Article:
Garcia-Papani, F, Leiva, V, Uribe-Opazo, MA et al. (1 more author) (2018) Birnbaum-Saunders spatial regression models: Diagnostics and application to chemical data. Chemometrics and Intelligent Laboratory Systems, 177. pp. 114-128. ISSN 0169-7439

https://doi.org/10.1016/j.chemolab.2018.03.012

Birnbaum-Saunders spatial regression models: Diagnostics and application to chemical data

Fabiana Garcia-Papani[1], **Víctor Leiva**[2,3], **Miguel A. Uribe-Opazo**[1], **Robert G. Aykroyd**[4]

[1]Postgraduate Program in Agricultural Engineering, and
Center of Exact Sciences and Technology, Universidade Estadual do Oeste do Paraná, Brazil
[2]School of Industrial Engineering, Pontificia Universidad Católica de Valparaíso, Chile
[3]Faculty of Administration, Accounting and Economics, Universidade Federal de Goiás, Brazil
[4]Department of Statistics, University of Leeds, UK

Abstract

Geostatistical modelling is widely used to describe data with spatial dependence structure. Such modelling often assumes a Gaussian distribution, an assumption which is frequently violated due to the asymmetric nature of variables in diverse applications. The Birnbaum-Saunders distribution is asymmetrical and has several appealing properties, including theoretical arguments for describing chemical data. This work examines a Birnbaum-Saunders spatial regression model and derives global and local diagnostic methods to assess the influence of atypical observations on the maximum likelihood estimates of its parameters. Modelling and diagnostic methods are then applied to experimental data describing the spatial distribution of magnesium and calcium in the soil in the Parana state of Brazil. This application shows the importance of such a diagnostic analysis in spatial modelling with chemical data.

Keywords Chemical data analysis; global and local influence; Matérn model; maximum likelihood methods; non-normality; R software.

1 Introduction

Statistical distributions are largely used to model variables studied across a wide range of applications. These distributions may help to determining the expectation and covariance of data with spatial dependence. Such data can be related to geochemical variables as fertilizer content for agricultural management, where deficiency and imbalance of nutrients in the soil are important aspects to be studied (De Bastiani et al., 2015; Garcia-Papani et al., 2017). Due to their inherent variation, the geochemical variables are considered as random and often following asymmetric statistical distributions with positive skewness. Hence, the Gaussian distribution is inappropriate and should not be used to model this type of random variable, which must be checked (Barros et al., 2014; Stehlík et al., 2014).

There is little literature investigating the use of asymmetric distributions to analyse spatial data. Some of these few works are attributed to Allard and Naveau (2007) and Rimstad and Omre (2014), who employed the skew-normal distribution to model random fields. However, data to be modeled spatially often have support on the positive real line and then these distributions are inappropriate. Consequently, the Birnbaum-Saunders (BS), exponential, gamma, inverse Gaussian, log-normal and Weibull distributions might be more suitable (Johnson et al., 1994, 1995; Leiva, 2016). In particular, the BS distribution

has attractive properties, which are useful for modelling asymmetric data, specially in geochemistry. For example, BS and normal distributions have a close relationship, but the BS model is defined in an asymmetric framework, which is a good motivation for further investigations. Other properties of the BS distribution are related to proportionality, reciprocity, diverse shapes for its hazard rate, and its membership of the log-symmetry class. All of these properties provide wide flexibility for modelling different phenomena and types of data, as well as an attraction for the BS distribution. In particular, the log-symmetry class of distributions arises when a random variable has the same distribution as its reciprocal, such as the log-normal distribution, but the BS distribution has an associated logarithmic version (log-BS) which allows for bimodality and the log-normal distribution not; see more details of these properties in Johnson et al. (1995), Leiva (2016) and Leão et al. (2017). In addition, the BS distribution has been widely studied and applied across different fields, including geochemistry, which have been carried by an international, transdisciplinary group of researchers; see, example, Leiva et al. (2011, 2015, 2016a), Ferreira et al. (2012), Marchant et al. (2013) and Saulo et al. (2013). Despite its origins in material fatigue, Leiva et al. (2015) confirmed the BS distribution as an adequate model to describe data from chemical and environmental sciences using the proportionate-effect law. In particular, Xia et al. (2011) and Garcia-Papani et al. (2017) introduced BS spatial models. However, both of these works only deal with the spatial problem, whereas the extension to spatial regression models, allowing the inclusion of explanatory variables (covariates hereafter), has not been considered to date.

Fitting a distribution to spatial data is useful in several areas (Cressie, 2015). For example, Cambardella et al. (1994) indicated that, in the case of soil properties, it can be utilized to improve agricultural management practices. Specifically, modelling the expectation and covariance of data allows spatial dependence parameters to be estimated and this dependence to be quantified. Such a modelling may be characterized by the variogram, whereas the response variable (response hereafter), described by regression with covariates, may be modeled using Kriging (De Bastiani et al., 2015). For example, Hengl et al. (2004) reported successful results when regression-Kriging is considered.

An essential step in all statistical modelling is the diagnostic analysis employed to detect the influence of atypical cases on the parameter estimates (Green and Kalivas, 2002). Diagnostics can be carried out using global or local influence. Global influence is often conducted by case-deletion methods. Two well-know case-deletion methods are the Cook distance (CD) and the likelihood distance or displacement (LD) (Cook et al., 1988). It is important to note that Cook-type approaches do not always correctly identify influential points (Fung, 1995; Kim, 2013) and hence it is important to consider several contrasting measures, as we are proposing in this work. For spatial data, Militino et al. (2006) considered global influence for multivariate spatial linear models. However, single case-deletion cannot detect jointly influential cases. Instead, the local influence method may be used, which allows assessment of combined influence of cases (Cook, 1987) . Local influence typically examines the normal curvature of an LD after perturbing the model and/or data. Since the work of Cook (1987), many authors have considered the local influence method. Additive perturbations in local influence for Gaussian linear models have been extended to more general models by Galea et al. (2003) and Leiva et al. (2016b). Studies of local influence in BS models were conducted by a number of authors; see, for example, Santana et al. (2011), Leiva et al. (2014a,b), Marchant et al. (2016b), Garcia-Papani et al. (2017), Saulo et al. (2017) and Desousa et al. (2018). For spatial models, Uribe-Opazo et al. (2012) derived local influence measures in Gaussian spatial models, whereas Assumpção et al. (2014) considered the Student-t case. For non-additive perturbations, Zhu et al. (2007) proposed a method to find the most appropriate perturbation for a specific model, whereas Gimenez and Galea (2013) applied this method to functional heteroscedastic measurement-error models. Although the normal curvature approach proposed by Cook (1987) has been extensively used, the conformal curvature proposed by Poon and Poon (1999) is more efficient, because

the conformal approach is normalized and invariant under certain reparameterizations. De Bastiani et al. (2015) utilized conformal curvature to study local influence in non-normal spatial models.

The main objective of this article is to derive diagnostic methods for BS spatial regression models. On the one hand, assessment of global influence uses the single case-deletion method proposed by Pan et al. (2014). On the other hand, the local influence method is also considered, making perturbations in the model by means of the response and some continuous covariate, employing the scheme suggested by Zhu et al. (2007). The potential influence of a case is quantified using conformal curvature and its associated measures. Furthermore, a generalised leverage (GL) study is conducted, which evaluates the influence of the observed response on its estimated value (Leiva et al., 2014b). The diagnostic study is applied to experimental spatial data collected by the authors related to contents of magnesium and calcium in the soil.

The paper is organized as follows. In Section 2, background to the BS and log-BS distributions is provided. In this section, the BS spatial regression model is also introduced and the maximum likelihood (ML) method is considered for parameter estimation. Section 3 presents concepts related to diagnostic analysis and discusses how to select the most appropriate perturbation scheme for the BS spatial regression model. Evaluation of the model by analysing experimental data is carried out in Section 4. Finally, Section 5 gives some conclusions and possible future research.

2 The Birnbaum-Saunders spatial model

2.1 The Birnbaum-Saunders distribution

As mentioned, the BS distribution has its origins in material fatigue analysis based on the Miner law. Details of its technical derivation using this law are available in Leiva (2016, pp. 5-11). However, as also mentioned, despite its origins in material fatigue, the BS distribution can be utilized to model chemical data. Details of its justification and theoretical derivation based on the proportionate-effect law are available in Leiva et al. (2015).

The BS distribution is unimodal and has shape ($\alpha > 0$) and scale ($\sigma > 0$) parameters, in addition to asymmetry to the right with positive support. Note that σ is also a location parameter, because it corresponds to the median of the distribution. The random variable

$$T = \sigma \left(\alpha Z_1/2 + \sqrt{(\alpha Z_1/2)^2 + 1} \right)^2$$

is said to follow a BS distribution with parameters α and σ if $Z_1 = (1/\alpha)\xi(T/\sigma) \sim \mathrm{N}(0,1)$, with $\xi(u) = \sqrt{u} - 1/\sqrt{u} = 2\sinh(\log(\sqrt{u}))$, which is denoted by $T \sim \mathrm{BS}(\alpha, \sigma)$. The corresponding cumulative distribution function (CDF) is given by

$$F_T(t; \alpha, \sigma) = \mathrm{P}(T \leq t) = \Phi\left(\frac{1}{\alpha}\xi(t/\sigma)\right), \quad t > 0, \tag{1}$$

where Φ is the standard normal CDF. The qth quantile of T is $t(q; \alpha, \sigma) = \sigma(\alpha z(q)/2 + \sqrt{(\alpha z(q)/2)^2 + 1})^2$, for $0 < q < 1$, where $z(q)$ is the qth quantile of the standard normal distribution. Thus, if $q = 0.5$, then $t(0.5) = \sigma$ and, as mentioned, σ is the median of the BS distribution. Note that the probability density function (PDF) of the BS distribution can be directly obtained from its CDF defined in (1). The BS PDF has increasing positive skewness as α increases and is approximately symmetrical around σ as α goes to zero; see examples of its diverse shapes in Figure 1 (left).

Some properties of $T \sim \text{BS}(\alpha, \sigma)$ are: (P1) $W = (1/\alpha^2)\xi^2(T/\sigma) \sim \chi^2(1)$; (P2) $r\,T \sim \text{BS}(\alpha,\, r\sigma)$, for $r > 0$; (P3) $1/T \sim \text{BS}(\alpha, 1/\sigma)$, that is, the BS distribution is related to the normal and chi-square distributions, as well as belongs to the scale (proportionality) and closed under reciprocation (reciprocity) families. In addition, (P4) the failure rate of the BS distribution admits several shapes, including the unimodal and bimodal cases; see details in Azevedo et al. (2012) and Athayde et al. (2018). Furthermore, (P5) the mean, variance and coefficients of variation (CV), skewness (CS) and kurtosis (CK) of $T \sim \text{BS}(\alpha, \sigma)$ are, respectively, defined as

$$\begin{aligned} \text{E}(T) &= \frac{\sigma}{2}(2+\alpha^2), \quad \text{Var}(T) = \frac{\sigma^2}{4}(4\alpha^2+5\alpha^4), \quad \text{CV}(T) = \frac{\sqrt{4\alpha^2+5\alpha^4}}{2+\alpha^2}, \\ \text{CS}(T) &= \frac{24\alpha+44\alpha^3}{\sqrt{(4+5\alpha^2)^3}}, \quad \text{CK}(T) = 3 + \frac{240\alpha^2+558\alpha^4}{(4+5\alpha^2)^2}. \end{aligned}$$

Modelling based on the BS distribution is often described in terms of the log-BS distribution. Note that a continuous random variable Y follows a log-BS distribution with shape parameter $\alpha > 0$ and location parameter $\mu \in \mathbb{R}$, which is denoted by log-BS(α, μ), if and only if

$$Z_2 = \frac{2}{\alpha}\sinh\left(\frac{Y-\mu}{2}\right) \sim \text{N}(0,1).$$

Thus, the CDF of Y is expressed as

$$F_Y(y;\alpha,\mu) = \Phi\left(\frac{2}{\alpha}\sinh\left(\frac{y-\mu}{2}\right)\right), \quad -\infty < y < +\infty, -\infty < \mu < +\infty, \alpha > 0. \tag{2}$$

As in the BS distribution, note that the PDF of the log-BS distribution can be directly obtained from its CDF defined in (2). If $Y \sim$ log-BS(α, μ), then the following properties hold: (P6) $T = \exp(Y) \sim \text{BS}(\alpha, \sigma)$, with $\log(\sigma) = \mu$, which means that the log-BS PDF obtained from (2) may be established from the standard normal PDF or from the BS PDF defined in (1); (P7) $\text{E}\,(Y) = \mu$, but no closed form for the variance of Y is possible, although using an asymptotic approximation for its moment generating function, we have that, if $\alpha \to 0$, $\text{Var}(T) = \alpha^2 - \alpha^4/4$, whereas that if $\alpha \to +\infty$, $\text{Var}(T) = 4(\log^2(\sqrt{2}\alpha) + 2 - 2\log(\sqrt{2}\alpha))$; (P8) if $X = \pm Y + d$, then $X \sim$ log-BS$(\alpha, \pm\mu + d)$; (P9) the log-BS distribution is symmetric around μ, unimodal for $\alpha \leq 2$ and bimodal for $\alpha > 2$; and (P10) it belongs to the class of log-symmetric distributions, which also includes the log-normal distribution. This class contains all random variables which have the same distribution as their reciprocal; for more details, see Leão et al. (2017). Figure 1 (right) shows some shapes for the log-BS PDF.

2.2 Formulation of the spatial model

In order to model a set of spatially correlated data, consider a stochastic process $\boldsymbol{T} \equiv \{T(\boldsymbol{s}), \boldsymbol{s} \in D\}$, defined on some region $D \subset \mathbb{R}^2$. It is assumed that the stochastic process $\boldsymbol{T}$ is stationary and isotropic, and that, for given spatial locations $\boldsymbol{s}_i$, with $i = 1, \ldots, n$, the response can be modeled by

$$T(\boldsymbol{s}_i) = \exp(\mu(\boldsymbol{s}_i))\,\eta(\boldsymbol{s}_i), \quad i = 1, \ldots, n, \tag{3}$$

where $\mu(\boldsymbol{s}_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta}$, with $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})^\top$ being a $p \times 1$ vector which contains the values of the covariates. Here, $x_{i1} = 1$, for $i = 1, \ldots, n$, and $x_{ij} = x_j(\boldsymbol{s}_i)$, for $j = 2, \ldots, p$, that is, x_{ij} is the value of the covariate X_j at the location $\boldsymbol{s}_i$. In addition, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ is a $p \times 1$ vector of unknown regression

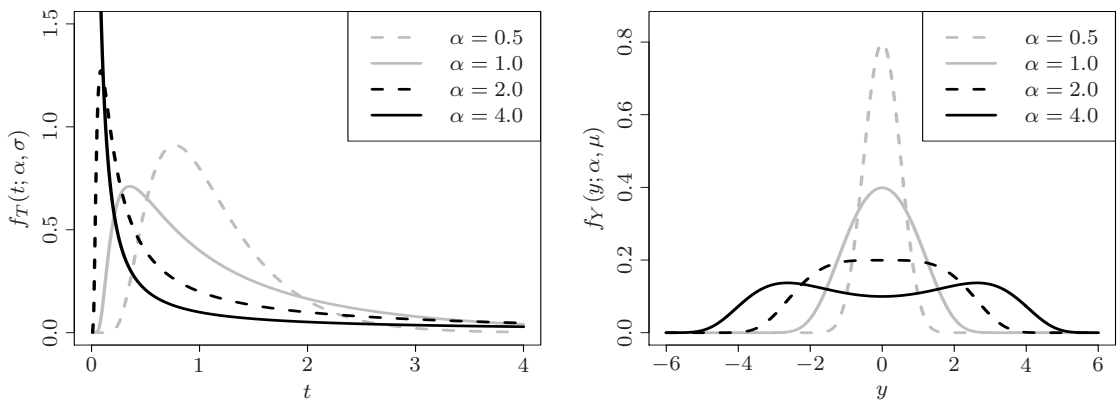

Figure 1: PDF of BS$(\alpha, 1)$ (left) and log-BS$(\alpha, 0)$ (right) distributions for the indicated value of α.

coefficients to be estimated. Furthermore, $\eta(\boldsymbol{s}_i)$ is the model random error at $\boldsymbol{s}_i$, for which we assume that $\boldsymbol{\eta} = (\eta(\boldsymbol{s}_1), \ldots, \eta(\boldsymbol{s}_n))^\top \sim \text{BS}_n(\alpha \mathbf{1}_{n\times 1}, \mathbf{1}_{n\times 1}; \boldsymbol{\Sigma})$, with $\alpha > 0$, where $\mathbf{1}_{n\times 1}$ is an $n \times 1$ vector of ones and $\boldsymbol{\Sigma}$ is an $n \times n$ (non-singular) scale-dependence matrix. More details about the multivariate BS distribution can be found in Kundu et al. (2013). The spatial model defined in (3) may be linearized by applying a logarithmic transformation as

$$Y(\boldsymbol{s}_i) = \log(T(\boldsymbol{s}_i)) = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \varepsilon(\boldsymbol{s}_i), \quad i = 1, \ldots, n, \tag{4}$$

where $\varepsilon(\boldsymbol{s}_i) = \log(\eta(\boldsymbol{s}_i))$. In matrix notation, the BS spatial log-linear regression model given in (4) can be written as

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{5}$$

where $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top$ is an $n \times 1$ vector of responses, with $Y_i = \log(T(\boldsymbol{s}_i))$; $\boldsymbol{X} = (\boldsymbol{x}_1^\top, \ldots, \boldsymbol{x}_n^\top)^\top$ is an $n \times p$ design matrix; and $\boldsymbol{\varepsilon} = (\varepsilon(\boldsymbol{s}_1), \ldots, \varepsilon(\boldsymbol{s}_n))^\top$ is an $n \times 1$ vector of random errors of the stationary process. Then, $\boldsymbol{\varepsilon} \sim \text{log-BS}_n(\alpha \mathbf{1}_{n\times 1}, \mathbf{0}_{n\times 1}; \boldsymbol{\Sigma})$, where $\mathbf{0}_{n\times 1}$ is an $n \times 1$ vector of zeros, and therefore $\text{E}(\boldsymbol{\varepsilon}) = \mathbf{0}_{n\times 1}$. More details about the multivariate log-BS distribution can be found in Marchant et al. (2016a, 2018) and Garcia-Papani et al. (2017).

It is also assumed that the spatial dependence is determined by an $n \times n$ scale matrix $\boldsymbol{\Sigma}$, which is symmetric, non-singular and positive definite. Note that $\boldsymbol{\Sigma}$ is proportional to $\text{Cov}\,(\varepsilon(\boldsymbol{s}_i), \varepsilon(\boldsymbol{s}_j))$, which depends only on the Euclidean distance between the locations $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$, that is,

$$\text{Cov}\,(\varepsilon(\boldsymbol{s}_i), \varepsilon(\boldsymbol{s}_j)) = \rho(h_{ij}), \tag{6}$$

where $h_{ij} = \|\boldsymbol{s}_i - \boldsymbol{s}_j\|$. In addition, it is assumed that the function ρ defined in (6) is expressed in terms of the spatial dependence parameter vector $\boldsymbol{\varphi} = (\varphi_1, \varphi_2, \varphi_3)^\top$ established in the relation

$$\boldsymbol{\Sigma} = \varphi_1 \boldsymbol{I}_n + \varphi_2 \boldsymbol{R}(\varphi_3), \tag{7}$$

where φ_1 is a nugget effect, φ_2 is a scale parameter (sill), φ_3 is a function of the spatial dependence radius, $\boldsymbol{I}_n$ is the $n \times n$ identity matrix, and $\boldsymbol{R}(\varphi_3) = (r_{ij})$ is an $n \times n$ symmetric matrix, with main diagonal elements equal to one. Note that $\boldsymbol{R}(\varphi_3)$ depends on the theoretical covariance function adopted to describe the spatial dependence (Mardia and Marshall, 1984). In the Matérn family model (Diggle

and Ribeiro, 2007), r_{ij} is given by

$$r_{ij} = \begin{cases} 1, & i = j; \\ \dfrac{1}{2^{\delta-1}\Gamma(\delta)} \left(\dfrac{h_{ij}}{\varphi_3}\right)^{\delta} K_{\delta}\left(\dfrac{h_{ij}}{\varphi_3}\right), & i \neq j; \end{cases} \tag{8}$$

where δ is a shape parameter, Γ is the standard gamma function and K_δ is the modified Bessel function of third kind with order δ. Particular cases of the Matérn family model include the exponential and Gaussian members, when $\delta = 0.5$ and $\delta \to +\infty$, respectively.

2.3 Parameter estimation

The parameters of the model defined in (5) are summarized in the vector $\boldsymbol{\theta} = \left(\boldsymbol{\beta}^\top, \boldsymbol{\varphi}^\top, \alpha\right)^\top$, which are unknown but can be estimated by the ML method as follows. From (5), we have $\boldsymbol{\varepsilon} = \left(\boldsymbol{Y} - \boldsymbol{X}^\top\boldsymbol{\beta}\right) \sim$ log-BS$_n$ $(\alpha\mathbf{1}_{n\times 1}, \mathbf{0}_{n\times 1}; \boldsymbol{\Sigma})$. Then, the logarithm of the likelihood (log-likelihood) function for $\boldsymbol{\theta}$, ignoring constant terms, is given by

$$\ell(\boldsymbol{\theta}) = -\frac{1}{2}\log\left(|\boldsymbol{\Sigma}|\right) - n\log(\alpha) - \frac{2}{\alpha^2}\boldsymbol{V}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{V} + \sum_{i=1}^{n}\log\left(\cosh\left(\frac{y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta}}{2}\right)\right), \tag{9}$$

where $\boldsymbol{V} = (V_1, \ldots, V_n)^\top$ is an $n \times 1$ vector, with elements $V_i = \sinh((y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta})/2)$, for $i = 1, \ldots, n$, and $\boldsymbol{\Sigma}$ is given in (7). Taking the derivative of (9), with respect to the corresponding parameters, leads to the $(p+4) \times 1$ score vector defined as

$$\dot{\boldsymbol{\ell}}\left(\boldsymbol{\theta}\right) = \left(\left(\frac{\partial\ell(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}}\right)^\top, \left(\frac{\partial\ell(\boldsymbol{\theta})}{\partial\boldsymbol{\varphi}}\right)^\top, \frac{\partial\ell(\boldsymbol{\theta})}{\partial\alpha}\right)^\top = \left(\dot{\ell}(\beta_1), \ldots, \dot{\ell}(\beta_p), \dot{\ell}(\varphi_1), \dot{\ell}(\varphi_2), \dot{\ell}(\varphi_3), \dot{\ell}(\alpha)\right)^\top. \tag{10}$$

For details of the elements of the score vector given in (10), see Appendix A. In order to find the ML estimate $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$, the non-linear system $\dot{\boldsymbol{\ell}}(\boldsymbol{\theta}) = \mathbf{0}_{(p+4)\times 1}$, obtained from (10), must be solved. Since this system does not provide a closed analytical solution, $\widehat{\boldsymbol{\theta}}$ must be computed using an iterative procedure for non-linear systems (Nocedal and Wright, 1999). The `optim` function of the `R` software can be employed to solve such a system (`www.R-project.org`, R Core Team, 2016). By default, the `optim` command carries out a minimization, so that the instruction `control = list(fnscale = -1)` must be added to this command for maximization. Note that such an iterative procedure does not necessary converge to a maximum, which can be an inflection point or a minimum. We compute the eigenvalues of the Hessian matrix in order to assure that a maximum has been found.

2.4 Information matrix

Observe that the Hessian matrix $\ddot{\boldsymbol{\ell}}(\boldsymbol{\theta})$, for the BS spatial regression model presented in (5), is a $(p+4) \times (p+4)$ diagonal block matrix. The Hessian matrix is obtained by taking the second derivative of (9), with respect to the corresponding parameters, and is given by

$$\ddot{\boldsymbol{\ell}}(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} & \frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\varphi}^\top} & \frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\alpha} \\ \frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\boldsymbol{\varphi}\partial\boldsymbol{\beta}^\top} & \frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\boldsymbol{\varphi}\partial\boldsymbol{\varphi}^\top} & \frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\boldsymbol{\varphi}\partial\alpha} \\ \frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\alpha\partial\boldsymbol{\beta}^\top} & \frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\alpha\partial\boldsymbol{\varphi}^\top} & \frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\alpha^2} \end{pmatrix} = \begin{pmatrix} \ddot{\boldsymbol{\ell}}(\boldsymbol{\beta}) & \ddot{\boldsymbol{\ell}}(\boldsymbol{\beta}\boldsymbol{\varphi}) & \ddot{\boldsymbol{\ell}}(\boldsymbol{\beta}\alpha) \\ \ddot{\boldsymbol{\ell}}(\boldsymbol{\varphi}\boldsymbol{\beta}) & \ddot{\boldsymbol{\ell}}(\boldsymbol{\varphi}) & \ddot{\boldsymbol{\ell}}(\boldsymbol{\varphi}\alpha) \\ \ddot{\boldsymbol{\ell}}(\alpha\boldsymbol{\beta}) & \ddot{\boldsymbol{\ell}}(\alpha\boldsymbol{\varphi}) & \ddot{\ell}(\alpha) \end{pmatrix}, \tag{11}$$

where the $p \times p$, $p \times 3$ and 3×3 sub-matrices $\ddot{\boldsymbol{\ell}}(\boldsymbol{\beta})$, $\ddot{\boldsymbol{\ell}}(\boldsymbol{\beta\varphi}) = (\ddot{\boldsymbol{\ell}}(\boldsymbol{\varphi\beta}))^\top$ and $\ddot{\boldsymbol{\ell}}(\boldsymbol{\varphi})$, respectively, have elements detailed in Appendix A. Therefore, for the BS spatial regression model, the $(p+4) \times (p+4)$ expected Fisher information matrix, obtained from (11), is expressed as

$$\boldsymbol{K}(\boldsymbol{\theta}) = \mathrm{E}(-\ddot{\boldsymbol{\ell}}(\boldsymbol{\theta})) = \begin{pmatrix} \boldsymbol{K}(\boldsymbol{\beta}) & \boldsymbol{K}(\boldsymbol{\beta\varphi}) & \boldsymbol{K}(\boldsymbol{\beta}\alpha) \\ \boldsymbol{K}(\boldsymbol{\varphi\beta}) & \boldsymbol{K}(\boldsymbol{\varphi}) & \boldsymbol{K}(\boldsymbol{\varphi}\alpha) \\ \boldsymbol{K}(\alpha\boldsymbol{\beta}) & \boldsymbol{K}(\alpha\boldsymbol{\varphi}) & K(\alpha) \end{pmatrix}, \tag{12}$$

where $\boldsymbol{K}(\boldsymbol{\beta}) = \mathrm{E}(-\ddot{\boldsymbol{\ell}}(\boldsymbol{\beta}))$, $\boldsymbol{K}(\boldsymbol{\varphi}) = \mathrm{E}(-\ddot{\boldsymbol{\ell}}(\boldsymbol{\varphi}))$ and $\boldsymbol{K}(\boldsymbol{\beta\varphi}) = (\boldsymbol{K}(\boldsymbol{\varphi\beta}))^\top = \mathrm{E}(-\ddot{\boldsymbol{\ell}}(\boldsymbol{\beta\varphi})) = \mathbf{0}_{p\times 3}$ are $p \times p$, 3×3 and $p \times 3$ sub-matrices, whereas $\boldsymbol{K}(\boldsymbol{\beta}\alpha) = (\boldsymbol{K}(\alpha\boldsymbol{\beta}))^\top = \mathrm{E}(-\ddot{\boldsymbol{\ell}}(\boldsymbol{\beta}\alpha)) = \mathbf{0}_{p\times 1}$ and $\boldsymbol{K}(\boldsymbol{\varphi}\alpha) = (\boldsymbol{K}(\alpha\boldsymbol{\varphi}))^\top = \mathrm{E}(-\ddot{\boldsymbol{\ell}}(\boldsymbol{\varphi}\alpha))$ are $p \times 1$ and 3×1 vectors, respectively, whose elements also are detailed in Appendix A.

2.5 Inference and asymptotic frameworks

Recalling that $\boldsymbol{\theta} = \left(\boldsymbol{\beta}^\top, \boldsymbol{\varphi}^\top, \alpha\right)^\top$ and as usual for ML estimators, note that

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \quad \stackrel{D}{\to} \quad \mathrm{N}_{p+4}(\mathbf{0}_{(p+4)\times 1}, \boldsymbol{J}(\boldsymbol{\theta})^{-1}), \tag{13}$$

as $n \to +\infty$, where $\stackrel{D}{\to}$ denotes convergence in distribution and $\boldsymbol{J}(\boldsymbol{\theta}) = \lim_{n\to+\infty}(1/n)\boldsymbol{K}(\boldsymbol{\theta})$, with $\boldsymbol{K}(\boldsymbol{\theta})$ being the expected Fisher information matrix given in (12). Details of the asymptotic behaviour and performance of ML estimators will, of course, depend on the optimal design. This has been studied in the case of Gaussian models (Baran et al., 2015), but it is an open question, worthy of further investigations, for the case of BS models. The results presented in (12) and (13) can be used to find asymptotic standard errors of the estimators from the inverse of the expected Fisher information matrix and to carry out asymptotic inference on the model parameters. However, to infer on the spatial parameters, asymptotic properties of their estimators must be known, primarily because of their approximations in the case of finite-samples. Nevertheless, applicability of asymptotic frameworks to spatial data is not an easy aspect, due to there being at least two relevant frameworks, which can behave quite differently when estimating the spatial dependence parameters (Zhang and Zimmerman, 2005). One of these asymptotic frameworks is called "increasing-domain", which requires D, defined in Section 2.2, to tend to $\mathbb{R}^2$ in an anisotropic fashion (same speed in all directions), with a constant density of sample points. The other asymptotic framework is called "fixed-domain" (or "in-fill"), where the domain D is fixed, and the density of points tends to $+\infty$. The increasing-domain and fixed-domain asymptotic frameworks were derived to obtain limiting distributions of ML estimators of the spatial dependence parameters in Gaussian spatial models. As mentioned, the asymptotic properties of the ML estimators for these parameters may be different in the two frameworks. In general, according to Zhang and Zimmerman (2005), the spatial dependence parameters are not consistently estimated upon a fixed-domain asymptotic framework (Zhang, 2004; Stein, 2012), whereas upon an increasing-domain asymptotic framework, the parameters are consistently estimated and their ML estimators are asymptotically normal distributed, subject to regularity conditions (Mardia and Marshall, 1984). In addition, parameters can be consistently estimated in both asymptotic frameworks for some cases, but their convergence rates are different. Zhang and Zimmerman (2005) suggested to use the fixed-domain asymptotic framework.

3 Influence diagnostics

3.1 Global influence

Cook et al. (1988) proposed case-deletion based on the LD corresponding to the influence measure

$$\mathrm{LD}_i(\boldsymbol{\theta}) = 2(\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widehat{\boldsymbol{\theta}}_{(i)})), \quad i = 1, \ldots, n, \tag{14}$$

where ℓ is the log-likelihood function, whereas $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\theta}}_{(i)}$ are, respectively, the ML estimates of $\boldsymbol{\theta}$ considering the full data set and the data set with the case i removed. Note that (14) may be used to evaluate the global potential influence of the case i, that is, (14) measures the change in the LD with estimated parameters when the case i is deleted. Another case-deletion method alternative to the LD given in (14) is the CD, which has been generalised to several non-normal models (Desousa et al., 2018). To facilitate the calculations, the first order approximation defined as $\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_{(i)} \approx \ddot{\boldsymbol{\ell}}_{(i)}^{-1}(\widehat{\boldsymbol{\theta}})\dot{\boldsymbol{\ell}}_{(i)}(\widehat{\boldsymbol{\theta}})$, based on a Taylor expansion around $\widehat{\boldsymbol{\theta}}$ until the second order term, and the one-step-late Newton-Raphson estimate, can be employed for the CD. Thus, an alternative measure of global influence (Pan et al., 2014) based on the CD is given by

$$\mathrm{CD}_i(\boldsymbol{\theta}) = (\dot{\boldsymbol{\ell}}_{(i)}(\widehat{\boldsymbol{\theta}}))^\top(-\ddot{\boldsymbol{\ell}}_{(i)}(\widehat{\boldsymbol{\theta}}))^{-1}(\dot{\boldsymbol{\ell}}_{(i)}(\widehat{\boldsymbol{\theta}})), \quad i = 1, \ldots, n,$$

where $\dot{\boldsymbol{\ell}}_{(i)}(\boldsymbol{\theta}) = \partial\ell_{(i)}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ and $\ddot{\boldsymbol{\ell}}_{(i)}(\boldsymbol{\theta}) = \partial\ell^2_{(i)}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$, with $\ell_{(i)}$ being the log-likelihood function obtained after deleting the case i. Pan et al. (2014) showed that it is possible to replace $\ddot{\boldsymbol{\ell}}_{(i)}(\widehat{\boldsymbol{\theta}})$ by $\ddot{\boldsymbol{\ell}}(\widehat{\boldsymbol{\theta}})$ or by $\boldsymbol{K}(\boldsymbol{\theta}) = -\mathrm{E}(\ddot{\boldsymbol{\ell}}(\boldsymbol{\theta}))$, obtaining, respectively, the measures of influence

$$\mathrm{CD}_i(\boldsymbol{\theta}) = (\dot{\boldsymbol{\ell}}_{(i)}(\widehat{\boldsymbol{\theta}}))^\top(-\ddot{\boldsymbol{\ell}}(\widehat{\boldsymbol{\theta}}))^{-1}(\dot{\boldsymbol{\ell}}_{(i)}(\widehat{\boldsymbol{\theta}})) = (\dot{\boldsymbol{\ell}}_{(i)}(\widehat{\boldsymbol{\theta}}))^\top\boldsymbol{K}(\widehat{\boldsymbol{\theta}})^{-1}(\dot{\boldsymbol{\ell}}_{(i)}(\widehat{\boldsymbol{\theta}})), \quad i = 1, \ldots, n.$$

Note that $\mathrm{CD}_i(\boldsymbol{\theta})$ is often preferred to $\mathrm{LD}(\widehat{\boldsymbol{\theta}}_{(i)})$, because it reduces the computational burden. In the case of the BS spatial regression model, the matrix $\boldsymbol{K}(\boldsymbol{\theta})$ given in (12) has a diagonal block structure and then the CD for the vector $\boldsymbol{\theta}$ is given by

$$\begin{aligned}
\mathrm{CD}_i(\boldsymbol{\theta}) &= (\dot{\boldsymbol{\ell}}_{(i)}(\widehat{\boldsymbol{\theta}}))^\top\boldsymbol{K}(\widehat{\boldsymbol{\theta}})^{-1}(\dot{\boldsymbol{\ell}}_{(i)}(\widehat{\boldsymbol{\theta}})) \\
&= (\dot{\boldsymbol{\ell}}_{(i)}(\widehat{\boldsymbol{\beta}}))^\top\boldsymbol{K}(\widehat{\boldsymbol{\beta}})^{-1}(\dot{\boldsymbol{\ell}}_{(i)}(\widehat{\boldsymbol{\beta}})) + (\dot{\boldsymbol{\ell}}_{(i)}(\widehat{\boldsymbol{\psi}}))^\top\boldsymbol{K}(\widehat{\boldsymbol{\psi}})^{-1}(\dot{\boldsymbol{\ell}}_{(i)}(\widehat{\boldsymbol{\psi}})) \\
&= \mathrm{CD}_i(\boldsymbol{\beta}) + \mathrm{CD}_i(\boldsymbol{\psi}), \quad i = 1, \ldots, n,
\end{aligned} \tag{15}$$

where $\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\psi}}$ are the ML estimates of $\boldsymbol{\beta}, \boldsymbol{\psi} = \left(\boldsymbol{\varphi}^\top, \alpha\right)^\top$, whereas $\dot{\boldsymbol{\ell}}_{(i)}(\widehat{\boldsymbol{\beta}}), \dot{\boldsymbol{\ell}}_{(i)}(\widehat{\boldsymbol{\psi}})$ are subvectors of $\dot{\boldsymbol{\ell}}_{(i)}(\widehat{\boldsymbol{\theta}})$ related to the vectors $\boldsymbol{\beta}, \boldsymbol{\psi}$, respectively. In addition, $\boldsymbol{K}(\widehat{\boldsymbol{\beta}}), \boldsymbol{K}(\widehat{\boldsymbol{\psi}})$ are the blocks of the expected Fisher information matrix related to the vectors $\boldsymbol{\beta}, \boldsymbol{\psi}$ and evaluated at $\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\psi}}$, respectively. If the value $\mathrm{CD}_i(\boldsymbol{\theta})$ given in (15) is large, then the case i is potentially influential. There is no consensus about what values are considered large, but Cook and Weisberg (1982) stated that the definition of large depends on the problem. Analogously, a large value of $\mathrm{CD}_i(\boldsymbol{\beta})$ indicates that the case i is potentially influential in the estimation of $\boldsymbol{\beta}$, and similarly for $\mathrm{CD}_i(\boldsymbol{\psi})$.

3.2 Local Influence

Local Influence consists of studying the changes in the estimated parameters when making small perturbations in the data and/or the model assumptions. Cook (1987) evaluated local influence by examining

$$\mathrm{LD}(\boldsymbol{\theta}_{\boldsymbol{\omega}}) = 2(\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}})), \tag{16}$$

where $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$ are, respectively, the ML estimates of $\boldsymbol{\theta}$ in the proposed model and the model perturbed by $\boldsymbol{\omega}$. Specifically, Cook (1987) studied the normal curvature of the influence graph $\mathrm{LD}(\boldsymbol{\theta}_{\boldsymbol{\omega}})$, in the neighbourhood of the non-perturbation point, $\boldsymbol{\omega}_0$ namely, in the direction of a unit vector $\boldsymbol{d}$. Using differential geometry, it was shown that the normal curvature in the direction vector $\boldsymbol{d}$ takes the form $C_{\boldsymbol{d}} = 2|\boldsymbol{d}^\top \boldsymbol{B}\boldsymbol{d}|$, where

$$\boldsymbol{B} = -\boldsymbol{\Delta}^\top \ddot{\boldsymbol{\ell}}(\widehat{\boldsymbol{\theta}})^{-1}\boldsymbol{\Delta}, \tag{17}$$

with $\ddot{\boldsymbol{\ell}}(\widehat{\boldsymbol{\theta}})$ being the Hessian matrix, evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\Delta} = \partial^2\ell(\boldsymbol{\theta}|\boldsymbol{\omega})/\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^\top$ being the perturbation matrix, evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\omega} = \boldsymbol{\omega_0}$. Further details of the perturbation matrix are given in Appendix B. In addition, Cook (1987) stated that an important direction to consider is $\boldsymbol{d} = \boldsymbol{d}_{\max}$, which is related to the maximum normal curvature, $C_{\boldsymbol{d}_{\max}}$ namely, given by the largest absolute eigenvalue of the matrix $\boldsymbol{B}$, where $\boldsymbol{d}_{\max}$ is the eigenvector associated with this eigenvalue. Thus, the plot of the ith element (in absolute value) of $\boldsymbol{d}_{\max}$ versus the index i can detect points with the largest influence in the neighbourhood of $\mathrm{LD}(\boldsymbol{\theta}_{\boldsymbol{\omega}})|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0}$. Hence, such a potentially influential case may be responsible for considerable changes in the estimated parameters, under small perturbations of the data. Another important direction to assume is $\boldsymbol{d} = \boldsymbol{e}_i$, where $\boldsymbol{e}_i$ is a basis vector of $\mathbb{R}^n$, whose ith coordinate is one and the others are zero. In this case, the normal curvature is given by $C_i = 2|b_{ii}|$, where b_{ii} is the ith element on the diagonal of the matrix $\boldsymbol{B}$ defined in (17), for $i = 1, \ldots, n$. Similarly, the plot of C_i versus the index i can be used to identify potentially influential points.

Although the normal curvature of Cook (1987) is often employed, other measures of local influence have been studied. The conformal curvature of Poon and Poon (1999) is defined by

$$B_i = \frac{C_i}{\mathrm{tr}(\boldsymbol{B})}, \quad i = 1, \ldots, n, \tag{18}$$

whose calculation requires no more effort than the calculation of C_i. Note that (18) is invariant under conformal reparametrization. Then, the conformal curvature given in (18) is a standardized measure, making it easy to establish a cut-off point. As the curvatures C_i and B_i differ only by a positive constant, the eigenvector $\boldsymbol{d}_{\max}$ also provides the maximum conformal curvature $B_{\boldsymbol{d}_{\max}}$. Poon and Poon (1999) suggested that the ith element (in absolute value) associated with the vector $B_{\boldsymbol{d}_{\max}}$ presenting a value greater than $1/\sqrt{n}$ indicates it is a potentially influential point. With respect to conformal curvature in the direction of B_i defined in (18), Poon and Poon (1999) mentioned that the case i is potentially influential if $B_i > 2\overline{B}$, where $\overline{B}$ is the arithmetic mean of the basic conformal curvatures, that is, of $B_1, \ldots, B_n$. Thus, the case i is potentially influential if $B_i > \overline{B} + 2\mathrm{SD}(B)$, where $\mathrm{SD}(B)$ is the standard deviation (SD) of $B_1, \ldots, B_n$.

Another measure used for local influence, formally defined by Billor and Loynes (1993), is given by

$$\mathrm{LD}^\star(\boldsymbol{\theta}_{\boldsymbol{\omega}}) = -2(\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}|\boldsymbol{\omega})), \tag{19}$$

where $\ell(\boldsymbol{\theta}_{\boldsymbol{\omega}}|\boldsymbol{\omega})$ is the log-likelihood function perturbed by $\boldsymbol{\omega}$. Note that $\mathrm{LD}(\boldsymbol{\theta}_{\boldsymbol{\omega}})$ defined in (16) and $\mathrm{LD}^\star(\boldsymbol{\theta}_{\boldsymbol{\omega}})$ in (19) are different, because $\mathrm{LD}^\star(\boldsymbol{\theta}_{\boldsymbol{\omega}})$ has the first derivative with respect to $\boldsymbol{\omega}$, evaluated at $\boldsymbol{\omega}_0$, equal to zero. Cook (1987) used the normal curvature (second order derivative) to evaluate the local influence. Thus, the main advantage of $\mathrm{LD}^\star(\boldsymbol{\theta}_{\boldsymbol{\omega}})$, compared to $\mathrm{LD}(\boldsymbol{\theta}_{\boldsymbol{\omega}})$, is that the first derivative of $\mathrm{LD}^\star(\boldsymbol{\theta}_{\boldsymbol{\omega}})$ is not zero at $\boldsymbol{\omega}_0$. Therefore, the slope can be employed to analyse the local change. Notice that $\mathrm{LD}^\star(\boldsymbol{\theta}_{\boldsymbol{\omega}})$ considers the log-likelihood function based on the perturbed model and a negative sign multiplying it to assure its positivity. Thus, a potentially influential point is detected when studying the influence graph of the surface $a^\star(\boldsymbol{\omega}) = (\boldsymbol{w}^\top, \mathrm{LD}^\star(\boldsymbol{\theta}_{\boldsymbol{\omega}}))^\top$. Observe that its first derivative (slope) $S_{\boldsymbol{d}}$ in the direction vector $\boldsymbol{d}$ does not vanish, except in trivial cases. Particularly, the maximum slope and

the corresponding maximum direction vector, $S_{\max}$ and $\boldsymbol{d}_{\max}$, respectively, are useful in detecting local influence. The maximum direction of $\mathrm{LD}^{\star}(\boldsymbol{\theta}_{\boldsymbol{\omega}})$, evaluated at $\boldsymbol{\omega}_0$, is the direction of the gradient vector $\nabla\mathrm{LD}^{\star}(\boldsymbol{\theta}_{\boldsymbol{\omega}})$. Consequently, the maximum slope (Longford, 2005) is given by

$$S_{\max} = \|\nabla\mathrm{LD}^{\star}(\boldsymbol{\theta}_{\boldsymbol{\omega}})\| = 2\left\|\frac{\partial\ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\boldsymbol{\omega}}\right\|,$$

which must be evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\omega} = \boldsymbol{\omega}_0$. The approach of Billor and Loynes (1993) allows the individual components of the gradient vector $\nabla\mathrm{LD}^{\star}(\boldsymbol{\theta}_{\boldsymbol{\omega}})$ to be considered, and it can be used to identify those cases which have the largest contribution to the maximum slope. Thus, the plot of the ith element (in absolute value) associated with $S_{\max}$ versus the index i can also be employed to identify whether the case i is potentially influential. Because the method based on (19) only involves first derivatives, and not second derivatives, it is easier to utilize than the method based on (16).

Next, local influence results for the BS spatial regression model are provided, which consider perturbation schemes in the response and in one continuous covariate. Each scheme assumes the most appropriate perturbation, according to the methodology proposed by Zhu et al. (2007). Such perturbation schemes are used to assess the sensitivity of estimated parameters utilizing the influence measures: (i) conformal curvature in the direction of B_i, corresponding to maximum conformal curvature ($B_{\boldsymbol{d}_{\max}}$), and (ii) slope displacement of the modified likelihood function ($S_{\max}$). The perturbation matrices, for each case, are provided in Appendix B.

***Perturbation in the response*:** Consider the perturbation $\boldsymbol{Y}_{\boldsymbol{\omega}}(\boldsymbol{s}) = \boldsymbol{Y}(\boldsymbol{s}) + \boldsymbol{A}\boldsymbol{\omega}$, where $\boldsymbol{A}$ is a symmetric, non-singular matrix and $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^{\top} \in \mathbb{R}^n$ is the perturbation vector. Note that $\boldsymbol{\omega}_0 = \boldsymbol{0}_{n\times 1}$ is the non-perturbation vector. Then, the corresponding perturbed log-likelihood function, ignoring constant terms, is given by

$$\ell(\boldsymbol{\theta}|\boldsymbol{\omega}) = -\frac{1}{2}\log\left(|\boldsymbol{\Sigma}|\right) - n\log(\alpha) - \frac{2}{\alpha^2}\boldsymbol{V}_{\boldsymbol{\omega}}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{V}_{\boldsymbol{\omega}} + \sum_{i=1}^{n}\log\left(\cosh\left(\frac{y_i + \boldsymbol{A}_i\boldsymbol{\omega} - \boldsymbol{x}_i^{\top}\boldsymbol{\beta}}{2}\right)\right),$$

where $\boldsymbol{V}_{\boldsymbol{\omega}} = (V_{\omega_1}, \ldots, V_{\omega_n})^{\top}$, with $V_{\omega_i} = \sinh((y_i + \boldsymbol{A}_i\boldsymbol{\omega} - \boldsymbol{x}_i^{\top}\boldsymbol{\beta})/2)$, and $\boldsymbol{A}_i$ is the ith row of the matrix $\boldsymbol{A}$, for $i = 1\ldots, n$. According to Zhu et al. (2007), the perturbation $\boldsymbol{\omega}$ is appropriate if and only if $\boldsymbol{G}(\boldsymbol{\theta}|\boldsymbol{\omega}_0) = c\boldsymbol{I}_n$, where $c > 0$ and $\boldsymbol{G}(\boldsymbol{\theta}|\boldsymbol{\omega}) = \mathrm{E}(\dot{\boldsymbol{\ell}}(\boldsymbol{\theta}|\boldsymbol{\omega})\dot{\boldsymbol{\ell}}^{\top}(\boldsymbol{\theta}|\boldsymbol{\omega}))$, with $\dot{\boldsymbol{\ell}}(\boldsymbol{\theta}|\boldsymbol{\omega}) = \partial\ell(\boldsymbol{\theta}|\boldsymbol{\omega})/\partial\boldsymbol{\omega}$. For the BS spatial regression model, we have

$$\dot{\boldsymbol{\ell}}(\boldsymbol{\theta}|\boldsymbol{\omega}) = -\frac{2}{\alpha^2}\boldsymbol{A}\boldsymbol{\Sigma}^{-1}\boldsymbol{V}_{\boldsymbol{\omega}} + \frac{1}{2}\boldsymbol{A}\boldsymbol{V}_{\boldsymbol{\omega}}.$$

Therefore, it conducts to

$$\boldsymbol{G}(\boldsymbol{\theta}|\boldsymbol{\omega}) = \boldsymbol{A}\left(\frac{\alpha}{4}\boldsymbol{\Sigma}^{1/2} - \frac{1}{\alpha}\boldsymbol{\Sigma}^{-1/2}\right)^2\boldsymbol{A}.$$

Now, in order to use the approach by Zhu et al. (2007), consider

$$\boldsymbol{A} = \left(\frac{\alpha}{4}\boldsymbol{\Sigma}^{\frac{1}{2}} - \frac{1}{\alpha}\boldsymbol{\Sigma}^{-\frac{1}{2}}\right)^{-1}. \tag{20}$$

Thus, an appropriate perturbation scheme for the response is given by

$$\boldsymbol{Y}_{\boldsymbol{\omega}}(\boldsymbol{s}) = \boldsymbol{Y}(\boldsymbol{s}) + \left(\frac{\alpha}{4}\boldsymbol{\Sigma}^{1/2} - \frac{1}{\alpha}\boldsymbol{\Sigma}^{-1/2}\right)^{-1}\boldsymbol{\omega}.$$

***Perturbation in a continuous covariate*:** Next, consider perturbations in a single continuous covariate, which is assumed to be labelled X_t. The other covariates are not perturbed, that is, the perturbation scheme of the covariate is $\boldsymbol{x}_{t,\boldsymbol{\omega}}(\boldsymbol{s}) = \boldsymbol{x}_t(\boldsymbol{s}) + \boldsymbol{A}\boldsymbol{\omega}$ and $\boldsymbol{x}_{j,\boldsymbol{\omega}}(\boldsymbol{s}) = \boldsymbol{x}_j(\boldsymbol{s})$, for $j \neq t, j = 1, \ldots, p$, where $\boldsymbol{\omega} \in \mathbb{R}^n$ and $\boldsymbol{\omega}_0 = \mathbf{0}_{n\times 1}$ are such as in the response perturbation scheme. In this case, the perturbed log-likelihood function, ignoring constant terms, is given by

$$\ell(\boldsymbol{\theta}|\boldsymbol{\omega}) = -\frac{1}{2}\log\left(|\boldsymbol{\Sigma}|\right) - n\log(\alpha) - \frac{2}{\alpha^2}\boldsymbol{V}_{\boldsymbol{\omega}}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{V}_{\boldsymbol{\omega}} + \sum_{i=1}^{n}\log\left(\cosh\left(\frac{y_i - \boldsymbol{x}_{i,\boldsymbol{\omega}}^\top\boldsymbol{\beta}}{2}\right)\right),$$

where $\boldsymbol{V}_{\boldsymbol{\omega}} = (V_{\boldsymbol{\omega}_1}, \ldots, V_{\boldsymbol{\omega}_n})^\top$, with $V_{\boldsymbol{\omega}_i} = \sinh((y_i - \boldsymbol{x}_{i,\boldsymbol{\omega}}^\top\boldsymbol{\beta})/2)$, for $i = 1, \ldots n$. Thus, we have

$$\dot{\boldsymbol{\ell}}(\boldsymbol{\theta}|\boldsymbol{\omega}) = \frac{\partial\ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\boldsymbol{\omega}} = \frac{2\beta_t}{\alpha^2}\boldsymbol{A}\boldsymbol{\Sigma}^{-1}\boldsymbol{V}_{\boldsymbol{\omega}} - \frac{\beta_t}{2}\boldsymbol{A}\boldsymbol{V}_{\boldsymbol{\omega}},$$

and consequently

$$\boldsymbol{G}(\boldsymbol{\theta}|\boldsymbol{\omega}) = \mathrm{E}\left(\dot{\boldsymbol{\ell}}(\boldsymbol{\theta}|\boldsymbol{\omega})\dot{\boldsymbol{\ell}}(\boldsymbol{\theta}|\boldsymbol{\omega})^\top\right) = \beta_t^2\boldsymbol{A}\left(\frac{\alpha}{4}\boldsymbol{\Sigma}^{\frac{1}{2}} - \frac{1}{\alpha}\boldsymbol{\Sigma}^{-\frac{1}{2}}\right)^2\boldsymbol{A}$$

is obtained. Now, as in the case of the response perturbation with $\boldsymbol{A}$ given in (20), the most appropriate covariate perturbation scheme is given by

$$\boldsymbol{x}_{t,\boldsymbol{\omega}}(\boldsymbol{s}) = \boldsymbol{x}_t(\boldsymbol{s}) + \left(\frac{\alpha}{4}\boldsymbol{\Sigma}^{\frac{1}{2}} - \frac{1}{\alpha}\boldsymbol{\Sigma}^{-\frac{1}{2}}\right)^{-1}\boldsymbol{\omega}.$$

3.3 Generalised leverage

The leverage in linear regression models can be used to measure the influence that individual cases have on their predicted values (Leiva et al., 2014b). The GL matrix has the form

$$\mathrm{GL}(\boldsymbol{\theta}) = \frac{\partial\widehat{\boldsymbol{Y}}}{\partial\boldsymbol{Y}^\top} = \dot{\boldsymbol{D}}(\boldsymbol{\theta})\big(-\ddot{\boldsymbol{\ell}}(\boldsymbol{\theta})\big)^{-1}\ddot{\boldsymbol{\ell}}(\boldsymbol{\theta}\boldsymbol{Y}), \tag{21}$$

where $\dot{\boldsymbol{D}}(\boldsymbol{\theta}) = \partial\boldsymbol{\mu}/\partial\boldsymbol{\theta}^\top$, with $\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta}$ being the expected value of $\boldsymbol{Y}$, $-\ddot{\boldsymbol{\ell}}(\boldsymbol{\theta})$ is the observed information matrix and $\ddot{\boldsymbol{\ell}}(\boldsymbol{\theta}\boldsymbol{Y}) = \partial^2\ell(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{Y}^\top$. The element GL_{ii} of the matrix $\mathrm{GL}(\widehat{\boldsymbol{\theta}})$ is the instantaneous rate of change in the predicted value i with respect to its observed value. Thus, the main diagonal elements of the matrix $\mathrm{GL}(\widehat{\boldsymbol{\theta}})$ with large values indicate the leverage points, that is, points whose observed value has high influence on its predicted value. Here, it is proposed that the case i is potentially influence if $\mathrm{GL}_{ii} > \overline{\mathrm{GL}} + 2\mathrm{SD}(\mathrm{GL})$, where $\overline{\mathrm{GL}}$ and $\mathrm{SD}(\mathrm{GL})$ are the mean and SD of $\mathrm{GL}_{11}, \ldots, \mathrm{GL}_{nn}$, respectively. For the BS spatial regression model, the GL given in (21) defines its components as

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_{12} & \ldots & x_{1p} \\ 1 & x_{22} & \ldots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n2} & \ldots & x_{np} \end{pmatrix}; \quad \boldsymbol{\theta} = \left(\boldsymbol{\beta}^\top, \boldsymbol{\varphi}^\top, \alpha\right)^\top,$$

recalling that $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ and $\boldsymbol{\varphi} = (\varphi_1, \varphi_2, \varphi_3)^\top$, whereas the $n \times (p+4)$ matrix $\dot{\boldsymbol{D}}(\boldsymbol{\theta})$ is

$$\dot{\boldsymbol{D}}(\boldsymbol{\theta}) = \begin{pmatrix} 1 & x_{12} & \ldots & x_{1p} & 0 & 0 & 0 & 0 \\ 1 & x_{22} & \ldots & x_{2p} & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n2} & \ldots & x_{np} & 0 & 0 & 0 & 0 \end{pmatrix} = \left(\boldsymbol{X}, \mathbf{0}_{n\times 4}\right).$$

In addition, $\ddot{\boldsymbol{\ell}}(\boldsymbol{\theta}\boldsymbol{Y}) = (\ddot{\boldsymbol{\ell}}(\boldsymbol{\beta}\boldsymbol{Y}), \ddot{\boldsymbol{\ell}}(\boldsymbol{\varphi}\boldsymbol{Y}), \ddot{\boldsymbol{\ell}}(\alpha\boldsymbol{Y}))^\top$, where $\ddot{\boldsymbol{\ell}}(\boldsymbol{\beta}\boldsymbol{Y})$ is the $p \times n$ matrix with elements

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_j \partial y_i} = \frac{1}{\alpha^2}\left(\left(\boldsymbol{S}_i^\top \odot \boldsymbol{x}_j^\top\right)\boldsymbol{\Sigma}^{-1}\boldsymbol{V} + \left(\dot{\boldsymbol{\ell}}^\top \odot \boldsymbol{x}_j\right)\boldsymbol{\Sigma}^{-1}\boldsymbol{F}_i\right) - \frac{1}{4}\left(\text{sech}\left(\frac{y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta}}{2}\right)\right)^2 x_{ij},$$

where $\boldsymbol{S}_i$ and $\boldsymbol{F}_i$ are $n \times 1$ vectors, with elements $\sinh((y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta})/2)$ and $\cosh((y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta})/2)$, respectively, at the position i and zero in the other positions, for $i = 1, \ldots, n$ and $j = 1, \ldots, p$. Furthermore, $\ddot{\boldsymbol{\ell}}(\boldsymbol{\varphi}\boldsymbol{Y})$ is a $3 \times n$ matrix with elements defined as

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \varphi_j \partial y_i} = \frac{2}{\alpha^2}\boldsymbol{F}_i^\top\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j}\boldsymbol{\Sigma}^{-1}\boldsymbol{V}, \quad i = 1, \ldots, n, j = 1, 2, 3,$$

whereas $\ddot{\boldsymbol{\ell}}(\alpha\boldsymbol{Y})$ is a $1 \times n$ vector with elements expressed by

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \alpha \partial y_i} = \frac{4}{\alpha^3}\boldsymbol{F}_i^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{V}, \quad i = 1, \ldots, n.$$

4 Application to chemometrical data

4.1 Background

An important problem in agricultural management is the identification of imbalances and deficiencies of key nutrients in the soils. One such nutrient is magnesium (Mg), which is an elemental component of the chlorophyll molecule allowing plants to absorb energy from light and to combine water (H_2O) and carbon dioxide (CO_2), to produce sugar molecules. In turn, these are used to synthesise starch, protein, fat and vitamins. Deficiency in Mg also inhibits development of the root system, which reduces the absorption for other nutrients. Hence, low levels of Mg have a significant adverse effect on plant growth and vitality. In addition to decreased crop yield, inadequate levels of Mg also effects key properties, such as protein content in grains and flavour, colour, sweetness and tenderness in fruits and vegetables. When these deficiencies become visible, it is usually too late to make corrections, meaning that the whole crop year is effected dramatically reducing farming profitability. Thus, regular soil analysis is essential to allow intervention before symtoms are visible (Wolter, 2007). Although the controlling of individual soil nutrients is very important, monitoring the relationship between nutrients has been identified as equally important. For example, and to be considered here, it is well known that calcium (Ca) competes with Mg for absorption in the root system. That is, excess in levels of one may inhibit the absorption of the other and then restrict plant growth and vitality (Lopes, 1998).

4.2 Description and exploratory analysis of the spatial data

The data set corresponds to measurements taken at 82 locations during the crop year 2014/2015 within an area of approximately 167 ha. This area is located at Cascavel city, in the west of the state of Parana, Brazil. In general, the Brazilian soils are poor in nutrients and have Ca and Mg at very low levels. The response (T) is the content of Mg in the soil (cmolc/dm^3) and the covariate (X) is the content of Ca in the soil (cmolc/dm^3). The locations were georeferenced and the regular sampling grid can be seen in Figure 2 (left). The box-plot of Mg content data is displayed in Figure 2 (right), where five outliers are detected. Figure 3 (left) displays the sampled points divided by quartiles, with the locations of the outliers being again identified. The directional variogram in Figure 3 (right) shows that there is

no preferred direction, that is, an omni-directional semi-variogram is appropriate. Thus, the associated stochastic process can be considered as isotropic. To estimate the spatial dependence parameters, a variogram model in the Matérn family is assumed with $\delta = 0.25$, a value selected by cross-validation. Note that, although the variogram in Figure 3 (right) seems to be flat, indicating a possible lack of spatial structure, we must have in mind the following. Once the spatial parameters are estimated, it is possible to calculate the value of the relative nugget effect (RNE) as RNE $= \varphi_1/(\varphi_1 + \varphi_2)$, which indicates the degree of spatial dependence (Cambardella et al., 1994): if RNE < 0.25, the data present a strong spatial dependence; if $0.25 \leq$ RNE < 0.75, the data indicate an average spatial dependence; and if RNE ≥ 0.75, the data show a weak spatial dependence. In the case of the example considered here, we have $\widehat{\varphi}_1 = 0.0301$ and $\widehat{\varphi}_2 = 0.0156$. Therefore. $\widehat{\text{RNE}} = 0.6586$, indicating that the data present an average spatial dependence, which supports the use of the spatial model proposed in this study.

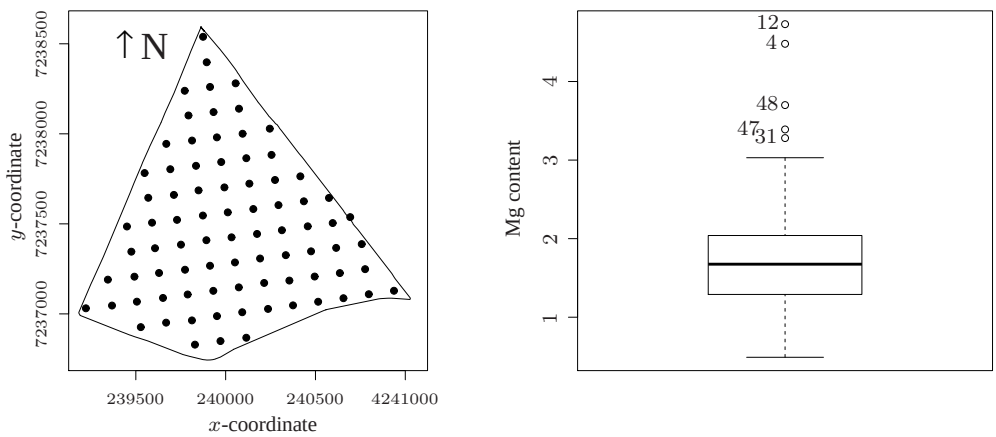

Figure 2: grid of sampled data (left) and box-plot of Mg content data (right).

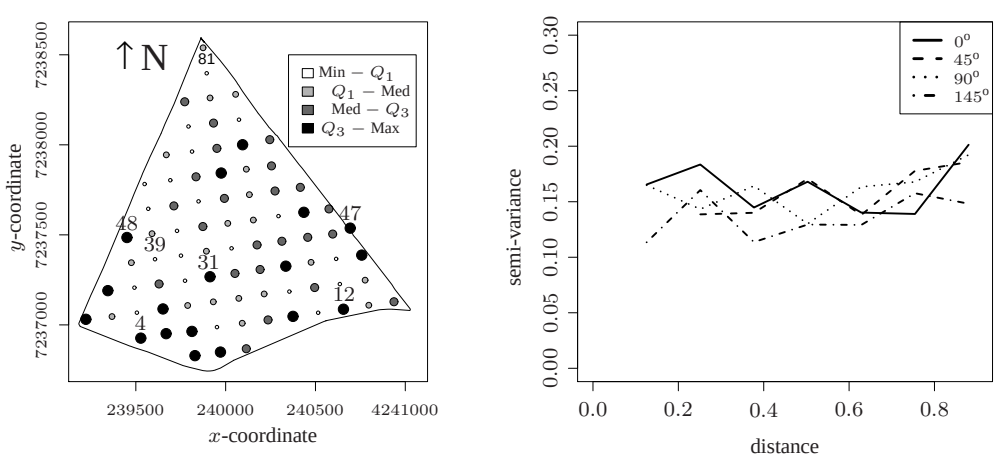

Figure 3: sampled points divided by quartiles and location of outliers (left) and directional semi-variograms (right) for Mg content data.

Figure 4 provides the QQ plot of the residuals, transformed by the Wilson-Hilferty approximation (Marchant et al., 2016b), which shows that the model fits the Mg content data reasonably well. In particular, when the four lower values of the QQ plot (left) are removed, an adequate fit is detected, with

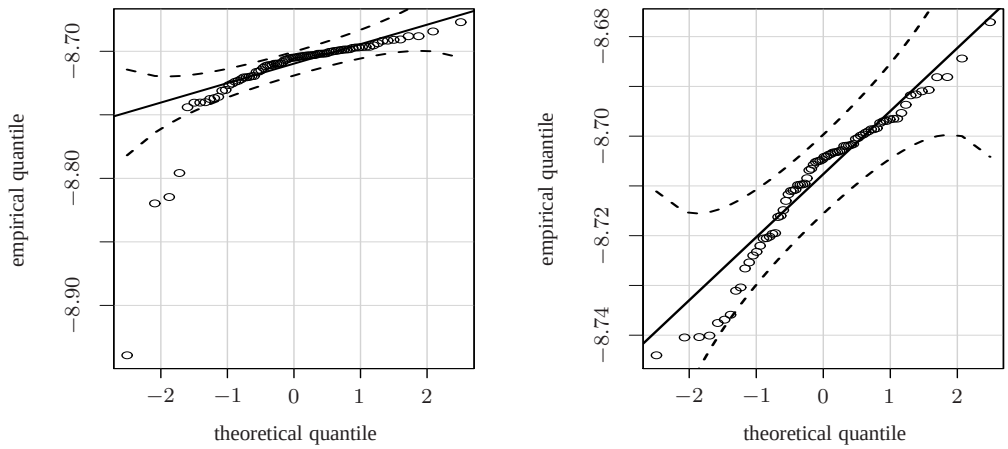

Figure 4: QQ plots of residuals using the Wilson-Hilferty approximation for Mg content data.

most points closely scattered around the y = x line, and all of them inside the envelope. When using the robust Jarque-Beran test (Gel and Gastwirth, 2008; Stehlík et al., 2014), we obtain a p-value less than 0.0001 for the full data set and a p-value of 0.03 for the data set with the mentioned four values removed, that is, we pass from a highly significant result to a non-significant result at 1%, favoring normality of the residuals, which is coherent with the graphical analysis based on the QQ plots. This suggests that a BS spatial log-linear regression model with heavier tails, such as a BS spatial Student-t log-linear regression model, might improve the fit to the full data set—this is out of the scope of this work but will be considered for future research. Observe in Figure 4 (left) that the cases #2, #12, #62 and #66 produce this bad fit in the QQ plot. First, the case #12 is an outlier with a high value and identified as potentially influential by practically all diagnostic plots. Second, the case #62 is identified as potentially influential by local influence plots. Thus, there exist cases that are identified as potentially influential by some graphs but not by others. Therefore, third, the case #2, although it has not been identified as potentially influential nor an outlier, this could be identified as potentially influential by the plot of the CD, because it is very close to the cutoff. Recall that there is no consensus in the literature about the most suitable cutoff. Fourth, the case #66 can be assessed as potentially influential as well, although it is not an outlier. These four extreme residuals are associated with Mg values higher than those surrounding and where the Ca value also does not support a raised Mg value.

The ML estimates of the model parameters, with the corresponding estimated asymptotic standard errors shown in brackets, are: $\widehat{\alpha} = 0.9185(1.0439)$, $\widehat{\beta}_0 = -0.9619(0.1046)$, $\widehat{\beta}_1 = 0.2731(0.0144)$, $\widehat{\varphi}_1 = 0.0301(0.0656)$, $\widehat{\varphi}_2 = 0.0156(0.0403)$ and $\widehat{\varphi}_3 = 1.1375(0.0120)$. Thus, the estimated model is

$$\widehat{\text{Mg}}(\boldsymbol{s}_i) = \exp(-0.9619 + 0.2731\text{Ca}(\boldsymbol{s}_i)), \quad i = 1, \ldots, 82,$$

where the scale-dependence matrix is estimated as $\widehat{\boldsymbol{\Sigma}} = 0.0301\boldsymbol{I}_{82} + 0.0156\boldsymbol{R}(1.1375)$, with $\boldsymbol{R}(\varphi_3)$ given from (8) for $\delta = 0.25$ and evaluated at $\varphi_3 = \widehat{\varphi}_3 = 1.1375$.

4.3 Comparison of BS and Gaussian models

We compare the BS and Gaussian spatial regression models using the Akaike information criterion (AIC) and the Schwarz Bayesian information criterion (BIC). The AIC and BIC are given respectively by AIC $= -2\ell(\widehat{\boldsymbol{\theta}}) + 2d$ and BIC $= -2\ell(\widehat{\boldsymbol{\theta}}) + d\log(n)$, where $\ell(\widehat{\boldsymbol{\theta}})$ is the log-likelihood function for the

parameter $\boldsymbol{\theta}$ associated with the model evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$, d is the dimension of the parameter space, and n the size of the data set. Both criteria are based on the log-likelihood function and penalize the model with more parameters. A model whose information criterion has a smaller value is better (Ferreira et al., 2012). Thus, according to Table 1, we detect that the BS spatial log-linear regression model outperforms the Gaussian spatial linear regression model.

Table 1: comparison of BS and Gaussian models for Mg content data.

Model	$\ell(\widehat{\boldsymbol{\theta}})$	AIC	BIC
BS	94.79	-177.58	-163.14
Gaussian	-43.89	97.78	109.81

4.4 Diagnostic analysis

Figure 5 shows the potentially influential cases in the ML estimates of the parameter vector $\boldsymbol{\theta} = \left(\boldsymbol{\beta}^\top, \boldsymbol{\psi}^\top\right)^\top$, according to the CD as criterion of global influence. Note that the cases #12 and #27 influence both estimates of the vectors $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$ and $\boldsymbol{\psi} = (\boldsymbol{\varphi}^\top, \alpha)^\top$, whereas the case #15 is potentially influential only for the estimate of $\boldsymbol{\beta}$. The only potentially influence case detected also as an outlier is the case #12. Note that the case #12 is an outlier in the crop year 2014/2015, presenting a high Mg level in relation to the other Mg levels during the same period. However, in the crop year 2013/2014, the case #12 is not an outlier, which is located at the first quartile of the data distribution. This fact can be explained as follows. The farmer responsible for a particular agricultural area obtains a chemical analysis of the soil each year. Then, the farmer makes a chemical correction of the soil in an appropriate way, that is, a corrective application is carried out depending on the need. During 2013/2014, the farmer could have observed that the case #12 needed to be corrected. Once this correction was made, this case had a high value due to an over correction, leading to the case # 12 appearing as an outlier in the next period 2014/2015.

The local influence study is conducted assuming two types of scheme: perturbation in the response and perturbation in the covariate Ca. As mentioned, three measures of influence are considered: conformal curvature in the direction of basis vectors (B_i), maximum conformal curvature ($B_{\boldsymbol{d}_{\max}}$) and slope displacement of the modified likelihood function ($S_{\max}$). The local influence graphs for perturbations in the response are shown in Figure 6, whereas Figure 7 presents local influence graphs when the covariate Ca is perturbed. The cases #12 and #27, identified by the global influence plots, are also identified as locally influential by the plots for B_i, $B_{\boldsymbol{d}_{\max}}$ and $S_{\max}$, when the response is perturbed. The case #15, identified as globally influential, is also indicated by the local influence plots for B_i and $S_{\max}$. For the other cases identified, at least two of the three graphs detect the cases #62 and #67. For covariate perturbation, the three graphs identify as potentially influential points the same cases #12, #27 and #62.

From Figure 8, observe that the cases #14, #15, #81 and #82 are detected as potentially influential by their predicted values (leverage points). The case #12, identified as an outlier, is also identified as influential in almost all influence index plots. Cases identified as locally influential are not identified as outliers, which agrees with other studies, such as Assumpção et al. (2014). The cases #14, #15, #81 and #82 are possible points that exert a potential influence on the parameter estimates, that is, their individual or join removal from the data set can lead to different conclusions and decisions. Note that an analysis of mineral content in the soil is carried out to decide whether it is necessary to correct some mineral or not, since the mineral content can affect the agricultural productivity. Then, the removal of these

potentially influential cases from the data set may lead to a change of decision, that is, whether minerals must be added to the soil or not. Therefore, we need to detect potential influential points. Observe that, in spatial statistics, an influential point is not necessarily an outlier, as well as an outlier is not necessarily an influential point. The concept of influence depends, in addition to the value of the variable, on its location and value of the variable in neighboring locations. Note that, in Figure 2 (left), the case #12 presents a value in the fourth quartile of the data distribution. However, neighboring locations have values in the first or second quartile. Thus, the case #12 is an outlier, because it is an atypical value within the data set, but it also is an influential point, since it has a very high value, but it is surrounded by points with low values. It is very noticeable that most of the points identified as influential are peripheral to the study region. The cases #12 and #14 are in the far right and also have high Ca values, with the case #12 having the highest Mg value. Similarly, the cases #15 and #27 are in the far left, with high Mg values. Also, the cases #81 and #82 are at the very top with almost the lowest Ca and Mg value. The remaining influential points, the cases #62 and #67, are not peripheral but they correspond to high values surround by lower values, with the highest Ca value for the case #67.

As mentioned, we have presented several measures of global (CD, GL) and local (B_i, $B_{d_{\max}}$, $S_{\max}$) influence, considering response and covariate perturbations. In order to conduct a study about the relative change (RC) when the case detected as potentially influential is removed, we select the points that were detected as influential by most of these measures, that is, the cases #12, #15, #27 and #62. We consider individual and joint removal of these cases. The impact of the influential cases on the parameter estimates is checked by computing $\mathrm{RC}_{\theta_{j(I)}} = |(\widehat{\theta}_j - \widehat{\theta}_{j(i)})/\widehat{\theta}_j| \times 100\%$, where $\widehat{\theta}_{j(I)}$ is the ML estimate of θ_j after removing the set of case(s) I, for $j = 1, \ldots, 6$ and $I = 1, \ldots, 15$, with $\theta_1 = \beta_0$, $\theta_2 = \beta_1$, $\theta_3 = \varphi_1$, $\theta_4 = \varphi_2$, $\theta_5 = \varphi_3$ and $\theta_6 = \alpha$. Table 2 reports the RCs in the parameter estimates obtained by considering the data with removed cases. Note that, in general, the RCs for the parameters β_0, β_1, φ_1 and φ_2 are large, with the RC for φ_2 being the largest. From the results obtained, it is verified that the removal of the potentially influential cases greatly modifies the spatial dependence of the data — though the radius of spatial dependence, φ_3, is not modified — recalling that the variance of the data is given by $\varphi_1 + \varphi_2$, whereas φ_1 is the nugget effect of the semi-variogram. Note that, with the exception of the case #62, which only influences the ML estimate of the parameter φ_2, all other cases influence both spatial dependence and model mean. Remind that the model prediction is carried out by Kriging, which depends on β_0, β_1, ϕ_1 and ϕ_2, all which are affected by the influential cases, and therefore, the model prediction is also affected.

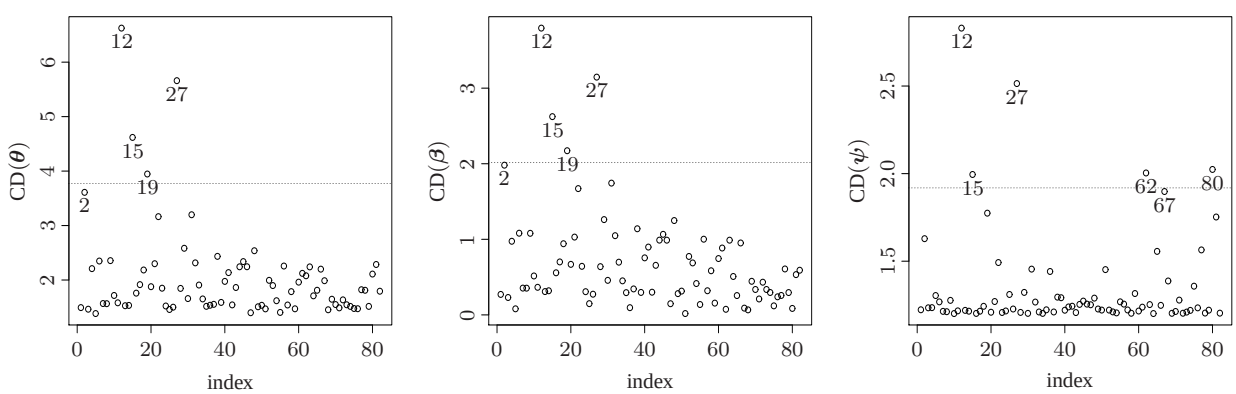

Figure 5: CD for $\boldsymbol{\theta}$ (left), $\boldsymbol{\beta}$ (center) and $\boldsymbol{\varphi}$ (right) with Mg content data.

Table 2: relative change (RC) in % of ML estimates for the indicated parameter and removed cases.

Removed case(s)	β_0	β_1	φ_1	φ_2	φ_3	α
#12	5.8530	4.5771	10.2990	19.8718	0.0000	0.0762
#15	3.1084	2.6730	7.3090	14.1026	0.0000	0.0544
#27	1.9129	1.9407	8.9701	17.3077	0.0000	0.0544
#62	0.0000	0.1098	6.9767	13.4615	0.0000	0.0544
#12,#15	6.4872	5.3826	14.2857	27.5641	0.0000	0.0871
#12,#27	5.2396	4.6137	15.9468	30.7692	0.0000	0.0871
#12,#62	3.3371	2.5632	13.2890	25.6410	0.0000	0.0762
#15,#27	2.4639	2.6730	12.6246	57.0513	0.0000	0.0762
#15,#62	0.6134	0.6591	10.2990	19.8718	0.0000	0.0653
#27,#62	0.6030	0.0732	11.9601	23.0769	0.0000	0.0653
#12,#15,#27	5.8842	5.4559	19.9336	38.4615	0.0000	0.0871
#12,#15,#62	4.0129	3.4053	16.9435	32.6923	0.0000	0.1089
#12,#27,#62	2.7862	2.6730	18.6047	93.5897	0.0000	0.0871
#15,#27,#62	0.0104	0.6957	15.9468	30.7692	0.0000	0.0762
#12,#15,#27,#62	3.4723	3.5152	22.9236	44.2308	0.0000	0.0980

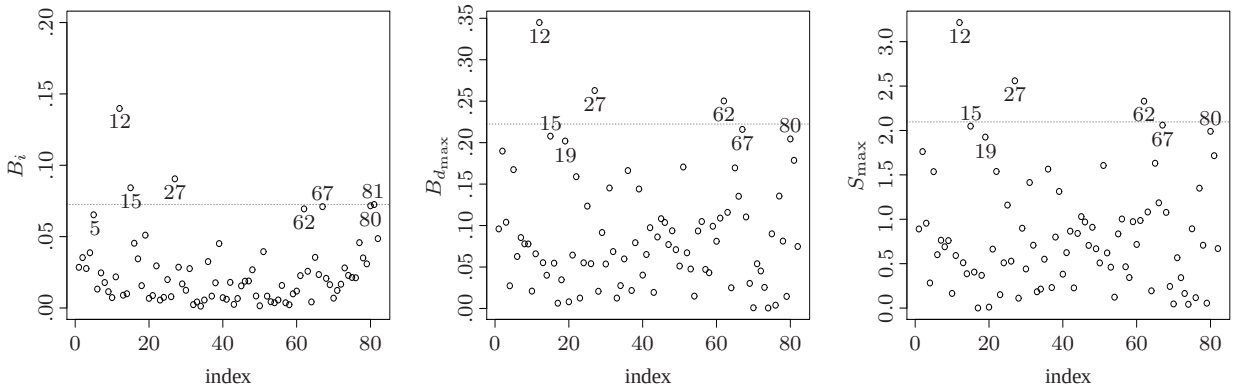

Figure 6: response perturbation for B_i (left), $B_{\boldsymbol{d}_{\max}}$ (center) and $S_{\max}$ (right) with Mg content data.

5 Conclusions and future work

The main contribution of this work is the derivation of diagnostic methods of global and local influence for Birnbaum-Saunders spatial regression models. These models make it possible to describe the spatial dependence of strictly positive data with a distribution which is skewed to the right. The proposed methods are used to analyse geochemical data. Specifically, an agricultural management problem in Brazil was addressed to evaluate its effects. Because agricultural systems are exposed to deficiency and imbalance of nutrients in the soil, the magnesium content in the soil and its relation to the calcium content has been studied. The relationship between these components was found to be significant statistically. Thus, a Birnbaum-Saunders spatial regression model has been fitted to predict the magnesium content in the soil using the calcium content at different spatial locations. The diagnostic measures de-

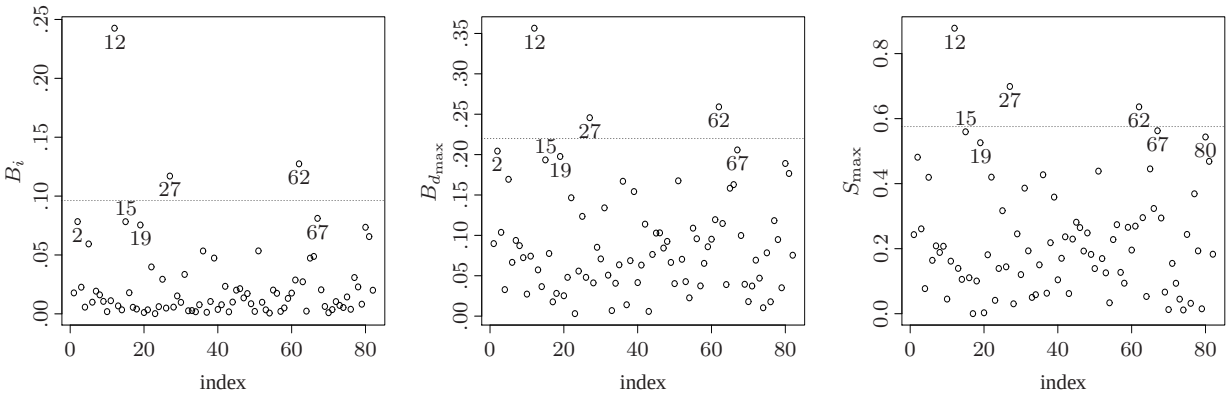

Figure 7: covariate perturbation for B_i (left), $B_{\boldsymbol{d}_{\max}}$ (center) and $S_{\max}$ (right) with Mg content data.

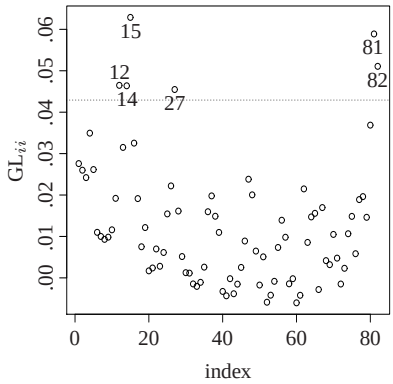

Figure 8: index plot of the GL for log(Mg) data.

rived in this work have been employed to evaluate the effect of atypical cases. Based on the application shown here, the individual or join removal of atypical cases from the data set can lead to different decisions compared to when they are not removed. This shows the importance of a diagnostic analysis in the statistical modelling.

Some possible issues to be addressed in future studies are the following. First, because the Birnbaum-Saunders distribution is based on the Gaussian distribution, parameter estimation can be influenced by atypical cases, which are known to have an adverse effect on spatial maps. Instead, an estimation procedure robust to atypical data, for example based on the Birnbaum-Saunders Student-t spatial regression model, can be considered to reduce their effects. This will allow comparison of the Birnbaum-Saunders spatial log-linear regression model with its analogous model based on the BS-Student-t distribution, and could also include comparison with the regular Student-t model (Azevedo et al., 2012). Second, in addition to fixed effects added to the spatial modelling by regression, random effects can also be added by mixed models, which can produce a more sophisticated model and close to reality (Villegas et al., 2011). Third, spatio-temporal models in the line of the model studied in this paper may be considered as well (Reich, 2016). Fourth, other parameters different to the mean can be described by spatial models, for example, some quantile, such as the median which is one of the parameters of the Birnbaum-Saunders distribution, or dispersion parameters (Noufaily and Jones, 2013). Fifth, other parameterizations of the

Birnbaum-Saunders distribution might be considered, which allow us to assume non-linear structures under a framework of generalized linear models (Leiva et al., 2014b; Santos-Neto et al., 2016). Sixth, we can assume more than one random variable in the spatial modelling using the multivariate Birnbaum-Saunders distribution (Marchant et al., 2016a,b, 2018). Seventh, we have not performed normality tests directly from the data since, in spatial statistics, we have one observation (trajectory of the process) in each spatial point, that is, no replicated data are available, which are needed in order to carry out any goodness-of-fit analysis. When replicated data are available, we can conduct goodness-of-fit tests directly with the data, for example, following the work by Barros et al. (2014) and Stehlík et al. (2014). As alternative approaches for goodness of fit, discrimination and model selection based on the available data, we have provided a comparison between the Gaussian and Birnbaum-Saunders models and used residual analysis to check goodness of fit of the model to the data. The comparison showed that the Birnbaum-Saunders model is more suitable to the studied data than the Gaussian model, whereas the residual analysis provided evidence about the distributional adequacy. Research on these seven issues is currently in progress and their findings will be reported in a future paper.

Appendix A: Score vector and Fisher information matrix

***Score vector*:** The elements of the $(p+4)\times 1$ score vector given in (10) are detailed as

$$\begin{aligned}
\dot{\ell}(\beta_j) &= \frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_j} = \frac{\partial}{\partial \beta_j}\left(-\frac{2}{\alpha^2}\boldsymbol{V}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{V} + \sum_{i=1}^{n}\log\left(\cosh\left(\frac{y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}}{2}\right)\right)\right) \\
&= \frac{2}{\alpha^2}\left(\boldsymbol{W}^\top \odot \boldsymbol{x}_j^\top\right)\boldsymbol{\Sigma}^{-1}\boldsymbol{V} - \frac{1}{2}\sum_{i=1}^{n}\tanh\left(\frac{y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}}{2}\right)x_{ij}, \quad j = 1,\ldots,p; \\
\dot{\ell}(\varphi_j) &= \frac{\partial \ell(\boldsymbol{\theta})}{\partial \varphi_j} = \frac{\partial}{\partial \varphi_j}\left(-\frac{1}{2}\log\left(|\boldsymbol{\Sigma}|\right) - \frac{2}{\alpha^2}\boldsymbol{V}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{V}\right) \\
&= -\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\varphi_j}\right) + \frac{2}{\alpha^2}\boldsymbol{V}^\top \boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\varphi_j}\boldsymbol{\Sigma}^{-1}\boldsymbol{V}, \quad j = 1,2,3; \\
\dot{\ell}(\alpha) &= \frac{\partial \ell(\boldsymbol{\theta})}{\partial \alpha} = \frac{\partial}{\partial \alpha}\left(-n\log(\alpha) - \frac{2}{\alpha^2}\boldsymbol{V}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{V}\right) = -\frac{n}{\alpha} + \frac{4}{\alpha^3}\boldsymbol{V}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{V};
\end{aligned}$$

where $\odot$ denotes the Hadamard product (Caro-Lopera et al., 2012); tr $(\boldsymbol{A})$ denotes the trace of $\boldsymbol{A}$; $\boldsymbol{W} = (W_1,\ldots,W_n)^\top$, with $W_i = \cosh((y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta})/2)$, for $i = 1,\ldots,n$; $\boldsymbol{V} = (V_1,\ldots,V_n)^\top$ is defined in (9); and $\boldsymbol{x}_j = (x_{1j},\ldots,x_{nj})^\top$, for $j = 1,\ldots p$. It is worth to remember that $x_{i1} = 1$, for $i = 1,\ldots,n$, and that

$$\frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_1} = \boldsymbol{I}_n, \quad \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_2} = \boldsymbol{R}(\varphi_3), \quad \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_3} = \varphi_2\frac{\partial \boldsymbol{R}(\varphi_3)}{\partial \varphi_3}.$$

For the Matérn model given in (8), $\boldsymbol{R}(\varphi_3)' = \partial\boldsymbol{R}(\varphi_3)/\partial\varphi_3 = (\partial r_{ij}/\partial\varphi_3)$, with elements defined as

$$\frac{\partial r_{ij}}{\partial \varphi_3} = -\left(\frac{1}{\varphi_3}\right)\left(\delta r_{ij} + \frac{1}{2^{\delta-1}\Gamma(\delta)}\left(\frac{h_{ij}}{\varphi_3}\right)^{\delta+1}K'_\delta\left(\frac{h_{ij}}{\varphi_3}\right)\right),$$

for $i \neq j$, $i,j = 1,\ldots,n$, and $K'_\delta(u) = -(1/2)(K_{\delta-1}(u) + K_{\delta+1}(u))$, recalling that K_δ is the modified Bessel function of the third kind and order δ.

***Information matrix*:** Note that the observed Fisher information matrix is defined by $-\ddot{\boldsymbol{\ell}}(\boldsymbol{\theta})$, which must be evaluated at $\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}$, where $\ddot{\boldsymbol{\ell}}(\boldsymbol{\theta})$ is the Hessian matrix. For the BS spatial regression model presented in (5), the Hessian matrix given in (11) has elements expressed as

$$\begin{aligned}
\ddot{\ell}(\beta_j\beta_l)=\frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\beta_j\partial\beta_l} &= -\frac{1}{\alpha^2}\left(\boldsymbol{V}^\top\odot\boldsymbol{x}_j^\top\odot\boldsymbol{x}_l^\top\right)\boldsymbol{\Sigma}^{-1}\boldsymbol{V}-\frac{1}{\alpha^2}\left(\boldsymbol{W}^\top\odot\boldsymbol{x}_j^\top\right)\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{W}\odot\boldsymbol{x}_l\right)\\
&\quad+\frac{1}{4}\sum_{i=1}^{n}\left(\text{sech}\left(\frac{y_i-\boldsymbol{x}_i^\top\boldsymbol{\beta}}{2}\right)\right)^2\boldsymbol{x}_{ij}\boldsymbol{x}_{il},\quad j,l=1,\ldots,p;\\
\ddot{\ell}(\beta_j\varphi_l)=\frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\beta_j\partial\varphi_l} &= -\frac{2}{\alpha^2}\left(\boldsymbol{W}^\top\odot\boldsymbol{x}_j^\top\right)\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\varphi_l}\boldsymbol{\Sigma}^{-1}\boldsymbol{V},\quad j=1,\ldots,p,l=1,2,3;\\
\ddot{\ell}(\varphi_j\varphi_l)=\frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\varphi_j\partial\varphi_l} &= -\frac{1}{2}\text{tr}\left(-\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\varphi_l}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\varphi_j}+\boldsymbol{\Sigma}^{-1}\frac{\partial^2\boldsymbol{\Sigma}}{\partial\varphi_l\partial\varphi_j}\right)+\frac{2}{\alpha^2}\boldsymbol{V}^\top\left(-\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\varphi_l}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\varphi_j}\boldsymbol{\Sigma}^{-1}\right.\\
&\quad\left.+\boldsymbol{\Sigma}^{-1}\left(\frac{\partial^2\boldsymbol{\Sigma}}{\partial\varphi_l\partial\varphi_j}\boldsymbol{\Sigma}^{-1}-\frac{\partial\boldsymbol{\Sigma}}{\partial\varphi_j}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\varphi_l}\boldsymbol{\Sigma}^{-1}\right)\right)\boldsymbol{V};\quad j,l=1,2,3,
\end{aligned}$$

with

$$\frac{\partial^2\boldsymbol{\Sigma}}{\partial\varphi_j\partial\varphi_l}=\frac{\partial^2\boldsymbol{\Sigma}}{\partial\varphi_l\partial\varphi_j},\frac{\partial^2\boldsymbol{\Sigma}}{\partial\varphi_1\partial\varphi_l}=\frac{\partial^2\boldsymbol{\Sigma}}{\partial\varphi_2^2}=0,\frac{\partial^2\boldsymbol{\Sigma}}{\partial\varphi_2\partial\varphi_3}=\boldsymbol{R}'(\varphi_3),\frac{\partial^2\boldsymbol{\Sigma}}{\partial\varphi_3^2}=\varphi_2\boldsymbol{R}''(\varphi_3),l=1,2,3,$$

and $\boldsymbol{R}(\varphi_3)''=\partial^2\boldsymbol{R}(\varphi_3)/\partial\varphi_3^2=(\partial^2 r_{ij}/\partial\varphi_3^2)$. Note that the elements of $\boldsymbol{R}(\varphi_3)''$ are defined by

$$\frac{\partial^2 r_{ij}}{\partial\varphi_3^2}=\frac{\delta(\delta+1)r_{ij}}{\varphi_3^2}+\frac{1}{\varphi_3^2 2^{\delta-1}\Gamma(\delta)}\left(\frac{k_{ij}}{\varphi_3}\right)^{\delta+1}\left(2(\delta+1)K_\delta'\left(\frac{k_{ij}}{\varphi_3}\right)+\left(\frac{k_{ij}}{\varphi_3}\right)K_\delta''\left(\frac{k_{ij}}{\varphi_3}\right)\right),$$

for $i\neq j,i,j=1,\ldots,n$, with $K_\delta'(u)=-(1/2)(K_{\delta-1}(u)+K_{\delta+1}(u))$, $K_\delta''(u)=(1/4)(K_{\delta-2}(u)+2K_\delta(u)+K_{\delta+2}(u))$. In addition, the $p\times 1$ and 3×1 vectors $\ddot{\boldsymbol{\ell}}(\boldsymbol{\beta}\alpha)=(\ddot{\boldsymbol{\ell}}(\alpha\boldsymbol{\beta}))^\top$ and $\ddot{\boldsymbol{\ell}}(\boldsymbol{\varphi}\alpha)=(\ddot{\boldsymbol{\ell}}(\alpha\boldsymbol{\varphi}))^\top$, respectively, have elements given by

$$\begin{aligned}
\ddot{\ell}(\beta_j\alpha) &= \frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\beta_j\partial\alpha}=-\frac{4}{\alpha^3}\left(\boldsymbol{W}^\top\odot\boldsymbol{x}_j^\top\right)\boldsymbol{\Sigma}^{-1}\boldsymbol{V},\quad j=1,\ldots,p,\\
\ddot{\ell}(\varphi_j\alpha) &= \frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\varphi_j\partial\alpha}=-\frac{4}{\alpha^3}\boldsymbol{V}^\top\left(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\varphi_j}\boldsymbol{\Sigma}^{-1}\right)\boldsymbol{V},\quad j=1,2,3.
\end{aligned}$$

Furthermore, the scalar $\ddot{\ell}(\alpha)$ is expressed as

$$\ddot{\ell}(\alpha)=\frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\alpha\partial\alpha}=\frac{n}{\alpha^2}-\frac{12}{\alpha^4}\boldsymbol{V}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{V}.$$

Therefore, for the BS spatial regression model, the $(p+4)\times(p+4)$ expected Fisher information matrix, obtained from (11) and given in (12), is specified as

$$\boldsymbol{K}(\boldsymbol{\theta})=\text{E}(-\ddot{\boldsymbol{\ell}}(\boldsymbol{\theta}))=\begin{pmatrix}\boldsymbol{K}(\boldsymbol{\beta}) & \boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{K}(\boldsymbol{\varphi}) & \boldsymbol{K}(\boldsymbol{\varphi}\alpha)\\ \boldsymbol{0} & \boldsymbol{K}(\alpha\boldsymbol{\varphi}) & K(\alpha)\end{pmatrix}=\begin{pmatrix}\boldsymbol{K}(\boldsymbol{\beta}) & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{K}(\boldsymbol{\psi})\end{pmatrix},$$

where $\boldsymbol{K}(\boldsymbol{\beta})=\text{E}(-\ddot{\boldsymbol{\ell}}(\boldsymbol{\beta}))=(K(\beta_j\beta_l))$ is a $p\times p$ matrix, with elements given by

$$\begin{aligned}
K(\beta_j\beta_l) &= \text{E}\left(\frac{1}{\alpha^2}\left(\boldsymbol{V}^\top\odot(\boldsymbol{x}_j\odot\boldsymbol{x}_l)^\top\right)\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\right)+\text{E}\left(\frac{1}{\alpha^2}\left(\boldsymbol{W}^\top\odot\boldsymbol{x}_j^\top\right)\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{W}\odot\boldsymbol{x}_l\right)\right)\\
&\quad+\text{E}\left(-\frac{1}{4}\sum_{i=1}^{n}\left(\text{sech}\left(\frac{Y_i-\boldsymbol{x}_i^\top\boldsymbol{\beta}}{2}\right)\right)^2\boldsymbol{x}_{ij}\boldsymbol{x}_{il}\right)=\frac{1}{\alpha^2}\boldsymbol{x}_j^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{x}_l,\quad j,l=1,\ldots,p;
\end{aligned}$$

which are obtained using the approximations $\text{sech}(x) = 1/\cosh(x) \approx 1$, with $\cosh(x) \approx 1$, and standard results of the matrix differential calculus (Magnus and Neudecker, 2007), as well as assuming that $(2/\alpha)\boldsymbol{V} \sim \text{N}_n(\boldsymbol{0}_{n\times 1}, \boldsymbol{\Sigma})$, for $\boldsymbol{V} = (V_1, \ldots, V_n)^\top$, with $V_i = \sinh((Y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta})/2)$, that is, $(2/\alpha)V_i \sim \text{N}(0,1)$. Approximations employed for the cosh and sinh functions are based on a Taylor expansion of first order, around zero, similar to that employed by Rieck and Nedelman (1991) for sinh. Note that $\boldsymbol{K}(\boldsymbol{\beta\varphi}) = (\boldsymbol{K}(\boldsymbol{\varphi\beta}))^\top = \text{E}(-\ddot{\boldsymbol{\ell}}(\boldsymbol{\beta\varphi})) = (K(\beta_j\varphi_l))$ is a $p \times 3$ matrix, with elements defined as

$$K(\beta_j\varphi_l) = \text{E}\left(\frac{2}{\alpha^2}\left(\boldsymbol{W}^\top \odot \boldsymbol{x}_j^\top\right)\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\varphi_l}\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\right) = \frac{1}{\alpha}\boldsymbol{x}_j^\top\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\varphi_l}\boldsymbol{\Sigma}^{-1}\text{E}\left(\frac{2}{\alpha}\boldsymbol{V}\right) = 0, j = 1, \ldots, p, l = 1, 2, 3.$$

In addition, we have that $\boldsymbol{K}(\boldsymbol{\varphi}) = \text{E}(-\ddot{\boldsymbol{\ell}}(\boldsymbol{\varphi})) = (K(\varphi_j\varphi_l))$ is a symmetric 3×3 matrix, with elements expressed as

$$\begin{aligned} K(\varphi_j\varphi_l) &= \text{E}\left(\frac{1}{2}\text{tr}\left(-\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\varphi_l}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\varphi_j} + \boldsymbol{\Sigma}^{-1}\frac{\partial^2\boldsymbol{\Sigma}}{\partial\varphi_l\partial\varphi_j}\right)\right) + \text{E}\left(\frac{2}{\alpha^2}\boldsymbol{V}^\top\left(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\varphi_l}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\varphi_j}\boldsymbol{\Sigma}^{-1}\right)\boldsymbol{V}\right) \\ &\quad -\text{E}\left(\frac{2}{\alpha^2}\boldsymbol{V}^\top\left(\boldsymbol{\Sigma}^{-1}\frac{\partial^2\boldsymbol{\Sigma}}{\partial\varphi_l\partial\varphi_j}\boldsymbol{\Sigma}^{-1}\right)\boldsymbol{V}\right) + \text{E}\left(\frac{2}{\alpha^2}\boldsymbol{V}^\top\left(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\varphi_j}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\varphi_l}\boldsymbol{\Sigma}^{-1}\right)\boldsymbol{V}\right) \\ &= \frac{1}{2}\text{tr}\left(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\varphi_j}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\varphi_l}\right), \quad j,\ l = 1, 2, 3. \end{aligned}$$

Also, observe that $\boldsymbol{K}(\boldsymbol{\beta}\alpha) = (\boldsymbol{K}(\alpha\boldsymbol{\beta}))^\top = \text{E}(-\ddot{\boldsymbol{\ell}}(\boldsymbol{\beta}\alpha)) = (K(\beta_j\alpha))$ and $\boldsymbol{K}(\boldsymbol{\varphi}\alpha) = (\boldsymbol{K}(\alpha\boldsymbol{\varphi}))^\top = \text{E}(-\ddot{\boldsymbol{\ell}}(\boldsymbol{\varphi}\alpha)) = (K(\varphi_j\alpha))$ are $p \times 1$ and 3×1 vectors, respectively, with elements given by

$$\begin{aligned} K(\beta_j\alpha) &= \text{E}\left(\frac{4}{\alpha^3}\left(\boldsymbol{W}^\top \odot \boldsymbol{x}_j^\top\right)\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\right) = \text{E}\left(\frac{4}{\alpha^3}\boldsymbol{x}_j^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\right) = 0, \quad j = 1, \ldots, p; \\ K(\varphi_j\alpha) &= \text{E}\left(\frac{4}{\alpha^3}\boldsymbol{V}\left(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\varphi_j}\boldsymbol{\Sigma}^{-1}\right)\boldsymbol{V}\right) = \frac{1}{\alpha}\text{tr}\left(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\varphi_j}\right), \quad j = 1, 2, 3; \end{aligned}$$

since $(2/\alpha)\boldsymbol{V} \sim \text{N}_n(\boldsymbol{0}_{n\times 1}, \boldsymbol{\Sigma})$. Furthermore, the scalar $K(\alpha)$ is expressed as

$$K(\alpha) = \text{E}(-\ddot{\ell}(\alpha)) = \text{E}\left(-\frac{n}{\alpha^2} + \frac{12}{\alpha^4}\boldsymbol{V}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\right) = \frac{2n}{\alpha^2},$$

because $((2/\alpha)\boldsymbol{V}^\top)\boldsymbol{\Sigma}^{-1}((2/\alpha)\boldsymbol{V}) \sim \chi^2(n)$.

Appendix B: Perturbation matrices for the BS spatial model

***Perturbation in the response*:** For the model defined in (5) and its log-likelihood function expressed in (9), perturbing the response, we have

$$\frac{\partial\ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\boldsymbol{\omega}^\top} = \boldsymbol{V}_{\boldsymbol{\omega}}^\top\left(-\frac{2}{\alpha^2}\boldsymbol{\Sigma}^{-1} + \frac{1}{2}\boldsymbol{I}_n\right)\boldsymbol{A},$$

where $\boldsymbol{V}_{\boldsymbol{\omega}} = (\boldsymbol{V}_{\boldsymbol{\omega}_1}, \ldots, \boldsymbol{V}_{\boldsymbol{\omega}_n})$, with $\boldsymbol{V}_{\omega_i} = \sinh((y_i + \boldsymbol{A}_i\boldsymbol{\omega} - \boldsymbol{x}_i^\top\boldsymbol{\beta})/2)$, for $i = 1, \ldots, n$, and $\boldsymbol{A}_i$ is the ith row of the matrix $\boldsymbol{A}$ defined in (20). Thus, the corresponding $(p+4) \times n$ perturbation matrix is given by

$$\boldsymbol{\Delta} = \frac{\partial\ell^2(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^\top} = (\boldsymbol{\Delta}(\boldsymbol{\beta}), \boldsymbol{\Delta}(\boldsymbol{\varphi}), \Delta(\alpha))^\top,$$

where

$$\boldsymbol{\Delta}(\boldsymbol{\beta}) = \frac{\partial\ell^2(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\omega}^\top} = \left(\frac{\partial\ell^2(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\beta_1\partial\boldsymbol{\omega}^\top}, \ldots, \frac{\partial\ell^2(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\beta_p\partial\boldsymbol{\omega}^\top}\right)^\top,$$

whose elements are specified as

$$\frac{\partial \ell^2(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \beta_j \partial \boldsymbol{\omega}^\top} = \frac{\partial \boldsymbol{V}_{\boldsymbol{\omega}}^\top}{\partial \beta_j}\left(-\frac{2}{\alpha^2}\boldsymbol{\Sigma}^{-1} + \frac{1}{2}\boldsymbol{I}_n\right)\boldsymbol{A} = -\frac{1}{2}\boldsymbol{W}_{\boldsymbol{\omega}}^\top \odot \boldsymbol{x}_j^\top\left(-\frac{2}{\alpha^2}\boldsymbol{\Sigma}^{-1} + \frac{1}{2}\boldsymbol{I}_n\right)\boldsymbol{A}, j = 1, \ldots, p, \qquad (22)$$

with $\boldsymbol{W}_{\boldsymbol{\omega}} = (W_{\omega_1}, \ldots, W_{\omega_n})^\top$ and $W_{\omega_i} = \cosh((y_i + \boldsymbol{A}_i\boldsymbol{\omega} - \boldsymbol{x}_i^\top\boldsymbol{\beta})2)$, for $i = 1, \ldots, n$. In addition, as $W_{\omega_i} = \cosh((y_i + \boldsymbol{A}_i\boldsymbol{\omega} - \boldsymbol{x}_i^\top\boldsymbol{\beta})/2) \approx 1$, (22) reduces to

$$\frac{\partial \ell^2(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\beta_j} \partial \boldsymbol{\omega}^\top} = \boldsymbol{x}_j^\top\left(\frac{1}{\alpha^2}\boldsymbol{\Sigma}^{-1} - \frac{1}{4}\boldsymbol{I}_n\right)\boldsymbol{A}.$$

Furthermore, note that

$$\boldsymbol{\Delta}(\boldsymbol{\varphi}) = \frac{\partial \ell^2(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\omega}^\top} = \left(\frac{\partial \ell^2(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \varphi_1 \partial \boldsymbol{\omega}^\top}, \frac{\partial \ell^2(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \varphi_2 \partial \boldsymbol{\omega}^\top}, \frac{\partial \ell^2(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \varphi_3 \partial \boldsymbol{\omega}^\top}\right)^\top,$$

whose elements are defined as

$$\begin{aligned}
\frac{\partial \ell^2(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \varphi_i \partial \boldsymbol{\omega}^\top} &= \frac{\partial \boldsymbol{V}_{\boldsymbol{\omega}}^\top}{\partial \varphi_i}\left(-\frac{2}{\alpha^2}\boldsymbol{\Sigma}^{-1} + \frac{1}{2}\boldsymbol{I}_n\right)\boldsymbol{A} + \boldsymbol{V}_{\boldsymbol{\omega}}^\top\left(\frac{2}{\alpha^2}\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_i}\boldsymbol{\Sigma}^{-1}\boldsymbol{A} + \left(-\frac{2}{\alpha^2}\boldsymbol{\Sigma}^{-1} + \frac{1}{2}\boldsymbol{I}_n\right)\frac{\partial \boldsymbol{A}}{\partial \varphi_i}\right) \\
&= \boldsymbol{M}^\top\left(-\frac{1}{\alpha^2}\boldsymbol{\Sigma}^{-1} + \frac{1}{4}\boldsymbol{I}_n\right)\boldsymbol{A} + \boldsymbol{V}_{\boldsymbol{\omega}}^\top\left(\frac{2}{\alpha^2}\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_i}\boldsymbol{\Sigma}^{-1}\right)\boldsymbol{A} \\
&\quad + \boldsymbol{V}_{\boldsymbol{\omega}}^\top\left(-\frac{2}{\alpha^2}\boldsymbol{\Sigma}^{-1} + \frac{1}{2}\boldsymbol{I}_n\right)\boldsymbol{A}\left(\frac{1}{\alpha}\boldsymbol{\Sigma}^{-1/2}\frac{\partial \boldsymbol{\Sigma}^{1/2}}{\partial \varphi_i}\boldsymbol{\Sigma}^{-1/2} + \frac{\alpha}{4}\frac{\partial \boldsymbol{\Sigma}^{1/2}}{\partial \varphi_i}\right)\boldsymbol{A}, \quad i = 1, 2, 3,
\end{aligned}$$

with $\boldsymbol{M} = (M_1, \ldots, M_n)^\top$, $M_j = \boldsymbol{l_j}\boldsymbol{\omega}$ and $\boldsymbol{l_j}$ being the jth row of the matrix

$$\boldsymbol{L} = \boldsymbol{A}\left(-\frac{1}{\alpha}\boldsymbol{\Sigma}^{-1/2}\frac{\partial \boldsymbol{\Sigma}^{1/2}}{\partial \varphi_i}\boldsymbol{\Sigma}^{-1/2} + \frac{\alpha}{4}\frac{\partial \boldsymbol{\Sigma}^{1/2}}{\partial \varphi_i}\right)\boldsymbol{A}, \quad j = 1, \ldots, n.$$

Additional details about $\partial \boldsymbol{\Sigma}^{1/2}/\partial \varphi_i$ can be found in De Bastiani et al. (2015). Moreover, we have

$$\begin{aligned}
\Delta(\alpha) &= \frac{\partial \ell^2(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \alpha \partial \boldsymbol{\omega}^\top} \\
&= \frac{\partial \boldsymbol{V}_{\boldsymbol{\omega}}^\top}{\partial \alpha}\left(-\frac{2}{\alpha^2}\boldsymbol{\Sigma}^{-1} + \frac{1}{2}\boldsymbol{I}_n\right)\boldsymbol{A} + \boldsymbol{V}_{\boldsymbol{\omega}}^\top\left(\frac{4}{\alpha^3}\boldsymbol{\Sigma}^{-1}\boldsymbol{A} + \left(-\frac{2}{\alpha^2}\boldsymbol{\Sigma}^{-1} + \frac{1}{2}\boldsymbol{I}_n\right)\frac{\partial \boldsymbol{A}}{\partial \alpha}\right) \\
&= \boldsymbol{D}^\top\left(\frac{-1}{\alpha^2}\boldsymbol{\Sigma}^{-1} + \frac{1}{4}\boldsymbol{I}_n\right)\boldsymbol{A} + \boldsymbol{V}_{\boldsymbol{\omega}}^\top\left(\frac{4}{\alpha^3}\boldsymbol{\Sigma}^{-1}\boldsymbol{A} + \left(-\frac{2}{\alpha^2}\boldsymbol{\Sigma}^{-1} + \frac{1}{2}\boldsymbol{I}_n\right)\boldsymbol{A}\left(\frac{1}{\alpha^2}\boldsymbol{\Sigma}^{-1/2} + \frac{1}{4}\boldsymbol{\Sigma}^{1/2}\right)\boldsymbol{A}\right),
\end{aligned}$$

where $\boldsymbol{D} = (D_1, \ldots, D_n)^\top$, with $D_j = \boldsymbol{l}_j\boldsymbol{\omega}$ and $\boldsymbol{l}_j$ being the jth row of the matrix

$$\boldsymbol{L} = \boldsymbol{A}\left(\frac{1}{\alpha^2}\boldsymbol{\Sigma}^{-1/2} + \frac{1}{4}\boldsymbol{\Sigma}^{1/2}\right)\boldsymbol{A}, \quad j = 1, \ldots, n.$$

***Perturbation in a continuous covariate*:** As in the perturbation scheme of the response, but now perturbing the covariate X_t, we have

$$\frac{\partial \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\omega}^\top} = \beta_t \boldsymbol{V}_{\boldsymbol{\omega}}^\top\left(\frac{2}{\alpha^2}\boldsymbol{\Sigma}^{-1} - \frac{1}{2}\boldsymbol{I}_n\right)\boldsymbol{A},$$

where $\boldsymbol{V}_{\boldsymbol{\omega}} = (\boldsymbol{V}_{\omega_1}, \ldots, \boldsymbol{V}_{\omega_n})^\top$, with $\boldsymbol{V}_{\boldsymbol{\omega_i}} = \sinh((y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta} - \beta_t\boldsymbol{A}_i\boldsymbol{\omega})/2)$, for $i = 1, \ldots, n$ and $\boldsymbol{A}$ defined in (20). Thus, the corresponding $(p+4) \times n$ perturbation matrix is given by

$$\boldsymbol{\Delta} = \frac{\partial \ell^2\left(\boldsymbol{\theta}|\boldsymbol{\omega}\right)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\omega}^\top} = \left(\boldsymbol{\Delta}(\boldsymbol{\beta}), \boldsymbol{\Delta}(\boldsymbol{\varphi}), \Delta(\alpha)\right)^\top,$$

where

$$\boldsymbol{\Delta}(\boldsymbol{\beta}) = \frac{\partial \ell^2(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^\top} = \left(\frac{\partial \ell^2(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \beta_1 \partial \boldsymbol{\omega}^\top}, \ldots, \frac{\partial \ell^2(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \beta_t \partial \boldsymbol{\omega}^\top}, \ldots, \frac{\partial \ell^2(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \beta_p \partial \boldsymbol{\omega}^\top} \right)^\top,$$

whose elements are defined as

$$\begin{aligned}
\frac{\partial \ell^2(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \beta_t \partial \boldsymbol{\omega}^\top} &= \boldsymbol{V}_{\boldsymbol{\omega}}^\top \left(\frac{2}{\alpha^2} \boldsymbol{\Sigma}^{-1} - \frac{1}{2} \boldsymbol{I}_n \right) \boldsymbol{A} + \beta_t \frac{\partial \boldsymbol{V}_{\boldsymbol{\omega}}^\top}{\partial \beta_t} \left(\frac{2}{\alpha^2} \boldsymbol{\Sigma}^{-1} - \frac{1}{2} \boldsymbol{I}_n \right) \boldsymbol{A} \\
&= \left(\boldsymbol{V}_{\boldsymbol{\omega}}^\top - \frac{\beta_t}{2} (\boldsymbol{x}_t + \boldsymbol{A}\boldsymbol{\omega}) \right) \left(\frac{2}{\alpha^2} \boldsymbol{\Sigma}^{-1} - \frac{1}{2} \boldsymbol{I}_n \right) \boldsymbol{A}; \\
\frac{\partial \ell^2(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \beta_j \partial \boldsymbol{\omega}^\top} &= \beta_t \frac{\partial \boldsymbol{V}_{\boldsymbol{\omega}}^\top}{\partial \beta_j} \left(\frac{2}{\alpha^2} \boldsymbol{\Sigma}^{-1} - \frac{1}{2} \boldsymbol{I}_n \right) \boldsymbol{A} = -\beta_t \boldsymbol{x}_j^\top \left(\frac{2}{\alpha^2} \boldsymbol{\Sigma}^{-1} - \frac{1}{2} \boldsymbol{I}_n \right) \boldsymbol{A}, \quad j = 1, \ldots, p, j \neq t.
\end{aligned}$$

In addition, we get

$$\boldsymbol{\Delta}(\boldsymbol{\varphi}) = \frac{\partial \ell^2(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\omega}^\top} = \left(\frac{\partial \ell^2(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \varphi_1 \partial \boldsymbol{\omega}^\top}, \frac{\partial \ell^2(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \varphi_2 \partial \boldsymbol{\omega}^\top}, \frac{\partial \ell^2(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \varphi_3 \partial \boldsymbol{\omega}^\top} \right)^\top,$$

whose elements, for $j = 1, 2, 3$, are expressed as

$$\begin{aligned}
\frac{\partial \ell^2(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \varphi_j \partial \boldsymbol{\omega}^\top} &= \beta_t \frac{\partial \boldsymbol{V}_{\boldsymbol{\omega}}^\top}{\partial \varphi_j} \left(\frac{2}{\alpha^2} \boldsymbol{\Sigma}^{-1} - \frac{1}{2} \boldsymbol{I}_n \right) \boldsymbol{A} + \beta_t \boldsymbol{V}_{\boldsymbol{\omega}}^\top \left(-\frac{2}{\alpha^2} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} \boldsymbol{\Sigma}^{-1} \boldsymbol{A} + \left(\frac{2}{\alpha^2} \boldsymbol{\Sigma}^{-1} - \frac{1}{2} \boldsymbol{I}_n \right) \frac{\partial \boldsymbol{A}}{\partial \varphi_j} \right) \\
&= -\frac{\beta_t^2}{2} \boldsymbol{\omega}^\top \boldsymbol{A} \left(\frac{1}{\alpha} \boldsymbol{\Sigma}^{-\frac{1}{2}} \frac{\partial \boldsymbol{\Sigma}^{\frac{1}{2}}}{\partial \varphi_j} \boldsymbol{\Sigma}^{-\frac{1}{2}} + \frac{\alpha}{4} \frac{\partial \boldsymbol{\Sigma}^{\frac{1}{2}}}{\partial \varphi_j} \right) \boldsymbol{A} \left(\frac{2}{\alpha^2} \boldsymbol{\Sigma}^{-1} - \frac{1}{2} \boldsymbol{I}_n \right) \boldsymbol{A} \\
&\quad + \beta_t \boldsymbol{V}_{\boldsymbol{\omega}}^\top \left(\frac{2}{\alpha^2} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} \boldsymbol{\Sigma}^{-1} \boldsymbol{A} \left(\frac{2}{\alpha^2} \boldsymbol{\Sigma}^{-1} - \frac{1}{2} \boldsymbol{I}_n \right) \boldsymbol{A} \left(\frac{1}{\alpha} \boldsymbol{\Sigma}^{-\frac{1}{2}} \frac{\partial \boldsymbol{\Sigma}^{\frac{1}{2}}}{\partial \varphi_j} \boldsymbol{\Sigma}^{-\frac{1}{2}} + \frac{\alpha}{4} \frac{\partial \boldsymbol{\Sigma}^{\frac{1}{2}}}{\partial \varphi_j} \boldsymbol{A} \right) \right).
\end{aligned}$$

Finally, we get

$$\begin{aligned}
\Delta(\alpha) &= \frac{\partial \ell^2(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \alpha \partial \boldsymbol{\omega}^\top} = \beta_t \frac{\partial \boldsymbol{V}_{\boldsymbol{\omega}}^\top}{\partial \alpha} \left(\frac{2}{\alpha^2} \boldsymbol{\Sigma}^{-1} - \frac{1}{2} \boldsymbol{I}_n \right) \boldsymbol{A} + \beta_t \boldsymbol{V}_{\boldsymbol{\omega}}^\top \left(-\frac{4}{\alpha^3} \boldsymbol{\Sigma}^{-1} \boldsymbol{A} + \left(\frac{2}{\alpha^2} \boldsymbol{\Sigma}^{-1} - \frac{1}{2} \boldsymbol{I}_n \right) \frac{\partial \boldsymbol{A}}{\partial \alpha} \right) \\
&= -\frac{\beta_t^2}{2} \boldsymbol{\omega}^\top \boldsymbol{A} \left(\frac{-1}{\alpha^2} \boldsymbol{\Sigma}^{-\frac{1}{2}} - \frac{1}{4} \boldsymbol{\Sigma}^{\frac{1}{2}} \right) \boldsymbol{A} \left(\frac{2}{\alpha^2} \boldsymbol{\Sigma}^{-1} - \frac{1}{2} \boldsymbol{I}_n \right) \boldsymbol{A} \\
&\quad + \beta_t \boldsymbol{V}_{\boldsymbol{\omega}}^\top \left(\frac{-4}{\alpha^3} \boldsymbol{\Sigma}^{-1} \boldsymbol{A} + \left(\frac{2}{\alpha^2} \boldsymbol{\Sigma}^{-1} - \frac{1}{2} \boldsymbol{I}_n \right) \boldsymbol{A} \left(\frac{-1}{\alpha^2} \boldsymbol{\Sigma}^{-\frac{1}{2}} - \frac{1}{4} \boldsymbol{\Sigma}^{\frac{1}{2}} \right) \boldsymbol{A} \right) \\
&= \left(\frac{\beta_t^2}{2} \boldsymbol{\omega}^\top \boldsymbol{A} - \beta_t \boldsymbol{V}_{\boldsymbol{\omega}}^\top \right) \boldsymbol{A} \left(\frac{1}{\alpha^2} \boldsymbol{\Sigma}^{-\frac{1}{2}} + \frac{1}{4} \boldsymbol{\Sigma}^{\frac{1}{2}} \right) \boldsymbol{A} \left(\frac{2}{\alpha^2} \boldsymbol{\Sigma}^{-1} - \frac{1}{2} \boldsymbol{I}_n \right) - \frac{4\beta_t}{\alpha^3} \boldsymbol{V}_{\boldsymbol{\omega}}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{A}.
\end{aligned}$$

Acknowledgements

The authors thank the Editors and three anonymous referees for their constructive comments on an earlier version of this manuscript which resulted in this improved version. This research work was partially supported by CNPq and CAPES grants from the Brazilian government, and by FONDECYT 1160868 grant from the Chilean government.

References

Allard, D. and Naveau, P. (2007). A new spatial skew-normal random field model. *Communications in Statistics: Theory and Methods*, 36:1821–1834.

Assumpção, R. A. B., Uribe-Opazo, M. A., and Galea, M. (2014). Analysis of local influence in geostatistics using Student-t distribution. *Journal of Applied Statistics*, 41:2323–2341.

Athayde, E., Azevedo, A., Barros, M., and Leiva, V. (2018). Failure rate of Birnbaum-Saunders distributions: shape, change-point, estimation and robustness. *Brazilian Journal of Probability and Statistics*, pages in press.

Azevedo, C., Leiva, V., Athayde, E., and Balakrishnan, N. (2012). Shape and change point analyses of the Birnbaum-Saunders-t hazard rate and associated estimation. *Computational Statistics and Data Analysis*, 56:3887–3897.

Baran, S., Sikolya, S., and Stehlik, M. (2015) Optimal designs for the methane flux in troposphere. *Chemometrics and Intelligent Laboratory Systems*, 146:407–417

Barros, M., Leiva, V., Ospina, R., and Tsuyuguchi, A. (2014). Goodness-of-fit tests for the Birnbaum-Saunders distribution with censored reliability data. *IEEE Transactions on Reliability*, 63:543–554.

Billor, N. and Loynes, R. (1993). Local influence: A new approach. *Communications in Statistics: Theory and Methods*, 22:1595–1611.

Caro-Lopera, F., Leiva, V. and Balakrishnan, N. (2012). Connection between the Hadamard and matrix products with an application to matrix-variate Birnbaum-Saunders distributions. *Journal of Multivariate Analysis*, 104:126–139.

Cambardella, C., Moorman, T., Novak, J., Parkin, T., Karlen, D., Turco, R., and Konopka, A. (1994). Field-scale variability of soil properties in central Iowa soils. *Soil Science Society of America Journal*, 58:1501–1511.

Cook, R. D. (1987). Influence assessment. *Journal of Applied Statistics*, 14:117–131.

Cook, R. D., Peña, D., and Weisberg, S. (1988). The likelihood displacement: A unifying principle for influence measures. *Communications in Statistics: Theory and Methods*, 17:623–640.

Cook, R. D. and Weisberg, S. (1982). *Residuals and Influence in Regression*. Chapman and Hall, London, UK.

Cressie, N. (2015). *Statistics for Spatial Data*. Wiley, New York, US.

De Bastiani, F., Cysneiros, A. H. M. A., Uribe-Opazo, M. A., and Galea, M. (2015). Influence diagnostics in elliptical spatial linear models. *TEST*, 24:322–340.

Desousa, M. F., Saulo, H., Leiva, V., and Scalco, P. (2018). On a tobit-Birnbaum-Saunders model with an application to antibody response to vaccine. *Journal of Applied Statistics*, pages in press.

Diggle, P. J. and Ribeiro, P. J. (2007). *Model-Based Geoestatistics*. Springer, New York, US.

Ferreira, M., Gomes, M. I., and Leiva, V. (2012). On an extreme value version of the Birnbaum-Saunders distribution. *REVSTAT Statistical Journal*, 10:181–210.

Fung, W. K. (1995). A cautionary note on the use of generalized Cook-type measures.. *Computational Statistics & Data Analysis*, 19:321–326.

Galea, M., Paula, G. A., and Uribe-Opazo, M. A. (2003). On influence diagnostic in univariate elliptical linear regression models. *Statistical Papers*, 44:23–45.

Garcia-Papani, F., Uribe-Opazo, M. A., Leiva, V., and Aykroyd, R. G. (2017). Birnbaum-Saunders spatial modelling and diagnostics applied to agricultural engineering data. *Stochastic Environmental Research and Risk Assessment*, 31:105–124.

Gel, Y. R. and Gastwirth, J. L. (2008). A robust modification of the Jarque-Bera test of normality. *Economics Letters*, 99:30–32.

Gimenez, P. and Galea, M. (2013). Influence measures on corrected score estimators in functional heteroscedastic measurement error models. *Journal of Multivariate Analysis*, 114:1–15.

Green, R. L. and Kalivas, J. H. (2002). Graphical diagnostics for regression model determinations with consideration of the bias/variance trade-off. *Chemometrics and Intelligent Laboratory Systems*, 60:173–188.

Hengl, T., Heuvelink, G., and Stein, A. (2004). A generic framework for spatial prediction of soil variables based on regression-Kriging. *Geoderma*, 120:75–93.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994). *Continuous Univariate Distributions*, volume 1. Wiley, New York, US.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1995). *Continuous Univariate Distributions*, volume 2. Wiley, New York, US.

Kim, M. G. (2017). A cautionary note on the use of Cook's distance. *Communications for Statistical Applications and Methods*, 24:317–324.

Kundu, D., Balakrishnan, N., and Jamalizadeh, A. (2013). Generalized multivariate Birnbaum-Saunders distributions and related inferential issues. *Journal of Multivariate Analysis*, 116:230–244.

Leão, J., Leiva, V., Saulo, H., and Tomazella, V. (2017). Birnbaum-Saunders frailty regression models: Diagnostics and application to medical data. *Biometrical Journal*, 59:291–314.

Leiva, V. (2016). *The Birnbaum-Saunders Distribution*. Academic Press, New York, US.

Leiva, V., Athayde, E., Azevedo, C., and Marchant, C. (2011). Modeling wind energy flux by a Birnbaum-Saunders distribution with unknown shift parameter. *Journal of Applied Statistics*, 38:2819–2838.

Leiva, V., Ferreira, M., Gomes, M. I., and Lillo, C. (2016a). Extreme value Birnbaum-Saunders regression models applied to environmental data. *Stochastic Environmental Research and Risk Assessment*, 30:1045–1058.

Leiva, V., Liu, S., Shi, L., and Cysneiros, F. J. A. (2016b). Diagnostics in elliptical regression models with stochastic restrictions applied to econometrics. *Journal of Applied Statistics*, 43:627–642.

Leiva, V., Marchant, C., Ruggeri, F., and Saulo, H. (2015). A criterion for environmental assessment using Birnbaum-Saunders attribute control charts. *Environmetrics*, 26:463–476.

Leiva, V., Rojas, E., Galea, M., and Sanhueza, A. (2014a). Diagnostics in Birnbaum-Saunders accelerated life models with an application to fatigue data. *Applied Stochastic Models in Business and Industry*, 30:115–131.

Leiva, V., Santos-Neto, M., Cysneiros, F. J. A., and Barros, M. (2014b). Birnbaum-Saunders statistical modelling: A new approach. *Statistical Modelling*, 14:21–48.

Longford, N. T. (2005). *Missing Data and Small Area Estimation*. Springer, New York, US.

Lopes, A. S. (1998). *International Soil Fertility Manual (in Portuguese)*. Potafos, Piracicaba, Brazil.

Magnus, J. and Neudecker, H. (2007). *Matrix Differential Calculus with applications in statistics and econometrics*. Wiley, Chichester, UK.

Marchant, C., Leiva, V., Cavieres, M. F., and Sanhueza, A. (2013). Air contaminant statistical distributions with application to PM10 in Santiago, Chile. *Reviews of Environmental Contamination and Toxicology*, 223:1–31.

Marchant, C., Leiva, V., and Cysneiros, F. J. A. (2016a). A multivariate log-linear model for Birnbaum-Saunders distributions. *IEEE Transactions on Reliability*, 65:816–827.

Marchant, C., Leiva, V., Cysneiros, F. J. A., and Liu, S. (2018). Robust multivariate control charts based on Birnbaum-Saunders distributions. *Journal of Statistical Computation and Simulation*, 88:182–202.

Marchant, C., Leiva, V., Cysneiros, F. J. A., and Vivanco, J. F. (2016b). Diagnostics in multivariate generalized Birnbaum-Saunders regression models. *Journal of Applied Statistics*, 43:2829–2849.

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

Militino, A., Palacius, M., and Ugarte, M. (2006). Outliers detection in multivariate spatial linear models. *Journal of Statistical Planning and Inference*, 136:125–146.

Nocedal, J. and Wright, S. (1999). *Numerical Optimization*. Springer, New York, US.

Noufaily, A. and Jones, M. (2013). Parametric quantile regression based on the generalized gamma distribution. *Journal of the Royal Statistical Society C*, 62:723–740.

Pan, J., Fei, Y., and Foster, P. (2014). Case-deletion diagnostics for linear mixed models. *Technometrics*, 56:269–281.

Poon, W. Y. and Poon, Y. S. (1999). Conformal normal curvature and assessment of local influence. *Journal of the Royal Statistical Society B*, 61:51–61.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Reich, B. R. (2011). Spatio-temporal quantile regression for detecting distributional changes in environmental processes. *Journal of the Royal Statistical Society C*, 61:535–553.

Rieck, J. R. and Nedelman, J. R. (1991). A log-linear model for the Birnbaum-Saunders distribution. *Technometrics*, 3:51–60.

Rimstad, K. and Omre, H. (2014). Skew-Gaussian random fields. *Spatial Statistics*, 10:43–62.

Santana, L., Vilca, F., and Leiva, V. (2011). Influence analysis in skew-Birnbaum-Saunders regression models and applications. *Journal of Applied Statistics*, 38:1633–1649.

Santos-Neto, M., Cysneiros, F. J. A., Leiva, V., and Barros, M. (2016). Reparameterized Birnbaum-Saunders regression models with varying precision. *Electronic Journal of Statistics*, 10:2825–2855.

Saulo, H., Leão, J., Leiva, V., and Aykroyd, R. G. (2017). Birnbaum-Saunders autoregressive conditional duration models applied to high-frequency financial data. *Statistical Papers*, pages in press.

Saulo, H., Leiva, V., Ziegelmann, F. A., and Marchant, C. (2013). A nonparametric method for estimating asymmetric densities based on skewed Birnbaum-Saunders distributions applied to environmental data. *Stochastic Environmental Research and Risk Assessment*, 27:1479–1491.

Stehlík, M., Strelec, L., and Thulin, M. (2014). On robust testing for normality in chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 130:98–108.

Stein, M. L. (2012). *Interpolation of spatial data: some theory for kriging*. Springer, New York, US.

Uribe-Opazo, M. A., Borssoi, J. A., and Galea, M. (2012). Influence diagnostics in Gaussian spatial linear models. *Journal of Applied Statistics*, 39:615–630.

Villegas, C., Paula, G. A. and Leiva, V. (2011). Birnbaum-Saunders mixed models for censored reliability data analysis. *IEEE Transactions on Reliability*, 60:748–758.

Wolter, K. M. (2007). *Introduction to Variance Estimation*. Springer-Verlag, New York, US.

Xia, J., Zeephongsekul, P., and Packer, D. (2011). Spatial and temporal modelling of tourist movements using semi-Markov processes. *Tourism Management*, 51:844–851.

Zhang, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*, 99:250–261.

Zhang, H. and Zimmerman, D. L. (2005). Towards reconciling two asymptotic frameworks in spatial statistics. *Biometrika*, 92:921–936.

Zhu, H., Ibrahim, J. G., Lee, S., and Zhang, H. (2007). Perturbation selection and influence measures in local influence analysis. *The Annals of Statistics*, 35:2565–2588.