Influence Diagnostics in Spatial Models with Censored Response

Thais S. Barbosa[a], Victor H. Lachos[a*] and Dipak K. Dey[b]

[a] Department of Statistics, Campinas State University, Campinas, São Paulo, Brazil

[b] Department of Statistics, University of Connecticut, Storrs, CT, USA

Abstract

Environmental data is often spatially correlated and sometimes include below detection limit observations (i.e., censored values reported as less than a level of detection). Existing work mainly concentrates on parameter estimation using Gibbs sampling, and work conducted from a frequentist perspective in spatial censored models are elusive. In this paper, we propose an exact estimation procedure to obtain the maximum likelihood estimates of the fixed effects and variance components, using a stochastic approximation of the EM (SAEM) algorithm (Delyon *et al.*, 1999). This approach permits estimation of the parameters of spatial linear models when censoring is present in an easy and fast way. As a by-product, predictions of unobservable values of the response variable are possible. Motivated by this algorithm, we develop local influence measures on the basis of the conditional expectation of the complete-data log-likelihood function which eliminates the complexity associated with the approach of Cook (1977, 1986) for spatial censored models. Some useful perturbation schemes are discussed. The newly developed methodology is illustrated using data from a dioxin contaminated site in Missouri. In addition, a simulation study is presented, which explores the accuracy of the proposed measures in detecting influential observations under different perturbation schemes.

Key words: Censored data, Geostatistical data, SAEM Algorithm, Influential observations, Limit of detection (LOD).

1. Introduction

Spatial data are common in ecology, environmental health, mining, hydrology and epidemiology, where sampling units are geographical areas or spatially located individuals. Analysis of spatial data is challenged by the spatial correlation among the observations such as the conditional autoregressive (CAR) structure or the Matèrn correlation structure. An additional complication is that spatial data are subjected to upper or lower detection limits below and above which they are not quantifiable. For example, environmental (spatial) monitoring of different variables often involves left-censored observations falling below the minimum limit of detection (LOD) of the instruments used to quantify them. The proportion of censored data in these studies may not be small and so the use of crude/ad hoc methods, such as substituting a threshold value or some arbitrary point like a mid-point between zero and cut-off for detection, might lead to biased estimates of fixed effects and variance components (De Oliveira, 2005; Fridley & Dixon, 2007).

Address for correspondence: Víctor Hugo Lachos Dávila, Departamento de Estatística, IMECC, Universidade Estadual de Campinas, CEP 13083-859, Campinas, São Paulo, Brazil. E-mail: hlachos@ime.unicamp.br.

As an alternative to crude imputation methods Militino & Ugarte (1999) develop an EM algorithm for maximum likelihood (ML) estimation in censored spatial data, however this approach suffers from several drawbacks that restrict its applicability. For instance, De Oliveira (2005) notes that this ML approach does not provide a means to estimate the correlation structure in the data and hence assume it is known. In addition, when estimating variables at non-sampled locations the observed and imputed data are not differentiated, thus underestimating the predictive uncertainty. Due to the complexity of the likelihood-based methods, that involves computationally intractable integrals, De Oliveira (2005) and Fridley & Dixon (2007) adopt a Bayesian approach to inference and prediction for spatially correlated censored observations. In both papers data augmentation and Markov chain Monte Carlo (MCMC) algorithms are utilized. In this paper, we first propose a stochastic versions of the EM algorithm for ML estimation, the so-called SAEM algorithm proposed by Delyon *et al.* (1999). Then, the diagnostic measures for assessing the local influence in spatial censored linear (SCL) models are developed and presented. In the framework of spatial models, Jank (2006) showed that the computational effort of SAEM is much smaller and reach the convergence in just a fraction of the simulation size when compared to Monte Carlo EM (MCEM) algorithm. This is due to the memory effect contained in the SAEM method, in which the previous simulations are considered in the computation of the posterior ones.

The study of influence analysis is an important and key step in data analysis subsequent to parameter estimation. This can be carried out by conducting an influence analysis for detecting influential observations. There are two primary approaches for detecting influential observations. The first approach is the case-deletion approach (Cook, 1977) and it is an intuitively appealing method (see also Cook and Weisberg, 1982). Deletion diagnostics such as Cook's distance or the likelihood distance have been applied to many statistical models. The second approach, which is a general statistical technique used to assess the stability of the estimation outputs with respect to the model inputs, is the local influence approach of Cook (1986). Following the pioneering work of Cook (1986), this method has received considerable attention recently in the statistical literature of spatial models; see, for example, Assumpção *et al.* (2014) and De Bastiani *et al.* (2014).

Although several diagnostic studies on spatial models have appeared in the literature, to the best of our knowledge, no study seems to have been made on influence diagnostics for cesored spatial data and certainly not on the local influence analysis. The main difficulty is due to the fact that the observed log-likelihood function of censored spatial models involves intractable integrals (for instance, the pdfs of truncated multinormal distributions), rendering the direct application of Cook's approach (Cook, 1986) to this model to be very difficult if not impossible, since the measures involve the first and second partial derivatives of this function. Zhu & Lee (2001) developed an approach for performing local influence analysis for general statistical models with missing data, and it is based on the Q-displacement function that is closely related to the conditional expectation of the complete-data log-likelihood in the E-step of the EM algorithm. This approach produces results very similar to those obtained from Cook's method. Moreover, the case-deletion can be studied by Q-displacement function following the approach of Zhu *et al.* (2001). So, we develop here methods to obtain case-deletion measures and local influence measures by using the method of Zhu *et al.* (2001) (see also Lee & Xu, 2004; Zhu & Lee, 2001) in the context of SCL models.

The rest of the paper is organized as follows. Section 2 gives a brief description of the spatial linear model, including an outline of the SAEM algorithm. Section 3 proposes the SCL model and shows how to get the ML estimates through the SAEM algorithm. In Section 4, we provide a brief sketch of the local influence approach for models with incomplete data, and also develop a methodology pertinent to the SCL model. Three different perturbation schemes are considered. The methodology is illustrated in Section 5 with the analysis of a data set from a dioxin contaminated site in Missouri and by empirical studies in Section 6. Section 7 concludes with a short

discussion of issues raised by our study and some possible directions for a future research.

2. Preliminaries

2.1. The Spatial linear model

As in De Bastiani *et al.* (2014), we consider a Gaussian stochastic process $\{Z(\mathbf{s}), \mathbf{s} \in \mathbb{D}\}$, where $\mathbb{D}$ is a subset of $\mathbb{R}^d$, the d-dimensional Euclidean space. It supposes that data $Z(\mathbf{s}_1), ..., Z(\mathbf{s}_n)$ of this process are observed at known sites (locations) $\mathbf{s}_i$, for $i = 1, \ldots, n$, where $\mathbf{s}_i$ is a d-dimensional vector of spatial site coordinates, and generated from the model,

$$Z(\mathbf{s}_i) = \mu(\mathbf{s}_i) + \epsilon(\mathbf{s}_i), \tag{1}$$

where both the deterministic term $\mu(\mathbf{s}_i)$ and the stochastic term $\epsilon(\mathbf{s}_i)$ may depend on the spatial location at which $Z(\mathbf{s}_i)$ is observed. We assume that the stochastic errors have zero mean, $E\{\epsilon(\mathbf{s}_i)\} = 0$, and that variation between spatial points is determined by a covariance function $C(\mathbf{s}_i, \mathbf{s}_j) = \text{Cov}\{\epsilon(\mathbf{s}_i), \epsilon(\mathbf{s}_j)\}$. Suppose that for some known functions of $\mathbf{s}_i$, $x_1(\mathbf{s}_i), \ldots, x_p(\mathbf{s}_i)$, the mean of the stochastic process is

$$\mu(\mathbf{s}_i) = \sum_{j=1}^{p} x_j(\mathbf{s}_i)\beta_j, \tag{2}$$

where $\beta_1, \ldots, \beta_p$ are unknown parameters to be estimated. In addition, each family of covariance functions $C(\mathbf{s}_i, \mathbf{s}_j)$, is fully specified by a q-dimensional parameter vector $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_q)^\top$. We use the following notations: $Z(\mathbf{s}_i) = Z_i$, $\mathbf{Z} = (Z_1, \ldots, Z_n)^\top$, $x_{ij} = x_j(\mathbf{s}_i)$, $\mathbf{x}_i^\top = (x_{i1}, \ldots, x_{ip})$, $\mathbf{X}$ as the $n \times p$ matrix with ith row $\mathbf{x}_i^\top$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$, $\epsilon_i = \epsilon(\mathbf{s}_i)$, and $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^\top$, with $i = 1, \ldots, n$ and $j = 1, \ldots, p$. Thus, $\mu(\mathbf{s}_i) = \mathbf{x}_i^\top \boldsymbol{\beta}$ and then $Z_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i$, $i = 1, \ldots, n$. Equivalent, in matrix notation, we have the spatial linear model

$$\mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{3}$$

where $E\{\boldsymbol{\epsilon}\} = \mathbf{0}$ and the scale matrix of $\boldsymbol{\epsilon}$, is $\boldsymbol{\Sigma} = [C(\mathbf{s}_i, \mathbf{s}_j)] = \tau^2\mathbf{I}_n + \sigma^2\mathbf{R}(\rho)$. We assume that $\boldsymbol{\Sigma}$ is non singular and that $\mathbf{X}$ has a full rank. τ^2 can be viewed as a measurement error variance or a nugget effect, σ^2 is defined as sill, $\mathbf{R} = \mathbf{R}(\rho) = [r_{ij}]$, is an $n \times n$ symmetric matrix with diagonal elements $r_{ii} = 1$, for $i = 1, \ldots, n$ and ρ is a function of the range of the model. In general $\mathbf{R}$ depends on the Euclidean distance $d_{ij} = ||\mathbf{s}_i - \mathbf{s}_j||$ between the points $\mathbf{s}_i$ and $\mathbf{s}_j$. This parametric form occurs for several isotropic processes, for instance, the Matérn family of correlation functions $\mathbf{R}$ is defined by

$$\mathbf{R}(\rho) = \mathbf{R}(\rho, d_{ij}) = \begin{cases} \dfrac{1}{2^{\kappa-1}\Gamma(\kappa)} \left(\dfrac{d_{ij}}{\rho}\right)^{\kappa} K_\kappa(d_{ij}/\rho), & d_{ij} > 0, \\ 1, & d_{ij} = 0, \end{cases}$$

where $\rho > 0$; $K_\kappa(u) = \frac{1}{2}\int_0^\infty x^{\kappa-1}e^{-\frac{1}{2}u(x+x^{-1})}dx$ is the modified Bessel function of the third kind of order κ; (see, Gradshtejn & Ryzhik, 1965), with $\kappa > 0$ fixed. The Gaussian covariance function is a special case when $\kappa \to \infty$ and its correlation function is given by

$$\mathbf{R}(\rho) = \mathbf{R}(\rho, d_{ij}) = \begin{cases} \exp\left\{-\left(\dfrac{d_{ij}}{\rho}\right)^2\right\}, & d_{ij} > 0, \\ 1, & d_{ij} = 0. \end{cases}$$

The exponential covariance is also a special case of Matérn Family, which corresponds to $\kappa = 1/2$ and can be written more simply as

$$\mathbf{R}(\rho) = \mathbf{R}(\rho, d_{ij}) = \begin{cases} \exp\left\{-\left(\frac{d_{ij}}{\rho}\right)\right\}, & d_{ij} > 0, \\ 1, & d_{ij} = 0. \end{cases}$$

In classical geostatistics, the spherical family is also widely used. This has correlation function given by

$$\mathbf{R}(\rho) = \mathbf{R}(\rho, d_{ij}) = \begin{cases} 1 - \frac{3}{2}(d_{ij}/\rho) + \frac{1}{2}(d_{ij}/\rho)^3, & 0 \leq d_{ij} \leq \rho, \\ 0, & d_{ij} > \rho. \end{cases}$$

where $\rho > 0$ is a single parameter with the dimensions of distance. One qualitative difference between this and the Matérn family is that it has a finite range i.e., $\mathbf{R}(\rho) = 0$ for sufficiently large d_{ij}, namely $d_{ij} > \rho$. The spherical family lacks flexibility by comparison with the two-parameter Matérn class. Also, $\mathbf{R}(\rho)$ is only once differentiable at $d_{ij} = \rho$, which causes technical difficulties with maximum likelihood estimation.

2.2. The EM and SAEM algorithms

In models with missing and censored data, the EM algorithm (Dempster *et al.*, 1977) has established itself as the most popular tool for obtaining the ML estimates of model parameters. Let define $\mathbf{z}_{\rm com} = (\mathbf{z}_{\rm m}, \mathbf{z}_{\rm obs})$, where $\mathbf{z}_{\rm m}$ denotes the missing data and $\mathbf{z}_{\rm obs}$ the observed data, this iterative algorithm maximizes the complete log-likelihood function $\ell_c(\boldsymbol{\theta}; \mathbf{z}_{\rm com})$ at each step, converging quickly to a stationary point of the observed likelihood $(\ell(\boldsymbol{\theta}; \mathbf{z}_{\rm obs}))$ under mild regularity conditions (Wu, 1983; Vaida, 2005). The EM algorithm proceeds in two simple steps:

E-Step: Replace the observed likelihood by the complete likelihood and compute its conditional expectation $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) = \mathrm{E}\left\{\ell_c(\boldsymbol{\theta}; \mathbf{z}_{\rm com})|\widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{z}_{\rm obs}\right\}$, where $\widehat{\boldsymbol{\theta}}^{(k)}$ is the estimate of $\boldsymbol{\theta}$ at the *k*-th iteration;

M-Step: Maximize $Q(\theta|\widehat{\boldsymbol{\theta}}^{(k)})$ with respect to $\boldsymbol{\theta}$ to obtain $\widehat{\boldsymbol{\theta}}^{(k+1)}$.

However, in some applications of the EM algorithm, the E-step cannot be obtained analytically and has to be calculated using simulations. Wei & Tanner (1990) proposed the Monte Carlo EM (MCEM) algorithm in which the E-step is replaced by a Monte Carlo approximation based on a large number of independent simulations of the missing data. This simple solution is infact computationally expensive, given the need to generate a large number of independent simulations of the missing data for a good approximation. Thus, in order to reduce the amount of required simulations compared to the MCEM algorithm, the SAEM algorithm proposed by Delyon *et al.* (1999) replaces the E-step of the EM algorithm by a stochastic approximation procedure, while the Maximization step remains unchanged. Besides having good theoretical properties, the SAEM estimates the population parameters accurately, converging to the global maxima of the ML estimates under quite general conditions (Allassonnière *et al.*, 2010; Delyon *et al.*, 1999; Kuhn & Lavielle, 2004). At each iteration, the SAEM algorithm successively simulates missing data with the conditional distribution, and updates the unknown parameters of the model. Thus, at iteration *k*, the SAEM proceeds as follows:

E-Step:

- Simulation: Draw $(\mathbf{q}^{(\ell,k)})$, $\ell = 1, \ldots, m$ from the conditional distribution $f(\mathbf{q}|\boldsymbol{\theta}^{(k-1)}, \mathbf{z}_i)$.
- Stochastic Approximation: Update the $Q(\theta|\widehat{\theta}^{(k)})$ function as

$$Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) \approx Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k-1)}) + \delta_k \left[\frac{1}{m}\sum_{\ell=1}^{m} \ell_c(\theta\,; \mathbf{z}_{\rm obs}, \mathbf{q}^{(\ell,k)})|\widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{z}_{\rm obs} - Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k-1)})\right]. \tag{4}$$

M-Step:

- Maximization: Update $\widehat{\boldsymbol{\theta}}^{(k)}$ as $\widehat{\boldsymbol{\theta}}^{(k+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$,

where δ_k is a smoothness parameter (Kuhn & Lavielle, 2004), i.e., a decreasing sequence of positive numbers such that $\sum_{k=1}^{\infty} \delta_k = \infty$ and $\sum_{k=1}^{\infty} \delta_k^2 < \infty$. Note that, for the SAEM algorithm, the E-Step coincides with the MCEM algorithm, however a small number of simulations m (suggested to be $m \leq 20$) is necessary. This is possible because unlike the traditional EM algorithm and its variants, the SAEM algorithm uses not only the current simulation of the missing data at the iteration *k* denoted by $(\mathbf{q}^{(\ell,k)})$, $\ell = 1, \ldots, m$ but some or all previous simulations, where this 'memory' property is set by the smoothing parameter δ_k.

Note, in Equation (4), if the smoothing parameter δ_k is equal to 1 for all k, the SAEM algorithm will have 'no memory', and will be equivalent to the MCEM algorithm. The SAEM with no memory will converge quickly (convergence in distribution) to a solution neighborhood, however the algorithm with memory will converge slowly (almost sure convergence) to the ML solution. We suggested the following choice of the smoothing parameter:

$$\delta_k = \begin{cases} 1, & \text{for} \quad 1 \leq k \leq cW, \\ \frac{1}{k-cW}, & \text{for} \quad cW+1 \leq k \leq W, \end{cases}$$

where W is the maximum number of iterations, and c a cut point ($0 \leq c \leq 1$) which determines the percentage of initial iterations with no memory. For example, if $c = 0$, the algorithm will have memory for all iterations, and hence will converge slowly to the ML estimates. If $c = 1$, the algorithm will have no memory, and so will converge quickly to a solution neighborhood. For the first case, W would need to be large in order to achieve the ML estimates. For the second, the algorithm will output a Markov Chain where after applying a *burn in* and *thining*, the mean of the chain observations can be a reasonable estimate.

A number between 0 and 1 ($0 < c < 1$) will assure an initial convergence in distribution to a solution neighborhood for the first cW iterations and an almost sure convergence for the rest of the iterations. Hence, this combination will lead us to a fast algorithm with good estimates. To implement SAEM, the user must fix several constants matching the number of total iterations W and the cut point c that defines the starting of the smoothing step of the SAEM algorithm, however those parameters will vary depending on the model and the data. To determine those constants, a graphical approach is recommended to monitor the convergence of the estimates for all the parameters, and, if possible, to monitor the difference (relative difference) between two successive evaluations of the log-likelihood $\ell(\boldsymbol{\theta}|\mathbf{z}_{obs})$, given by $||\ell(\boldsymbol{\theta}^{(k+1)}|\mathbf{y}_{obs}) - \ell(\boldsymbol{\theta}^{(k)}|\mathbf{z}_{obs})||$ or $||\ell(\boldsymbol{\theta}^{(k+1)}|\mathbf{z}_{obs})/\ell(\boldsymbol{\theta}^{(k)}|\mathbf{z}_{obs}) - 1||$, respectively.

3. The spatial linear model for censored responses

We will consider the Gaussian model with a linear specification for the spatial trend, which allows the inclusion of a polynomial trend surface or, more generally, spatially referenced covariates. From (3), the spatial linear model is defined by

$$\mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{5}$$

where $\mathbf{X}$ and $\boldsymbol{\beta}$ are as defined in (3) and $\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \boldsymbol{\Sigma})$. Moreover, we assume that the response Z_i is not fully observed for all i, $i = 1, \ldots, n$. Let the observed data for the ith area be (V_i, C_i), where V_i represents the vector

of uncensored reading or censoring level, and C_i is the vector of censoring indicators such that

$$\begin{aligned} Z_i &\leq V_i \quad \text{if} \quad C_i = 1, \\ Z_i &= V_i \quad \text{if} \quad C_i = 0. \end{aligned} \tag{6}$$

Note that since the observed response Z_i, $i = 1, \ldots, n$, is defined over the real line, extensions to right censored data are straightforward. In fact, the right censored problem can be represented by a left censored problem by simultaneously transforming the response Z_i and censoring level V_i to $-Z_i$ and $-V_i$. The model defined in (5)-(6), will be called spatial censored linear (SCL) model.

3.1. The log-likelihood function

Classical inference on the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\alpha}^\top)^\top$, with $\boldsymbol{\alpha} = (\tau^2, \sigma^2, \rho)^\top$, is based on the marginal distribution of $\mathbf{Z}$. For complete data, we have marginally that $\mathbf{Z} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \tau^2\mathbf{I}_n + \sigma^2\mathbf{R}(\rho)$. For responses with censoring pattern as in (6), we have $\mathbf{Z} \sim TN_n(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}; \mathbb{A})$, where $TN_n(.; \mathbb{A})$ denotes the truncated normal distribution on the interval $\mathbb{A}$, where $\mathbb{A} = A_1 \times \ldots \times A_n$, with A_i being the interval $(-\infty, \infty)$ if $C_i = 0$ and the interval $(-\infty, V_i]$ if $C_i = 1$. For computing the likelihood function associated with model (5)-(6), the first step is to treat separately the observed and censored components of Z_i.

Let $\mathbf{Z}^o$ be the n^o-vector of observed outcomes and $\mathbf{Z}^c$ be the n^c-vector of censored observations with $(n = n^o + n^c)$ such that $C_i = 0$ for all elements in $\mathbf{Z}^o$, and 1 for all elements in $\mathbf{Z}^c$. After reordering, $\mathbf{Z}$, $\mathbf{V}$, $\mathbf{X}$, and $\boldsymbol{\Sigma}$ can be partitioned as follows:

$$\mathbf{Z} = vec(\mathbf{Z}^o, \mathbf{Z}^c),\ \mathbf{V} = vec(\mathbf{V}^o, \mathbf{V}^c),\ \mathbf{X}^\top = (\mathbf{X}^o, \mathbf{X}^c) \text{ and } \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}^{oo} & \boldsymbol{\Sigma}^{oc} \\ \boldsymbol{\Sigma}^{co} & \boldsymbol{\Sigma}^{cc} \end{pmatrix},$$

where $vec(.)$ denotes the function which stacks vectors or matrices of the same number of columns. Then, we have $\mathbf{Z}^o \sim N_{n^o}(\mathbf{X}^o\boldsymbol{\beta}, \boldsymbol{\Sigma}^{oo})$, $\mathbf{Z}^c|\mathbf{Z}^o \sim N_{n^c}(\boldsymbol{\mu}, \mathbf{S})$, where $\boldsymbol{\mu} = \mathbf{X}^c\boldsymbol{\beta} + \boldsymbol{\Sigma}^{co}(\boldsymbol{\Sigma}^{oo})^{-1}(\mathbf{Z}^o - \mathbf{X}^o\boldsymbol{\beta})$ and $\mathbf{S} = \boldsymbol{\Sigma}^{cc} - \boldsymbol{\Sigma}^{co}(\boldsymbol{\Sigma}^{oo})^{-1}\boldsymbol{\Sigma}^{oc}$. Now, let $\Phi_n(\mathbf{u}; \mathbf{a}, \mathbf{A})$ and $\phi_n(\mathbf{u}; \mathbf{a}, \mathbf{A})$ be the cdf (left tail) and pdf, respectively, of $N_n(\mathbf{a}, \mathbf{A})$ computed at vector $\mathbf{u}$. From Vaida & Liu (2009) and Matos *et al.* (2013), the likelihood function (using conditional probability arguments) is given by

$$\begin{aligned} L(\boldsymbol{\theta}) = f(\mathbf{Z}|\boldsymbol{\theta}) &= P(\mathbf{V}|\mathbf{C}, \boldsymbol{\theta}) = P(\mathbf{Z}^c \leq \mathbf{V}^c|\mathbf{Z}^o = \mathbf{V}^o, \boldsymbol{\theta})P(\mathbf{Z}^o = \mathbf{V}^o|\boldsymbol{\theta}), \\ &= P(\mathbf{Z}^c \leq \mathbf{V}^c|\mathbf{Z}^o, \boldsymbol{\theta})f(\mathbf{Z}^o|\boldsymbol{\theta}) \\ &= \phi_{n^o}(\mathbf{Z}^o; \mathbf{X}^o\boldsymbol{\beta}, \boldsymbol{\Sigma}^{oo})\Phi_{n^c}(\mathbf{V}^c; \boldsymbol{\mu}, \mathbf{S}), \end{aligned} \tag{7}$$

which can be evaluated without much computational burden through the routine *mvtnorm()* available in R (see, Genz *et al.*, 2008; R Development Core Team, 2015).

Straightforwardly, the log-likelihood function for the observed data is given by $\ell(\boldsymbol{\theta}|\mathbf{z}) = \log(L(\boldsymbol{\theta}))$. The log-likelihood function for the observed data is used to compute different model selection criteria, such as:

$$AIC = 2\,m - 2\,\ell_{max} \text{ and } BIC = m\,log\,n - 2\,\ell_{max},$$

where m is the number of model parameters and ℓ_{max} is the maximized log-likelihood value.

3.2. The SAEM algorithm

In this section, we propose the SAEM algorithm by considering $\mathbf{Z}$ as missing data to update (M-step) all the parameters involved in the model. First, we parameterise to $\nu^2 = \tau^2/\sigma^2$ and write $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{\Psi}$, with $\boldsymbol{\Psi} =$

$\nu^2\mathbf{I}_n + \mathbf{R}(\rho)$. Now, let $\mathbf{Z} = (Z_1, \ldots, Z_n)^\top$, $\mathbf{V} = (V_1, \ldots, V_n)$ and $\mathbf{C} = (C_1, \ldots, C_n)$, and that we observe (V_i, C_i) for the ith subject. In their estimation procedure, $\mathbf{V}$ and $\mathbf{C}$ are treated as hypothetical missing data, and augmented with the observed data set $\mathbf{Z}_c = (\mathbf{C}^\top, \mathbf{V}^\top, \mathbf{Z}^\top)^\top$. Hence, the complete-data log-likelihood function is given by

$$\ell_c(\boldsymbol{\theta}|\mathbf{Z}_c) \quad \propto \quad -\frac{1}{2}\left[\log(|\boldsymbol{\Psi}|) + n\log(\sigma^2) + \frac{1}{\sigma^2}(\mathbf{Z} - \mathbf{X}\boldsymbol{\beta})^\top \boldsymbol{\Psi}^{-1}(\mathbf{Z} - \mathbf{X}\boldsymbol{\beta})\right] + c, \tag{8}$$

with c being a constant independent of the parameter vector $\boldsymbol{\theta}$. Given the current estimate $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(k)}$, the E-step calculates the conditional expectation of the complete data log-likelihood function given by

$$Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) \quad = \quad E[\ell_c(\boldsymbol{\theta}|\mathbf{Z}_c)|\mathbf{V}, \mathbf{C}, \widehat{\boldsymbol{\theta}}^{(k)}] = -\frac{1}{2}\left[\log(\boldsymbol{\Psi}) + n\log(\sigma^2) + \frac{1}{\sigma^2}\widehat{A}^{(k)}\right],$$

where $\widehat{A}^{(k)} = tr\left(\widehat{\mathbf{Z}\mathbf{Z}^\top}^{(k)}\boldsymbol{\Psi}^{-1}\right) - 2\widehat{\mathbf{Z}}^{(k)\top}\Psi^{-1}\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}^\top\mathbf{X}^\top\boldsymbol{\Psi}^{-1}\mathbf{X}\boldsymbol{\beta}$.

It is clear that the E-step reduces only to the computation of

$$\widehat{\mathbf{Z}\mathbf{Z}^\top}^{(k)} = E\{\mathbf{Z}\mathbf{Z}^\top|\mathbf{V}, \mathbf{C}, \widehat{\boldsymbol{\theta}}^{(k)}\} \text{ and } \widehat{\mathbf{Z}}^{(k)} = E\{\mathbf{Z}|\mathbf{V}, \mathbf{C}, \widehat{\boldsymbol{\theta}}^{(k)}\}, \tag{9}$$

In the traditional EM algorithm, we should now evaluate the conditional expectations. As there are no closed form for them, we have to introduce two intermediate steps, the simulation and approximation steps. In the simulation, we generate samples from the conditional distributions through the Gibbs sampler algorithm, according to the following scheme:

Step E-1 (Sampling). Sample $\mathbf{Z}^c$ from a truncated normal distribution of the form $TN_{n^c}(\boldsymbol{\mu}, \mathbf{S}; \mathbb{A}^c)$, with $\mathbb{A}^c = \{\mathbf{Z}^c = (Z_1^c, \ldots, Z_{n^c}^c)^\top | Z_1^c \leq V_1^c, \ldots, Z_{n^c}^c \leq V_{n^c}^c\}$, $\boldsymbol{\mu} = \mathbf{X}^c\boldsymbol{\beta} + \boldsymbol{\Sigma}^{co}(\boldsymbol{\Sigma}^{oo})^{-1}(\mathbf{Z}^o - \mathbf{X}^o\boldsymbol{\beta})$ and $\mathbf{S} = \boldsymbol{\Sigma}^{cc} - \boldsymbol{\Sigma}^{co}(\boldsymbol{\Sigma}^{oo})^{-1}\boldsymbol{\Sigma}^{oc}$.

Thus, the new observation $\mathbf{Z}^{(k,l)} = (Z_{i1}^{(k,l)}, \ldots, Z_{in^c}^{(k,l)}, Z_{n_i^c+1}, \ldots, Z_n)$ is a sample generated for the n^c censored cases and the observed values (uncensored cases), for $l = 1, \ldots, M$.

Step E-2 (Stochastic Approximation). Since we have the sequence $\mathbf{Z}^{(k,l)}$, at the k-th iteration, we replace the conditional expectations in (9) by the following stochastic approximations:

$$\widehat{\mathbf{Z}\mathbf{Z}^\top}^{(k)} \quad = \quad \widehat{\mathbf{Z}\mathbf{Z}^\top}^{(k-1)} + \delta_k\left[\frac{1}{M}\sum_{l=1}^{M}\widehat{\mathbf{Z}}^{(k,l)}\widehat{\mathbf{Z}}^{(k,l)\top} - \widehat{\mathbf{Z}\mathbf{Z}^\top}^{(k-1)}\right], \tag{10}$$

$$\widehat{\mathbf{Z}}^{(k)} \quad = \quad \widehat{\mathbf{Z}}^{(k-1)} + \delta_k\left[\frac{1}{M}\sum_{l=1}^{M}\widehat{\mathbf{Z}}^{(k,l)} - \widehat{\mathbf{Z}}^{(k-1)}\right]. \tag{11}$$

An advantage of SAEM is that, even though it performs a Monte Carlo E-step, it requires a small and fixed Monte Carlo sample size, making it much faster than MCEM. Some authors claim that $m \leq 10$ is large enough, but to be more conservative, we chose $m = 20$. Then, the Monte Carlo sample is combined in a smooth way with the previous step of the algorithm. The conditional maximization (CM) then conditionally maximizes $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ with respect to $\boldsymbol{\theta}$ and obtains a new estimate $\widehat{\boldsymbol{\theta}}^{(k+1)}$, as follows:

Step CM

$$\widehat{\boldsymbol{\beta}}^{(k+1)} = \left(\mathbf{X}^\top \widehat{\boldsymbol{\Psi}}^{-1(k)} \mathbf{X}\right)^{-1} \mathbf{X}^\top \widehat{\boldsymbol{\Psi}}^{-1(k)} \widehat{\mathbf{Z}}^{(k)}, \tag{12}$$

$$\widehat{\sigma^2}^{(k+1)} = \frac{1}{n}\left[tr\left(\widehat{\mathbf{Z}\mathbf{Z}^\top}^{(k)} \widehat{\boldsymbol{\Psi}}^{-1(k)}\right) - 2\widehat{\mathbf{Z}^\top}^{(k)} \widehat{\Psi}^{-1(k)} \mathbf{X}\widehat{\boldsymbol{\beta}}^{(k+1)} + \widehat{\boldsymbol{\beta}}^{\top(k+1)} \mathbf{X}^\top \widehat{\boldsymbol{\Psi}}^{-1(k)} \mathbf{X}\widehat{\boldsymbol{\beta}}^{(k+1)}\right], \tag{13}$$

$$\begin{aligned}(\widehat{\nu}^2, \widehat{\rho})^{(k+1)} = \underset{(\nu^2,\rho)\in\mathbb{R}^+\times\mathbb{R}^+}{\operatorname{argmax}} \Bigg(&-\frac{1}{2}\log(|\boldsymbol{\Psi}|) - \frac{1}{2\widehat{\sigma^2}^{(k+1)}}\left[tr\left(\widehat{\mathbf{Z}\mathbf{Z}^\top}^{(k)} \boldsymbol{\Psi}^{-1}\right)\right. \\ &\left.-2\widehat{\mathbf{Z}^\top}^{(k)} \Psi^{-1}\mathbf{X}\widehat{\boldsymbol{\beta}}^{(k+1)} + \widehat{\boldsymbol{\beta}}^{\top(k+1)} \mathbf{X}^\top \boldsymbol{\Psi}^{-1}\mathbf{X}\widehat{\boldsymbol{\beta}}^{(k+1)}\right]\Bigg).\end{aligned} \tag{14}$$

Note that $\widehat{\tau}^2$ can be recovered using that $\widehat{\tau}^{2(k+1)} = \widehat{\nu}^{2(k+1)}\widehat{\sigma}^{2(k+1)}$. The more efficient CM step (14) can be easily accomplished by using, for instance, the optim routine in R software. This process is iterated until some distance between two successive evaluations of the actual log-likelihood $\ell(\boldsymbol{\theta}|\mathbf{z})$ in Subsection 3.1, such as $|\ell(\widehat{\boldsymbol{\theta}}^{(k+1)}) - \ell(\widehat{\boldsymbol{\theta}}^{(k)})|$ or $|\ell(\widehat{\boldsymbol{\theta}}^{(k+1)})/\ell(\widehat{\boldsymbol{\theta}}^{(k)}) - 1|$, becomes small enough. The variance of the fixed effects in the SCL model is then given by (Hughes, 1999)

$$Var(\widehat{\boldsymbol{\beta}}) = \left(\mathbf{X}^\top \boldsymbol{\Sigma}^{-1}\mathbf{X} - \mathbf{X}^\top \boldsymbol{\Sigma}^{-1} Var(\mathbf{Z}|\mathbf{V}, \mathbf{C})\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}. \tag{15}$$

4. Diagnostic analysis

Influence diagnostics is widely used in statistical modeling to identify and evaluate aberrant and influential points which may cause unwanted effects on estimation and goodness of fit. There are several tools to proceed with influence diagnostics and here we used two approaches, case deletion, also known as global influence, and local influence diagnostics. The case-deletion approach (Cook, 1977) is the most popular one for identifying influential observations. To assess the impact of influential observations on parameter estimates some metrics have been used for measuring the distance between $\widehat{\boldsymbol{\theta}}_{[i]}$ and $\widehat{\boldsymbol{\theta}}$, such as the likelihood distance and Cook's distance. The second approach is a general statistical technique used to assess the stability of the estimation outputs with respect to the model inputs (Cook, 1986). By using the results of Zhu *et al.* (2001), we introduce here the case-deletion measures and the local influence diagnostics for the censored data on the basis of Q-function (see, Zhu & Lee, 2001). We first present the Hessian matrix used by both diagnostics measures, after, we consider the case-deletion measures and finally the perturbation schemes employed to obtain local influence measures.

The Hessian matrix

Following Zhu & Lee (2001), for obtaining the diagnostic measures for case-deletion diagnostics and for local influence of a particular perturbation scheme, it is necessary to compute $\ddot{Q}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = \partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$, where $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\alpha}^\top)^\top$, with $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)^\top$ and $\alpha_1 = \sigma^2$, $\alpha_2 = \tau^2$, $\alpha_3 = \rho$. Since $\dfrac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\alpha_k} = \mathbf{0},\ k = 1, 2, 3$, after evaluating this derivative at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$, the Hessian matrix $\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$ is block-diagonal of the form

$$\ddot{Q}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = BlockDiag(\ddot{Q}_{11}(\boldsymbol{\beta}), \ddot{Q}_{22}(\boldsymbol{\alpha})), \tag{16}$$

where

$$\ddot{Q}_{11}(\boldsymbol{\beta}) = \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} = -\mathbf{X}^\top\boldsymbol{\Sigma}^{-1}\mathbf{X}, \tag{17}$$

$$\ddot{Q}_{22}(\boldsymbol{\alpha}) = \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\alpha_k\partial\alpha_l} = -\frac{1}{2}\text{tr}\left[\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\alpha_l}\frac{\partial\boldsymbol{\Sigma}}{\partial\alpha_k}\right] - \frac{1}{2}\text{tr}\left[(\boldsymbol{\Sigma}^{-1} + \widehat{\mathbf{ZZ}^\top})\frac{\partial}{\partial\alpha_l}\left(\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\alpha_k}\right)\right], \tag{18}$$

with

$$\begin{aligned}
\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\alpha_k} &= -\boldsymbol{\Sigma}^{-1}\left(\frac{\partial\boldsymbol{\Sigma}}{\partial\alpha_k}\right)\boldsymbol{\Sigma}^{-1}, \\
\frac{\partial^2\boldsymbol{\Sigma}^{-1}}{\partial\alpha_k\partial\alpha_l} &= \boldsymbol{\Sigma}^{-1}\left(\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\alpha_l}\right)\boldsymbol{\Sigma}^{-1}\left(\frac{\partial\boldsymbol{\Sigma}}{\partial\alpha_k}\right)\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}\left(\frac{\partial\boldsymbol{\Sigma}}{\partial\alpha_k}\right)\boldsymbol{\Sigma}^{-1}\left(\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\alpha_l}\right)\boldsymbol{\Sigma}^{-1} \\
&\quad -\boldsymbol{\Sigma}^{-1}\left[\frac{\partial}{\partial\alpha_l}\left(\frac{\partial\boldsymbol{\Sigma}}{\partial\alpha_k}\right)\right]\boldsymbol{\Sigma}^{-1},
\end{aligned}$$

for $k, l = 1, 2, 3$, where $\dfrac{\partial\boldsymbol{\Sigma}}{\partial\alpha_1} = \mathbf{R}(\rho)$, $\dfrac{\partial\boldsymbol{\Sigma}}{\partial\alpha_2} = \mathbf{I}_n$, $\dfrac{\partial\boldsymbol{\Sigma}}{\partial\alpha_3} = \sigma^2\mathbf{R}'(\rho)$ and $\mathbf{R}'(\rho)$ and $\mathbf{R}''(\rho)$ are the first and second derivatives of $\mathbf{R}(\cdot)$ with respect to ρ. For each covariance function considered in this work, $\mathbf{R}'(\rho)$ and $\mathbf{R}''(\rho)$ are given in Appendix.

4.1. Case-deletion measures

Case-deletion is a common approach for studying the effects of dropping the ith case from the data set. In the following, a quantity with a subscript "$[i]$" denotes the original quantity with the ith case deleted; for example, $\mathbf{Z}_{[i]}$, denotes the complete-data with the ith case deleted. The log-likelihood function of $\boldsymbol{\theta}$, based on the data with the ith case deleted, is then denoted by $\ell(\boldsymbol{\theta}|\mathbf{Z}_{[i]})$. Let $\widehat{\boldsymbol{\theta}}_{[i]} = (\widehat{\boldsymbol{\beta}}_{[i]}^\top, \widehat{\boldsymbol{\alpha}}_{[i]}^\top)^\top$ with $\hat{\boldsymbol{\alpha}} = (\hat{\sigma^2}, \hat{\tau^2}, \hat{\rho})^\top$, be the maximizer of the function $Q_{[i]}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = E\{\ell(\boldsymbol{\theta}|\mathbf{Z}_{[i]})|\mathbf{V}, \mathbf{C}, \widehat{\boldsymbol{\theta}}\}$, where $\widehat{\boldsymbol{\theta}}$ is the ML estimate of $\boldsymbol{\theta}$. To assess the influence of the ith case on the ML estimate $\widehat{\boldsymbol{\theta}}$, we compare the difference between $\widehat{\boldsymbol{\theta}}_{[i]}$ and $\widehat{\boldsymbol{\theta}}$. If the deletion of a case seriously influences the estimates, more attention should be paid to that case. Hence, if $\widehat{\boldsymbol{\theta}}_{[i]}$ is far from $\widehat{\boldsymbol{\theta}}$ in some sense, then the ith case is regarded as influential.

Case-deletion measures can be developed for assessing influential observations, such as the generalized Cook distance and the likelihood distance (Zhu & Lee, 2001). The generalized Cook distance is defined as

$$GD_i(\widehat{\boldsymbol{\theta}}) = (\widehat{\boldsymbol{\theta}}_{[i]} - \widehat{\boldsymbol{\theta}})^\top\{-\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\}(\widehat{\boldsymbol{\theta}}_{[i]} - \widehat{\boldsymbol{\theta}}),\ i = 1, \ldots, n. \tag{19}$$

where $\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \dfrac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$ is the Hessian matrix. Using the fact that the Hessian matrix is block-diagonal, $GD_i(\widehat{\boldsymbol{\theta}})$ can be decomposed into two parts that correspond to the generalized Cook distance for the parameter subsets $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ which are denoted, respectively, by $GD_i(\boldsymbol{\beta})$ and $GD_i(\boldsymbol{\alpha})$, as follows:

$$GD_i = GD_i(\widehat{\boldsymbol{\beta}}) + GD_i(\widehat{\boldsymbol{\alpha}}),$$

where

$$GD_i(\widehat{\boldsymbol{\beta}}) = (\widehat{\boldsymbol{\beta}}_{[i]} - \widehat{\boldsymbol{\beta}})^\top\{-\ddot{Q}_{11}(\widehat{\boldsymbol{\beta}})\}(\widehat{\boldsymbol{\beta}}_{[i]} - \widehat{\boldsymbol{\beta}}) \text{ and } GD_i(\widehat{\boldsymbol{\alpha}}) = (\widehat{\boldsymbol{\alpha}}_{[i]} - \widehat{\boldsymbol{\alpha}})^\top\{-\ddot{Q}_{22}(\widehat{\boldsymbol{\alpha}})\}(\widehat{\boldsymbol{\alpha}}_{[i]} - \widehat{\boldsymbol{\alpha}}).$$

and $\ddot{Q}_{11}(\widehat{\boldsymbol{\beta}})$ and $\ddot{Q}_{22}(\widehat{\boldsymbol{\alpha}})$ are as defined in (16).

Another measure for the influence of the ith case is the likelihood displacement QD_i, which is similar to the likelihood distance LD_i (Cook & Weisberg, 1982) and is defined by

$$QD_i = 2\{Q(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) - Q(\widehat{\boldsymbol{\theta}}_{[i]}|\widehat{\boldsymbol{\theta}})\}. \tag{20}$$

4.2. Local influence

In this subsection, we derive the normal curvature of the local influence (Cook, 1986) for some common perturbation schemes either in the model or in the data. We will consider the response perturbation scheme, the explanatory variable perturbation and the matrix scale perturbation for this purpose.

Consider a perturbation vector $\boldsymbol{\omega} = (\omega_1, ..., \omega_g)^\top$ varying in an open region $\boldsymbol{\Omega} \subset \mathbb{R}^g$. Let $\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{y}_c)$ be the complete-data log-likelihood to the perturbed model. We assume that there is a $\boldsymbol{\omega}_0$ in $\boldsymbol{\Omega}$ such that $\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}_0|\mathbf{y}_c) = \ell_c(\boldsymbol{\theta}|\mathbf{y}_c)$ for all $\boldsymbol{\theta}$. Let $\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})$ denote the maximum of the function $Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}) = \mathrm{E}[\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{y}_c)|\mathbf{V}, \mathbf{C}, \widehat{\boldsymbol{\theta}}]$. The influence graph is then defined as $\boldsymbol{\delta}(\boldsymbol{\omega}) = (\boldsymbol{\omega}^\top, f_Q(\boldsymbol{\omega}))^\top$, where $f_Q(\boldsymbol{\omega})$ is the Q-displacement function defined as

$$f_Q(\boldsymbol{\omega}) = 2\left[Q\left(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}\right) - Q\left(\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})|\widehat{\boldsymbol{\theta}}\right)\right].$$

Following the approach of Cook (1986) and Zhu & Lee (2001), the normal curvature $C_{f_Q,\mathbf{d}}$ of $\boldsymbol{\delta}(\boldsymbol{\omega})$ at $\boldsymbol{\omega}_0$ in the direction of some unit vector $\mathbf{d}$ can be used to summarize the local behavior of the Q-displacement function. It can be shown that

$$C_{f_Q,\mathbf{d}} = -2\mathbf{d}^\top \ddot{Q}_{\boldsymbol{\omega}_o}\mathbf{d} \text{ and } -\ddot{Q}_{\boldsymbol{\omega}_0} = \boldsymbol{\Delta}_{\boldsymbol{\omega}_0}^\top \left\{-\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\right\}^{-1} \boldsymbol{\Delta}_{\boldsymbol{\omega}_0},$$

where $\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \dfrac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$ and $\boldsymbol{\Delta}_{\boldsymbol{\omega}} = \dfrac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^\top}|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})}$.

Following the same procedure as in Cook (1986), the quantity $-\ddot{Q}_{\boldsymbol{\omega}_0}$ is useful for detecting influential observations. From the spectral decomposition of a symmetric matrix $-2\ddot{Q}_{\boldsymbol{\omega}_0} = \sum_{k=1}^g \zeta_k \boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_k^\top$, where $\{(\zeta_k, \boldsymbol{\varepsilon}_k), k = 1, \ldots, g\}$ are eigenvalue–eigenvector pairs of $-2\ddot{Q}_{\boldsymbol{\omega}_0}$ with $\zeta_1 \geq \ldots \geq \zeta_r > \zeta_{r+1} = \ldots = 0$ and orthonormal eigenvectors $\{\boldsymbol{\varepsilon}_k, k = 1, \ldots, g\}$, Zhu & Lee (2001) proposed to inspect all eigenvectors corresponding to nonzero eigenvalues for capturing more information. Following the work of Zhu & Lee (2001), we consider the following aggregated contribution vector of all eigenvectors that correspond to nonzero eigenvalues. Let $\tilde{\zeta}_k = \zeta_k/(\zeta_1 + \ldots + \zeta_r)$, $\boldsymbol{\varepsilon}_k^2 = (\boldsymbol{\varepsilon}_{k1}^2, \ldots, \boldsymbol{\varepsilon}_{kg}^2)^\top$ and $M(0) = \sum_{k=1}^r \tilde{\zeta}_k \boldsymbol{\varepsilon}_k^2$. The lth component of $M(0)$, $M(0)_l$, is equal to $\sum_{k=1}^r \tilde{\zeta}_k \boldsymbol{\varepsilon}_{kl}^2$. The assessment of influential cases is based on the visual inspection of the $\{M(0)_l, l = 1, \ldots, g\}$ plotted against the index l. The lth case may be regarded as influential if $M(0)_l$ is larger than the benchmark value.

The inconvenience in the use of the normal curvature is in deciding about the influence of the observations, since $C_{f_Q,\mathbf{d}}(\boldsymbol{\theta})$ may assume any value and it is not invariant under a uniform change of scale. Based on the work of Poon & Poon (1999) in using a conformal normal curvature, Zhu & Lee (2001) considered the following conformal normal curvature $B_{f_Q,\mathbf{d}}(\boldsymbol{\theta}) = C_{f_Q,\mathbf{d}}(\boldsymbol{\theta})/tr[-2\ddot{Q}_{\boldsymbol{\omega}_0}]$, whose computation is quite simple and also has the property that $0 \leq B_{f_Q,\mathbf{d}}(\boldsymbol{\theta}) \leq 1$. Let $\mathbf{d}_l$ be a basic perturbation vector with lth entry as 1 and all other entries as 0. Zhu & Lee (2001) then showed that for all l, $M(0)_l = B_{f_Q,\mathbf{d}_l}$. We can, therefore, obtain $M(0)_l$ via $B_{f_Q,\mathbf{d}_l}$.

So far, there is no general rule to judge how large is the influence of a specific case in the data. Let $\overline{M}(0)$ and $SM(0)$ denote, respectively, the mean and standard error of $\{M(0)_l : l = 1, \ldots, g\}$, where $\overline{M(0)} = 1/g$. Poon & Poon (1999) proposed to use $2\overline{M}(0)$ as a benchmark for $M(0)$. But, we may use different functions of $M(0)$. For instance, Zhu & Lee (2001) proposed to use $\overline{M}(0) + 2SM(0)$ as a benchmark to take into account the variance of

$\{M(0)_l : l = 1, \ldots, g\}$ as well. According to Lee & Xu (2004), the exact choice of the function of $\overline{M}(0)$ as the benchmark is subjective. Lee & Xu (2004) also proposed to use $\overline{M}(0) + c^* SM(0)$, where c^* is a selected constant, and depending on the specific application, c^* may be chosen suitably.

4.3. Perturbation scheme

In this section, we will evaluate the $\mathbf{\Delta}$ matrix under the following perturbation scheme for SCL models. *Perturbation of response variables* is made on the response values, which may indicate observations with large influence on their own predicted values (in our case, the response variables are $\mathbf{V}'s$); *Scale perturbation* is made on the scale matrix $\mathbf{\Sigma} = \tau^2 \mathbf{I}_n + \sigma^2 R(\rho)$, which may reveal individuals that are most influential on the scale structure and consequently on the $\boldsymbol{\alpha}$ estimate and finally perturbation of explanatory variables.

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\alpha}^\top)^\top$, where $\boldsymbol{\alpha} = (\sigma^2, \tau^2, \rho)^\top$. Given that, the matrix $\ddot{Q}(\widehat{\boldsymbol{\theta}})$ is block-diagonal with blocks $\ddot{Q}_{11}(\widehat{\boldsymbol{\beta}})$ and $\ddot{Q}_{22}(\widehat{\boldsymbol{\alpha}})$, then we have, for any unit vector $\mathbf{d}$,

$$C_{f_Q,\mathbf{d}} = C_{1,\mathbf{d}}(\hat{\boldsymbol{\beta}}) + C_{2,\mathbf{d}}(\hat{\boldsymbol{\alpha}}),$$

where

$$C_{1,\mathbf{d}}(\hat{\boldsymbol{\beta}}) = 2\mathbf{d}^\top \Delta_{1\boldsymbol{\omega}_0}^\top (-\ddot{Q}_{11})^{-1} \Delta_{1\boldsymbol{\omega}_0} \mathbf{d} \text{ and } C_{2,\mathbf{d}}(\hat{\boldsymbol{\alpha}}) = 2\mathbf{d}^\top \Delta_{2\boldsymbol{\omega}_0}^\top (-\ddot{Q}_{22})^{-1} \Delta_{2\boldsymbol{\omega}_0} \mathbf{d},$$

with $\mathbf{\Delta}_{1\boldsymbol{\omega}_0} = \mathbf{\Delta}_{\boldsymbol{\beta}}$ and $\mathbf{\Delta}_{2\boldsymbol{\omega}_0} = \mathbf{\Delta}_{\boldsymbol{\alpha}}$. Furthermore, $\mathbf{\Delta}_{\boldsymbol{\beta}} = \dfrac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^\top}|_{\boldsymbol{\omega}_0}$, $\mathbf{\Delta}_{\boldsymbol{\alpha}} = (\mathbf{\Delta}_{\alpha 1}^\top, \mathbf{\Delta}_{\alpha 2}^\top, \mathbf{\Delta}_{\alpha 3}^\top)^\top$, with $\mathbf{\Delta}_{\alpha k} = \dfrac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})}{\partial \alpha_k \partial \boldsymbol{\omega}^\top}|_{\boldsymbol{\omega}_0}$ for $k = 1, 2, 3$ and $\ddot{Q}_{11}$ and $\ddot{Q}_{22}$ as in (16).

Response perturbation

A perturbation of the response variables $V_i, i = 1, \ldots, n$, can be introduced by replacing V_i by $V_i(\omega) = V_i + \omega_i$. Now, substituting $V_i(\omega_i)$ into (6), we can write the perturbed model as

$$\begin{aligned} y_i(\omega_i) &\leq V_i(\omega_i) \quad \text{if} \quad C_i = 1, \\ y_i(\omega_i) &= V_i(\omega_i) \quad \text{if} \quad C_i = 0, \end{aligned}$$

where $\mathbf{y}(\boldsymbol{\omega}) = \mathbf{y} - \boldsymbol{\omega}$. Hence, the perturbed Q-function $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}, \boldsymbol{\omega})$ is as in Subsection 2.2, with $\widehat{\mathbf{y}}$ and $\widehat{\mathbf{y}\mathbf{y}^\top}$ being replaced by $\widehat{\mathbf{y}_\omega} = \widehat{\mathbf{y}} - \boldsymbol{\omega}$ and $\widehat{\mathbf{y}_\omega \mathbf{y}_\omega^\top} = \widehat{\mathbf{y}\mathbf{y}^\top} - \widehat{\mathbf{y}}\boldsymbol{\omega}^\top - \boldsymbol{\omega}\widehat{\mathbf{y}}^\top + \boldsymbol{\omega}\boldsymbol{\omega}^\top$, respectively. Under this perturbation scheme, the vector $\boldsymbol{\omega}_0$, representing no perturbation, is given by $\boldsymbol{\omega}_0 = \mathbf{0}$ and $\mathbf{\Delta}_{\boldsymbol{\omega}_0}$ has the following elements:

$$\mathbf{\Delta}_{\boldsymbol{\beta}} = \mathbf{X}^\top \mathbf{\Sigma}^{-1} \quad \text{and} \quad \mathbf{\Delta}_{\alpha_k} = -\hat{\mathbf{Z}}^\top (\mathbf{I}_n - 2\mathbf{P})^\top \frac{\partial \mathbf{\Sigma}^{-1}}{\partial \alpha_k}, \quad k = 1, 2, 3, \tag{21}$$

where $\hat{\mathbf{Z}} = (\widehat{Z}_1, \ldots, \widehat{Z}_n)^\top$, $\mathbf{P} = \mathbf{X}(\mathbf{X}^\top \mathbf{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{\Sigma}^{-1}$ and $\dfrac{\partial \mathbf{\Sigma}^{-1}}{\partial \alpha_k}$ is as defined in (17).

Scale matrix perturbation

In order to study the effects of perturbation over the scale matrix, we consider the scheme perturbation of the form $\mathbf{\Sigma}(\boldsymbol{\omega}) = \mathbf{D}(\boldsymbol{\omega})\mathbf{\Sigma}$, where $\mathbf{D}(\boldsymbol{\omega})$ is an $n \times n$ diagonal matrix with value ω_i on ith diagonal element. Under this scheme, the non-perturbed model is obtained when $\omega_i = 1$ for $i = 1, \ldots, n$. Considering this perturbation scheme, $\mathbf{\Delta}_{\boldsymbol{\omega}_0}$ is a $(p + 3) \times n$ matrix and has components given by:

$$\mathbf{\Delta}_{\boldsymbol{\beta}_i} = -\mathbf{X}^\top \mathbf{\Sigma}^{-1} \mathbf{d}(i)(\mathbf{I}_n - \mathbf{P})\mathbf{Z}$$

and $\mathbf{\Delta}_{\boldsymbol{\alpha}} = [\Delta_{\alpha_{ki}}]$, where

$$\Delta_{\alpha_{ki}} = -\frac{1}{2}\left\{\text{tr}\left[\widehat{\mathbf{Z}\mathbf{Z}^\top}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\alpha_k}\boldsymbol{\Sigma}^{-1}\mathbf{d}(i)\right] - \widehat{\mathbf{Z}}^\top\boldsymbol{\Sigma}^{-1}\mathbf{P}\frac{\partial\boldsymbol{\Sigma}}{\partial\alpha_k}\boldsymbol{\Sigma}^{-1}\mathbf{d}(i)(2\mathbf{I}_n - \mathbf{P})\widehat{\mathbf{Z}}\right\},$$

for $k = 1, 2, 3$ and $i = 1, \ldots, n$, where $\mathbf{d}(i)$ is a $n \times n$ matrix with *i*th diagonal element equal to one and the others equal to zero, $\dfrac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\alpha_k}$ and $\mathbf{P}$ are as defined in (17) and (21), respectively.

Explanatory variable perturbation

Similar to response variable perturbation, to evaluate influential points on the explanatory variable, we replaced $\mathbf{X}$ by $\mathbf{X}(\boldsymbol{\omega}) = \mathbf{X} + \mathbf{W}$ on the perturbed Q-function, with $\mathbf{W} = \boldsymbol{\omega}\mathbf{1}^\top$ where $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^\top$ and $\mathbf{1}$ is a $p \times 1$ vector of ones, then $\mathbf{W}$ is a $n \times p$ matrix. Thus, taking $\boldsymbol{\omega} = \mathbf{0}$, the $\mathbf{\Delta}_{\boldsymbol{\omega}_0}$ has the following elements:

$$\mathbf{\Delta}_{\boldsymbol{\beta}i} = \hat{\mathbf{Z}}^\top(\mathbf{I}_n - 2\mathbf{P})^\top\boldsymbol{\Sigma}^{-1}\mathbf{W}_i$$

and $\mathbf{\Delta}_{\boldsymbol{\alpha}} = [\Delta_{\alpha_{ki}}]$, where

$$\Delta_{\alpha_{ki}} = \widehat{\mathbf{Z}}^\top(\mathbf{I}_n - 2\mathbf{P})^\top\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\alpha_k}\mathbf{W}_i^{(1)}(\mathbf{X}^\top\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\boldsymbol{\Sigma}^{-1}\widehat{\mathbf{Z}},$$

for $k = 1, 2, 3$ and $i = 1, \ldots, n$, where $\mathbf{W}_i^{(1)}$ is a $n \times p$ matrix with the *i*th row equal to one and the others equal to zero, $\dfrac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\alpha_k}$ and $\mathbf{P}$ are as defined in (17) and (21), respectively.

5. Application: Missouri Dioxin Contamination Site

5.1. Description of data

In order to exemplify the developed methodology showed on this work, we considered a dataset reported in Zirschky & Harris (1986) with 127 observations distributed in an area of $3600m \times 65m$ on the shoulders of a country road located on Missouri. The observations correspond to a level of contamination by dioxin (2,3,7,8-tetrachlorodibenzo-p-dioxin or TCDD) on sampled points around the road. The spatial directions are the X-direction (measured in $1/100$ feet), representing direction parallel to the road, and the Y-direction (measured in feet), representing the direction perpendicular to or away from the road. The road is located at the Y coordinate of 30. The shoulder of the road was divided into long transects in the X direction, most 200 feet, in which eight samples were taken. The eight samples were aggregated together to give one measurement per transect. For illustration purposes, we will treat the values reported as coming from one sampled location, with the X coordinate indicating the start of the transect (see Figure 1(left panel)). Forty-three percent of the observations (55 sites) were censored, falling below some LOD. The level of detections range from 0.10 to 0.79 mg/kg.

5.2. Model specification and results

To illustrate our methods, we propose to fit the model

$$\log\{Z_i\} = \mu + \epsilon_i,$$

with different covariance function for the stochastic errors ϵ_i, $i = 1, \ldots, 127$. This application is based on left censoring, and the SAEM algorithm for censored data was implemented as explained in Subsection 3.2, we choose a Monte Carlo sample size of $m = 20$, a number maximum of iterations $W = 150$ and a cut point $c = 0.2$. These

Table 1: Missouri data. ML estimates under different covariance functions

	Spherical	Exponential	Matérn ($\kappa = 0.75$)	Matérn ($\kappa = 1.00$)
$\widehat{\mu}$	-1.4137	-2.0122	-1.8044	-1.7030
$\widehat{\sigma^2}$	3.7597	4.8016	4.4306	4.1682
$\widehat{\rho}$	0.2	14.0557	8.4704	6.0301
$\widehat{\tau^2}$	3.4526	0.2445	0.3751	0.4276
loglik	-216.2290	-143.8896	-144.0814	-144.5400
AIC	440.4580	295.7793	296.1627	297.0800
BIC	451.8348	307.1560	307.5395	308.4567

computational procedures were implemented using the R software (R Core Team, 2015). The results of the ML estimates, using the spherical, exponential and Matérn (with $\kappa = 0.75,\ 1.00$) covariance functions are presented in Table 1. Notice that, although the spherical covariance function is broadly used in classical geostatics, this structure does not present a good fit, showing the worst value for the log-likelihood and information criteria.

The Gaussian and exponential covariance functions are special cases of the Matérn family of covariance functions. In Figure 1 (right panel) we show the values of the log-likelihood using the Matérn covariance after applying the SAEM algorithm fixing κ at values $0.1, 0.2, 0.3, \cdots, 2.0$. Thus, we choose $\kappa = 0.5$, which maximizes the profile log-likelihood and corresponding to the exponential covariance function. In the following we proceed with diagnostics analysis using the exponential covariance function (or Matérn ($\kappa = 0.5$)). Because we currently focus on exploring influence diagnostics, details on the estimation and interpretation of the parameter estimates are omitted for brevity.

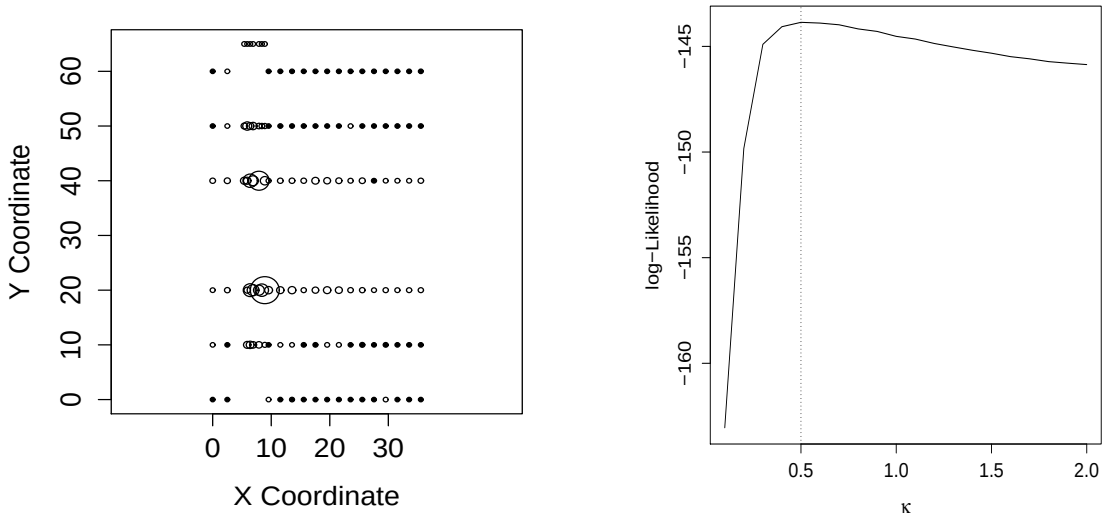


Figure 1: Missouri data. (left) Proportional TCDD observed on each location. ∘ represents an observed value and • represents a censored value. (right) Plot of the profile log-likelihood versus κ using the Matérn family.

Local influence

Now we focus on the local influence analysis for the Missouri data, based on $M(0)$, with interest focussing on $\boldsymbol{\theta}$. We consider the following perturbation schemes: response perturbation, explanatory variables perturbation and scale matrix perturbation. We use the criterion $M(0)_i > \overline{M}(0)+3SM(0)$, $i = 1, \ldots, 127$, to discriminate whether an observation is influential or not.

Figure 2 shows the index plot of $M(0)$ under the three perturbation schemes. We find that subjects $\#40$, $\#42$, $\#45$, $\#47$ and $\#48$ appears as influential under response perturbation an explanatory variable schemes. Since we do not have explanatory variable, the influential points showed by this scheme can be viewed as influential points in the expected average of the variable response μ. For perturbation on $\boldsymbol{\Sigma}$, we find that observations $\#40$ and $\#45$ appears as influential, which may indicate a more significant impact on its neighbors than the others points.

It is important to stress that one of the points indicated as influential, the observation ($\#40$), is the maximum value of dioxin observed and it is located on the boarder of the road, where the dioxin was dumped, then it is expected to be an influential point. The others points are located in a perpendicular line (see Figure 4) to the road that contains the observation $\#40$ and considering the response and explanatory perturbations, which compare the observed data in relation its neighborhood, this may represent the direction where the dioxin was spread.

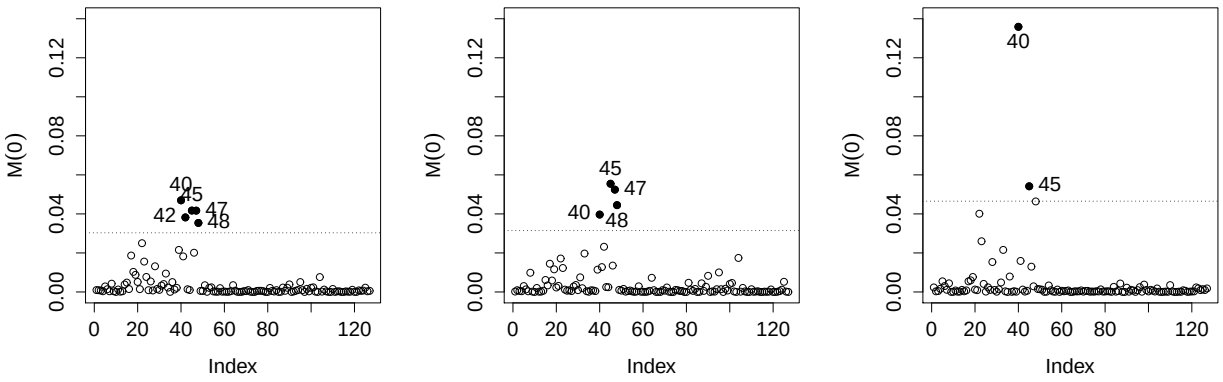


Figure 2: Missouri data. Index plot of $M(0)$ using exponential covariance function for the TCDD (left) Response perturbation; (middle) exploratory variable perturbation and (right) scale matrix perturbation.

Global influence

In order to evaluate the effect on the ML estimates when some observations are deleted, we analyze the $GD_i(\boldsymbol{\theta})$ distance, which is depicted in Figure 3 (left panel). The plot reveals that once again cases $(\#40, \#45, \#47, \#48)$ are potentially influential on the parameter estimates. Figures 3 (panels middle and right) present the index plot of $GD_i(\boldsymbol{\beta})$ and $GD_i(\boldsymbol{\alpha})$, respectively. From these figures, we infer that the same observations are influential for $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$.

5.3. The impact of the detected influential observations

The diagnostic analysis (global influence and local influence) indicated the five observations $(\#40, \#42, \#45, \#47, \#48)$ as potentially influential. In order to reveal the impact of these five observations on the parameter estimates,

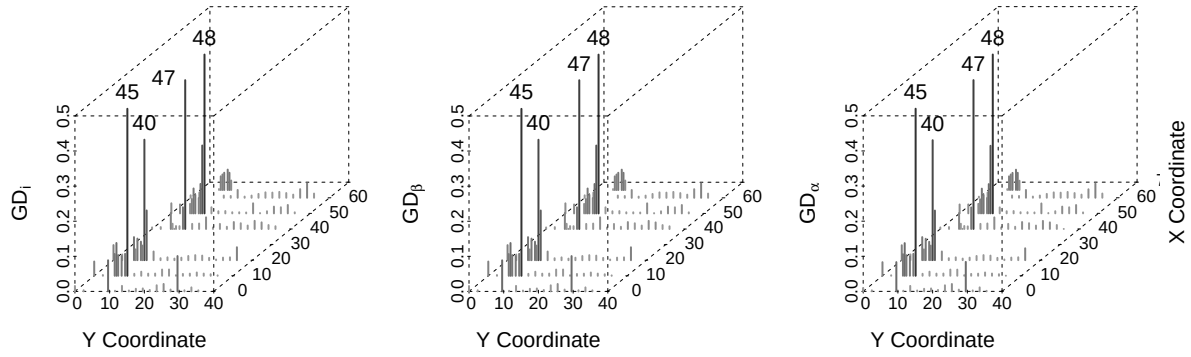


Figure 3: Missouri data. Index plot of GD_i (left), $GD(\boldsymbol{\beta})$ (middle) and $GD(\boldsymbol{\alpha})$ (right) for the TCDD data using exponential covariance function.

we refitted the model individually eliminating each of these five cases. In Table 2 we show the relative changes (in percentage) of each parameter estimate, defined by

$$RC_{\widehat{\gamma}} \quad = \quad \left| \frac{\widehat{\gamma} - \widehat{\gamma}_{[i]}}{\widehat{\gamma}} \right|,$$

where $\hat{\gamma} = \hat{\beta}, \hat{\sigma^2}, \hat{\rho}$ or $\hat{\tau^2}$ and $\widehat{\gamma}_{[i]}$ denotes the ML estimate of $\widehat{\gamma}$ after the ith observation of $\mathbf{Z}$ is removed. Note from this Table that significant changes are observed only for the nugget effect τ^2, indicating the needed of special attention on the estimation of the variance.

Table 2: Relative changes [RC (in %)] for the Missouri data.

Dropped	$RC_{\hat{\mu}}$	$RC_{\hat{\sigma^2}}$	$RC_{\hat{\rho}}$	$RC_{\hat{\tau^2}}$
[#40]	1.7675	1.6641	2.8503	18.7987
[#42]	1.1246	0.5709	4.3562	15.9953
[#45]	0.1962	2.0432	5.0879	22.5714
[#47]	0.3180	0.7147	5.3426	24.7738
[#48]	1.0575	2.3936	2.3200	22.5531

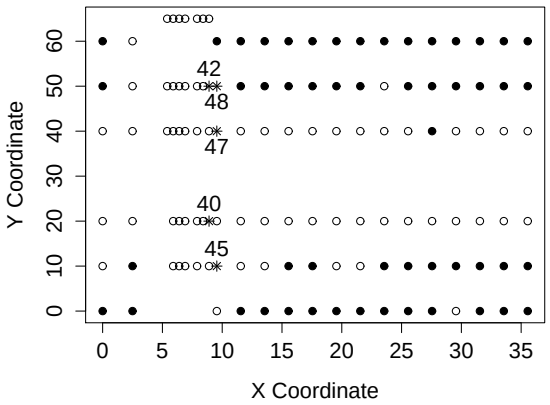


Figure 4: Missouri data. TCDD observed on each location. ○ represents an observed value, • represents a censored value. The influential points (∗) are numbered.

6. Simulation Studies

In order to examine the performance of the proposed methods, we present two simulation studies. We performed the simulation with a left censored spatial linear model as defined in (5)-(6) with different covariance structures. We set $\boldsymbol{\beta}^\top = (\beta_0, \beta_1, \beta_2) = (1, 3, -1)$, $\sigma^2 = 3$, $\rho = 3$, $\tau^2 = 2$, and $\mathbf{x}_i^\top = (1, x_{i1}, x_{i2})$, where $x_{i1} \sim U(0, 1)$ and $x_{i2} \sim U(2, 5)$, for $i = 1, \ldots, 100$.

6.1. First experiment

This study explores the accuracy of the proposed diagnostics measures in detecting a single outlier at different levels of censoring $(8\%, 16\%, 32\%)$ and different covariance functions (exponential, Gaussian and Matérn($\kappa = 0.75$). Here, we generated only one sample under three setting of covariance functions and the simulated samples were perturbed by replacing the maximum value by $z_{max} = z_{max} + 5sd(\mathbf{z})$. The atypical points corresponds to the points $\#52$, $\#53$ and $\#30$ under 8%, 16% and 32% of censoring, respectively.

Following the approach described in Section 4, Figure 5 depicts the index plots of $M(0)$ for the response perturbation, explanatory variable perturbation and scale perturbation, respectively, along with the Lee & Xu (2004) benchmark computed for $c = 3$. For these perturbation schemes, we notice the influence of the observations $\#52$, $\#53$ and $\#30$, that is, the perturbed observations. This confirms the efficiency of the local influence measures in detecting influential observations. For the others covariance functions the result were similar and they are not presented here to save space.

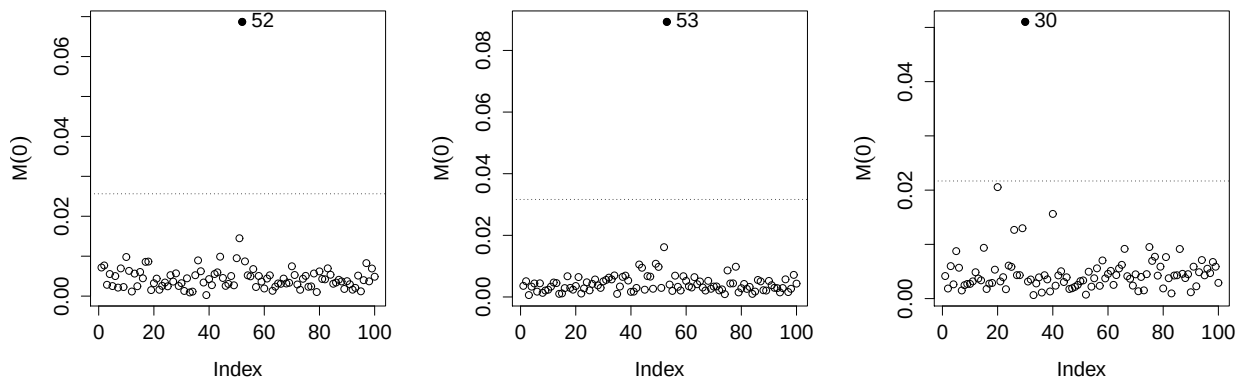
Table 3: Simulation study 2: The values in the table denotes the % of correctly identifying the influential observations using response perturbation ($\mathbf{y}$), explanatory variable perturbation ($\mathbf{X}$) and scale matrix perturbation ($\boldsymbol{\Sigma}$) from 100 simulated datasets under the SCL model with exponential and Matérn correlation structures.

	% of censoring								
	8%			16%			32%		
	$\mathbf{y}$	$\mathbf{X}$	$\boldsymbol{\Sigma}$	$\mathbf{y}$	$\mathbf{X}$	$\boldsymbol{\Sigma}$	$\mathbf{y}$	$\mathbf{X}$	$\boldsymbol{\Sigma}$
Exponential									
Pert. $2.5\beta_1$	90	71	86	71	75	71	66	65	72
Pert. $5\beta_1$	97	97	95	99	50	97	79	79	79
Pert. $7.5\beta_1$	99	99	99	99	88	99	68	68	68
Pert. $10\beta_1$	100	100	100	100	98	100	70	70	70
Matérn ($\kappa = 0.75$)									
Pert. $2.5\beta_1$	90	94	93	72	75	73	66	16	68
Pert. $5\beta_1$	84	84	84	97	65	90	96	63	90
Pert. $7.5\beta_1$	88	88	88	99	91	97	98	81	94
Pert. $10\beta_1$	89	89	89	100	99	100	93	89	92

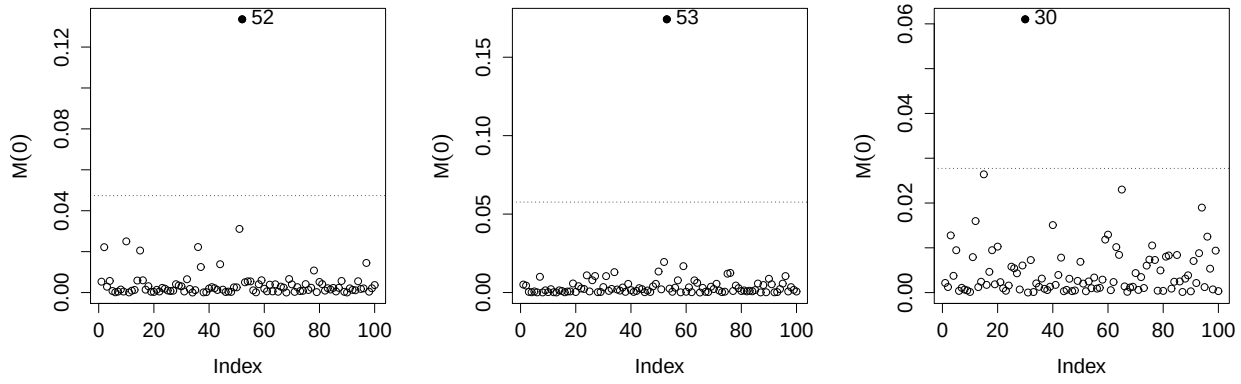
In Figure 6 we present the index plot of the global influence measures, GD_i and QD_i. As expected, once again the perturbed points were detected as influential. Note however that the generalized Cook distance GD_i detected the point $\#51$ as influential. This behavior is expected, since this point is a neighbor of the perturbed point $\#52$.

6.2. Second Study

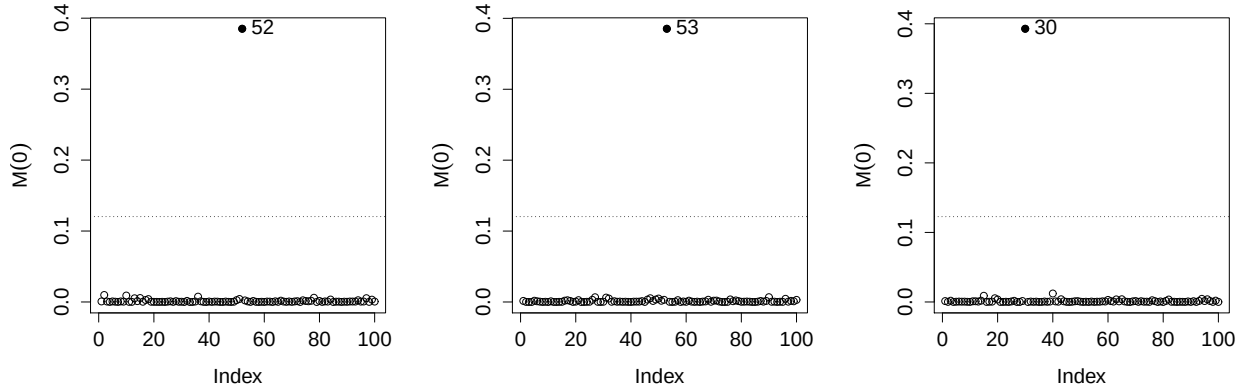
The second study is a Monte Carlo experiment, that shows the capacity of the methodology to detect atypical points. Here, to generate an atypical point, we replaced β_1 by $2.5\beta_1, 5\beta_1, 7.5\beta_1$ and $10\beta_1$ to generate the perturbed



(a)

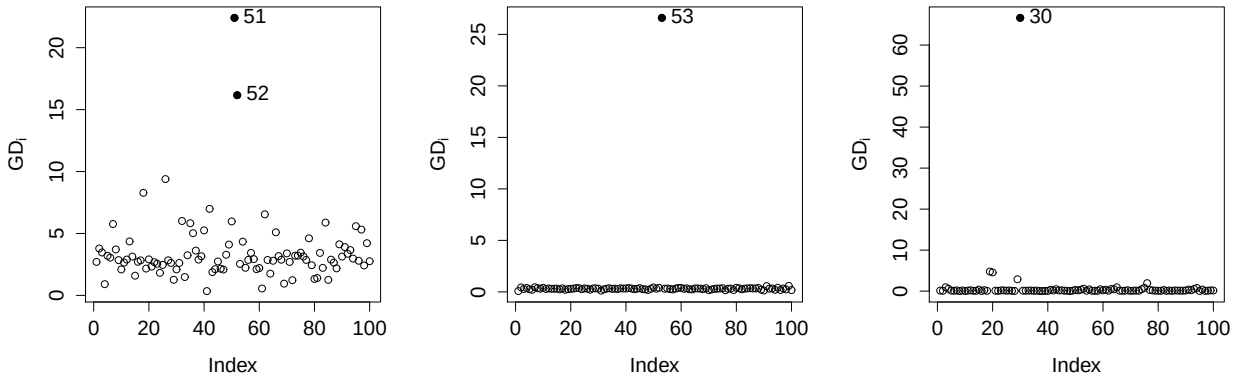


(b)

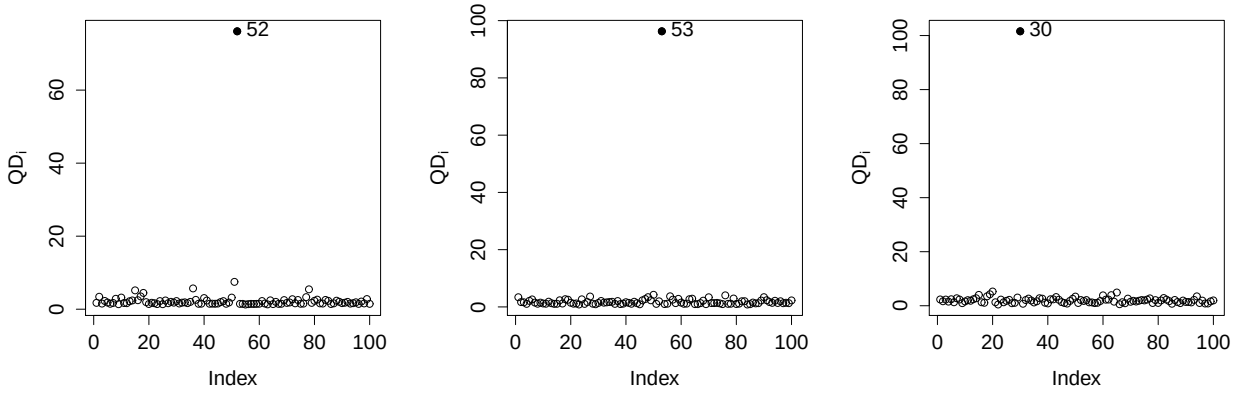


(c)

Figure 5: Simulation Study 1. Index plot of $M(0)$ for (a) response perturbation; (b) explanatory variable perturbation and (c) scale matrix perturbation, using Matérn Covariance function with 8% (left), 16% (middle) and 32% (right) of censoring.



(a)



(b)

Figure 6: Simulation Study 1. Index plot of (a) the generalized Cook distance, GD_i and (b) likelihood displacement, QD_i, using Matérn Covariance function with 8% (left), 16% (middle) and 32% (right) of censoring levels.

observation $\#50$ (z_{50}). Since we have a benchmark to decide which point is influential or not, the diagnostic measure $M(0)$ were computed for 100 simulated datasets under three censoring proportions $(8\%, 16\%, 32\%)$ and using exponential and Matérn (with $\kappa = 0.75$) covariance functions.

Table 3 shows the results of this experiment, in all cases the capacity to detect influential points is reasonable, especially when the percentage of censoring is moderate or low. The explanatory variable perturbation appears to be less sensitive to detect atypical observations, showing the worst percentage of detection in almost all cases.

7. Conclusions

This article proposes influence diagnostic tools for detecting influential observations in the context of spatial censored linear models. It extends the recently published works by Assumpção *et al.* (2014) and De Bastiani *et al.* (2014) which considers estimation and diagnostics of spatial linear models. Our proposed method relies on the Q function, the conditional expectation of the logarithm of the complete-data likelihood, which facilitates the theoretical development of the stochastic approximation of the EM algorithm (SAEM) to obtain the maximum likelihood

estimates of model parameters and the development of diagnostic influence measures. Explicit expressions are obtained for the Hessian matrix $\ddot{\mathbf{Q}}$ and for the matrix Δ under different perturbation schemes. A simulation study compares the outlier detection accuracy under different censoring and perturbation schemes. For practical demonstration, the methodology is applied to a data from dioxin contaminated sites in Missouri and the proposed methods are implemented using the R software (codes available upon request from the first author), providing practitioners with a convenient tool for further applications in their domain.

Future extensions of the work include the use of scale mixtures of normal distributions to accommodate heavy-tailed feature, or the development of some diagnostics and tests for the model. Bayesian influence diagnostics, in the context of spatial censored linear models, can be treated via the Kullback–Leibler divergence, as proposed by Cancho *et al.* (2011). Other extensions of the current work include, for example, diagnostics analysis in censored spatial data with measurement errors (Li *et al.*, 2009).

Acknowledgements

The research of Víctor H. Lachos was supported by Grant 305054/2011-2 from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq-Brazil) and by Grant 2014/02938-9 from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP-Brazil). The research of Thais S. Barbosa was supported by CAPES. We also thank Dr. Aldo William Medina Garay from Federal University of Pernanbuco for his help on the earlier version of the paper.

Appendix: Derivatives for some covariance functions

In this appendix we obtain the first and second derivatives of $\mathbf{\Sigma} = \tau^2\mathbf{I} + \sigma^2\mathbf{R}(\rho)$ with respect to ρ for the exponential, Gaussian and Matérn covariance functions.

Exponential covariance function

$$\mathbf{R}'(\rho) = \begin{cases} \dfrac{d_{ij}}{\rho^2}\exp\left\{-\left(\dfrac{d_{ij}}{\rho}\right)\right\}, & d_{ij} > 0, \\ 0, & d_{ij} = 0 \end{cases}$$

and

$$\mathbf{R}''(\rho) = \begin{cases} \dfrac{d_{ij}(d_{ij}-2\rho)}{\rho^4}\exp\left\{-\left(\dfrac{d_{ij}}{\rho}\right)\right\}, & d_{ij} > 0, \\ 0, & d_{ij} = 0. \end{cases}$$

Gaussian covariance function

$$\mathbf{R}'(\rho) = \begin{cases} \dfrac{2d_{ij}^2}{\rho^3}\exp\left\{-\left(\dfrac{d_{ij}}{\rho}\right)^2\right\}, & d_{ij} > 0, \\ 0, & d_{ij} = 0 \end{cases}$$

and

$$\mathbf{R}''(\rho) = \begin{cases} \dfrac{(4d_{ij}^4 - 6d_{ij}^2\rho^2)}{\rho^6}\exp\left\{-\left(\dfrac{d_{ij}}{\rho}\right)^2\right\}, & d_{ij} > 0, \\ 0, & d_{ij} = 0. \end{cases}$$

Matérn covariance function

$$\mathbf{R}'(\rho) = \begin{cases} \dfrac{1}{2^{\kappa}\rho^2\Gamma(\kappa)} \left(\dfrac{d_{ij}}{\rho}\right)^{\kappa} \left[2\kappa\rho K_{\kappa}\left(\dfrac{d_{ij}}{\rho}\right) + d_{ij}\left(A_K\right)\right], & d_{ij} > 0, \\ 0, & d_{ij} = 0 \end{cases}$$

and

$$\mathbf{R}''(\rho) = \begin{cases} \dfrac{2^{1-\kappa}}{\Gamma(\kappa)} \left\{ -\dfrac{d_{ij}^2}{4\rho^6} \left(\dfrac{d_{ij}}{\rho}\right)^{\kappa-2} \left[-d_{ij}^2\left(B_K\right)\right] - 4\rho(\kappa+1)\left[d_{ij}\left(A_K\right) + \kappa\rho K_k\left(\dfrac{d_{ij}}{\rho}\right)\right]\right\}, & d_{ij} > 0, \\ 0, & d_{ij} = 0, \end{cases}$$

where $\rho > 0$; $K_{\kappa}(u) = \frac{1}{2}\int_0^{\infty} x^{\kappa-1} e^{-\frac{1}{2}u(x+x^{-1})} dx$ is the modified Bessel function of the third kind of order κ (see, Gradshtejn & Ryzhik, 1965), with $\kappa > 0$ fixed and

$$A_K = K_{\kappa-1}\left(\tfrac{d_{ij}}{\rho}\right) - K_{\kappa+1}\left(\tfrac{d_{ij}}{\rho}\right); \quad B_K = K_{\kappa-2}\left(\tfrac{d_{ij}}{\rho}\right) - 2K_{\kappa}\left(\tfrac{d_{ij}}{\rho}\right) + K_{\kappa+2}\left(\tfrac{d_{ij}}{\rho}\right).$$

References

Allassonnière, S., Kuhn, E., Trouvé, A. *et al.* (2010). Construction of Bayesian deformable models via a stochastic approximation algorithm: a Convergence study. *Bernoulli*, **16**(3), 641–678.

Assumpção, R., Uribe-Opazo, M. & Galea, M. (2014). Analysis of local influence in geostatistics using student's t-distribution. *Journal of Applied Statistics*, **41**(11), 2323–2341.

Cancho, V., Dey, D., Lachos, V. & Andrade, M. (2011). Bayesian nonlinear regression models with scale mixtures of skew normal distributions: Estimation and case influence diagnostics. *Computational Statistics & Data Analysis*, **55**(1), 588–602.

Cook, R. (1977). Detection of influential observation in linear regression. *Technometrics*, **19**, 15–18.

Cook, R. D. (1986). Assessment of local influence. *Journal of the Royal Statistical Society, Series B,*, **48**, 133–169.

Cook, R. D. & Weisberg, S. (1982). *Residuals and Influence in Regression*. Chapman & Hall/CRC, Boca Raton, FL.

De Bastiani, F., de Aquino Cysneiros, A. H. M., Uribe-Opazo, M. A. & Galea, M. (2014). Influence diagnostics in elliptical spatial linear models. *TEST*, **24**(2), 322–340.

De Oliveira, V. (2005). Bayesian inference and prediction of Gaussian random fields based on censored data. *Journal of Computational and Graphical Statistics*, **14**(1), 95–115.

Delyon, B., Lavielle, M. & Moulines, E. (1999). Convergence of a stochastic approximation version of the EM algorithm. *Annals of Statistics*, **27**(1), 94–128.

Dempster, A., Laird, N. & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B,*, **39**, 1–38.

Fridley, B. L. & Dixon, P. (2007). Data augmentation for a bayesian spatial model involving censored observations. *Environmetrics*, **18**, 107–123.

Genz, A., Bretz, F., Hothorn, T., Miwa, T., Mi, X., Leisch, F. & Scheipl, F. (2008). mvtnorm: Multivariate normal and t distribution. *R package version 0.9-2, URL http://CRAN. R-project. org/package= mvtnorm*.

Gradshtejn, I. S. & Ryzhik, I. M. (1965). *Table of integrals, series and products*. Academic Press.

Hughes, J. (1999). Mixed effects models with censored data with application to HIV RNA levels. *Biometrics*, **55**, 625–629.

Jank, W. (2006). Implementing and diagnosing the stochastic approximation EM algorithm. *Journal of Computational and Graphical Statistics*, **15**(4), 803–829.

Kuhn, E. & Lavielle, M. (2004). Coupling a stochastic approximation version of EM with an MCMC procedure. *ESAIM: Probability and Statistics*, **8**, 115–131.

Lee, S. Y. & Xu, L. (2004). R influence analysis of nonlinear mixed-effects models. *Computational Statistics & Data Analysis*, **45**, 321–341.

Li, Y., Tang, H. & Lin, X. (2009). Spatial linear mixed models with covariate measurement errors. *Statistica Sinica*, **19**(3), 1077–1093.

Matos, L. A., Lachos, V. H., Balakrishnan, N. & Labra, F. V. (2013). Influence diagnostics in linear and nonlinear mixed-effects models with censored data. *Computational Statistics & Data Analysis*, **57**(1), 450–464.

Militino, A. F. & Ugarte, M. D. (1999). Analyzing censored spatial data. *Mathematical Geology*, **31**(5), 551–561.

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

R Development Core Team (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Vaida, F. (2005). Parameter convergence for EM and MM algorithms. *Statistica Sinica*, **15**(3), 831–840.

Vaida, F. & Liu, L. (2009). Fast implementation for normal mixed Effects models with censored response. *Journal of Computational and Graphical Statistics*, **18**, 797–817.

Wei, G. C. & Tanner, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association*, **85**(411), 699–704.

Wu, C. J. (1983). On the convergence properties of the EM algorithm. *The Annals of Statistics*, **11**(1), 95–103.

Zhu, H. & Lee, S. (2001). Local influence for incomplete-data models. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, **63**, 111–126.

Zhu, H., Lee, S., Wei, B. & Zhou, J. (2001). Case-deletion measures for models with incomplete data. *Biometrika*, **88**, 727–737.

Zirschky, J. H. & Harris, D. J. (1986). Geostatistical analysis of hazardous waste site data. *Journal of Environmental Engineering*, **112**(4), 770–784.