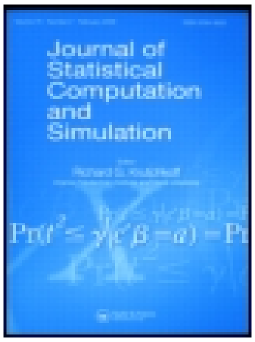

Journal of Statistical Computation and Simulation

ISSN: 0094-9655 (Print) 1563-5163 (Online) Journal homepage: http://www.tandfonline.com/loi/gscs20

On the effective geographic sample size

Jonathan Acosta, Ronny Vallejos & Daniel Griffth

To cite this article: Jonathan Acosta, Ronny Vallejos & Daniel Griffth (2018): On the effective geographic sample size, Journal of Statistical Computation and Simulation, DOI: 10.1080/00949655.2018.1428977

To link to this article: https://doi.org/10.1080/00949655.2018.1428977

Published online: 30 Jan 2018.

JOURNAL OF STATISTICAL COMPUTATION AND SIMULATION, 2018
https://doi.org/10.1080/00949655.2018.1428977

Taylor & Francis
Taylor & Francis Group

Check for updates

On the effective geographic sample size

Jonathan Acosta[a,b], Ronny Vallejos[a] and Daniel Griffth[c]

[a]Departamento de Matemática, Universidad Técnica Federico Santa María, Valparaíso, Chile; [b]Insituto de Estadística, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile; [c]School of Economic, Political and Policy Sciences, University of Texas at Dallas, Richardson, TX, USA

ABSTRACT

Griffith [Effective geographic sample size in the presence of spatial autocorrelation. Ann Assoc Amer Geogr. 2005;95:740–760] suggested a formula to compute the effective sample size, say n^*, for georeferenced data. In this article, we provide mathematical support that enhances the use of this definition in practice. We prove that $n^* \in [1, n]$ and that n^* is increasing in n. We also prove the asymptotic normality of the maximum likelihood estimate of n^* for an increasing domain sampling framework. Asymptotic normality leads to an approximate hypothesis testing that establishes whether redundant information exists in a sample.

ARTICLE HISTORY
Received 23 June 2017
Accepted 14 January 2018

KEYWORDS
Covariance function; Gaussian process; spatial regression model; increasing domain; hypothesis testing

1. Introduction

This note provides theoretical properties of the effective geographic sample size defined in [1]. This coefficient is based on the variance inflation factor (VIF) of a linear model of the form $\boldsymbol{Y} = \mu\mathbf{1} + \boldsymbol{\epsilon}$, $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{V})$, where the vectors $\boldsymbol{Y}$, $\mathbf{1}$ and $\boldsymbol{\epsilon}$ are of size n, $\mathbf{1}$ is a vector with entries equal to one, and $\boldsymbol{V}$ is a matrix of size $n \times n$, which describes the spatial dependence. The effective geographic sample size was defined as follows:

$$n^* = n\frac{\mathrm{tr}(\boldsymbol{V})}{\mathbf{1}^\top \boldsymbol{V}\mathbf{1}}, \tag{1}$$

where $\mathrm{tr}(\cdot)$ and $(\cdot)^\top$ denote the trace and transpose operators, respectively. Subsequently, Griffith [2] applied Equation (1) to soil samples in Syracuse. Telford and Birks [3] argued that the estimation of the effective sample size merits further investigation. Vallejos and Osorio [4] proposed an alternative definition based on the Fisher information quantity about μ, and derived a function of $\boldsymbol{V}^{-1}$. Although both definitions are based on different arguments, for several patterned correlation matrices, both coefficients coincide. Alternatively, Faes et al. [5] proposed another way of computing the effective sample size based on the inverse of the Fisher information quantity, which is applicable to linear mixed models with replicates. In a Bayesian context, the effective sample size is required to extend

CONTACT Ronny Vallejos ronny.vallejos@usm.cl Departamento de Matemática, Universidad Técnica Federico Santa María, Avenida España 1680, Valparaíso, Chile

the Bayesian information criteria (BIC) [6]. Some applications involving the computation of the effective sample size can be found in [7–9]. Griffith [10] established qualitative geographic sample size in the presence of autocorrelation [11,12].

Definition (1) was accompanied by the derivation of the effective sample size for several spatial models, Monte Carlo simulations, and applications of the methodology to real data, including remotely sensed image examples. However, some mathematical aspects related to definition (1) have not yet been addressed. For example, it is well known that $n^* = 1$ if $\boldsymbol{V} = \mathbf{1}\mathbf{1}^\top$ (perfect correlation), and $n^* = n$ if $\boldsymbol{V} = \mathbf{I}_n$ (uncorrelated), where $\mathbf{I}_n$ is the matrix identity of size $n \times n$. Here, we find conditions on matrix $\boldsymbol{V}$ for which $n^* \in [1, n]$. We also characterize the behaviour of n^* for the extremes of the interval $[1, n]$. In addition, we prove that n^* is an increasing function of n. The estimation problem is addressed via maximum likelihood (ML) under an increasing domain sampling framework. Illustrative examples accompany the discussion of the limiting results, including some cases where the asymptotic variance has a closed form. Asymptotic normality leads to an approximate hypothesis test that establishes whether redundant information exists in a sample due to the effect of spatial autocorrelation. Although there are apparent similarities between the asymptotics for n^* and the large sample inferences of the (principal component analysis) PCA eigenvalues of a correlation matrix [13], the later technique is applicable to normal models with replicates, which will not be considered in this work.

The rest of the paper is organized as follows. In Section 2, we state some preliminary facts. Later in Section 3, we establish the asymptotic normality of the ML estimate of n^*. In Section 4, we present the hypothesis testing for $\mathrm{H}_0 : n^* = n_0$. In Section 5, a numerical example is presented to illustrate the applicability of the methodology. Section 6 concludes the paper with a short discussion.

2. Preliminary results

Denote the ijth element of $\boldsymbol{V}$ as v_{ij}. From definition (1), the following properties for n^* can be obtained.

Theorem 2.1: *Let $\boldsymbol{V}$ be a positive semi-definite matrix with $v_{ij} \geq 0$. Then, $n^* \in [1, n]$.*

Proof: By definition, $\operatorname{tr}(\boldsymbol{V}) = \sum_i v_{ii}$ and $\mathbf{1}^\top \boldsymbol{V}\mathbf{1} = \sum_i \sum_j v_{ij}$; i.e. $\mathbf{1}^\top \boldsymbol{V}\mathbf{1} = \operatorname{tr}(\boldsymbol{V}) + \sum_{i\neq j} v_{ij}$. Because $v_{ij} \geq 0$, $\operatorname{tr}(\boldsymbol{V}) \leq \mathbf{1}^\top \boldsymbol{V}\mathbf{1}$. Now, $\boldsymbol{z}^\top \boldsymbol{V}\boldsymbol{z} \geq 0$, for all $\boldsymbol{z} \in \mathbb{R}^n$, because $\boldsymbol{V}$ is positive semi-definite. Let $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ be the canonical vectors in $\mathbb{R}^n$. Let $\boldsymbol{z}_{ij} = \boldsymbol{e}_i - \boldsymbol{e}_j$, for all $i, j = 1, \dots, n$. Then $\boldsymbol{z}_{ij}^\top \boldsymbol{V}\boldsymbol{z}_{ij} = v_{ii} + v_{jj} - 2v_{ij} \geq 0$. Summing on i and j yields

$$\sum_i \sum_j v_{ii} + v_{jj} - 2v_{ij} \geq 0.$$

This result implies that $2n \operatorname{tr}(\boldsymbol{V}) - 2\sum_i \sum_j v_{ij} \geq 0$. Then $n\operatorname{tr}(\boldsymbol{V}) \geq \mathbf{1}^\top \boldsymbol{V}\mathbf{1}$. Thus, $\operatorname{tr}(\boldsymbol{V}) \leq \mathbf{1}^\top \boldsymbol{V}\mathbf{1} \leq n\operatorname{tr}(\boldsymbol{V})$, yielding

$$1 = \frac{n \operatorname{tr}(\boldsymbol{V})}{n \operatorname{tr}(\boldsymbol{V})} \leq n^* = \frac{n \operatorname{tr}(\boldsymbol{V})}{\mathbf{1}^\top \boldsymbol{V}\mathbf{1}} \leq \frac{n \operatorname{tr}(\boldsymbol{V})}{\operatorname{tr}(\boldsymbol{V})} = n. \qquad \blacksquare$$

Proposition 2.2: *Let $\boldsymbol{V}$ be a positive semi-definite matrix with $v_{ij} \geq 0$. Then $n^* = n$ if only if $v_{ij} = 0$, for all $i \neq j$.*

Proof: ($\Leftarrow$) Suppose that $v_{ij} = 0$. Then $\mathrm{tr}(\boldsymbol{V}) = \sum_i v_{ii} = \sum_i \sum_j v_{ij} = \mathbf{1}^\top \boldsymbol{V}\mathbf{1}$. Replacing $\mathrm{tr}(\boldsymbol{V})$ in Equation (1), with this result we obtain $n^* = n$.

($\Rightarrow$) Suppose that $n^* = n$. This implies that $\mathbf{1}^\top \boldsymbol{V}\mathbf{1} = \mathrm{tr}(\boldsymbol{V})$. Moreover, if $\mathbf{1}^\top \boldsymbol{V}\mathbf{1} = \mathrm{tr}(\boldsymbol{V}) + \sum_{i\neq j} v_{ij}$, then $\sum_{i\neq j} v_{ij} = 0$. Because $v_{ij} \geq 0$, we have that $v_{ij} = 0$. ■

Proposition 2.3: *Let $\boldsymbol{V}$ be a positive semi-definite matrix with $v_{ij} \geq 0$. Then $n^* = 1$ if only if $\boldsymbol{V} = k\mathbf{1}\mathbf{1}^\top$, where k is a constant.*

Proof: ($\Leftarrow$) Suppose that $\boldsymbol{V} = k\mathbf{1}\mathbf{1}^\top$. Replacing $\boldsymbol{V}$ in Equation (1), with this result we have that

$$n^* = \frac{n\,\mathrm{tr}(\boldsymbol{k}\mathbf{1}\mathbf{1}^\top)}{\mathbf{1}^\top(k\mathbf{1}\mathbf{1}^\top)\mathbf{1}} = 1.$$

($\Rightarrow$) Suppose that $n^* = 1$. Then $\mathbf{1}^\top \boldsymbol{V}\mathbf{1} = n\,\mathrm{tr}(\boldsymbol{V})$. This happens only if $\boldsymbol{V} = k\mathbf{1}\mathbf{1}^\top$, because $v_{ij} \geq 0$ and $\boldsymbol{V}$ is a positive semi-definite matrix. ■

Lemma 2.4: *If $a, b, c > 0$, then $(a + c)/(b + c) > a/b$.*

Proposition 2.5: *Let $\boldsymbol{V}$ be a positive semi-definite matrix with $v_{ij} \geq 0$. Then n^* is increasing in n.*

Proof: It is enough to prove that $n^*_{n+1} \geq n^*_n$.

Assume that $\boldsymbol{V}_n$ and $\boldsymbol{V}_{n+1}$ are the covariance matrices associated with n^*_n and n^*_{n+1}, respectively, and consider the following partition:

$$\boldsymbol{V}_{n+1} = \begin{bmatrix} \boldsymbol{V}_n & \boldsymbol{v} \\ \boldsymbol{v}^\top & v_{n+1,n+1} \end{bmatrix}$$

where $v_{i,n+1}$ is the ith element of $\boldsymbol{v}$. Then $\mathrm{tr}(\boldsymbol{V}_{n+1}) = \mathrm{tr}(\boldsymbol{V}_n) + v_{n+1,n+1}$, and $\mathbf{1}^\top \boldsymbol{V}_{n+1}\mathbf{1} = \mathbf{1}^\top \boldsymbol{V}_n\mathbf{1} + 2\mathbf{1}^\top \boldsymbol{v} + v_{n+1,n+1}$. Substituting this result into Equation (1) yields

$$\begin{aligned} n^*_{n+1} &= \frac{(n+1)\mathrm{tr}(\boldsymbol{V}_{n+1})}{\mathbf{1}^\top \boldsymbol{V}_{n+1}\mathbf{1}} \\ &= \frac{n\mathrm{tr}(\boldsymbol{V}_n) + \mathrm{tr}(\boldsymbol{V}_n) + n v_{n+1,n+1} + v_{n+1,n+1}}{\mathbf{1}^\top \boldsymbol{V}_n\mathbf{1} + 2\mathbf{1}^\top \boldsymbol{v} + v_{n+1,n+1}}. \end{aligned} \tag{2}$$

Denoting $\boldsymbol{e}_i$, $i = 1, \ldots, n+1$, as the canonical vectors of $\mathbb{R}^{n+1}$, and defining $\boldsymbol{z}_i = \boldsymbol{e}_i - \boldsymbol{e}_{n+1}$ for $i = 1, \ldots, n$, we have that $\boldsymbol{z}_i^\top \boldsymbol{V}_{n+1}\boldsymbol{z}_i = v_{ii} + v_{n+1,n+1} - 2v_{i,n+1} \geq 0$, for all $i = 1, \ldots, n$. This result yields $\mathrm{tr}(\boldsymbol{V}_n) + n v_{n+1,n+1} - 2\mathbf{1}^\top \boldsymbol{v} \geq 0$. Substituting this outcome into

Equation (2), yields

$$\begin{aligned} n^*_{n+1} &= \frac{n\,\mathrm{tr}(\boldsymbol{V}_n) + \mathrm{tr}(\boldsymbol{V}_n) + n v_{n+1,n+1} + v_{n+1,n+1}}{\mathbf{1}^\top \boldsymbol{V}_n \mathbf{1} + 2\mathbf{1}^\top \boldsymbol{v} + v_{n+1,n+1}} \\ &\geq \frac{n\,\mathrm{tr}(\boldsymbol{V}_n) + 2\mathbf{1}^\top \boldsymbol{v} + v_{n+1,n+1}}{\mathbf{1}^\top \boldsymbol{V}_n \mathbf{1} + 2\mathbf{1}^\top \boldsymbol{v} + v_{n+1,n+1}} \\ &> \frac{n\,\mathrm{tr}(\boldsymbol{V}_n)}{\mathbf{1}^\top \boldsymbol{V}_n \mathbf{1}} \\ &= n^*_n, \end{aligned}$$

where the last inequality is obtained by using Lemma 2.4. ■

Proposition 2.6: *Let $\boldsymbol{V}$ be a positive semi-definite matrix with $v_{ij} \geq 0$, and assume that all elements of the diagonal of $\boldsymbol{V}$ are equal. Suppose that $\boldsymbol{V} = \boldsymbol{V}(\rho)$, and $\rho \in [0,1]$ is a parameter such that the correlation increases when ρ increases. Then, n^* is decreasing in ρ.*

Proof: Without loss of generality, suppose that $\boldsymbol{V}(\rho) = \boldsymbol{R}(\rho)$, where $\boldsymbol{R}(\rho)$ is a correlation matrix and $\rho \in [0,1]$. Then $n^*(\rho) = n^2/\mathbf{1}^\top \boldsymbol{R}(\rho)\mathbf{1}$. Now, let $r_{ij}(\rho)$ be the ijth element of $\boldsymbol{R}(\rho)$, and let $\rho_1, \rho_2 \in [0,1]$ be such that $\rho_1 < \rho_2$ and $r_{ij}(\rho_1) \leq r_{ij}(\rho_2)$. This implies that $\mathbf{1}^\top \boldsymbol{R}(\rho_1)\mathbf{1} \leq \mathbf{1}^\top \boldsymbol{R}(\rho_2)\mathbf{1}$. Hence,

$$\frac{1}{\mathbf{1}^\top \boldsymbol{R}(\rho_2)\mathbf{1}} \leq \frac{1}{\mathbf{1}^\top \boldsymbol{R}(\rho_2)\mathbf{1}} \Leftrightarrow n^*(\rho_2) \leq n^*(\rho_1).$$ ■

3. Estimation and asymptotics under increasing domain sampling

Let $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in D \subset \mathbb{R}^d\}$ be a Gaussian random field such that $Y(\cdot)$ is observed on $D_n \subset D$. Denote $\boldsymbol{Y} = (Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n))^\top$, and assume that $\mathbb{E}[\boldsymbol{Y}] = \mu\mathbf{1}$ and $\mathrm{cov}(Y(\boldsymbol{t}), Y(\boldsymbol{s})) = \sigma(\boldsymbol{t}, \boldsymbol{s}; \boldsymbol{\theta})$, where $\mu \in \mathbb{R}^1$ and $\boldsymbol{\theta} \in \Theta$, with Θ an open subset of $\mathbb{R}^q$. Let $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ be the covariance matrix of $\boldsymbol{Y}$ such that the ijth element of $\boldsymbol{\Sigma}$ is $\sigma_{ij} = \sigma(\boldsymbol{s}_i, \boldsymbol{s}_j; \boldsymbol{\theta})$. The estimation of $\boldsymbol{\theta}$ and μ can be made with ML techniques, maximizing the log-likelihood function

$$L = L(\mu, \boldsymbol{\theta}) = \mathrm{Conts} - \tfrac{1}{2}\ln \boldsymbol{\Sigma}| - \tfrac{1}{2}(\boldsymbol{Y} - \mu\mathbf{1})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \mu\mathbf{1}). \tag{3}$$

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

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

be the gradient vector and Hessian matrix, respectively, obtained from Equation (3). Let $\boldsymbol{B}_n = -\mathbb{E}[\boldsymbol{L}_n^{(2)}]$ be the Fisher information matrix with respect to μ and $\boldsymbol{\theta}$. Then, $\boldsymbol{B}_n = \mathrm{diag}(B_\mu, \boldsymbol{B}_{\boldsymbol{\theta}})$, where $B_\mu = -\mathbb{E}[L_{\mu\mu}]$ 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 [14] considered a Gaussian random field as before, but $\mathbb{E}[\boldsymbol{Y}] = \boldsymbol{X}\boldsymbol{\beta}$, and 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 3.1: *Let $\lambda_1 \leq \cdots \leq \lambda_n$ be the eigenvalues of $\boldsymbol{\Sigma}$, and let those of $\boldsymbol{\Sigma}_i = \partial\boldsymbol{\Sigma}/\partial\theta_i$ and $\boldsymbol{\Sigma}_{ij} = \partial^2\boldsymbol{\Sigma}/\partial\theta_i\partial\theta_j$ 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 $i, j = 1, \ldots, q$. 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* $i, j = 1, \ldots, q$, $a_{ij} = \lim[t_{ij}/(t_{ii}t_{jj})^{1/2}]$ *exists, where*

$$t_{ij} = \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j)$$

and $\boldsymbol{A} = (a_{ij})$ *is nonsingular; and,*
(iv) $\lim(\boldsymbol{X}^\top\boldsymbol{X})^{-1} = 0$.

Then, $\boldsymbol{B}_n^{-1/2}(\widehat{\boldsymbol{\beta}}^\top, \widehat{\boldsymbol{\theta}}^\top)^\top \xrightarrow{\mathcal{L}} \mathcal{N}((\boldsymbol{\beta}^\top, \boldsymbol{\theta}^\top)^\top, \boldsymbol{I})$ *as* $n \to \infty$, *in an increasing domain sampling sense.*

In this case, $\boldsymbol{X} = \mathbf{1}$ and $\boldsymbol{\beta} = \mu \in \mathbb{R}^1$; therefore, the condition (iv) is satisfied because $\lim(\mathbf{1}^\top\mathbf{1})^{-1} = \lim n^{-1} = 0$. For the application of Theorem 3.1, when data points are sampled on $\mathbb{R}^d$, D_n is assumed to be a non-random set satisfying $\|\boldsymbol{s} - \boldsymbol{t}\| \geq \gamma > 0$ for all $\boldsymbol{s}, \boldsymbol{t} \in D_n$. 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 $i, j = 1, \ldots, q$, 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 $\sigma_{\boldsymbol{k},ij} = \partial^2\sigma_{\boldsymbol{k}}/\partial\theta_i\partial\theta_j$ [14]. These conditions were established by Acosta and Vallejos [15] for a particular parametrization of the Matérn covariance function with smoothing parameter $\nu = m + \frac{1}{2}$, $m \in \mathbb{N}_0$, given by

$$\sigma_m(h; \sigma^2, \phi) = \sigma^2 \exp\left(-\frac{h}{\phi}\right) P_m\left(\frac{h}{\phi}\right), \tag{4}$$

where $h = \|\boldsymbol{s} - \boldsymbol{t}\|$, $\boldsymbol{s}, \boldsymbol{t} \in D_n$, and $P_m(x) = \sum_{i=0}^m a_i x^i$, with $a_i = (2^i/i!)\binom{m}{i}/\binom{2m}{i}$.

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

$$\boldsymbol{B}_{\boldsymbol{\theta}} = \frac{1}{2\sigma^4\phi^4}\begin{pmatrix} \phi^4 n & \sigma^2\phi^2\mathrm{tr}(\boldsymbol{U}) \\ \sigma^2\phi^2\mathrm{tr}(\boldsymbol{U}) & \sigma^4\mathrm{tr}(\boldsymbol{U}^2) \end{pmatrix}, \tag{5}$$

$\boldsymbol{U} = (\boldsymbol{Q} \odot \boldsymbol{P})^{-1/2}(\boldsymbol{Q} \odot \boldsymbol{H} \odot [\boldsymbol{P} - \dot{\boldsymbol{P}}])(\boldsymbol{Q} \odot \boldsymbol{P})^{-1/2}$, $\boldsymbol{H} = (h_{ij})$ 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 $p'_{ij} = \mathrm{d}P_m(h_{ij}/\phi)/\mathrm{d}\phi$, for all $i, j = 1, \ldots, n$. Here $\odot$ denotes the Hadamard product.

Alternatively, Kyung and Ghosh [16] proved that the conditions of Theorem 3.1 hold for (conditional autoregressive) CAR and (directional autoregressive) DCAR processes. Precisely, let $\boldsymbol{Y} \sim \mathcal{N}_n(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$ be a DCAR process with

$$\boldsymbol{\Sigma} = \tau^2(\mathbf{I} - \delta_1\widetilde{\boldsymbol{W}}^{(1)} - \delta_2\widetilde{\boldsymbol{W}}^{(2)})^{-1}\boldsymbol{D}^{-1}, \tag{6}$$

where $\boldsymbol{D} = \mathrm{diag}(w_{i+})$, $w_{i+} = \sum_j w_{ij}$, $\widetilde{\boldsymbol{W}}^{(1)}$ and $\widetilde{\boldsymbol{W}}^{(2)}$ are the normalized directional contiguity matrices satisfying $\widetilde{\boldsymbol{W}} = \widetilde{\boldsymbol{W}}^{(1)} + \widetilde{\boldsymbol{W}}^{(2)}$. For $\boldsymbol{\theta} = (\tau^2, \delta_1, \delta_2)$, the Fisher information matrix is

$$\boldsymbol{B}_{\theta} = \frac{1}{2\tau^4}\begin{pmatrix} n & \mathrm{tr}(\boldsymbol{G}_1) & \mathrm{tr}(\boldsymbol{G}_2) \\ \mathrm{tr}(\boldsymbol{G}_1) & \mathrm{tr}(\boldsymbol{G}_1^2) & \mathrm{tr}(\boldsymbol{G}_1\boldsymbol{G}_2) \\ \mathrm{tr}(\boldsymbol{G}_2) & \mathrm{tr}(\boldsymbol{G}_1\boldsymbol{G}_2) & \mathrm{tr}(\boldsymbol{G}_2^2) \end{pmatrix}, \tag{7}$$

where $\boldsymbol{G}_k = \widetilde{\boldsymbol{W}}^{(k)}\boldsymbol{D}\boldsymbol{\Sigma}$, $k=1,2$. For a CAR process, $\rho = \delta_1 + \delta_2$, and the Fisher information matrix of $\boldsymbol{\theta} = (\tau^2, \rho)$ is

$$\boldsymbol{B}_{\theta} = \frac{1}{2\tau^4}\begin{pmatrix} n & \mathrm{tr}(\boldsymbol{G}) \\ \mathrm{tr}(\boldsymbol{G}) & \mathrm{tr}(\boldsymbol{G}^2) \end{pmatrix}, \tag{8}$$

where $\boldsymbol{G} = \widetilde{\boldsymbol{W}}\boldsymbol{D}\boldsymbol{\Sigma}$.

Next, we establish the asymptotic normality of the ML estimate of n^*.

Theorem 3.2: *Suppose that* $\mathrm{var}[\boldsymbol{Y}] = \boldsymbol{\Sigma}(\boldsymbol{\theta})$. *Let* $\widehat{\boldsymbol{\theta}}$ *be the ML estimator of* $\boldsymbol{\theta}$ *as in Theorem* 3.1. *If* $\boldsymbol{B}_{\boldsymbol{\theta}}^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{\mathcal{L}} \mathcal{N}_q(\boldsymbol{0}, \boldsymbol{I})$, *as* $n \to \infty$, *and* $n^* = g(\boldsymbol{\theta}) = n\mathrm{tr}(\boldsymbol{V}(\boldsymbol{\theta}))/\boldsymbol{1}^\top \boldsymbol{V}(\boldsymbol{\theta})\boldsymbol{1}$, *then*

$$(\nabla^\top g(\boldsymbol{\theta})\boldsymbol{B}_{\boldsymbol{\theta}}^{-1}\nabla g(\boldsymbol{\theta}))^{-1/2}(g(\widehat{\boldsymbol{\theta}}) - g(\boldsymbol{\theta})) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1), \quad \text{as } n \to \infty.$$

Proof: $g(\widehat{\boldsymbol{\theta}})$ is the ML estimator of $g(\boldsymbol{\theta})$, and $g(\boldsymbol{\theta})$ is a differentiable function such that $\nabla g(\boldsymbol{\theta}) \neq 0$ for all $\boldsymbol{\theta} \in \Theta$. Then the mean value theorem implies that

$$g(\widehat{\boldsymbol{\theta}}) - g(\boldsymbol{\theta}) = \nabla^\top g(\widetilde{\boldsymbol{\theta}})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}),$$

where $\widetilde{\boldsymbol{\theta}} = \boldsymbol{\theta} + \alpha(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})$, and $0 < \alpha < 1$. By consistency of $\widetilde{\boldsymbol{\theta}}$ and Slutsky's theorem

$$\begin{aligned}&(\nabla^\top g(\boldsymbol{\theta})\boldsymbol{B}_{\boldsymbol{\theta}}^{-1}\nabla g(\boldsymbol{\theta}))^{-1/2}(g(\widehat{\boldsymbol{\theta}}) - g(\boldsymbol{\theta})) \\ &\quad = (\nabla^\top g(\boldsymbol{\theta})\boldsymbol{B}_{\boldsymbol{\theta}}^{-1}\nabla g(\boldsymbol{\theta}))^{-1/2}\nabla^\top g(\widetilde{\boldsymbol{\theta}})\boldsymbol{B}_{\boldsymbol{\theta}}^{-1/2}\boldsymbol{B}_{\boldsymbol{\theta}}^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).\end{aligned}$$ ■

We remark based on Theorem 3.2 that $\sigma_g^2 = \mathrm{var}[g(\widehat{\boldsymbol{\theta}})] = \nabla^\top g(\boldsymbol{\theta})\boldsymbol{B}_{\boldsymbol{\theta}}^{-1}\nabla g(\boldsymbol{\theta})$. The following three examples illustrate the computation of σ_g^2 for a Gaussian random field with the exponential covariance function, and for CAR and DCAR processes.

Example 3.3: Let $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{Z}^2\}$ be a stationary Gaussian process having an exponential covariance function, obtained by putting $m=0$ in (4). Here $\boldsymbol{\theta} = (\sigma^2, \phi)$, $P_m(h_{ij}/\phi) = 1$, $dP_m(h_{ij}/\phi)/d\phi = 0$, and $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{V}(\phi)$ is the covariance matrix of $\boldsymbol{Y} = (Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n))^\top$, where the ijth element of $\boldsymbol{V}$ is $v_{ij} = \exp(-\|\boldsymbol{s}_i - \boldsymbol{s}_j\|/\phi)$. Then the

Fisher information matrix of $\boldsymbol{\theta}$ is

$$\boldsymbol{B}_{\boldsymbol{\theta}} = \frac{1}{2\sigma^4\phi^4}\begin{pmatrix} \phi^4 n & \sigma^2\phi^2\mathrm{tr}(\boldsymbol{V}^{-1}(\boldsymbol{V}\odot\boldsymbol{H})) \\ \sigma^2\phi^2\mathrm{tr}(\boldsymbol{V}^{-1}(\boldsymbol{V}\odot\boldsymbol{H})) & \sigma^4\mathrm{tr}((\boldsymbol{V}^{-1}(\boldsymbol{V}\odot\boldsymbol{H}))^2)\end{pmatrix}, \tag{9}$$

which is equivalent to Equation (5) with $\boldsymbol{U} = \boldsymbol{V}^{-1/2}(\boldsymbol{V}\odot\boldsymbol{H})\boldsymbol{V}^{-1/2}$. In this case, $n^* = g(\phi) = n^2/(\mathbf{1}^\top \boldsymbol{V}(\phi)\mathbf{1})$, and hence $\sigma_g^2 = [g'(\phi)]^2\mathrm{var}[\widehat{\phi}]$, where

$$g'(\phi) = -\frac{n^2\mathbf{1}^\top(\boldsymbol{V}\odot\boldsymbol{H})\mathbf{1}}{\phi^2(\mathbf{1}^\top\boldsymbol{V}\mathbf{1})^2},$$

and $\mathrm{var}[\widehat{\phi}]$ is the element associated with entry $(2,2)$ of the inverse of Equation (9); i.e. $\mathrm{var}[\widehat{\phi}] = \phi^4/2nS^2$, where S^2 is given by

$$S^2 = \frac{1}{n}\mathrm{tr}((\boldsymbol{V}^{-1}(\boldsymbol{V}\odot\boldsymbol{H}))^2) - \left(\frac{1}{n}\mathrm{tr}(\boldsymbol{V}^{-1}(\boldsymbol{V}\odot\boldsymbol{H}))\right)^2.$$

Therefore,

$$\sigma_g^2 = \frac{n^3(\mathbf{1}^\top(\boldsymbol{V}\odot\boldsymbol{H})\mathbf{1})^2}{2S^2(\mathbf{1}^\top\boldsymbol{V}\mathbf{1})^4}. \tag{10}$$

Because the ML estimate of the effective sample size is increasing in n, its variance is also. To illustrate the consistency of n^*, we plotted the ratio n^*/n versus n for the exponential covariance function (see Figure 1). We observe that for both sampling schemes the variance of the ratio n^*/n rapidly goes to zero as n increases. This is in agreement with, and expected, for consistent estimates.

Example 3.4: Let $\boldsymbol{Y} \sim \mathcal{N}_n(\mu\mathbf{1}, \boldsymbol{\Sigma})$ be a CAR process, where $\boldsymbol{\Sigma} = \tau^2\boldsymbol{V}(\rho)$ with $\boldsymbol{V}(\rho) = (\mathbf{I} - \rho\widetilde{\boldsymbol{W}})^{-1}\boldsymbol{D}^{-1}$, and $\widetilde{\boldsymbol{W}}$ and $\boldsymbol{D}^{-1}$ as in (6). Then

$$n^* = g(\rho) = \frac{n\mathrm{tr}((\mathbf{I}-\rho\widetilde{\boldsymbol{W}})^{-1}\boldsymbol{D}^{-1})}{\mathbf{1}^\top((\mathbf{I}-\rho\widetilde{\boldsymbol{W}})^{-1}\boldsymbol{D}^{-1})\mathbf{1}}.$$

Moreover,

$$g'(\rho) = \frac{n\mathrm{tr}(\boldsymbol{V}\boldsymbol{D}\widetilde{\boldsymbol{W}}\boldsymbol{V})\mathbf{1}^\top\boldsymbol{V}\mathbf{1} - n\mathrm{tr}(\boldsymbol{V})\mathbf{1}^\top(\boldsymbol{V}\boldsymbol{D}\widetilde{\boldsymbol{W}}\boldsymbol{V})\mathbf{1}}{(\mathbf{1}^\top\boldsymbol{V}\mathbf{1})^2}. \tag{11}$$

Because $\boldsymbol{\theta} = (\tau^2, \rho)$, and $\boldsymbol{B}_\theta$ is given by Equation (8),

$$\mathrm{var}[\widehat{\rho}] = \frac{2n}{n\mathrm{tr}((\widetilde{\boldsymbol{W}}\boldsymbol{D}\boldsymbol{\Sigma})^2) - \mathrm{tr}(\widetilde{\boldsymbol{W}}\boldsymbol{D}\boldsymbol{\Sigma})^2}.$$

Finally,

$$\sigma_g^2 = \frac{2n[g'(\rho)]^2}{n\mathrm{tr}((\widetilde{\boldsymbol{W}}\boldsymbol{D}\boldsymbol{V})^2) - \mathrm{tr}^2(\widetilde{\boldsymbol{W}}\boldsymbol{D}\boldsymbol{V})},$$

where $g'(\rho)$ is given by Equation (11).

Example 3.5: Let $\boldsymbol{Y} \sim \mathcal{N}_n(\mu\mathbf{1}, \boldsymbol{\Sigma})$ be a DCAR process with $\boldsymbol{\Sigma} = \tau^2\boldsymbol{V}(\delta_!, \delta_2)$, and $\boldsymbol{V}(\delta_1, \delta_2) = (\mathbf{I} - \delta_1\widetilde{\boldsymbol{W}}^{(1)} - \delta_2\widetilde{\boldsymbol{W}}^{(2)})^{-1}\boldsymbol{D}^{-1}$, where $\widetilde{\boldsymbol{W}}^{(1)}$, $\widetilde{\boldsymbol{W}}^{(2)}$, and $\boldsymbol{D}$ are as in Equation (6).

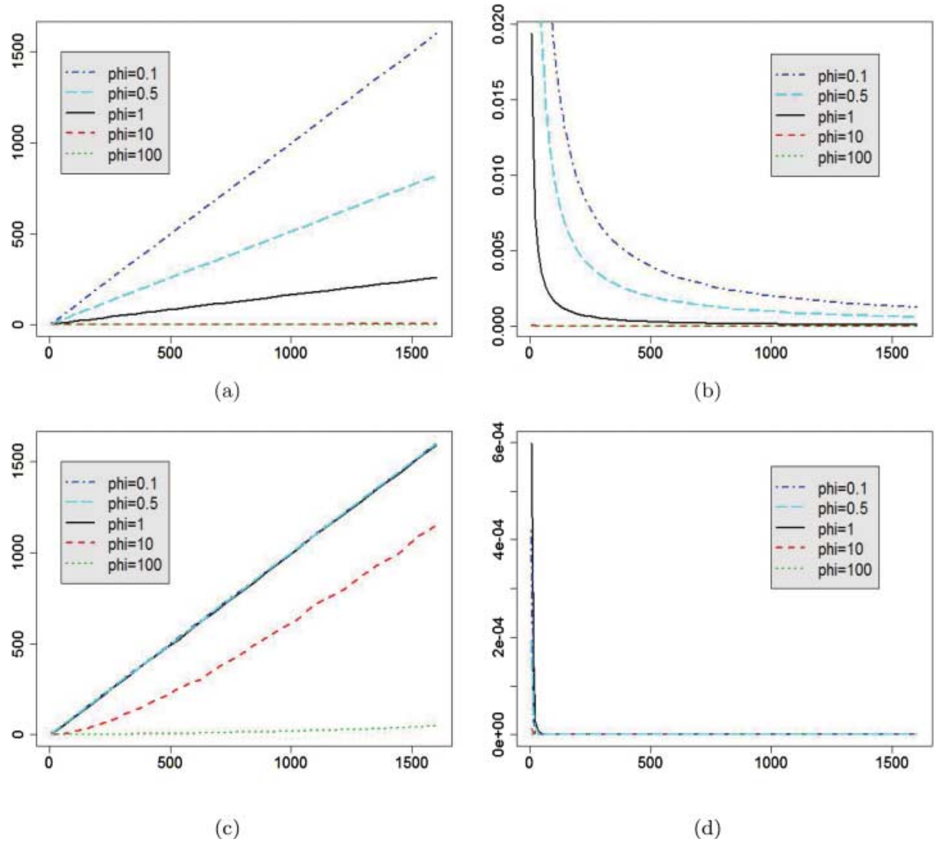

Figure 1. (a) n^* versus n for a regular grid; (b) var(n^*/n) versus n for a regular grid; (c) n^* versus n for a non-regular grid; (d) var(n^*/n) versus n for a non-regular grid. All figures were generated for an exponential covariance function with $\sigma^2 = 1$, and $\phi = \{0.1, 0.5, 1, 10, 100\}$.

In this case, $n^* = g(\delta_1, \delta_2) = n\text{tr}(\boldsymbol{V}(\delta_1, \delta_2))/\mathbf{1}^\top \boldsymbol{V}(\delta_1, \delta_2)\mathbf{1}$, and $\nabla g(\delta_1, \delta_2) = (g_{\delta_1}, g_{\delta_2})$, where

$$g_{\delta_k} = \frac{n\text{tr}(\boldsymbol{V}\boldsymbol{D}\widetilde{\boldsymbol{W}}^{(k)}\boldsymbol{V})\mathbf{1}^\top \boldsymbol{V}\mathbf{1} - n\text{tr}(\boldsymbol{V})\mathbf{1}^\top(\boldsymbol{V}\boldsymbol{D}\widetilde{\boldsymbol{W}}^{(k)}\boldsymbol{V})\mathbf{1}}{(\mathbf{1}^\top \boldsymbol{V}\mathbf{1})^2}, \tag{12}$$

with $k=1,2$. Because $\boldsymbol{\theta} = (\tau^2, \delta_1, \delta_2)$, $\boldsymbol{B_\theta}$ is given by Equation (7). Here, $\text{var}[\widehat{\tau}^2, \widehat{\delta}_1, \widehat{\delta}_2] = \boldsymbol{B}_{\boldsymbol{\theta}}^{-1}$, where

$$\boldsymbol{B}_{\boldsymbol{\theta}}^{-1} = \frac{2\tau^4 \boldsymbol{M}}{a}, \quad \boldsymbol{M} = \begin{pmatrix} M_{11} & \boldsymbol{M}_{12} \\ \boldsymbol{M}_{12}^\top & \boldsymbol{M}_{22} \end{pmatrix},$$

and $\text{var}[\widehat{\delta}_1, \widehat{\delta}_2] = 2\tau^4 \boldsymbol{M}_{22}/a$, with $a = n\text{tr}(\boldsymbol{G}_1^2)\text{tr}(\boldsymbol{G}_2^2) + 2\text{tr}(\boldsymbol{G}_1)\text{tr}(\boldsymbol{G}_2)\text{tr}(\boldsymbol{G}_1\boldsymbol{G}_2) - n\text{tr}^2(\boldsymbol{G}_1\boldsymbol{G}_2) - \text{tr}(\boldsymbol{G}_1^2)\text{tr}^2(\boldsymbol{G}_2) - \text{tr}(\boldsymbol{G}_2^2)\text{tr}^2(\boldsymbol{G}_1)$, and

$$\boldsymbol{M}_{22} = \begin{pmatrix} n\text{tr}(\boldsymbol{G}_2^2) - \text{tr}^2(\boldsymbol{G}_2) & \text{tr}(\boldsymbol{G}_1)\text{tr}(\boldsymbol{G}_2) - n\text{tr}(\boldsymbol{G}_1\boldsymbol{G}_2) \\ \text{tr}(\boldsymbol{G}_1)\text{tr}(\boldsymbol{G}_2) - n\text{tr}(\boldsymbol{G}_1\boldsymbol{G}_2) & n\text{tr}(\boldsymbol{G}_1^2) - \text{tr}^2(\boldsymbol{G}_1) \end{pmatrix}.$$

Consequently,

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

where

$$\begin{aligned} b &= g_{\delta_1}^2[n\mathrm{tr}(\boldsymbol{G}_2^2) - \mathrm{tr}^2(\boldsymbol{G}_2)] + g_{\delta_2}^2[n\mathrm{tr}(\boldsymbol{G}_1^2) - \mathrm{tr}^2(\boldsymbol{G}_1)] \\ &\quad - 2g_{\delta_1}g_{\delta_2}[\mathrm{tr}(\boldsymbol{G}_1)\mathrm{tr}(\boldsymbol{G}_2) - n\mathrm{tr}(\boldsymbol{G}_1\boldsymbol{G}_2)], \end{aligned}$$

with g_{δ_1} and g_{δ_2} as in (12), with $k = 1,2$, respectively.

Mardia [17] introduced an example to illustrate when the conditions of Theorem 3.1 are not satisfied. Indeed, the spherical covariance function

$$\sigma(\boldsymbol{h}, \alpha) = \begin{cases} 1 - \dfrac{3}{2}\dfrac{|\boldsymbol{h}|}{\alpha} + \dfrac{1}{2}\dfrac{|\boldsymbol{h}|^3}{\alpha^3}, & |\boldsymbol{h}| < \alpha \\ 0, & \text{otherwise}, \end{cases}$$

commonly used in geostatistics is not twice diferentiable. Then the standard error computed from the information matrix may not be valid for this scheme.

4. Hypothesis testing

As a consequence of the limiting distribution of $g(\widehat{\boldsymbol{\theta}})$ established in Theorem 3.2, an approximate hypothesis test for $g(\boldsymbol{\theta})$ is constructed. Consider the null hypothesis

$$\mathrm{H}_0 : g(\boldsymbol{\theta}) = n_0, \tag{13}$$

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 $1 < n_0 \leq n$. Proposition 2.2 implies that $n^* = n \Leftrightarrow \boldsymbol{\Sigma} = \sigma^2\mathrm{diag}(\boldsymbol{V})$. Then, when $n_0 = n$, the hypothesis testing problem is relevant because it leads to the case when $\boldsymbol{\Sigma} = \sigma^2\mathrm{diag}(\boldsymbol{V})$. For the Matérn covariance function, the parameter ϕ controls the range of spatial association such that if $\phi = 0$, then $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}$. Thus, the hypothesis test for ϕ, $\mathrm{H}_0 : \phi = 0$ versus $\mathrm{H}_1 : \phi > 0$, is equivalent to $\mathrm{H}_0 : g(\boldsymbol{\theta}) = n$ 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, $\mathcal{C}_\alpha = \{g(\widehat{\boldsymbol{\theta}}) : |g(\widehat{\boldsymbol{\theta}}) - n_0| > z_{\alpha/2}\sqrt{\sigma_g^2}\}$, $\mathcal{C}_\alpha = \{g(\widehat{\boldsymbol{\theta}}) : g(\widehat{\boldsymbol{\theta}}) - n_0 < z_{1-\alpha}\sqrt{\sigma_g^2}\}$, and $\mathcal{C}_\alpha = \{g(\widehat{\boldsymbol{\theta}}) : g(\widehat{\boldsymbol{\theta}}) - n_0 > z_\alpha\sqrt{\sigma_g^2}\}$, where z_α denotes the upper quantile of the standard normal distribution, and $\sigma_g^2 = \mathrm{var}(g(\widehat{\boldsymbol{\theta}}))$.

The applicability of the hypothesis testing problem (13) relies on the feasibility of computing the asymptotic variance σ_g^2. This computation depends on the sampling scheme designed for collecting a spatial dataset, as illustrated by the following examples

Example 4.1: Consider a stationary Gaussian spatial process $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{Z}^2\}$ with the Matérn covariance function defined in (4). Then

$$n^* = \frac{n^2}{\boldsymbol{1}^\top\boldsymbol{V}(\phi)\boldsymbol{1}} = g(\phi),$$

with $\boldsymbol{V}(\phi) = \boldsymbol{Q} \odot \boldsymbol{P}$ such that $\boldsymbol{Q}$ and $\boldsymbol{P}$ are as in Equation (5). To specify the hypothesis testing about ϕ it is enough to provide the asymptotic variance $\sigma_g^2 = (g'(\phi))^2\mathrm{var}(\widehat{\phi})$. In this case

$$g'(\phi) = \frac{-n^2}{\mathbf{1}^\top \boldsymbol{V}(\phi)\mathbf{1}} \frac{\mathbf{1}^\top(\boldsymbol{Q} \odot \boldsymbol{H}(\boldsymbol{P} - \dot{\boldsymbol{P}}))\mathbf{1}}{\phi^2},$$

$$\text{and} \quad \text{var}(\widehat{\phi}) = \frac{2n\phi^4}{n\text{tr}(\boldsymbol{U}^2) - \text{tr}(\boldsymbol{U})^2} = \frac{2\phi^4}{nS^2},$$

with

$$S^2 = \frac{1}{n}\text{tr}(\boldsymbol{U}^2) - \left(\frac{1}{n}\text{tr}(\boldsymbol{U})\right)^2.$$

Thus,

$$\sigma_g^2 = \frac{2n^3(\mathbf{1}^\top(\boldsymbol{Q} \odot \boldsymbol{H}(\boldsymbol{P} - \dot{\boldsymbol{P}}))\mathbf{1})^2}{S^2(\mathbf{1}^\top \boldsymbol{V}\mathbf{1})^2}.$$

If $m = 0$, the Matérn covariance function reduces to the exponential case. As a consequence, $\boldsymbol{P} = \mathbf{1}\mathbf{1}^\top$, $\dot{\boldsymbol{P}} = \mathbf{0}$, $\boldsymbol{V} = \boldsymbol{Q}$ y $\boldsymbol{U} = \boldsymbol{V}^{-1}(\boldsymbol{V} \odot \boldsymbol{H})$, recovering the results obtained in Example 3.3.

The following example illustrates how explicit the calculations can be for the exponential covariance function when the sampling points are located on a one-dimensional grid.

Example 4.2: Let $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{Z}^2\}$ be the process considered in Example 3.3. Suppose that the process $Y(\cdot)$ is sampled on a regular lattice of size $n = 1 \times M$. Consider the reparametrized covariance function $\rho = \exp(-1/\phi)$, such that $v_{ij} = \rho^{|i-j|}$, and $h_{ij} = |i - j|$, $i, j = 1, \ldots, n$. Then

$$\mathbf{1}^\top \boldsymbol{V}\mathbf{1} = \frac{n - 2\rho - n\rho^2 + 2\rho^{n+1}}{(1-\rho)^2},$$

$$\text{and} \quad \mathbf{1}^\top(\boldsymbol{V} \odot \boldsymbol{H})\mathbf{1} = \frac{2\rho(n(1-\rho)(1+\rho^n) - (1+\rho)(1-\rho^n))}{(1-\rho)^3}.$$

Because $\boldsymbol{V}$ is a Toeplitz matrix, its inverse is

$$\boldsymbol{V}^{-1} = \frac{1}{1-\rho^2}(\boldsymbol{I}_n - \rho\boldsymbol{B} - \rho\boldsymbol{B}^\top + \rho^2\boldsymbol{A}),$$

where the elements of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ satisfy $a_{11} = a_{nn} = 0$, $a_{ii} = 1$ if $i = 2, \ldots, n-1$, and $a_{ij} = 0$ if $i \neq j$, and $b_{ij} = 1$ if $j = i+1$ and $b_{ij} = 0$ otherwise. After simple but tedious algebra (see the appendix), we obtained

$$\text{tr}(\boldsymbol{V}^{-1}\boldsymbol{U}) = \frac{-2(n-1)\rho^2}{1-\rho^2},$$

$$\text{and} \quad \text{tr}((\boldsymbol{V}^{-1}\boldsymbol{U})^2) = \frac{2\rho^2\kappa}{(1-\rho^2)^6},$$

where

$$\begin{aligned}\kappa &= (n-1)(1-3\rho^2+2\rho^4)+\rho^6(2(n+3)-3(2n-1)\rho^2+8\rho^4)\\&\quad+\rho^{2n}(-(3n^2-2n+2)+(5n^2-8n-45)\rho^2-(n-3)(n+1)\rho^4\\&\quad+(5n^2-6n+5)\rho^6).\end{aligned}$$

Thus,

$$S^2 = \frac{2\rho^2}{n^2(1-\rho^2)^6}[n\kappa - 2(n-1)^2(1-\rho^2)^4],$$

and the asymptotic variance (10) is

$$\sigma_g^2 = \frac{n^5(1-\rho^2)^6(1-\rho)^2[n(1-\rho)(1+\rho^n)-(1+\rho)(1-\rho^n)]^2}{[n\kappa-2(n-1)^2(1-\rho^2)^4][n(1-\rho^2)-2\rho(1-\rho^n)]^4}. \tag{14}$$

5. A numerical experiment

In this section, we carry out a numerical experiment to illustrate one application in which the computation of the effective sample size can be useful in the construction of scatterplots for large size datasets. This problem has been addressed recently by Li et al. [18] in a spatial statistics context. First, we develop the effective sample size computations for the model to be considered, and then describe the numerical example. Assume that the spatial process $\boldsymbol{X}$ follows a simultaneous autoregressive (SAR) process with constant mean, defined through

$$\begin{aligned}\boldsymbol{X} &= \mu\mathbf{1}+\boldsymbol{v},\\ \boldsymbol{v} &= \boldsymbol{B}\boldsymbol{v}+\boldsymbol{\varepsilon},\end{aligned}$$

where $\boldsymbol{B} = \rho\boldsymbol{W}$, $\boldsymbol{W}$ is a row-standardized contiguity matrix, $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \tau^2\boldsymbol{I})$, $\mu \in \mathbb{R}$, $\boldsymbol{I}$ is the identity matrix and $0 < \rho < 1$. Then

$$\begin{aligned}\mathbb{E}[\boldsymbol{X}] &= \mu\mathbf{1},\\ \operatorname{var}[\boldsymbol{X}] &= \boldsymbol{\Sigma} = \tau^2(\boldsymbol{I}-\rho\boldsymbol{W})^{-1}(\boldsymbol{I}-\rho\boldsymbol{W}^\top)^{-1}.\end{aligned}$$

Acosta et al. [19] introduced an iterative method to obtain the maximum likelihood estimates of a Gaussian model with linear mean and partially linear covariance structure. An SAR process is a particular case, for which the estimation of the mean μ is

$$\widehat{\mu} = (\mathbf{1}^\top\boldsymbol{\Sigma}^{-1}\mathbf{1})^{-1}\mathbf{1}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{X} = \frac{\mathbf{1}^\top\boldsymbol{X} - \rho\,\mathbf{1}^\top\boldsymbol{W}\boldsymbol{X}}{n(1-\rho)}, \tag{15}$$

while the estimation of the covariance matrix satisfies the equation system

$$\operatorname{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i) = \operatorname{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{C}), \tag{16}$$

where $C = (\boldsymbol{I}-\mu\mathbf{1})(\boldsymbol{I}-\mu\mathbf{1})^\top$ and $\boldsymbol{\Sigma}_i$ denotes the derivative of $\boldsymbol{\Sigma}$ with respect to i. In this case, $i = \tau^2$ or $i = \rho$. Because the inverse of $\boldsymbol{\Sigma}$ is a linear function of ρ, Equation (16) can be rewritten in terms of the derivative of the inverse of $\boldsymbol{\Sigma}_i$ as

$$\operatorname{tr}(\boldsymbol{\Sigma}\boldsymbol{\Sigma}_i^{-1}) = \operatorname{tr}(\boldsymbol{\Sigma}_i^{-1}\boldsymbol{C}).$$

Then

$$\tau^2 = \frac{\operatorname{tr}((\boldsymbol{I} - \rho\boldsymbol{W}^\top)(\boldsymbol{I} - \rho\boldsymbol{W})\boldsymbol{C})}{n}, \tag{17}$$

$$\rho = \frac{\operatorname{tr}(\boldsymbol{W}(\boldsymbol{\Sigma} - \boldsymbol{C}))}{\operatorname{tr}(\boldsymbol{W}(\boldsymbol{\Sigma} - \boldsymbol{C})\boldsymbol{W}^\top)}. \tag{18}$$

Considering initial estimates, it is possible to solve iteratively Equations (15)–(18) for μ, τ^2 and ρ. One way to initialize the covariance function is by setting $\rho = 0$, which yields $\boldsymbol{\Sigma} = \boldsymbol{I}$. The effective sample size for an SAR process can be easily obtained using the definition,

$$n^* = g(\rho) = \frac{n\operatorname{tr}(\boldsymbol{D}\boldsymbol{D}^\top)}{\boldsymbol{1}^\top\boldsymbol{D}\boldsymbol{D}^\top\boldsymbol{1}},$$

where $\boldsymbol{D} = (\boldsymbol{I} - \rho\boldsymbol{W})^{-1}$. By the invariance of the maximum likelihood estimate, $\widehat{n^*} = g(\widehat{\rho})$. In order to derive the asymptotic variance, the inverse of the Fisher information matrix of $\theta = (\tau^2, \rho)^\top$ is

$$\boldsymbol{B}_{\boldsymbol{\theta}}^{-1} = \frac{\begin{pmatrix} 2\tau^4\operatorname{tr}([\boldsymbol{D}\boldsymbol{W}]^2) + 2\tau^4\operatorname{tr}(\boldsymbol{D}\boldsymbol{W}\boldsymbol{W}^\top\boldsymbol{D}^\top) & -2\tau^2\operatorname{tr}(\boldsymbol{D}\boldsymbol{W}) \\ -2\tau^2\operatorname{tr}(\boldsymbol{D}\boldsymbol{W}) & n \end{pmatrix}}{2n\operatorname{tr}((\boldsymbol{D}\boldsymbol{W})^2) + 2n\operatorname{tr}(\boldsymbol{D}\boldsymbol{W}\boldsymbol{W}^\top\boldsymbol{D}^\top) - 4\operatorname{tr}^2(\boldsymbol{D}\boldsymbol{W})}.$$

Thus,

$$\operatorname{var}[\widehat{\rho}] = \frac{n}{2n\operatorname{tr}((\boldsymbol{D}\boldsymbol{W})^2) + 2n\operatorname{tr}(\boldsymbol{D}\boldsymbol{W}\boldsymbol{W}^\top\boldsymbol{D}^\top) - 4\operatorname{tr}^2(\boldsymbol{D}\boldsymbol{W})}.$$

In addition,

$$g'(\rho) = 2n^*\left(\frac{\operatorname{tr}(\boldsymbol{D}^\top\boldsymbol{W})}{\operatorname{tr}(\boldsymbol{D}\boldsymbol{D}^\top)} - \frac{\boldsymbol{1}^\top\boldsymbol{D}^\top\boldsymbol{W}\boldsymbol{1}}{\boldsymbol{1}^\top\boldsymbol{D}\boldsymbol{D}^\top\boldsymbol{1}}\right).$$

Finally,

$$\operatorname{var}[\widehat{n^*}] = (g'(\widehat{\rho}))^2\operatorname{var}[\widehat{\rho}]. \tag{19}$$

Here, we carry out a simulation experiment and show how an approximation of the effective sample size can help to extract a subset from an original dataset through spatial sampling, preserving much of the information contained in the original dataset. The reduced dataset can be used to construct a scatterplot as well as fitted lines. We generated an image X from an SAR process of size $n = 100 \times 100$ with $\mu = 0$, $\rho_X = 0.95$ and $\boldsymbol{W}$ being the row-standardized contiguity matrix associated with the nearest neighbour. Another image Y is generated as a nonlinear transformation of X, $Y = 0.5X + X^2 + X^3$. For each image, we compute the effective sample size $\widehat{n_X^*}$ and $\widehat{n_Y^*}$ and the corresponding variances $\operatorname{var}[\widehat{n_X^*}]$ and $\operatorname{var}[\widehat{n_Y^*}]$. In addition, we compute a joint effective sample size [18] given by

$$\widehat{n_{XY}^*} = 1 + \frac{n-1}{n}\left[\frac{n_X^* + n_Y^*}{2} + k\sqrt{n_X^* n_Y^*}\right], \tag{20}$$

where $k = \sqrt{2}(\widehat{\rho}_X + \widehat{\rho}_Y)^{1.16}(1 - \rho_{XY}^2)^{0.04}$, $\widehat{\rho}_X$ and $\widehat{\rho}_Y$ are the ML estimations of ρ_X and ρ_Y, respectively, and ρ_{XY} is the Pearson correlation coefficient between X and Y. Assuming

Table 1. Estimations of the effective sample size and its variances.

	$\widehat{\rho}$	var$[\widehat{\rho}]$	$\widehat{n^*}$	var$[\widehat{n^*}]$	$\widehat{n^*} \mp 3\sigma$
X	0.9499225	0.000004	237	0.003240	[236–238]
Y	0.5836380	0.000059	2870	62.134690	[2846–2894]
(X, Y)	0.4271218		3454	42.902470	[3434–3474]

that k is a constant, and $\widehat{n_X^*}$ and $\widehat{n_Y^*}$ are non-correlated, and linearizing the term $\sqrt{n_X^* n_Y^*}$, we obtain approximately that

$$\mathrm{var}[\widehat{n_{XY}^*}] = \left(\frac{n-1}{n}\right)^2 \left[\frac{\mathrm{var}[\widehat{n_X^*}] + \mathrm{var}[\widehat{n_Y^*}]}{4} + k^2 \mathrm{var}[\sqrt{\widehat{n_X^*}\widehat{n_Y^*}}] \right.$$
$$\left. + k\,\mathrm{cov}(\widehat{n_X^*} + \widehat{n_Y^*}, \sqrt{\widehat{n_X^*}\widehat{n_Y^*}})\right], \tag{21}$$

where

$$\mathrm{var}[\sqrt{\widehat{n_X^*}\widehat{n_Y^*}}] = \frac{\mathrm{var}[\widehat{n_X^*}]\mathrm{var}[\widehat{n_Y^*}] + \mathrm{var}[\widehat{n_X^*}]\mathbb{E}[\widehat{n_Y^*}]^2 + \mathbb{E}[\widehat{n_X^*}]^2\mathrm{var}[\widehat{n_Y^*}]}{4\mathbb{E}[\widehat{n_X^*}]\mathbb{E}[\widehat{n_Y^*}]},$$

$$\mathrm{cov}(\widehat{n_X^*}, \sqrt{\widehat{n_X^*}\widehat{n_Y^*}}) = \frac{\mathrm{var}[\widehat{n_X^*}]}{2}\sqrt{\frac{\mathbb{E}[\widehat{n_Y^*}]}{\mathbb{E}[\widehat{n_X^*}]}},$$

$$\mathrm{cov}(\widehat{n_Y^*}, \sqrt{\widehat{n_X^*}\widehat{n_Y^*}} = \frac{\mathrm{var}[\widehat{n_Y^*}]}{2}\sqrt{\frac{\mathbb{E}[\widehat{n_X^*}]}{\mathbb{E}[\widehat{n_Y^*}]}}.$$

According to the results shown in Table 1, the joint effective sample size, $\widehat{n_{XY}^*}$, is greater than that in [18]. One way to obtain a smaller value is to consider $\widehat{n_{XY}^*} \approx (\widehat{n_X^*} + \widehat{n_Y^*})/2$ and $\mathrm{var}[\widehat{n_{XY}^*}] \approx (\mathrm{var}[\widehat{n_X^*}] + \mathrm{var}[\widehat{n_Y^*}])/4$. In such a case, the variance of the joint effective sample size is 15.53448, and the corresponding confidence interval is [1542–1566]. Using the lower limit value (1542) of the confidence interval as the maximum reduction of the original sample size, we generate samples from the simulated images X and Y by two sampling schemes: random sampling and systematic sampling. The sample sizes were 1542 and 1600, respectively. The sampled points and the original images are shown in Figure 2. The spatially simplified scatterplots for the original dataset and for the sampled points are shown in Figure 3. We observe that in both cases the adjusted curves with the extracted samples provide a reasonable fit. The mean, standard deviation, skewness, kurtosis and correlation between the processes of the extracted samples (see Table 2) are close enough to the true values. This is in agreement with the results reported in [18]. In this case, the main discrepancy between the theoretical values and the estimates is in Moran's I, which we conjecture is affected by the type of sampling performed. Although the effective sample size used in this experiment corresponds to the lower limit of the confidence interval for n^*, the results do not differ too much from those that uses the ML estimate of n^*. In both cases, substantial reductions in sample size and an increase in spacing among the data points were observed; however, the autocorrelations in the sampled set are still strong, offering the opportunity for further reduction.

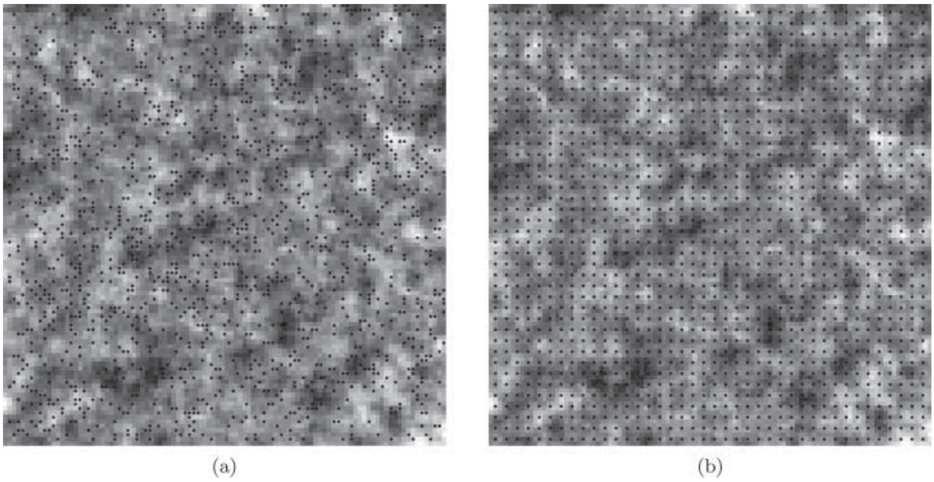

Figure 2. (a) Original image X generated from a SAR process, and the 1542 points sampled using a random sampling scheme; (b) Transformed image Y generated form image X, and the 1600 points sampled using a systematic sampling scheme.

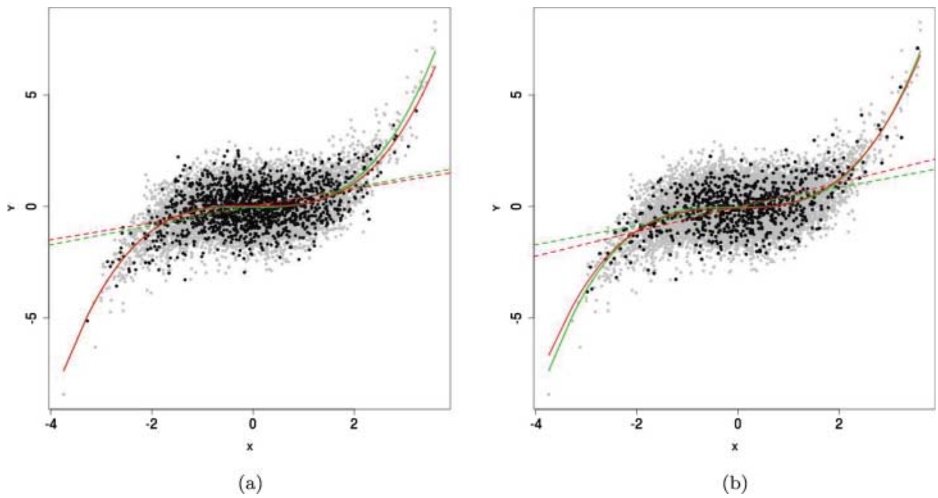

Figure 3. Spatially simplified scatterplots. (a) Scatterplot generated with the 1542 sampled points from a random sampling scheme; (b) Scatterplot generated with the 1600 sampled points from a systematic sampling scheme. The red lines represent the fitted lines using the sampled points while the green lines represent the fitted lines using the whole information. Fitted regression lines are also shown with both datasets. In both cases, the points coloured in grey represent the original dataset, and the points coloured in black are the sampled points.

6. Discussion

This article develops mathematical properties of the effective geographic sample size defined in Equation (1). Our results complement and enhance Griffith's proposal. This is possible in part because Equation (1) does not involve the inverse of the covariance matrix, contrary to the case studied in [4]. The Mardia and Marshall conditions are well illustrated

Table 2. Summary statistics for simulated variables from the resampled set for which $n^* = 10,000$, $n^*_{\text{random}} = 1542$, and $n^*_{\text{regular}} = 1600$.

Parameter	Mean	Std. deviation	Skewness	Kurtosis	Moran coeff.	ρ	ρ_{XY}
X	−0.18	2.97	0.05	0.05	0.82	0.94	0.42
Random	−0.14	2.98	0.08	0.10	0.83	0.44	0.46
Regular	−0.08	3.18	0.09	0.26	0.55	0.41	0.52
Y	6.75	186.97	0.27	3.78	0.37	0.58	0.42
Random	4.50	195.44	0.55	4.00	0.56	0.44	0.46
Regular	6.16	209.38	0.91	5.25	0.51	0.41	0.52

through several spatial models and covariance functions commonly used in spatial statistics. The extension of such conditions to more general models seems a challenging problem to be tackled in future research; e.g. the normality of the ML estimates for the covariance functions associated with models for counts or proportions data. We expect these results to apply normal approximation cases (i.e. non-normal RVs that essentially are indistinguishable from normal RVs), such as a Poisson RV with a relatively large mean, a binomial (e.g. percentage) RV with a relatively large number of trials, among others.

Although in some of the examples, the asymptotic variance does not have a closed form, in practical cases standard errors can be computed from the Fisher information matrix, allowing hypothesis testing about the parameters of the covariance function. For those models that the asymptotic variance is difficult to compute, resampling techniques might be considered in the light of García-Soidán et al.[20].

In Section 5, the numerical results obtained were based on the asymptotic variance provided by Theorem 3.2. The assumptions of that theorem must be satisfied. This could be done similarly to the conditional autoregressive (CAR) processes studied in [15]. There is still room to study the performance of curve fitting for larger image sizes. Exploration of the effective sample size for larger data-sizes remains an open problem to be addressed in future research.

Acknowledgements

Daniel A. Griffith is an Ashbel Smith Professor of Geospatial Information Sciences at the University of Texas at Dallas.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

AC3E stands for Advanced Center for Electrical and Electronic Engineering, UTFSM, grant FB-0008.

References

[1] Griffith DA. Effective geographic sample size in the presence of spatial autocorrelation. Ann Assoc Amer Geogr. 2005;95:740–760.

[2] Griffith DA. Geographic sampling of urban soils for contaminant mapping: how many samples and from where. Environ Geochem Health. 2008;30:495–509.

[3] Telford RJ, Birks HJB. Evaluation of transfer functions in spatially structured environments. Quat Sci Rev. 2009;28:1309–1316.
[4] Vallejos R, Osorio F. Effective sample size for spatial process models. Spat Stat. 2014;9:66–92.
[5] Faes C, Molenberghs G, Aerts M, et al. The effective sample size and an alternative small-sample degrees-of-freedom method. Am Stat. 2009;63:389–399.
[6] Berger J, Bayarri MJ, Pericchi LR. The effective sample size. Econ Lett. 2014;33:197–2014.
[7] Thiébaux HJ, Zwiers FW. Interpretation and estimation of effective sample size. J Climate Appl Meteor. 1984;23:800–811.
[8] de Gruijter JJ, Ter Braak CJF. Model-free estimation from spatial samples: a reappraisal of classical sampling theory. Math Geol. 1990;22:407–415.
[9] Cogley JG. Effective sample size for glacier mass balance. Geogr Ann A. 1999;81:497–507.
[10] Griffith DA. Establishing qualitative geographic sample size in the presence of spatial autocorrelation. Ann Assoc Amer Geogr. 2013;103:1107–1122.
[11] Lo JM, Zhu Z. Accounting for spatial correlation in the scan statistic. Ann Appl Stat. 2007;1:560–584.
[12] Pigoli D, Secchi P. Estimation of the mean for spatially dependent data belonging to a Riemannian manifold. Electron J Statist. 2012;6:1926–1942.
[13] Johnson RA, Wichern DW. Applied multivariate statistical analysis. Upper Saddle River (NJ): Prentice-Hall; 1998.
[14] Mardia KV, Marshall RJ. Maximum likelihood estimation of models for residual covariance in spatial regression. Biometrika. 1984;71:135–146.
[15] Acosta J, Vallejos R. Effective sample size for spatial regression models. Submitted 2017.
[16] Kyung M, Ghosh SK. Maximum likelihood estimation for directional conditionally autoregressive models. J Stat Plan Inference. 2010;140:3160–3179.
[17] Mardia K. Maximum likelihood estimation for spatial models. Proceedings from the Symposium Spatial Statistics: Past, Present, and Future, held on the campus of Syracuse University; from March to June; 1989, p. 203–253.
[18] Li B, Griffith DA, Becker B. Spatially simplified scatterplots for large raster datasets. Geo-spat Inf Sci. 2016;19:81–93.
[19] Acosta J, Osorio F, Vallejos R. Effective sample size for line transect sampling models with an application to marine macroalgae. J Agric, Biol Environ Stat. 2016;21:407–425.
[20] García-Soidán P, Menezes R, Rubiños O. Bootstrap approaches for spatial data. Stoch Environ Res Risk Assess. 2014;28:1207–1219.

Appendix. Derivations for Example 4.2

Let $\boldsymbol{U} = \boldsymbol{V} \odot \boldsymbol{H}$, (i.e. $u_{ij} = |i-j|\rho^{|i-j|}$), and recall that

$$S^2 = \frac{1}{n}\mathrm{tr}((\boldsymbol{V}^{-1}\boldsymbol{U})^2) - \left[\frac{1}{n}\mathrm{tr}(\boldsymbol{V}^{-1}\boldsymbol{U})\right]^2.$$

Given $\mathrm{tr}(\boldsymbol{B}^\top\boldsymbol{U}) = \mathrm{tr}(\boldsymbol{BU})$, $\mathrm{tr}(\boldsymbol{B}^\top\boldsymbol{U}^2) = \mathrm{tr}(\boldsymbol{BU}^2)$, $\mathrm{tr}((\boldsymbol{B}^\top\boldsymbol{U})^2) = \mathrm{tr}((\boldsymbol{BU})^2)$, and $\mathrm{tr}(\boldsymbol{B}^\top\boldsymbol{UAU}) = \mathrm{tr}(\boldsymbol{BUAU})$,

$$\begin{aligned}
\mathrm{tr}(\boldsymbol{V}^{-1}\boldsymbol{U}) &= \frac{1}{1-\rho^2}[\mathrm{tr}(\boldsymbol{U}) - 2\rho\,\mathrm{tr}(\boldsymbol{BU}) + \rho^2\mathrm{tr}(\boldsymbol{AU})],\\
\text{and}\quad \mathrm{tr}((\boldsymbol{V}^{-1}\boldsymbol{U})^2) &= \frac{1}{(1-\rho^2)^2}[\mathrm{tr}(\boldsymbol{U}^2) + 2\rho^2\mathrm{tr}((\boldsymbol{BU})^2) + \rho^4\mathrm{tr}((\boldsymbol{AU})^2) - 4\rho\,\mathrm{tr}(\boldsymbol{BU}^2)\\
&\quad + 2\rho^2\mathrm{tr}(\boldsymbol{AU}^2) + 2\rho^2\mathrm{tr}(\boldsymbol{BUB}^\top\boldsymbol{U}) - 4\rho^3\mathrm{tr}(\boldsymbol{AUBU})].
\end{aligned}$$

Moreover, $\mathrm{tr}(\boldsymbol{U}) = \sum_{i=1}^{n} u_{ii} = 0$ because $u_{ii} = 0$. Let $(\boldsymbol{BU})_{ij} = \sum_{k=1}^{n} b_{ik}u_{kj} = b_{i,i+1}u_{i+1,j} = |i + 1 - j|\rho^{|i+1-j|}$, $i = 1, \ldots, n-1$, $j = 1, \ldots, n$. If $i = j$, then $(\boldsymbol{BU})_{ii} = \rho$, and

$$\mathrm{tr}(\boldsymbol{BU}) = \sum_{i=1}^{n} (\boldsymbol{BU})_{ii} = \sum_{i=1}^{n-1} \rho = (n-1)\rho.$$

Similarly, $(\boldsymbol{AU})_{ij} = \sum_{k=1}^{n} a_{ik}u_{kj} = u_{ij}$, $i, j = 2, \ldots, n-1$. In particular, if $i = j$, and $u_{ii} = 0$, then $(\boldsymbol{AU})_{ii} = 0$ and $\mathrm{tr}(\boldsymbol{AU}) = \sum_{i=1}^{n} (\boldsymbol{AU})_{ii} = 0$. Therefore,

$$\mathrm{tr}(\boldsymbol{V}^{-1}\boldsymbol{U}) = \frac{-2(n-1)\rho^2}{1-\rho^2}.$$

Analogous calculations yield

$$\begin{aligned}
\mathrm{tr}(\boldsymbol{U}^2) &= 2\left(\frac{(n-1)\rho^2}{1-\rho^2} + \frac{(3n-7)\rho^4 + (n+1)^2\rho^{2(n+1)}}{(1-\rho^2)^2}\right.\\
&\quad \left. + \frac{2(n-6)\rho^6 + 2(2n+3)\rho^{2(n+2)}}{(1-\rho^2)^3} - \frac{6\rho^8 - 6\rho^{2(n+3)}}{(1-\rho^2)^4}\right),\\
\mathrm{tr}((\boldsymbol{AU})^2) &= 2\left(\frac{(n-3)\rho^2 + (n-1)^2\rho^{2(n-1)}}{1-\rho^2} + \frac{(3n-13)\rho^4 + (n+1)^2\rho^{2n}}{(1-\rho^2)^2}\right.\\
&\quad \left. + \frac{2(n-8)\rho^6 + 4(n+1)\rho^{2(n+1)}}{(1-\rho^2)^3} - \frac{6\rho^8 - 6\rho^{2(n+2)}}{(1-\rho^2)^4}\right),\\
\mathrm{tr}((\boldsymbol{BU})^2) &= \sum_{i=1}^{n-1}\sum_{j=1}^{n-1} |(i-j)^2 - 1|\rho^{|i-j+1|+|i-j-1|}\\
&= 2(n-1)(n-2)\rho^{2n-3} + 2\rho\left(\frac{2(n-2)\rho^2 - (n-1)(n-2)\rho^{2(n-2)}}{1-\rho^2}\right.\\
&\quad + \frac{2(2n-7)\rho^4 + (n-1)(n-2)\rho^{2(n-1)}}{(1-\rho^2)^2} + \frac{2(n-8)\rho^6 + 2(2n-1)\rho^{2n}}{(1-\rho^2)^3}\\
&\quad \left. - \frac{6\rho^8 - 6\rho^{2(n+1)}}{(1-\rho^2)^4}\right),\\
\mathrm{tr}(\boldsymbol{BU}^2) &= 2(n-1)(n-2)\rho^{2n-3} + 2\rho\left(\frac{2(n-2)\rho^2 - (n-1)(n-2)\rho^{2(n-2)}}{1-\rho^2}\right.\\
&\quad + \frac{2(2n-7)\rho^4 + (n-1)(n-2)\rho^{2(n-1)}}{(1-\rho^2)^2} + \frac{2(n-8)\rho^6 + 2(2n-1)\rho^{2n}}{(1-\rho^2)^3}\\
&\quad \left. - \frac{6\rho^8 - 6\rho^{2(n+1)}}{(1-\rho^2)^4}\right),\\
\mathrm{tr}(\boldsymbol{AU}^2) &= 2\left(\frac{(n-2)\rho^2 + n^2\rho^{2n}}{1-\rho^2} + \frac{(3n-10)\rho^4 + (n^2+4n+2)\rho^{2(n+1)}}{(1-\rho^2)^2}\right.\\
&\quad \left. + \frac{2(n-7)\rho^6 + 2(n+4)\rho^{2(n+2)}}{(1-\rho^2)^3} - \frac{6\rho^8 - 6\rho^{2(n+3)}}{(1-\rho^2)^4}\right),
\end{aligned}$$

$$\mathrm{tr}(\boldsymbol{BUB}^\top\boldsymbol{U}) = 2\left(\frac{(n-2)\rho^2}{1-\rho^2} + \frac{(3n-10)\rho^4 + n^2\rho^{2n}}{(1-\rho^2)^2}\right.$$
$$\left. + \frac{2(n-7)\rho^6 + 2(2n+1)\rho^{2(n+1)}}{(1-\rho^2)^3} - \frac{6\rho^8 - 6\rho^{2(n+2)}}{(1-\rho^2)^4}\right),$$

and $$\mathrm{tr}(\boldsymbol{AUBU}) = 2(n-2)(n-3)\rho^{2n-5} + 2\rho\left(\frac{2(n-3)\rho^2 - (n-2)(n-3)\rho^{2(n-3)}}{1-\rho^2}\right.$$
$$+ \frac{2(2n-9)\rho^4 + (n-2)(n-3)\rho^{2(n-2)}}{(1-\rho^2)^2}$$
$$\left. + \frac{2(n-9)\rho^6 + 2(2n-3)\rho^{2(n-1)}}{(1-\rho^2)^3} - \frac{6\rho^8 - 6\rho^{2n}}{(1-\rho^2)^4}\right).$$

Therefore,

$$\mathrm{tr}((\boldsymbol{V}^{-1}\boldsymbol{U})^2) = \frac{2}{(1-\rho^2)^2}(a_0 + a_1 + a_2 + a_3 + a_4),$$

where

$$a_0 = (n-1)\rho^4 - 8(n-2)^2\rho^{2(n-1)},$$
$$a_1 = \frac{1}{1-\rho^2}((n-1)\rho^2 - 4(n-2)\rho^4 - (n-3)\rho^6 + 8(n-2)^2\rho^{2(n-1)}$$
$$+ (3n^2 - 2n + 1)\rho^{2(n+1)}),$$
$$a_2 = \frac{1}{(1-\rho^2)^2}((3n-7)\rho^4 - 4(n-4)\rho^6 - (3n-13)\rho^8 - 8(n-2)^2\rho^{2n}$$
$$+ 2(n^2 + 2n + 2)\rho^{2(n+1)} + 2(n^2 + 4n + 2)\rho^{2(n+2)}),$$
$$a_3 = \frac{2}{(1-\rho^2)^3}((n-6)\rho^6 + 4\rho^8 - (n-8)\rho^{10} - 16(n-1)\rho^{2(n+1)}$$
$$+ (10n + 7)\rho^{2(n+2)} + 2(2n+5)\rho^{2(n+3)}),$$
and $$a_4 = \frac{-6}{(1-\rho^2)^4}(\rho^8 - \rho^12 + 14\rho^{2(n+2)} - 5\rho^{2(n+3)} - 3\rho^{2(n+4)}).$$

Putting the terms together yields the asymptotic variance (14).