Effective Sample Size for Line Transect Sampling Models with an Application to Marine Macroalgae

Jonathan Acosta[a], Felipe Osorio[b] and Ronny Vallejos[a]

[a]Departamento de Matemática
Universidad Técnica Federico Santa María, Chile

[b]Instituto de Estadística
Pontificia Universidad Católica de Valparaíso, Chile

Abstract

This paper provides a framework for estimating the effective sample size in a spatial regression model context when the data have been sampled using a line transect scheme and there is an evident serial correlation due to the chronological order in which the observations were collected. We propose a linear regression model with a partially linear covariance structure to address the computation of the effective sample size when spatial and serial correlations are present. A recursive algorithm is described to separately estimate the linear and nonlinear parameters involved in the covariance structure. The kriging equations are also presented to explore the kriging variance between our proposal and a typical spatial regression model. An application in the context of marine macroalgae, which motivated the present work, is also presented.

Keywords: Effective sample size; line transects; spatial association; ARMA processes; marine macroalgae.

1 Introduction

Lessonia trabeculata (Chilean kelp), a macroalgae that grows off the northern and central coasts of Chile, is important because it is related to the ecosystem. Vasquez and Santelices (1990) reported the negative effects of harvesting and kelp removal. Lessonia trabeculata became important in 1978, when it began to be regularly exported from Chile as a raw material for producing alginate. To maintain continuous surveillance over this species off the central coast of Chile and exploit it in an organized way, the government of Chile created protected areas reserved for artisanal fisheries (AMERB in Spanish), which can be exploited by a small group of previously registered fisheries (see `http://www.subpesca.cl/institucional/602/w3-article-79853.html`). The Chilean government determined that the Chilean Fisheries Research Institute (Instituto de Fomento Pesquero (IFOP)) is in charge of conducting studies to evaluate the abundance of the species that are present in the AMERB areas. The dataset that is analyzed in this paper is from one of the studies performed by IFOP in 2012, in which the main goal was to determine the equivalent number of independent observations to use in posterior studies

assuming that the spatial correlation associated with the new dataset is the same as that in the present case.

To address the quantification of the equivalent number of independent observations, it is necessary to extend the notion of spatial effective sample size for models that adequately represent the variable under study. This requires considering the sampling scheme used to collect the data and the type of correlation to be considered in the statistical model to be developed. Griffith (2005) proposed a formula for computing the spatial effective sample size for spatial regression models. This coefficient is based on the computation of the variation inflation factor associated with the regression model, but the formula depends solely on the covariance structure of the model, which can assume several patterned forms but is not sufficiently flexible to consider serial correlation, for example. Subsequently, Vallejos and Osorio (2014) developed a formula for computing the spatial effective sample size based on the Fisher information quantity for the mean of a spatial regression process with a constant mean. Although this approach has a different motivation, in practice, it is comparable to Griffith's proposal. Unfortunately, none of these approaches consider including serial correlation (Dale and Fortin, 2009) as an extra term in the correlation structure when a line transect sampling scheme has been used to collect the data (e.g., Hedley and Buckland, 2004). Moreover, we could not find a solution to this problem in the vast literature about the effective sample size in different contexts. For example, Faes et al. (2009) discussed the notion of effective sample size for mixed effects models with replicates. Seminal works treating the reduction of sample size or degrees of freedom can be found in Box (1954a,b), Clifford et al. (1989), and Dutilleul (1993), among others.

In this paper, we quantify the spatial effective sample size for a macroalgae dataset collected off the central coast of Chile. Because this dataset was collected using line transects and because a chronological order exists between each pair of observations, as described in Section 2, we propose quantifying the effective sample size for a regression model with a constant mean but with a partially linear covariance structure, which allows us to include the existing spatial and serial correlations. Three aspects related to the suggested model are developed in Section 3: the covariance structure, the estimation process, and prediction. The estimation is performed using a recursive algorithm developed for partially linear covariance structures, which is a variant of the well-known maximum likelihood method. The full analysis of the macroalgae dataset, including the estimation of the effective sample size for the density of the Chilean kelp in an AMERB study area, located off the central coast of Chile, will be discussed in Section 4. This includes results obtained under different spatial correlation models.

The broader impacts resulting from this research lie in its contributions to science, particularly in a country where climatic effects and disasters are constantly changing the conditions in which the species described in this paper can be studied.

2 The Macroalgae Dataset

To determine a procedure for the sampling scheme and the posterior methodological analysis of data collected from the AMERB areas, in 2012, the IFOP held a meeting with experts in the area to establish a protocol to follow in future investigations. After defining the area to be sampled and the species, they followed the recommendations of experts to conduct a study on Lessonia trabeculata (Chilean kelp) from the AMERB area located in Punta Lunes (Ventana, comuna de Puchuncavi), Chile. This area was chosen because the species are located at a depth of no more than 20 m 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.

To conduct the study, 26 line transects were defined perpendicular to the coastline. Each line transect was 100 or 200 m long and separated from each other by 100 m, as shown in Figure 1. The original line transects were designed to be perfectly perpendicular to the coastline and to have an exact length of 200 m; however, because of the weather conditions or the topography of the coast, the resulting line transects are those shown in Figure 1.

For each line transect and every 10 m within each transect, observations were collected by divers who followed red marks located on underwater ropes. The density of Lessonia trabeculata measured in several quadrants of 20 m^2 was georeferenced and recorded for posterior analysis. The total number of collected observations was 427. Note that there is a chronological order among the collected observations because the divers followed the line transects from north to south. The average time between two consecutive observations was 5 min. The 427 observations obtained for the Lessonia trabeculata used in this study

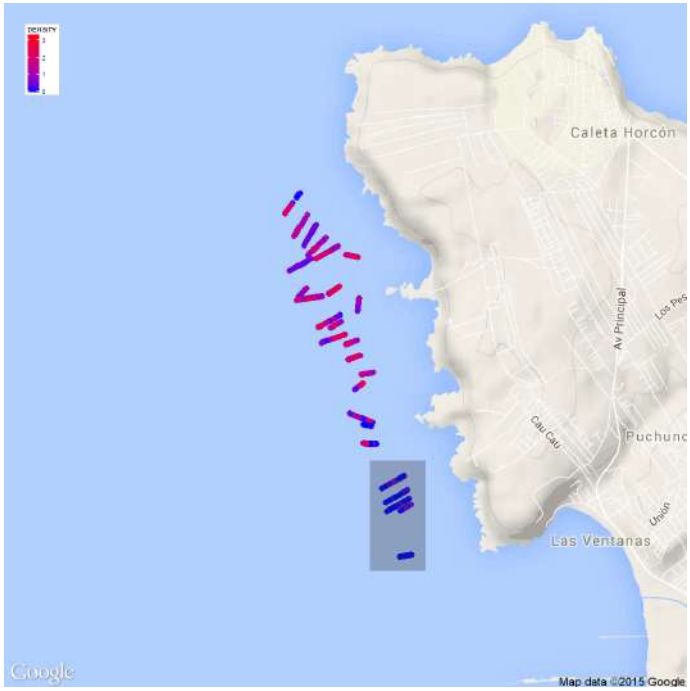


Figure 1: Study area including the 26 transect locations where the observations were collected. The study area is inside the rectangle $(264000, 266500) \times (6374500, 637700)$ considering UTM coordinates. The gray rectangle contains 5 transects belonging to a free access area.

are based on count data. Thus, alternatively to the treatment presented in this paper, a model for count data (e.g., Poisson) might be useful.

Although in this paper we deal with the raw and transformed data, the transformation that is introduced in Section 4.2 is used only to achieve normality. However, because the variable of interest is the original density of Lessonia trabeculata, the exploratory data analysis will be carried out for the raw dataset.

3 Material and Methods

In this section, we provide three aspects related to the computation of the effective sample size. First, a spatial model that includes spatial and serial autocorrelations is described. Second, the estimation of the effective sample size is briefly developed. Third, spatial prediction (kriging) for the macroalgae density will be presented.

3.1 Sampling and Modeling

Let us consider the spatial process $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in D \subset \mathbb{R}^2\}$. We assume that the process $Y(\cdot)$ has been measured at n distinct locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$ and that the locations are equispaced over m distinct line transects on D such that the measurements have been sampled chronologically in time starting from the observations located on transect 1 until the last observations located on transect m. For simplicity, we assume that the distance between locations in a given transect is fixed. In other words, the transect sampling scheme occurs chronologically in time such that when the last observation on a given transect is taken, the next observation is obtained from the next (closest) transect following the same orientation as in the previous one. To clarify the sampling scheme used in this work, in Figure 2, two typical sampling configurations are described. In both cases, m transects have been located on the two-dimensional plane. If the transects are not connected as in Figure 2 (b), the observations are taken sequentially, as shown by the dashed lines.

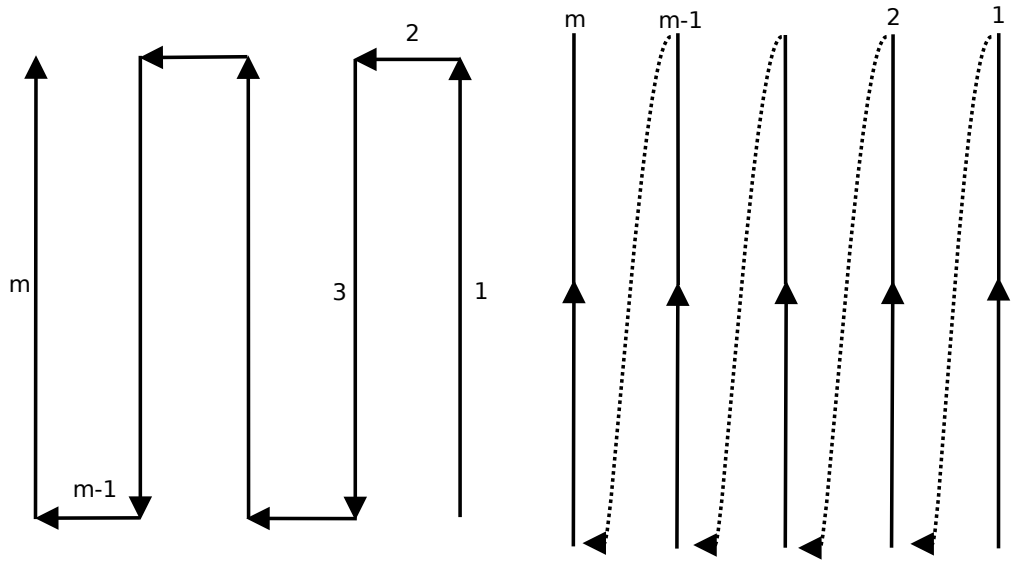


Figure 2: (a) Sampling scheme with m connected transects; (b) sampling scheme with m unconnected transects.

Let $\boldsymbol{Y} = (Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n))^\top$, where $\top$ denotes the transpose of a vector, and

$$\boldsymbol{Y} \sim \mathcal{N}(\mu \mathbf{1}, \boldsymbol{\Sigma}(\boldsymbol{\theta})), \tag{1}$$

where $\mathcal{N}(\cdot, \cdot)$ denotes the normal distribution, $\mu \in \mathbb{R}$, $\mathbf{1}$ is an $n \times 1$ vector of ones, $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is a non-singular covariance matrix, and $\boldsymbol{\theta} \in \mathbb{R}^k$ is a vector of unknown parameters. Denoting as $\boldsymbol{R}(\boldsymbol{\theta})$ the correlation matrix associated with $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, Vallejos and Osorio (2014) defined

the spatial effective sample size (ESS) of $\boldsymbol{Y}$ as the equivalent number of independent observations associated with vector $\boldsymbol{Y}$ (Cressie, 1993) and corresponding to the Fisher information number about μ.

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

$$\text{ESS} = \mathbf{1}^\top \boldsymbol{R}(\boldsymbol{\theta})^{-1}\mathbf{1}, \tag{2}$$

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

Because in many different cases, the effective sample size lies in the interval $[1, n]$, it can be regarded as a reduction of the information due to the spatial association present in the data. It is well known that as the spatial autocorrelation latent in georeferenced data increases, the amount of duplicated information in these data also increases. Hence, the effective sample size can be seen as a sample size compensation due to the spatial association.

Commonly in spatial statistics, $\boldsymbol{\theta} = (\tau^2, \sigma^2, \phi)^\top$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \tau^2\boldsymbol{I} + \sigma^2\boldsymbol{Q}_1(\phi)$, where τ^2 represents the nugget effect, σ^2 is the variance of the process, and ϕ is related to the range dependence of the spatial process. The spatial correlation structure can then be described by a number of correlation functions, e.g., the Matérn covariance function (Rasmussen and Williams, 2006, Chapter 4). A complete list of parametric models for $\boldsymbol{Q}_1(\phi)$ can be found in Banerjee et al. (2004, p. 27). In this article, we consider the serial correlation that could be present in the spatial data. This can be achieved by considering the model

$$\boldsymbol{Y} \sim \mathcal{N}(\mu\mathbf{1}, \boldsymbol{\Sigma}(\boldsymbol{\theta}, \boldsymbol{\rho}, \boldsymbol{\varphi})), \tag{3}$$

where

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}, \boldsymbol{\rho}, \boldsymbol{\varphi}) = \tau^2\boldsymbol{I} + \sigma^2\boldsymbol{Q}_1(\phi) + \lambda^2\boldsymbol{Q}_2(\boldsymbol{\rho}, \boldsymbol{\varphi}),$$

$\boldsymbol{\theta} = (\tau^2, \sigma^2, \phi, \lambda^2)^\top$ and $\boldsymbol{\rho}, \boldsymbol{\varphi}$ are parameter vectors containing information about the serial autocorrelation. For example, if the temporal autocorrelation can be modeled by an autoregressive moving average (ARMA) process, then $\boldsymbol{Q}_2(\cdot, \cdot)$ can be written as a function of the autoregressive and moving average parameters $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_p)^\top$ and $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_q)^\top$(Brockwell and Davis, 2006, Chapter 3). In particular, for an AR(1) process,

$$\boldsymbol{Q}_2(\rho) = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & \rho & \ldots & \rho^{n-1} \\ \rho & 1 & \ldots & \rho^{n-2} \\ \vdots & \vdots & \vdots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \ldots & 1 \end{pmatrix}. \tag{4}$$

Thus, the effective sample size associated with $\boldsymbol{Y}$ is given by

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

where $\boldsymbol{R}(\boldsymbol{\theta}, \boldsymbol{\rho}, \boldsymbol{\varphi})$ is the correlation matrix corresponding to $\boldsymbol{\Sigma}(\boldsymbol{\theta}, \boldsymbol{\rho}, \boldsymbol{\varphi})$.

The advantage of using (5) instead of (2) is the flexibility to include the serial correlation that originally was not considered in model (1). This approach can also be useful for other datasets when the spatial sampling has been planned such that the observations are obtained sequentially in time.

3.2 Estimation

We estimate the parameters of model (3) by maximizing the log-likelihood function

$$\ell(\mu, \boldsymbol{\theta}, \boldsymbol{\rho}, \boldsymbol{\varphi}) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\boldsymbol{\Sigma}| - \frac{1}{2}(\boldsymbol{Y} - \mu\mathbf{1})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \mu\mathbf{1}), \tag{6}$$

where $\boldsymbol{\Sigma} = \tau^2\boldsymbol{I} + \sigma^2\boldsymbol{Q}_1(\phi) + \lambda^2\boldsymbol{Q}_2(\boldsymbol{\rho}, \boldsymbol{\varphi})$; for notational convenience, we ignore the dependence of $\boldsymbol{\Sigma}$ on $\boldsymbol{\theta}$, $\boldsymbol{\rho}$ and $\boldsymbol{\varphi}$. Rather, we define $\boldsymbol{\omega} = (\boldsymbol{\theta}^\top, \boldsymbol{\rho}^\top, \boldsymbol{\varphi}^\top)^\top$, a vector of size $p + q + 4$. When $\boldsymbol{\omega}$ is known, the estimator of μ is

$$\widehat{\mu} = \left(\mathbf{1}^\top\boldsymbol{\Sigma}^{-1}\mathbf{1}\right)^{-1}\mathbf{1}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}. \tag{7}$$

Thus,

$$\frac{\partial\ell}{\partial\omega_i} = -\frac{1}{2}\operatorname{tr}\left\{\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}^{-1}\left(\tau^2\boldsymbol{Q}_0 + \sigma^2\boldsymbol{Q}_1(\phi) + \lambda^2\boldsymbol{Q}_2(\boldsymbol{\rho}, \boldsymbol{\varphi})\right)\right\} + \frac{1}{2}\operatorname{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{C}\}, \tag{8}$$

where $\boldsymbol{\Sigma}_i = \partial\boldsymbol{\Sigma}/\partial\omega_i$, $\boldsymbol{Q}_0 = \boldsymbol{I}$, and $\boldsymbol{C} = (\boldsymbol{Y} - \mu\mathbf{1})(\boldsymbol{Y} - \mu\mathbf{1})^\top$. We follow the estimation method proposed by Anderson (1973) to solve $\nabla\ell = 0$, where iterative equations are used, and the nonlinear term $\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}^{-1}$ in Equation (8) is assumed to be an estimation coming from a previous stage. In addition, μ and $\boldsymbol{\omega}$ can also be iteratively estimated. Given initial estimations $\mu^{(0)}$ and $\boldsymbol{\omega}^{(0)}$ for μ and $\boldsymbol{\omega}$, respectively, the estimation method is summarized in Algorithm 1 (see Appendix A). The nonlinear system stated in Equation (A.3) can be solved, for instance, using a Newton type algorithm.

One way to obtain $\boldsymbol{\omega}^{(0)}$ is by estimating the spatial parameters from the empirical variogram and the parameters $\lambda^{2(0)}$, $\boldsymbol{\rho}^{(0)}$ and $\boldsymbol{\varphi}^{(0)}$ by the usual time series techniques. In addition, $\mu^{(0)}$ can be initially estimated by $\mu^{(0)} = \overline{\boldsymbol{Y}}$ or by evaluating $\boldsymbol{\Sigma}$ at $\boldsymbol{\omega}^{(0)}$ in equation (7).

The main advantage of the estimation method summarized in Algorithm 1 is its flexibility and simplicity. One important feature noted by Szatrowski (1980) is that for certain covariance structures, the algorithm has an explicit solution. Other authors (e.g., Rubin and Szatrowski, 1982) explored the connection between this estimation method and the EM algorithm to simplify the iterative step in the estimation process. We stress the fact that the method described by Algorithm 1 and the maximum likelihood method are asymptotically equivalent.

3.3 Prediction

We provide the prediction equations for model (3); the kriging equations for model (1) are available in any spatial statistics textbook. See, for instance, Cressie (1993, Chapter 3)

Suppose for now that the parameters of model (3) are known. Let $Y_0 = Y(\boldsymbol{s}_0)$ be a new observation at $\boldsymbol{s}_0$, a non-observed site, and assume that

$$\begin{pmatrix} Y_0 \\ \boldsymbol{Y} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu \\ \mu\mathbf{1} \end{pmatrix}, \begin{pmatrix} v_0 & \mathbf{w}^\top \\ \mathbf{w} & \boldsymbol{\Sigma} \end{pmatrix}\right),$$

(Banerjee et al., 2004). Under the quadratic loss function, the best linear predictor is the conditional expectation of Y_0 given $\boldsymbol{Y}$. The distribution of $Y_0 \mid \boldsymbol{Y}$ is then normal with mean and variance defined through

$$\begin{aligned} \mathrm{E}[Y_0 \mid \boldsymbol{Y}] &= \mu + \mathbf{w}^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \mu\mathbf{1}), \\ \mathrm{var}[Y_0 \mid \boldsymbol{Y}] &= v_0 - \mathbf{w}^\top\boldsymbol{\Sigma}^{-1}\mathbf{w}, \end{aligned}$$

where $\mathbf{w}^\top = (\sigma^2\boldsymbol{Q}_1(\phi, h_{01}), \ldots, \sigma^2\boldsymbol{Q}_1(\phi, h_{0n})) + (\lambda^2\boldsymbol{Q}_2(\rho, \varphi, t_{01}), \ldots, \lambda^2\boldsymbol{Q}_2(\rho, \varphi, t_{0n}))$ and $v_0 = \tau^2 + \sigma^2 + \lambda^2$, with $h_{0i} = \|\boldsymbol{s}_0 - \boldsymbol{s}_i\|$ and $t_{0i} = |t_i - t_0|$, $i = 1, \ldots, n$.

Model (3) was constructed with a transect sampling scheme in mind. Thus, the observed sites are necessarily located on a transect. If t_i denotes a time index for a point located over a transect and t_0 represents the index time of the nearest neighbor observation to $\boldsymbol{s}_0$ for those points located over any transect, that is, $t_0 = t_{j^*}$ and

$$j^* = \underset{i \in \{1,\ldots,n\}}{\operatorname{argmin}} \|\boldsymbol{s}_0 - \boldsymbol{s}_i\|,$$

then the autocorrelation between a variable associated with an unobserved site and the variable associated with a point located on a given transect is $\rho^{|t_0 - t_i|}$. If the parameters are unknown, similar equations can be derived in the same spirit of the universal kriging using the estimates rather than the true parameters.

4 Analysis of Macroalgae Data

4.1 Exploratory Data Analysis

First, an exploratory data analysis was performed for the macroalgae dataset consisting of the 427 observations related to the density of Lessonia trabeculata per 20 m^2 in the AMERB area in Punta Lunes, Puchuncavi, Chile. Considering the UTM coordinates, the locations and the values of the density for the 427 sites are plotted in 3D in Figure 3. We observe that the values corresponding to the six transects belonging to the free access area are quite different compared with the remainder of the observations sampled from the transects belonging to the AMERB area. This is because in the free access area, there is human intervention without any restrictions. To make our study comparable with similar AMERB areas, we will not consider the observations sampled over the six transects belonging to the free access area in the computation of the effective sample size. However, in the exploratory data analysis, the entire raw dataset will be used.

Considering all observations, we constructed a table with double entries to explore the variance of the values as a function of the number of transects and the distance between the locations where the density was measured within each transect. The results are shown in Table 1. We observe that a simple exploratory data analysis allows us to

Table 1: Variance of the 427 observations for different numbers of line transects and distances between measurements.

Number of Transect Lines	10 [m]	20 [m]	30 [m]	40 [m]
26	307.43	311.99	308.36	334.51
13	302.57	**292.09**	298.88	308.97
9	314.58	332.40	338.77	418.17
7	311.99	304.75	294.06	306.59
6	330.18	356.27	330.40	371.59

reduce either the sample size or the number of transects to produce a smaller variance. The optimal combination occurs for 13 transects and 20 m, which represents a reduction in the sample size equivalent to 70%. Obviously, this analysis does not consider the existing spatial or temporal association in the data.

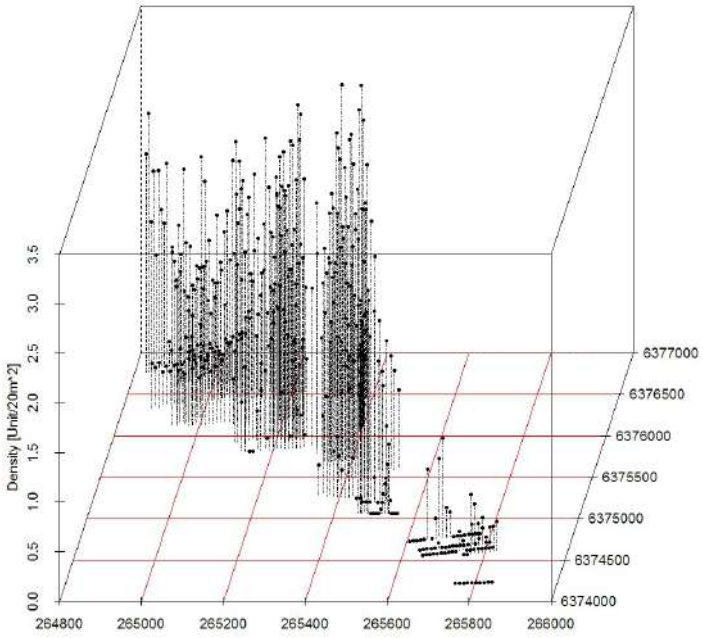


Figure 3: 3D plot for the density of Lessonia trabeculata in the study area. The total number of observations is 427.

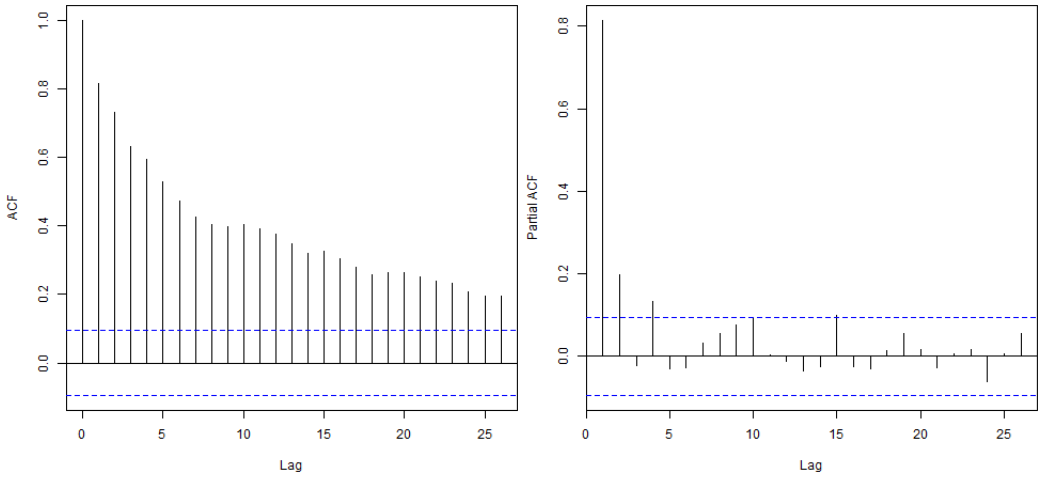


Figure 4: Autocorrelation function (left) and partial autocorrelation function (right) for the raw macroalgae data.

4.2 Building a Model

In Figure 4, the autocorrelation (ACF) and partial autocorrelation (PACF) functions of the raw data are plotted. There is a clear serial dependence because the ACF rapidly

decreases to zero, and the PACF has few lags for which the correlation is significant according to Bartlet's test. An autoregressive time series model appears to be suitable for representing the serial correlation that is present in the data. In Figure 5, empirical and theoretical variograms for the macroalgae dataset are plotted. There is evidence from the variogram plot that a spatial association exists among the observations. Thus, a suitable model for the density of the macroalgae should include both serial and spatial correlations.

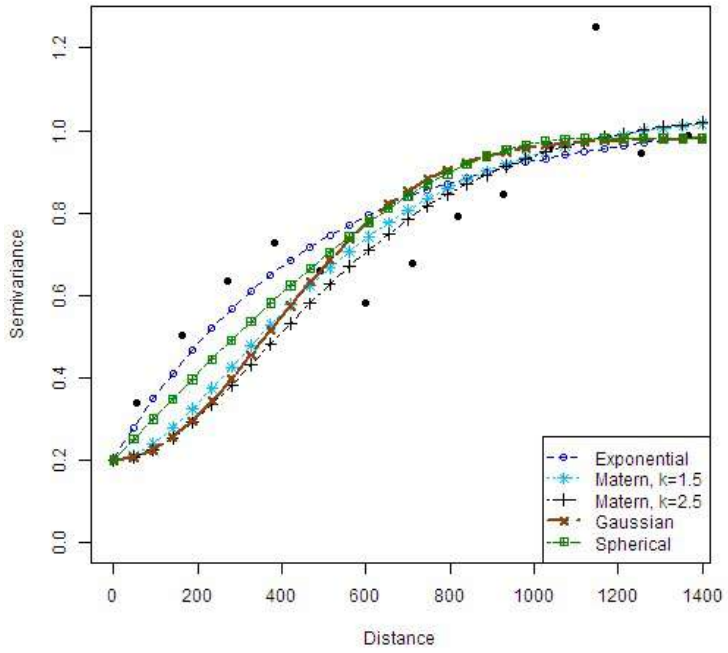


Figure 5: Empirical and some theoretical variograms for the macroalgae data. The distance in the x–axis is measured in meters.

Considering the foregoing comments, we propose using model (3) with covariance structures of the form

$$\boldsymbol{\Sigma} = \tau^2 \boldsymbol{I} + \sigma^2 \boldsymbol{Q}_1(\phi), \tag{9}$$

$$\boldsymbol{\Sigma} = \tau^2 \boldsymbol{I} + \sigma^2 \boldsymbol{Q}_1(\phi) + \lambda^2 \boldsymbol{Q}_2(\rho), \tag{10}$$

and $\boldsymbol{Q}_2(\rho)$ is given in (4). Before fitting model (3), it is necessary to explore the Gaussianity assumption. A preliminary analysis shows that normality in this case does not hold. The skewness and kurtosis are 0.5313 and 0.7947, respectively, highlighting a departure from normality. Moreover, the Shapiro–Wilk test rejects the normality hypothesis with a p-value of 3.49×10^{-14}. To overcome this drawback, we propose using a transformation to symmetrize the original data.

Consider the transformation

$$Z(\boldsymbol{s}_i) = \ln(Y(\boldsymbol{s}_i) + 2.586),$$

where 2.586 is the value that maximizes the correlation (0.9853) between the percentiles of the normal distribution and the order statistics $Z_{(i)}$ (see Appendix B). In the case

of using the Box–Cox transformation, this correlation is 0.9887, which is slightly higher than the previous value.

4.3 Estimation and Prediction

The parameters of model (1) with the covariance structure (9) were estimated by five different methods using the transformed data to achieve symmetry: the partially linear covariance (PLC) method described in Section 3.2, the maximum likelihood (ML) method, the restricted maximum likelihood (REML) method, the ordinary least squares (OLS) method, and the weighted least squares (WLS) method. Four isotropic covariance models were fit to the density of Lessonia trabeculata. The Matérn model defined through

$$C(h,\phi) = \frac{\left(\frac{h}{\phi}\right)^{\kappa} K_{\kappa}\left(\frac{h}{\phi}\right)}{2^{\kappa-1}\Gamma(\kappa)}$$

where $\Gamma(\cdot)$ is the gamma function, $K_{\kappa}(\cdot)$ is the modified Bessel function of the second kind, and κ, ϕ are nonnegative parameters were used with $\kappa = 0.5$ (exponential case), $\kappa = 1.5$, $\kappa = 2.5$, and $\kappa \to \infty$ (Gaussian case).

The results, including the computation of the effective sample size for the original and transformed variables and the estimations of the parameters, are summarized in Table 2. Here, the effective sample sizes range from 19.40 to 66.62 for the density of Lessonia trabeculata. In general, we observe that the five estimation methods are very comparable, yielding very similar estimations for the parameters of model (1). The estimations obtained using the ML, REML, OLS, and WLS methods, which are reported in Table 2, were computed using the R package `geoR`. In the estimation processes used to obtain Table 2, the serial correlation in the data was disregarded to compare the PLC method with other estimation methods that are currently available in R.

We use Algorithm 1 to estimate the parameters of model (1) with the covariance structure (10)(at the moment, the `geoR` (Ribeiro and Diggle, 2015) package does not have any routine available for estimating the ML parameters of the correlation structure (10)). The serial correlation here is part of the covariance structure, and it was characterized by an AR(1) process. Similar to the previous case, the effective sample size given in (5) was computed for the original and transformed variables. We also considered the approximations of first and second order of the effective sample size provided by Proposition 1 in Appendix B. The results are presented in Table 3. The effective sample sizes here range from 24.50 to 29.73 for the transformed variable and from 24.50 to 30.09 for the original variable. These results are in agreement with the results shown in Table 2 because there is a reduction in the sample size owing to the inclusion of the time dependence structure that is not present in model (1). From Proposition 1 in the Appendix, the effective sample size for the first-order approximation of the effective sample size coincides with the effective sample size of the transformed variable. This is valid only for the log transformation.

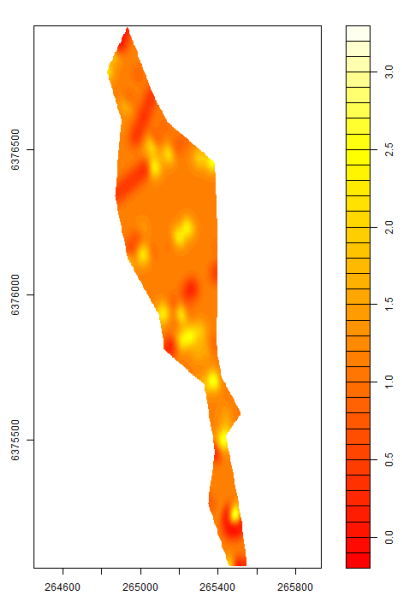
The kriging estimator and its variance defined in Section 3.3 were evaluated for the macroalgae dataset for models (1) and (3), and the maps are shown in Figure 6. We observe that the kriging estimator associated with model (1) is smoother than the kriging estimator associated with model (3). However, the kriging variance is smaller for model (3), which is one expected advantage that relies on the fact that the serial correlation has been considered. Finally, the R (R Core Team, 2015) code to perform the data analysis and modeling described in this article are available upon request.

Table 2: Effective sample sizes for the original and transformed variables associated with the PLC, ML, REML, OLS, and WLS estimation methods of model (1) for different covariance structures. $\mathrm{SS}(\widehat{Z})$ stands for the predicted sum of squares of model (1), whereas $\mathrm{ESS}_Y(1)$ stands for the first-order approximation of the effective sample size of Y, in accordance with the expansion given in Equation (B.4). The notation for $\mathrm{ESS}_Y(2)$ and $\mathrm{ESS}_Y(\infty)$ is similar.

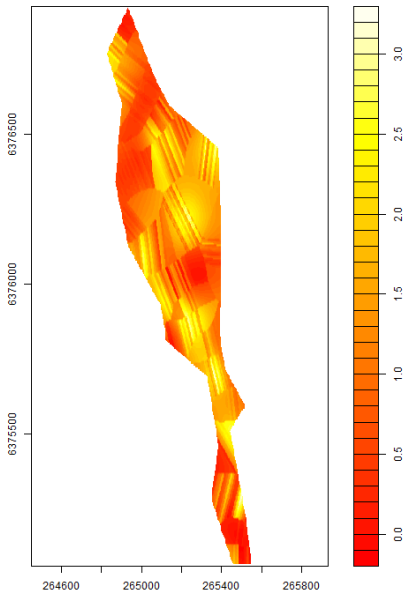
Model	Method	$\widehat{\mu}_Z$	$\widehat{\tau}^2_Z$	$\widehat{\sigma}^2_Z$	$\widehat{\phi}_Z$	$\mathrm{SS}(\widehat{Z})$	$\mathrm{ESS}_Y(1)$	$\mathrm{ESS}_Y(2)$	$\mathrm{ESS}_Y(\infty)$
Exponential	PLC	1.2927	0.0048	0.0430	64.64	16.40	29.18	29.53	29.54
	ML	1.2927	0.0048	0.0430	64.64	16.40	29.21	29.56	29.57
	REML	1.2917	0.0048	0.0450	69.24	16.42	26.99	27.33	27.34
	OLS	1.2877	0.0000	0.0502	82.40	16.50	20.45	20.70	20.70
	WLS	1.2864	0.0034	0.0470	91.87	16.53	19.14	19.39	19.40
Matérn $\kappa = 1.5$	PLC	1.2962	0.0085	0.0368	23.78	16.34	46.43	46.85	46.86
	ML	1.2962	0.0085	0.0368	23.78	16.34	46.46	46.89	46.89
	REML	1.2959	0.0085	0.0378	24.29	16.35	45.36	45.78	45.79
	OLS	1.2923	0.0124	0.0376	50.00	16.41	22.95	23.22	23.23
	WLS	1.2923	0.0123	0.0375	50.01	16.41	22.92	23.19	23.20
Matérn $\kappa = 2.5$	PLC	1.2952	0.0096	0.0397	19.57	16.36	43.50	43.91	43.92
	ML	1.2976	0.0090	0.0352	15.89	16.32	53.23	53.67	53.68
	REML	1.2974	0.0090	0.0359	16.13	16.32	52.36	52.80	52.81
	OLS	1.2922	0.0186	0.0319	50.00	16.41	19.71	19.97	19.98
	WLS	1.2944	0.0189	0.0315	50.00	16.37	26.81	27.18	27.19
Gaussian	PLC	1.3000	0.0095	0.0328	38.95	16.28	66.14	66.60	66.61
	ML	1.3000	0.0095	0.0328	38.95	16.28	66.13	66.59	66.60
	REML	1.2999	0.0095	0.0333	39.43	16.29	65.31	65.76	65.77
	OLS	1.2889	0.0000	0.0487	60.28	16.48	43.02	42.94	42.93
	WLS	1.2479	0.0000	0.0486	63.60	17.93	41.09	41.01	40.99

Table 3: Effective sample size for the original and transformed variables associated with the PLC estimation method of model (3) for different covariance structures. $\mathrm{SS}(\widehat{Z})$ represents the predicted sum of squares of model (3).

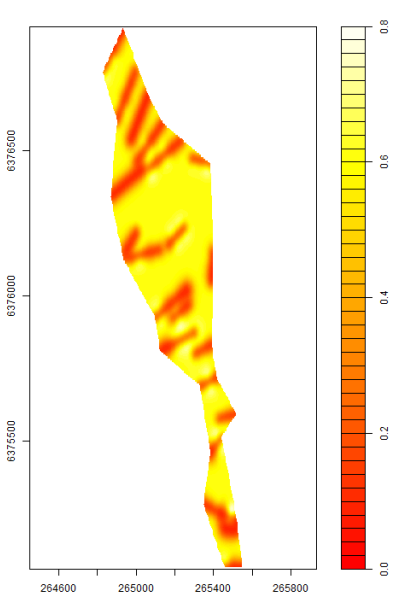
Model	$\widehat{\mu}_Z$	$\widehat{\tau}^2_Z$	$\widehat{\sigma}^2_Z$	$\widehat{\lambda}^2_Z$	$\widehat{\phi}_Z$	$\widehat{\rho}_Z$	$\mathrm{SS}(\widehat{Z})$	$\mathrm{ESS}_Y(1)$	$\mathrm{ESS}_Y(2)$	$\mathrm{ESS}_Y(\infty)$
Exponential	1.307	0.004	0.017	0.025	86.87	0.860	16.20	24.50	24.83	24.83
Matérn $\kappa = 1.5$	1.308	0.004	0.016	0.026	48.13	0.828	16.19	26.99	27.33	27.34
Matérn $\kappa = 2.5$	1.308	0.004	0.016	0.026	34.15	0.822	16.19	28.68	29.02	29.03
Gaussian	1.308	0.004	0.016	0.026	102.05	0.814	16.19	29.73	30.08	30.09



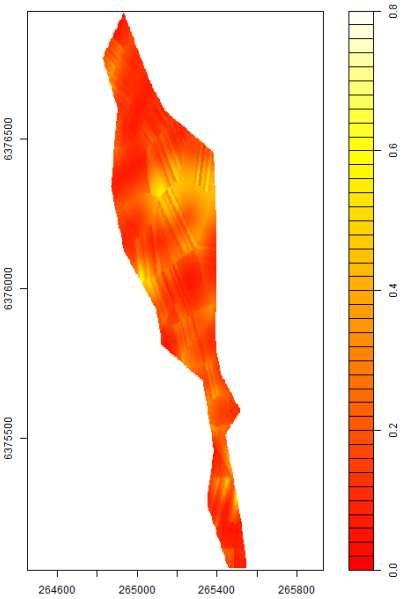
(a) Kriging predictor using model (1).



(b) Kriging predictor using model (3).



(c) Variance of the kriging predictor under model (1).



(d) Variance of the kriging predictor under model (3).

Figure 6: Kriging predictors and their variances under models (1) and (3).

5 Discussion

The method proposed in this paper offers a general way to estimate the effective sample size for models of the form (3) when the sampling scheme involves the use of line transects and the observations have been taken chronologically in time. The reduction of the sample size n in all cases reported in this article is severe owing to the inclusion of both spatial and serial correlations in model (3). After computing the effective sample size, new sampling strategies can be planned to obtain the equivalent number of independent observations. A review of these techniques can be found in Griffith (2005). From our experience with spatial data, a regular hexagonal tessellation sampling design is sufficient for ensuring the representativeness of the areas in space.

The effective sample size was developed for a linear regression model with a constant mean $\mu\mathbf{1}$. However, in practice, a constant mean could be a limitation because we might learn from an exploratory analysis that the mean is not constant for certain areas in space or that there are other covariates that have been measured for the same spatial sites, and this information must be used. In this framework, a more realistic model is

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{11}$$

where $\boldsymbol{X}$ is an $n \times k$ design matrix containing the information of the covariates, $\boldsymbol{\beta}$ is an unknown parameter vector, and $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$. To obtain a formula for the effective sample size of $\boldsymbol{Y}$ in model (11), several issues should be addressed in advance, for example, how to reduce the information contained in the Fisher information matrix about $\boldsymbol{\beta}$ to a single number and how to create a measure of the effective sample size including the information of the covariates but still lying on the interval $[1, n]$. One way to generalize (2) is by considering a weighted effective sample size of the form

$$\text{ESS} = \sum_{i=1}^{k} v_i \boldsymbol{x}_i^\top \boldsymbol{R}(\boldsymbol{\theta})^{-1} \boldsymbol{x}_i, \tag{12}$$

where v_i are weights to be specified, $\boldsymbol{x}_i$ is the ith column of matrix $\boldsymbol{X}$, and $\boldsymbol{R}(\boldsymbol{\theta})$ is the correlation matrix as before. This type of extension deserves attention in further research. However, the estimation of ESS in (12) can be obtained by using either maximum likelihood or restricted maximum likelihood estimation methods. These methods and their limiting distributions have been studied in a spatial statistics context in Mardia and Marshall (1984) and Cressie and Lahiri (1996), respectively. These methods represent two possible avenues for further research in the estimation of the effective sample size for spatial regression models with covariates.

Algorithm 1 was specifically constructed for the estimation of model (3). However, it can be generalized to estimate the parameters of models with a partially linear structure for the mean and for the covariance function. For instance, the models considered in the Monte Carlo simulation studies conducted by Crujeiras and Van Keilegom (2010) are special cases of models with partially linear components in the mean function. Moreover, regular techniques commonly used in numerical analysis can be employed to make this type of algorithm efficient.

One important aspect that should be stressed in the analysis of real datasets developed in Section 4 is the simplicity in computing the sample size associated with a transformed sample via the Box–Cox transformation. The result provided in the Appendix ensures that under a first-order Taylor expansion, the effective sample size before and after the logarithm transformation is the same. This result can be extended for the general Box–Cox function. In addition, the kriging analysis and the performed hypothesis testing provide empirical support for the adequacy of model (3). The reduction of information of the ESS when using model (3) provides good insights regarding the adequacy of the

correlation structure proposed in this article. An alternative to the Box–Cox transformation that allows one to handle the asymmetry and kurtosis is the proposal studied by Field and Genton (2006). These authors extended the family of distributions g-and-h to a multivariate context, considering the stochastic representation

$$\boldsymbol{Y} = \boldsymbol{\eta} + \boldsymbol{B}\boldsymbol{\tau}_{\boldsymbol{g},\boldsymbol{h}}(\boldsymbol{Z}),$$

where $\boldsymbol{\eta} \in \mathbb{R}^n$ and $\boldsymbol{\Sigma} = \boldsymbol{B}\boldsymbol{B}^\top$ are a position vector and an arbitrary scale matrix, respectively, $\boldsymbol{B}$ is a full rank matrix, and $\boldsymbol{\tau}_{\boldsymbol{g},\boldsymbol{h}}(\boldsymbol{Z})$ denotes the transformation of the random vector $\boldsymbol{Z} = (Z_1, \ldots, Z_n)^\top \sim \mathcal{N}_n(\boldsymbol{0}, \boldsymbol{I})$ defined as

$$\boldsymbol{\tau}_{\boldsymbol{g},\boldsymbol{h}}(\boldsymbol{Z}) = (\tau_{g_1,h_1}(Z_1), \ldots, \tau_{g_n,h_n}(Z_n))^\top, \tag{13}$$

for the vectors $\boldsymbol{g} = (g_1, \ldots, g_n)^\top$, $\boldsymbol{h} = (h_1, \ldots, h_n)^\top$ and

$$\tau_{g_i,h_i}(Z_i) = \frac{1}{g_i}(\exp(g_i Z_i) - 1)\exp(h_i Z_i^2/2), \qquad i = 1, \ldots, n.$$

For the vectors $\boldsymbol{g}$ and $\boldsymbol{h}$, the expected information matrix for μ is

$$\begin{aligned} \mathrm{E}\{W_\tau^2(\boldsymbol{Z})\}\boldsymbol{1}^\top \operatorname{cov}(\boldsymbol{Y})\boldsymbol{1} &= \mathrm{E}\{W_\tau^2(\boldsymbol{Z})\}\boldsymbol{1}^\top \boldsymbol{B} \operatorname{cov}\{\boldsymbol{\tau}_{\boldsymbol{g},\boldsymbol{h}}(\boldsymbol{Z})\}\boldsymbol{B}^\top \boldsymbol{1} \\ &= \mathrm{E}\{W_\tau^2(\boldsymbol{Z})\}\boldsymbol{1}^\top \boldsymbol{B}\boldsymbol{\Lambda}\boldsymbol{B}^\top \boldsymbol{1}, \end{aligned} \tag{14}$$

where $W_\tau(\boldsymbol{z}) = \dot{f}_\tau(\boldsymbol{z})/f_\tau(\boldsymbol{z})$ with $f_\tau(\boldsymbol{z})$ being the joint density function of the random vector $\boldsymbol{\tau}_{\boldsymbol{g},\boldsymbol{h}}(\boldsymbol{Z})$ given in (13), and $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ with $\lambda_i = \operatorname{var}(\tau_{g_i,h_i}(Z_i))$. Given the independence between the elements of $\boldsymbol{\tau}_{\boldsymbol{g},\boldsymbol{h}}(\boldsymbol{Z})$, it is straightforward to obtain the joint density of $f_\tau(\boldsymbol{z})$ and the variances $\operatorname{var}(\tau_{g_1,h_1}(Z_1)), \ldots, \operatorname{var}(\tau_{g_n,h_n}(Z_n))$ invoking the procedure proposed by Headrick et al. (2008). We recall that Equation (14) is similar to a result given in Definition 3 of Vallejos and Osorio (2014). Another alternative to handle asymmetric data with heavy tail distributions is the class of the variance-mean mixture of normal distributions (Barndorff-Nielsen et al., 1982), which has been applied in finance and physics. Although efforts have been made to develop suitable algorithms to estimate the parameters in this class of distributions (see, for instance Protassov, 2004) to the best of our knowledge, results regarding the Fisher information matrix have not yet been studied and represent an interesting matter for future research. Furthermore, the parameter estimation of the multivariate distributions g-and-h in the context of spatial modelling is a challenging problem to be tackled in further research. In particular, the algorithm proposed by Field and Genton (2006) must be adapted for use in a model such as (3).

The present work constitutes an important advance in the design of protocols in IFOP that can be applied in the future when they plan a new study in the same study area under similar conditions to those described in this paper.

Acknowledgments

Ronny Vallejos was partially supported by Fondecyt grant 1120048, Chile, AC3E grant FB-0008, and USM grant 12.15.09. Felipe Osorio was partially supported by FONDECYT grant 1140580. Jonathan Acosta was partially supported by PIIC at UTFSM, Chile. The authors are indebted to Luis Aris, Luis Figueroa, and Carlos Cortés from IFOP for providing the macroalgae dataset and for helpful discussions. The authors are also grateful to Diego Alvarez from UTFSM for providing preliminary computational results regarding the macroalgae dataset. In addition, the authors would like to thank Dr. Emilio Porcu at UTFSM for his constant support. The authors acknowledge the suggestions from two anonymous referees, an associate editor and the editor of JABES that improved the manuscript.

Appendix

A The Estimation Algorithm

Algorithm 1 Partially Linear Covariance (PLC) Algorithm.

1: Set $l = 0$.

2: Compute

$$\begin{aligned}
\boldsymbol{Q}_1^{(l)} &= \boldsymbol{Q}_1(\phi^{(l)}), \quad \boldsymbol{Q}_2^{(l)} = \boldsymbol{Q}_2(\boldsymbol{\rho}^{(l)}, \boldsymbol{\varphi}^{(l)}), \quad \dot{\boldsymbol{Q}}_{1,\phi}^{(l)} = \frac{\partial \boldsymbol{Q}_1}{\partial \phi}\left(\phi^{(l)}\right), \\
\dot{\boldsymbol{Q}}_{2,\rho_i}^{(l)} &= \frac{\partial \boldsymbol{Q}_2}{\partial \rho_i}\left(\boldsymbol{\rho}^{(l)}, \boldsymbol{\varphi}^{(l)}\right), \quad i = 1, \ldots, p, \\
\dot{\boldsymbol{Q}}_{2,\varphi_i}^{(l)} &= \frac{\partial \boldsymbol{Q}_2}{\partial \varphi_i}\left(\boldsymbol{\rho}^{(l)}, \boldsymbol{\varphi}^{(l)}\right), \quad i = 1, \ldots, q, \\
\boldsymbol{\Sigma}^{(l)} &= \boldsymbol{\Sigma}(\boldsymbol{\omega}^{(l)}), \quad \boldsymbol{C}^{(l)} = (\boldsymbol{Y} - \mu^{(l)}\mathbf{1})(\boldsymbol{Y} - \mu^{(l)}\mathbf{1})^\top.
\end{aligned}$$

3: Update the linear parameters of $\boldsymbol{\Sigma}$ through the linear system

$$\boldsymbol{A}^{(l)} \begin{pmatrix} \tau^{2^{(l+1)}} & \sigma^{2^{(l+1)}} & \lambda^{2^{(l+1)}} \end{pmatrix}^\top = \boldsymbol{b}^{(l)}, \tag{A.1}$$

where $\boldsymbol{A}^{(l)} = (a_{ij}^{(l)})$, $\boldsymbol{b}^{(l)} = (b_i^{(l)})$, $a_{ij}^{(l)} = \mathrm{tr}\{(\boldsymbol{\Sigma}^{(l)})^{-1}\boldsymbol{Q}_{i-1}^{(l)}(\boldsymbol{\Sigma}^{(l)})^{-1}\boldsymbol{Q}_{j-1}^{(l)}\}$, and $b_i^{(l)} = \mathrm{tr}\{(\boldsymbol{\Sigma}^{(l)})^{-1}\boldsymbol{Q}_{i-1}^{(l)}(\boldsymbol{\Sigma}^{(l)})^{-1}\boldsymbol{C}^{(l)}\}, i, j = 1, 2, 3$.

4: Solve the nonlinear equation $g(\phi) = 0$ to obtain $\phi^{(l+1)}$, where

$$g(\phi) = \mathrm{tr}\left\{(\boldsymbol{\Sigma}^{(l)})^{-1}\dot{\boldsymbol{Q}}_{1,\phi}^{(l)}(\boldsymbol{\Sigma}^{(l)})^{-1}\overset{\boldsymbol{h}}{\tau^{2^{(l+1)}}}\boldsymbol{I} + \sigma^{2^{(l+1)}}\boldsymbol{Q}_1(\phi) + \lambda^{2^{(l+1)}}\boldsymbol{Q}_2^{(l)} - \boldsymbol{C}^{(l)}\right]\right\}.$$

The Newton–Raphson algorithm provides a recursive expression for $\phi^{(l+1)}$ given by

$$\phi^{(l+1)} = \phi^{(l)} - \sigma^{2^{(l+1)}}\,\mathrm{tr}\left\{(\boldsymbol{\Sigma}^{(l)})^{-1}\dot{\boldsymbol{Q}}_{1,\phi}^{(l)}(\boldsymbol{\Sigma}^{(l)})^{-1}\dot{\boldsymbol{Q}}_{1,\phi}^{(l)}\right\} g(\phi^{(l)}). \tag{A.2}$$

5: Obtain $\boldsymbol{\rho}^{(l+1)}$ and $\boldsymbol{\varphi}^{(l+1)}$ by solving (e.g., using the Newton–Raphson algorithm) the nonlinear system

$$\begin{aligned}
\overset{\boldsymbol{\rho}}{\boldsymbol{h}}(\boldsymbol{\rho}, \boldsymbol{\varphi}) &= \mathbf{0}, \\
\boldsymbol{f}(\boldsymbol{\rho}, \boldsymbol{\varphi}) &= \mathbf{0},
\end{aligned} \tag{A.3}$$

where $\boldsymbol{h} = (h_1, \ldots, h_p)$, $\boldsymbol{f} = (f_1, \ldots, f_q)$,

$$\begin{aligned}
h_i(\boldsymbol{\rho}, \boldsymbol{\varphi}) &= \mathrm{tr}\left\{(\boldsymbol{\Sigma}^{(l)})^{-1}\dot{\boldsymbol{Q}}_{2,\rho_i}^{(l)}(\boldsymbol{\Sigma}^{(l)})^{-1}\overset{\boldsymbol{h}}{\tau^{2^{(l+1)}}}\boldsymbol{I} + \sigma^{2^{(l+1)}}\boldsymbol{Q}_1^{(l)} + \lambda^{2^{(l+1)}}\boldsymbol{Q}_2(\boldsymbol{\rho}, \boldsymbol{\varphi}) - \boldsymbol{C}^{(l)}\right]\right\}, \\
f_j(\boldsymbol{\rho}, \boldsymbol{\varphi}) &= \mathrm{tr}\left\{(\boldsymbol{\Sigma}^{(l)})^{-1}\dot{\boldsymbol{Q}}_{2,\varphi_j}^{(l)}(\boldsymbol{\Sigma}^{(l)})^{-1}\overset{\boldsymbol{h}}{\tau^{2^{(l+1)}}}\boldsymbol{I} + \sigma^{2^{(l+1)}}\boldsymbol{Q}_1^{(l)} + \lambda^{2^{(l+1)}}\boldsymbol{Q}_2(\boldsymbol{\rho}, \boldsymbol{\varphi}) - \boldsymbol{C}^{(l)}\right]\right\},
\end{aligned}$$

for $i = 1, \ldots, p$ and $j = 1, \ldots, q$.

6: Compute

$$\begin{aligned}
\boldsymbol{\Sigma}^{(l+1)} &= \boldsymbol{\Sigma}(\boldsymbol{\omega}^{(l+1)}), \\
\mu^{(l+1)} &= \left(\mathbf{1}^\top(\boldsymbol{\Sigma}^{(l+1)})^{-1}\mathbf{1}\right)^{-1}\mathbf{1}^\top(\boldsymbol{\Sigma}^{(l+1)})^{-1}\boldsymbol{Y}.
\end{aligned}$$

7: If $\|(\mu^{(l+1)}, \boldsymbol{\omega}^{\top(l+1)}) - (\mu^{(l)}, \boldsymbol{\omega}^{\top(l)})\| < \gamma$ or $|\ell(\mu^{(l+1)}, \boldsymbol{\omega}^{\top(l+1)}) - \ell(\mu^{(l)}, \boldsymbol{\omega}^{\top(l)})| < \gamma$, stop and set $\hat{\boldsymbol{\mu}} = \mu^{(l+1)}$ and $\hat{\boldsymbol{\omega}} = \boldsymbol{\omega}^{(l+1)}$; otherwise, $l = l + 1$ and return to Step 2. γ is a fixed tolerance constant. In general, typical values for the tolerance constant γ are 10^{-3} or 10^{-6}.

B The Box–Cox Transformation

The following function was proposed by Box and Cox (1964)

$$Z_{\boldsymbol{\delta}}(\boldsymbol{s}) = \begin{cases} \frac{(Y(\boldsymbol{s})+\delta_2)^{\delta_1}-1}{\delta_1} & ; \quad \delta_1 \neq 0, \\ \ln(Y(\boldsymbol{s})+\delta_2) & ; \quad \delta_1 = 0, \end{cases} \tag{B.1}$$

where $Y(\boldsymbol{s})$ is the original variable, and $\boldsymbol{\delta} = (\delta_1, \delta_2)$ is an unknown parameter vector to be estimated to achieve normality of the transformed variable $Z(\cdot)$. Given the vector of spatial observations $(Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_n))^\top$, $\boldsymbol{\delta}$ can be estimated (Box and Cox, 1964) by maximizing the likelihood function

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

An alternative way to estimate $\boldsymbol{\delta}$ is to find 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 $i = 1, \ldots, n$; (see Kutner et al., 2004).

Using the logarithm transformation ($\delta_1 = 0$), it is possible to obtain an approximated expression that relates the ESS for the original and transformed variables.

Proposition 1. Suppose that $\boldsymbol{Z} = (Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_n))^\top \sim \mathcal{N}(\boldsymbol{\mu}_Z, \boldsymbol{\Sigma}_Z)$ and consider the transformation

$$Z(\boldsymbol{s}_i) = \ln(Y(\boldsymbol{s}_i) + \delta_2), \quad i = 1, \ldots, n, \tag{B.3}$$

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 $(\boldsymbol{\mu_Z})_i = \mu$, and $(\boldsymbol{\Sigma_Z})_{ii} = \sigma_{ii} = \tilde{\sigma}^2, i = 1, \ldots, n$, then

$$(\boldsymbol{R_Y})_{ij} = c \cdot (\boldsymbol{R_Z})_{ij} k_{ij} + o\left(|\tilde{\sigma}^2|^m\right), \tag{B.4}$$

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

Proof. Because $Z(\boldsymbol{s}_i)$ is normally distributed, $Y(\boldsymbol{s}_i)$ has a lognormal distribution with

$$\begin{aligned} \mathrm{E}[Y(\boldsymbol{s}_i)] &= e^{\mu_i + 0.5\tilde{\sigma}^2} - \delta_2, \\ \mathrm{cov}(Y(\boldsymbol{s}_i), Y(\boldsymbol{s}_j)) &= e^{\mu_i + \mu_j + \tilde{\sigma}^2} \left(e^{\sigma_{ij}} - 1\right), \end{aligned}$$

for $i, j = 1, \ldots, n$.

Using a Taylor expansion for the function $e^{\sigma_{ij}}$, one obtains

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

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

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

□

References

Anderson, T. W. (1973), "Asymptotically efficient Estimation of the covariance matrices with linear structure," *The Annals of Statistics* 1: 135–141.

Banerjee, S., Carlin, B., and Gelfand, A. (2004), *Hierarchical Modeling and Analysis for Spatial Data*, Chapman & Hall, London.

Barndorff-Nielsen, O., Kent, J., Sørensen, M. (1982). Normal variance-mean mixtures and z distributions. *International Statistical Review* 50: 145–159.

Box, G. (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," *Annals of Mathematical Statistics* 25: 290–302.

Box, G. (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," *Annals of Mathematical Statistics* 25: 484–98.

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

Brockwell, P., and Davis, R. (2006), *Time Series: Theory and Methods*, Springer, New York.

Clifford, P., Richardson, S., and Hémon, D. (1989), "Assessing the significance of the correlation between two spatial processes," *Biometrics* 45: 123–34.

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

Cressie, N., and Lahiri, S. N. (1996), "Asymptotics for REML estimation of spatial covariance parameters," *Journal of Statistical Planning and Inference* 50: 327–341.

Crujeiras, R. M., and Van Keilegom I. (2010), "Least squares estimation of non linear spatial trends," *Computational Statistics and Data Analysis* 54: 452–465.

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

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

Faes, C., Molenberghs, G., Aerpts, M., Verbeke, G., and Kenward, M. (2009), "Effective sample size and an alternative small-sample degrees-of-freesom method," *The American Statistician* 63: 389–399.

Field, C., and Genton, M. G. (2006). " The multivariate g-and-h distribution," *Technometrics* 48: 104–111.

Griffith, D. (2005), "Effective geographic sample size in the presence of spatial autocorrelation," *Annals of the Association of American Geographers* 95: 740–760.

Headrick, T. C., Kowalchuk, R. K., and Sheng, Y. (2008). "Parametric probability densities and distribution functions for Tukey g-and-h transformations and their use for fitting data," *Applied Mathematical Sciences* 2: 449–462.

Hedley, S. L., and Buckland, S. T. (2004), "Spatial Models for line transect sampling," *Journal of the Agricultural, Biological, and Environmental Statistics* 9: 181–199.

Kutner, M., Nachtsheim, C., Neter, J., and Li, W. (2004), *Applied Linear Statistical Models*, McGraw-Hill/Irwin, Homewood, IL.

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

Protassov, R. S. (2004). "EM-based maximum likelihood parameter estimation for multivariate generalized hyperbolic distributions with fixed λ," *Statistics and Computing* 14: 67–77.

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

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

Ribeiro, P. J., Diggle, P. J. (2015). geoR: Analysis of Geostatistical Data. R package version 1.7-5.1. http://CRAN.R-project.org/package=geoR

Rubin, D. B., Szatrowski, T. H. (1982), "Finding maximum likelihood estimates of patterned covariance matrices by the EM algorithm," *Biometrika* 69: 657–660.

Szatrowski, T. H. (1980), "Necessary and sufficient conditions for explicit solutions in the multivariate normal estimation problem for patterned means and covariances," *The Annals of Statistics* 8: 802–810.

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

Vásquez, J., B., and Santelices, J. A. (1990), "Ecological effects of harvesting Lessonia (Laminariales, Phaeophyta) in central Chile," *Hydrobiología* 204/205: 41–47.