_1})$ for some $\delta_1 > 0$, and $\|\boldsymbol{\Sigma}_i^{(2)}\|^{-2} = \mathcal{O}(n^{-\frac{1}{2}-\delta_2})$ for some $\delta_2 > 0$, It is sufficient to set $\delta = \min\{\delta_1, \delta_2\} > 0$ such that $\|\boldsymbol{\Sigma}_i\|^{-2} = \mathcal{O}(n^{-\frac{1}{2}-\delta})$ for all $i = 1, \ldots, q_1 + q_2$. $\square$

References

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

BANERJEE, S., CARLIN B. and GELFAND A. (2004) *Hierarchical Modeling and Analysis for Spatial Data*, Boca Raton: Chapman Hall/CRC. MR3362184

BERGER, J., BAYARRI, M. J. and PERICCHI, L. R. (2014). The effective sample size. *Econometric Letters* **33** 197–2014. MR3170846

BOX, G.E.P. (1954a). Some theorems on quadratic forms applied in the study of analysis of variance problems: I. Effect of inequality of variance in the one way classification. *Ann. Math. Statist.*, **25**, 290–302. MR0061787

BOX, G.E.P. (1954b). Some theorems on quadratic forms applied in the study of analysis of variance problems: II. Effects of inequality of variance and of correlation between errors in the two way classification. *Ann. Math. Statist.*, **25**, 484–498. MR0064361

BOX, G. E. P. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B*, **26**, 211–252. MR0192611

CLIFFORD, P., RICHARDSON, S. and HÉMON, D. (1989). Assessing the significance of the correlation between two spatial processes. *Biometrics*, **45**, 123–134. MR0999441

COGLEY, J. G. (1999). Effective sample size for glacier mass balance. *Geogr. Ann. A* **81** 497–507.

CRESSIE, N. (1993). *Statistics for Spatial Data.* New York: Wiley. MR1239641

CRESSIE, N. AND LAHIRI, S. N. (1993). Asymptotic distribution of REML estimators. *Journal of Multivariate Analysis*, **45**, 217–233. MR1221918

CRUJEIRAS, R. M. and VAN KEILEGOM, I. (2010). Least squares estimation of nonlinear spatial trends. *Computational Statistics and Data Analysis*, **54**, 452–465. MR2756439

DALE, M. R. T. and FORTIN, M-J. (2009). Spatial autocorrelation and statistical tests: some solutions. *Journal of Agricultural, Biological, and Environmental Statistics*, **14**, 188–206. MR2750624

de GRUIJTER, J. J. and ter BRAAK, C. J. F. (1990). Model-free estimation from spatial samples: A reappraisal of classical sampling theory. *Mathematical Geology*, **22**, 407–415. MR1047605

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

DUTILLEUL, P., PELLETIER, B. and ALPARGU, G. (2008). Modified F tests for assessing the multiple correlation between one spatial process and several others. *Journal of Statistical Planning and Inference*, **138**, 1402–1415. MR2388019

FAES, C., MOLENBERGHS, G., AERTS, M., VERBEKE, G. and KENWARD, M. (2009). The effective sample size and an alternative small-sample degrees-of-freedom method. *The American Statistician* **63** 389–399. MR2751758

GELFAND A. and VOUATSOU, P. (2003). Peoper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics*, **4**, 11–25.

GELFAND, A. E., DIGGLE, P. J., FUENTES, M. and GUTTORP, P. (2010). *Handbook of Spatial Statistics*, Boca Raton, Fl: Chapman & Hall/CRC. MR2730964

GOLUB, G. H. and PEREYRA, V. (1973). The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate. *SIAM Journal of Numerical Analysis*, **22**, 413–431. MR0336980

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

GRIFFITH, D. (2008). Geographic sampling of urban soils for contaminant

mapping: how many samples and from where. *Environ. Geochem. Health.* **30** 495–509.

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

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

KUTNER, M., NACHTSHEIM, C., NETER, J. and LI, W. (2004) *Applied Linear Statistical Models*, Homewood, IL: McGraw-Hill/Irwin.

KYUNG, M. and GHOSH, S. K. (2010). Maximum likelihood estimation for directional conditionally autoregressive models. *Journal of Statistical Planning and Inference* **140** 3160–3179. MR2659845

LEDOIT, O. and WOLF, M. (2002). Some Hypothesis Tests for the covariance matrix when the dimension is large compared to the sample size. *The Annals of Statistics*, **30**, 1081–1102. MR1926169

LENTH, R. V. (2001). Some practical guidelines for effective sample size determination. *The American Statistician*, **55**, 187–193. MR1963395

MARDIA, K. and MARSHALL, R. (1984). Maximum likelihood of models for residual covariance in spatial regression. *Biometrika* **71** 135–146. MR0738334

RICHARDSON, S. (1990). Some remarks on the testing of association between spatial processes. In: Griffith, D. (Ed.), Spatial Statistics: Past, Present, and Future. *Institute of Mathematical Geography*, Ann Arbor, MI, 277?309. MR1076010

SCHABENBERGER, O. and GOTWAY, C. A. (2005). *Statistical Methods for Spatial Data Analysis.* Boca Raton, FL: Chapman & Hall/CRC. MR2134116

SOO, YUH-WEN. and BATES, D. M. (1992). Loosely coupled nonlinear least squares. *Computational Statistics and Data Analysis*, **14**, 249–259.

STEIN, M. (1999). *Interpolation of Spatial Data: Some Theory of Kriging.* New York: Springer. MR1697409

TELFORD, R. J. and BIRKS H. J. B. (2009). Evaluation of transfer functions in spatially structured environments. *Quat. Sci. Rev.* **28** 1309–1316.

THIÉBAUX, H. J. and ZWIERS, F. W. (1984). Interpretation and estimation of effective sample size. *J. Climate Appl. Meteor.* **23**, 800–811.

VALLEJOS, R. and OSORIO, F., (2014). Effective sample size for spatial process models. *Spat. Stat.* **9** 66–92. MR3326832