Published in final edited form as:

Spat Stat. 2018 March ; 23: 109–123. doi:10.1016/j.spasta.2017.12.001.

Geostatistical estimation and prediction for censored responses

José A. Ordoñez^a, Dipankar Bandyopadhyay^b, Victor H. Lachos^{c,*}, and Celso R. B. Cabral^d

^aDepartment of Statistics, Campinas State University, Campinas, São Paulo, Brazil

^bDepartment of Biostatistics, Virginia Commonwealth University, Richmond, VA, U.S.A

^cDepartment of Statistics, University of Connecticut, Storrs, CT 06269, U.S.A

^dDepartment of Statistics, Federal University of Amazonas, Manaus, Brazil

Abstract

Spatially-referenced geostatistical responses that are collected in environmental sciences research are often subject to detection limits, where the measures are not fully quantifiable. This leads to censoring (left, right, interval, etc), and various ad hoc statistical methods (such as choosing arbitrary detection limits, or data augmentation) are routinely employed during subsequent statistical analysis for inference and prediction. However, inference may be imprecise and sensitive to the assumptions and approximations involved in those arbitrary choices. To circumvent this, we propose an *exact* maximum likelihood estimation framework of the fixed effects and variance components and related prediction via a novel application of the Stochastic Approximation of the Expectation Maximization (SAEM) algorithm, allowing for easy and elegant estimation of model parameters under censoring. Both simulation studies and application to a real dataset on arsenic concentration collected by the Michigan Department of Environmental Quality demonstrate the advantages of our method over the available naïve techniques in terms of finite sample properties of the estimates, prediction, and robustness. The proposed methods can be implemented using the R package *CensSpatial*.

Keywords

Censored geostatistical data; Kriging; Limit of detection (LOD); SAEM algorithm

1. Introduction

Geostatistical data modeling has now virtually permeated all areas of epidemiology, hydrology, agriculture, environmental science, demographic studies, etc. Here, the prime objective is to account for the spatial correlation among the observations collected at various locations, and also to predict the values of interest for non-sampled sites. For most

* Address for correspondence: Víctor Hugo Lachos Dávila, Department of Statistics, University of Connecticut, 215 Glenbrook Rd. U-4120, Storrs, CT 06269-4120; hlachos@uconn.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

applications, the data are assumed to be fully observed. However, in many situations, the measured (spatial) responses are censored, i.e., the measurements are subjected to some upper or lower detection limits (depending on the measuring device), below or above which they are not quantifiable (Schelin and Sjöstedt-de Luna, 2014). In general, the proportion of such censored observations may be non-trivial, and employing crude/ad hoc methods, such as threshold substitution, or choosing some arbitrary point in the limit of detection (LOD) may lead to biased estimates of fixed effects and variance components (Fridley and Dixon, 2007).

The problem of inference and prediction for censored spatial data has, so far, received some attention in the literature. Militino and Ugarte (1999) developed an Expectation Maximization (EM) type algorithm for maximum likelihood (ML) estimation. However, this approach assumes a known correlation structure in the data. Additionally, this approach does not account for the difference in information contained in the exact and censored observations. As an improvement, De Oliveira (2005) developed a Bayesian approach using Gaussian random fields with applications to a dataset on depths of a geological horizon. Furthermore, Rathbun (2006) applied a Robbins-Monro stochastic approximation (Robbins and Monro, 1951) for parameter estimation, and employed weighted mean of kriging predictors for prediction at sampled sites. More recently, Schelin and Sjöstedt-de Luna (2014) proposed a semi-naïve method which determines imputed values at censored locations via an iterative algorithm, together with variogram estimation.

Models for censored data mostly consider the popular EM algorithm (Dempster et al., 1977) to obtain ML estimates of model parameters (see, for instance, Militino and Ugarte, 1999). This algorithm uses an iterative process, where each iteration consists of the E-step that computes the conditional expectation of the complete data log-likelihood (conditional on the observed, and censored data, and the current parameter estimates), and the M-step, which maximizes the complete likelihood, in particular the Q-function. Quite often, the conditional expectation in the E-step entertains multiple integrals, and is not available in closed form. This has led to considering analytical approximations, numerical quadratures, or using Monte Carlo methods that rely on empirical averages based on simulated data, such as the Monte Carlo EM (MCEM) algorithm (Wei and Tanner, 1990). However, MCEM experiences computational bottlenecks, often requiring a large number of simulations for convergence. As a remedy, the Stochastic Approximation of the EM (SAEM) algorithm proposed by Delyon et al. (1999) accelerates convergence by dividing the E-step into two parts: a simulation of the individual parameters using a Markov chain Monte Carlo (MCMC) algorithm (S-step), followed by a stochastic approximation (A-step) of the expected likelihood. The A-step (Robbins and Monro, 1951) computes a weighted mean of the approximation in the current and all previous iterations, and uses a decreasing sequence of step sizes to discard earlier iterations, thereby facilitating convergence with a fixed (and small) simulation size. In the framework of spatial models, Jank (2006) showed that the computational expense of SAEM is significantly smaller than the MCEM, often reaching convergence in just a fraction of the simulation size.

The central aim of this paper is to compare spatial predictions obtained via the SAEM algorithm with three competing prediction algorithms described in Schelin and Sjöstedt-de

Luna (2014), called the Naive 1, Naive 2 and the Seminaive methods. The rest of the paper is organized as follows. Section 2 starts with a brief introduction of the standard spatial linear model (SLM), related EM and SAEM algorithms, and then extends it to include censoring. The corresponding (log)-likelihood, details regarding the SAEM implementation, and prediction are presented in Section 3. Section 4 applies the model to a real dataset on arsenic concentration (Goovaerts et al., 2005), collected by the Michigan Department of Environmental Quality (MDEQ), and also compares it to available ad hoc methods in terms of prediction accuracy. The finite sample comparison of our proposal with competing techniques from synthetic data in terms of estimation and prediction under varying degrees of censoring are presented in Section 5. Finally, Section 6 presents some concluding remarks.

2. Statistical Model

2.1. Spatial linear model for censored responses

Consider a real-valued Gaussian random process $Z(\mathbf{s})$, $\mathbf{s} \in D$, where D is a subset of $\mathbb{R}^d$, the d -dimensional Euclidean space. Following Mardia and Marshall (1984) and De Bastiani et al. (2014), the realizations of this process $Z(\mathbf{s}) = (Z(\mathbf{s}_1), Z(\mathbf{s}_2), \dots, Z(\mathbf{s}_n))$ at known locations $\mathbf{s}_i$, $i = 1, \dots, n$, where $\mathbf{s}_i$ is a d -dimensional vector of spatial site coordinates, follows

$$Z(\mathbf{s}_i) = \mu(\mathbf{s}_i) + \varepsilon(\mathbf{s}_i), \quad (1)$$

where, both the deterministic term $\mu(\mathbf{s}_i)$ and the stochastic term $\varepsilon(\mathbf{s}_i)$ can depend on the observed spatial location for $Z(\mathbf{s}_i)$. We assume $E\{\varepsilon(\mathbf{s}_i)\} = \mathbf{0}$, and the variation between spatial points is determined by a covariance function $C(\mathbf{s}_i, \mathbf{s}_j) = \text{Cov}\{\varepsilon(\mathbf{s}_i), \varepsilon(\mathbf{s}_j)\}$. In particular, we define $\mu(\mathbf{s}_i)$, the mean of the stochastic process as

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

where $x_1(\mathbf{s}_i), \dots, x_p(\mathbf{s}_i)$ are known functions of $\mathbf{s}_i$ and $\beta_1, \dots, \beta_p$ 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 $\boldsymbol{\phi} = (\phi_1, \dots, \phi_q)^\top$, where $\top$ denotes transpose. Henceforth, for notational convenience, we redefine $\mathbf{s}_i = i$, which leads to $Z(\mathbf{s}_i) = Z_i$, $\mathbf{Z} = (Z_1, \dots, Z_n)^\top$, $x_j(\mathbf{s}_i) = x_{ij}$, $\mathbf{x}_i^\top = (x_{i1}, \dots, x_{ip})$ such that $\mathbf{X}$ is a $n \times p$ matrix with the i th row $\mathbf{x}_i^\top$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$, $\mu(\mathbf{s}_i) = \mu_i$, $\varepsilon(\mathbf{s}_i) = \varepsilon_i$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^\top$, with $i = 1, \dots, n$ and $j = 1, \dots, p$. This implies $\mu_i = \mathbf{x}_i^\top \boldsymbol{\beta}$ and $Z_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i$, $i = 1, \dots, n$. Equivalently, in matrix notation, we have our SLM as:

$$\mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (3)$$

where $E\{\boldsymbol{\varepsilon}\} = \mathbf{0}$ and $\boldsymbol{\Sigma} = [C(\mathbf{s}_i, \mathbf{s}_j)] = \tau^2 \mathbf{I}_n + \sigma^2 \mathbf{R}(\boldsymbol{\phi})$. We assume $\boldsymbol{\Sigma}$ is non-singular, and $\mathbf{X}$ has full rank. Using standard geostatistical terms, τ^2 is the nugget (or the measurement error variance), σ^2 the sill, and $\mathbf{R} = \mathbf{R}(\boldsymbol{\phi}) = [r_{ij}]$ is an $n \times n$ symmetric matrix with diagonal elements $r_{ii} = 1$, for $i = 1, \dots, n$ with $\boldsymbol{\phi}$ as 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 $\mathbf{s}_j$. Various choices of $\mathbf{R}$ in the literature exist. For example, the Matérn family for $\mathbf{R}$ is defined as

$$\mathbf{R}(\boldsymbol{\phi}) = \mathbf{R}(\boldsymbol{\phi}, d_{ij}) = \begin{cases} \frac{1}{2^{\kappa-1} \Gamma(\kappa)} \left(\frac{d_{ij}}{\boldsymbol{\phi}} \right)^{\kappa} K_{\kappa}(d_{ij}/\boldsymbol{\phi}), & d_{ij} > 0, \\ 1, & d_{ij} = 0, \end{cases} \quad (4)$$

where $\boldsymbol{\phi} > 0$; $K_{\kappa}(u) = \frac{1}{2} \int_0^{\infty} x^{\kappa-1} e^{-\frac{1}{2}u(x+x^{-1})} dx$ is the modified Bessel function of the third kind of order κ (see Gradshteyn and Ryzhik, 1965), with $\kappa > 0$ fixed. When $\kappa \rightarrow \infty$ and $\kappa = 0.5$, the Gaussian and exponential correlations respectively can be obtained from (4) (see, Diggle and Ribeiro, 2007).

We also assume that the response Z_i is not fully observed for all locations i . Hence, our observed data at the i th location is (V_i, C_i) , where V_i represents either an uncensored observation, or the LOD of the censoring level, with C_i the censoring indicator defined as

$$C_i = \begin{cases} 1 & \text{if } V_{1i} \leq Z_i \leq V_{2i}, \\ 0 & \text{if } Z_i = V_{0i}. \end{cases} \quad (5)$$

The model defined in (1)–(5), is called the spatial censored linear (SCL) model. When $C_i = 1$, we have a left-censored SCL model (Toscas, 2010) if $Z_i \in (-\infty, V_{2i}]$, and a right-censored SCL model if $Z_i \in [V_{1i}, \infty)$.

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 ML estimation of model parameters. Define $\mathbf{Z}_{\text{com}} = (\mathbf{Z}_m, \mathbf{Z}_{\text{obs}})$, where $\mathbf{Z}_m$ denotes the missing data, $\mathbf{Z}_{\text{obs}}$, the observed data and $\mathbf{Z}_{\text{com}}$, the complete data. This iterative algorithm maximizes the complete log-likelihood function $\ell_{\text{c}}(\boldsymbol{\theta}, \mathbf{Z}_{\text{com}})$ at each step, and converges to a stationary point of the observed likelihood ($\ell(\boldsymbol{\theta}, \mathbf{Z}_{\text{obs}})$) under mild regularity conditions (Wu, 1983; Vaida, 2005). The algorithm proceeds in two simple steps

E-Step: Replace the observed likelihood with the complete likelihood and compute its conditional expectation $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^k) = E\{\ell_{\text{c}}(\boldsymbol{\theta}, \mathbf{Z}_{\text{com}})|\hat{\boldsymbol{\theta}}^k, \mathbf{Z}_{\text{obs}}\}$, where $\hat{\boldsymbol{\theta}}^k$ is the estimate of $\boldsymbol{\theta}$ at the k -th iteration;

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

Note that one of the serious pitfalls of the EM algorithm is getting trapped at a local maximum for multi-modal likelihood functions (Zhou and Lange, 2010), for which various exit strategies has been suggested. As noted earlier, the E-step cannot be obtained

analytically. Hence, it is calculated via simulations, such as the MCEM (Wei and Tanner, 1990). As an alternative to the computationally intensive MCEM, the SAEM algorithm proposed by Delyon et al. (1999) replaces the E-step with 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 maximum of the ML estimates under quite general conditions (Delyon et al., 1999; Kuhn and Lavielle, 2005a; Allasonnière et al., 2010). 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 $(\mathbf{q}^{(l,k)})$, $l = 1, \dots, m$ from the conditional distribution $f(\mathbf{q}|\boldsymbol{\theta}^{(k-1)}, \mathbf{Z}_p)$.
- Stochastic Approximation: Update the $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)})$ function as

$$Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}) \approx Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k-1)}) + \delta_k \left[\frac{1}{m} \sum_{l=1}^m \ell_c(\boldsymbol{\theta}|\mathbf{Z}_{\text{obs}}, \mathbf{q}^{(l,k)}) - Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k-1)}) \right]. \quad (6)$$

M-Step:

- Maximization: Update $\hat{\boldsymbol{\theta}}^{(k)}$ as $\hat{\boldsymbol{\theta}}^{(k+1)} = \underset{\boldsymbol{\theta}}{\text{argmax}} Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)})$, where δ_k is a smoothness parameter (Kuhn and Lavielle, 2005a), 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, but only a small number of simulations m (suggested to be $m = 20$) 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 iteration k denoted by $(\mathbf{q}^{(l,k)})$, $l = 1, \dots, m$, but also some or all previous simulations, where this ‘memory’ property is set by the smoothing parameter δ_k .

Note that in Equation (6), 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, but the algorithm with memory will converge slowly (almost sure convergence) to the ML solution. We suggest the following choice of the smoothing parameter

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

where W is the maximum number of iterations, and c is a cutoff 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 be memory free, and hence will converge quickly to a solution neighborhood. For the first case, W needs to be large. For the second, the algorithm will initiate a Markov chain, leading to a reasonably estimated mean after applying the necessary *burn in* and *thinning* steps.

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 almost sure convergence for the rest of the iterations. Hence, this combination will lead to a fast algorithm with good estimates. To implement SAEM, the user must fix several constants that correspond to the number of total iterations W , and the cutoff c that defines the initiation of the smoothing step. However those parameters will vary depending on the model and the data. To determine those constants, a graphical approach is recommended that monitors the convergence of all parameter estimates, and if possible, the difference (relative difference) between two successive evaluations of the log-likelihood.

3. Likelihood and SAEM Implementation

To compute the likelihood function associated with the SCL model, the observed and censored components of $\mathbf{Z}$ must be treated separately. Let $\mathbf{Z}^o$ be the n^o -vector of observed responses 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 $C_i = 1$ for all elements in $\mathbf{Z}^c$. After reordering, $\mathbf{Z}$, $\mathbf{V}$, $\mathbf{X}$, and Σ can be partitioned as follows:

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

where $\text{vec}(\cdot)$ denotes the function which stacks vectors or matrices having the same number of columns. Consequently, $\mathbf{Z}^o \sim N_{n^o}(\mathbf{X}^o\boldsymbol{\beta}, \Sigma^{oo})$, $\mathbf{Z}^c|\mathbf{Z}^o \sim N_{n^c}(\boldsymbol{\mu}, \mathbf{S})$, where $\boldsymbol{\mu} = \mathbf{X}^c\boldsymbol{\beta} + \Sigma^{co}(\Sigma^{oo})^{-1}(\mathbf{Z}^o - \mathbf{X}^o\boldsymbol{\beta})$ and $\mathbf{S} = \Sigma^{cc} - \Sigma^{co}(\Sigma^{oo})^{-1}\Sigma^{oc}$. Now, let $\phi_n(\mathbf{u}; \mathbf{a}, \mathbf{A})$ be the *pdf* of $N_n(\mathbf{a}, \mathbf{A})$ evaluated at $\mathbf{u}$. From Vaida and Liu (2009) and Jacqmin-Gadda et al. (2000), the likelihood function (using conditional probability arguments) is given by

$$L(\boldsymbol{\theta}) = \phi_{n^o}(\mathbf{Z}^o; \mathbf{X}^o\boldsymbol{\beta}, \Sigma^{oo}) P(\mathbf{Z}^c \in \mathbb{V}^c | \mathbf{Z}^o), \quad (7)$$

where

$$\mathbb{V}^c = \{\mathbf{Z}^c = (Z_1^c, \dots, Z_{n^c}^c)^\top | V_{11}^c \leq Z_1^c \leq V_{21}^c, \dots, V_{1n^c}^c \leq Z_{n^c}^c \leq V_{2n^c}^c\}$$

and $P(\mathbf{u} \in \mathbb{A} | \mathbf{Z}^o)$ denotes the conditional probability of $\mathbf{u}$ being in the set $\mathbb{A}$ given the observed response. This function can be evaluated without much computational burden through the `mvtnorm` routine available in the R package `mvtnorm` (Genz et al., 2016). We can also compare models using likelihood based criteria such as the Akaike (AIC; Akaike, 1974) and Schwarz (BIC; Schwarz, 1978) information criteria.

3.1. SAEM algorithm for censored spatial data

We propose an EM-type (SAEM) algorithm by considering $\mathbf{Z}$ as missing data, or a latent variable. In the estimation, the reading $\mathbf{V}$ at the censored observations ($C_i = 1$) are treated as hypothetical missing data, and augmented with the observed dataset $\mathbf{Z}_c = (\mathbf{C}^\top, \mathbf{V}^\top, \mathbf{Z}^\top)^\top$. Hence, the complete-data log-likelihood function is given by:

$$\ell_c(\boldsymbol{\theta}) \propto -\frac{1}{2}[\log(|\sum|) + (\mathbf{Z} - \mathbf{X}\boldsymbol{\beta})^\top \sum^{-1} (\mathbf{Z} - \mathbf{X}\boldsymbol{\beta})] + K, \quad (8)$$

with K is a constant term, independent of the parameter vector $\boldsymbol{\theta}$. Given the current estimate $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^{(k)}$, the E-step computes the conditional expectation of the complete data log-likelihood function, i.e., $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}) = E\{\ell_c(\boldsymbol{\theta}|\mathbf{Z}_c)|\mathbf{V}, \mathbf{C}, \hat{\boldsymbol{\theta}}^{(k)}\}$. Denote the two conditional first moments for the response $\mathbf{Z}$ as $\hat{\mathbf{Z}}^{(k)} = E\{\mathbf{Z}|\mathbf{V}, \mathbf{C}, \hat{\boldsymbol{\theta}}^{(k)}\}$ and $\widehat{\mathbf{ZZ}}^{\top(k)} = E\{\mathbf{ZZ}^\top|\mathbf{V}, \mathbf{C}, \hat{\boldsymbol{\theta}}^{(k)}\}$. For the SCL model, we have

$$Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}) = -\frac{1}{2} \left[\log(|\sum|) + \hat{A}^{(k)} \right],$$

where

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

$\mathbf{Z}$ is not fully observed, hence the components of $\hat{\mathbf{Z}}^{(k)}$ and $\widehat{\mathbf{ZZ}}^{\top(k)}$ that correspond to $C_i = 1$ will be estimated by the first two moments of the truncated normal distribution respectively. When $C_i = 0$, these components can be obtained directly from the observed values, i.e., $\hat{\mathbf{Z}}^{(k)} = \mathbf{Z}^o$ and $\widehat{\mathbf{ZZ}}^{\top(k)} = \mathbf{Z}^o \mathbf{Z}^{o\top}$. Although these expectations exhibit closed forms as functions of multinormal probabilities (Arismendi, 2013), the calculation is computationally expensive requiring high-dimensional numerical integrations, often resulting in convergence issues when the proportion of censored observations is non-negligible. At each iteration, the SAEM algorithm successively simulates from the conditional distribution of the latent variable, and updates the unknown parameters of the model. Thus, at iteration k , the SAEM proceeds as follows

Step E-1 (Sampling). Sample $\mathbf{Z}^c$ from a truncated normal distribution, denoted by $TN_{n^c}(\boldsymbol{\mu}, \mathbf{S}; \mathbb{A})$, with

$\mathbb{A}^c = \{\mathbf{Z}^c = (Z_1^c, \dots, Z_{n^c}^c)^\top | V_{11}^c \leq Z_1^c \leq V_{21}^c, \dots, V_{1n^c}^c \leq Z_{n^c}^c \leq V_{2n^c}^c\}$, $\boldsymbol{\mu} = \mathbf{X}^c \boldsymbol{\beta} + \boldsymbol{\Sigma}^{co}(\boldsymbol{\Sigma}^{oo})^{-1}(\mathbf{Z}^o - \mathbf{X}^o \boldsymbol{\beta})$ and $\mathbf{S} = \boldsymbol{\Sigma}^{cc} - \boldsymbol{\Sigma}^{co}(\boldsymbol{\Sigma}^{oo})^{-1}\boldsymbol{\Sigma}^{oc}$. Here $TN_n(\cdot; \mathbb{A})$ denotes the n -variate truncated normal distribution on the interval $\mathbb{A}$, where $\mathbb{A} = A_1 \times \dots \times A_n$. The new observation $\mathbf{Z}^{(k,l)} = (Z_1^{(k,l)}, \dots, Z_{n^c}^{(k,l)}, Z_{n^c+1}, \dots, Z_n)$ is a sample generated for the n^c censored cases and the observed values (uncensored cases), for $l = 1, \dots, M$.

Step E-2 (Stochastic Approximation). Since we have the sequence $\mathbf{Z}^{(k,l)}$, we replace the conditional expectations $\hat{\mathbf{Z}}^{(k)}$ and $\widehat{\mathbf{Z}\mathbf{Z}^\top}^{(k)}$ at the k -th iteration with the following stochastic approximations

$$\hat{\mathbf{Z}}^{(k)} = \hat{\mathbf{Z}}^{(k-1)} + \delta_k \left[\frac{1}{M} \sum_{l=1}^M \mathbf{Z}^{(k,l)} - \hat{\mathbf{Z}}^{(k-1)} \right], \quad (9)$$

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

where δ_k is a smoothness parameter, i.e., a decreasing sequence of positive numbers (Kuhn and Lavielle, 2005a), such that $\sum_{k=1}^{\infty} \delta_k = \infty$ and $\sum_{k=1}^{\infty} \delta_k^2 < \infty$. For the SAEM, the E-Step coincides with the MCEM algorithm at the expense of a significantly smaller number of simulations $M = 20$. For more details, we refer the reader to Delyon et al. (1999). Finally, the conditional maximization (CM) step maximizes $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)})$ with respect to $\boldsymbol{\theta}$ to obtain a new estimate $\hat{\boldsymbol{\theta}}^{(k+1)}$ as follows

CM-Step (Conditional Maximization)

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

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

$$\hat{\boldsymbol{\alpha}}^{(k+1)} = \underset{\boldsymbol{\alpha} \in \mathbb{R}^+ \times \mathbb{R}^+}{\text{argmax}} \left(-\frac{1}{2} \log(|\widehat{\boldsymbol{\Sigma}}|) - \frac{1}{2} \left[\text{tr}(\widehat{\mathbf{Z}\mathbf{Z}^\top}^{(k)} \widehat{\boldsymbol{\Sigma}}^{-1}) - 2 \widehat{\mathbf{Z}}^{\top(k)} \widehat{\boldsymbol{\Sigma}}^{-1} \mathbf{X} \hat{\boldsymbol{\beta}}^{(k+1)} + \hat{\boldsymbol{\beta}}^{\top(k+1)} \mathbf{X}^\top \widehat{\boldsymbol{\Sigma}}^{-1} \mathbf{X} \hat{\boldsymbol{\beta}}^{(k+1)} \right] \right),$$

(11)

with $\boldsymbol{\alpha} = (\nu^2, \boldsymbol{\phi})^\top$. Note that $\hat{\boldsymbol{\tau}}^2$ can be recovered via $\hat{\boldsymbol{\tau}}^{2(k+1)} = \hat{\nu}^{2(k+1)} \widehat{\sigma^2}^{(k+1)}$. The CM-step (11) can be easily accomplished using the `optim` routine in R. This process is iterated until some absolute distance between two successive evaluations of the actual log-likelihood $\ell(\boldsymbol{\theta})$, such as $|\ell(\hat{\boldsymbol{\theta}}^{(k+1)}) - \ell(\hat{\boldsymbol{\theta}}^{(k)})|$ or $|\ell(\hat{\boldsymbol{\theta}}^{(k+1)})/\ell(\hat{\boldsymbol{\theta}}^{(k)}) - 1|$, becomes small enough.

Indeed, there are known results regarding the convergence of SAEM in the literature under various structured dependence settings. For example, Zhu et al. (2007) and Panhard and Samson (2008) present convergence results under a set of assumptions that were motivated by the primary results in Delyon et al. (1999), Kuhn and Lavielle (2005a) and Kuhn and Lavielle (2005b), for Markov random fields (for example, spatial random effects), and multilevel nonlinear mixed models, respectively. However, following Panhard and Samson (2008), we do not recommend exploring a deterministic convergence criterion here given that SAEM is a stochastic algorithm. Instead, we also recommend a graphical approach for convergence (see Subsection 2.2) checks by plotting the SAEM estimates against the iteration number.

3.2. Prediction

Following Diggle and Ribeiro (2007), let $\mathbf{Z}_{obs}$ denote a vector of random variables with observed values, and $\mathbf{Z}_{pred}$ denote another random variable whose realized values we would like to predict from the observed values of $\mathbf{Z}_{obs}$. A prediction for $\mathbf{Z}_{pred}$ can be any function of $\mathbf{Z}_{obs}$, which we denote by $\hat{\mathbf{Z}}_{pred}$. The mean square error (MSE) of $\hat{\mathbf{Z}}_{pred}$ is given by $MSE(\hat{\mathbf{Z}}_{pred}) = E\{(\hat{\mathbf{Z}}_{pred} - \mathbf{Z}_{pred})^2\}$, such that we can use $\hat{\mathbf{Z}}_{pred} = E\{\mathbf{Z}_{pred} | \mathbf{Z}_{obs}\}$ to find the predictor that minimizes $MSE(\mathbf{Z}_{pred})$.

For a Gaussian process, we write $\mathbf{Z}_{pred} = (Z_{pred,1}, \dots, Z_{pred,n})$ for the unobserved values of the signal at the sampling locations $i, i = 1, \dots, n$. We would like to predict the signal value at an arbitrary location; thus our target for prediction is $Z_{pred,i}$. Given that $(\mathbf{Z}_{obs}, \mathbf{Z}_{pred})$ is also a multivariate Gaussian process, we can use the results in Subsection 3 to minimize the $MSE(\hat{\mathbf{Z}}_{pred})$, i.e., if $\mathbf{X}^* = (\mathbf{X}_{obs}, \mathbf{X}_{pred})$ is the $(n_{obs} + n_{pred}) \times p$ design matrix corresponding to $\mathbf{Z}^* = (\mathbf{Z}_{obs}^\top, \mathbf{Z}_{pred}^\top)$, then $\mathbf{Z}^* \sim N_{n_{obs} + n_{pred}}(\mathbf{X}^* \boldsymbol{\beta}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sum_{obs,obs} & \sum_{obs,pred} \\ \sum_{pred,obs} & \sum_{pred,pred} \end{bmatrix}$$

and $\mathbf{Z}_{pred} | \mathbf{Z}_{obs} \sim N_{n_{pred}}(\boldsymbol{\mu}_p, \boldsymbol{\Sigma}_p)$, where

$$\boldsymbol{\mu}_p = \mathbf{X}_{pred} \boldsymbol{\beta} + \sum_{obs,pred} (\sum_{obs,obs})^{-1} (\mathbf{Z}_{obs} - \mathbf{X}_{obs} \boldsymbol{\beta}) \quad (12)$$

and

$$\boldsymbol{\Sigma}_p = \sum_{pred,pred} - \sum_{pred,obs} (\sum_{obs,obs})^{-1} \sum_{obs,pred}.$$

Then, the predictor that minimizes the MSE of prediction will be the conditional expectation in (12).

3.3. Competing prediction methods

In this paper, we consider three competing prediction algorithms as described in Schelin and Sjöstedt-de Luna (2014) – the Naive 1, Naive 2 and the Seminaive methods. As mentioned earlier, the central goal in this paper is to compare out-of-sample predictions at a new spatial location, essentially kriging. The first two methods simply consider two kinds of ad hoc imputation of the censored values based on the corresponding limit of detections to create a full dataset free of censoring, and the prediction (at a new location) is obtained via the plug-in simple kriging predictor. For the third (seminative) method, the imputed values of the censored observations are obtained within an iterative algorithm as proposed by Schelin and Sjöstedt-de Luna (2014). Using the previous results, we will now describe the algorithms for each of these three prediction methods

a. Naive 1 and Naive 2 algorithms

We proceed as follows

1. Impute the censored observations using LOD (for Naive 1), or $LOD/2$ (for Naive 2), depending on the type of censoring (left, or right).
2. Compute the least squares, or likelihood estimates for the mean (linear trend in our case), and covariance structure using the imputed data.
3. Evaluate the estimate $\hat{\theta}$ from step 2 in expression (12) to obtain the predicted values.

b. Seminaive algorithm

Let C_i denote the indicator variable, with $C_i = 1$ (presence), or 0 (absence) of a censored observation. We proceed with the following steps

1. Set $\hat{Z}_0 = (\mathbf{Z}_{\text{obs}}^T, \hat{\mathbf{v}}_0^T)$, where $\hat{\mathbf{v}}_{0i} = 0$ if $C_i = 1$.
2. Obtain $\hat{\theta}_0 = (\hat{\beta}, \hat{\sigma}^2, \hat{\phi})$ from $\hat{Z}_0$ by least squares
3. Set $k = 0$. Let $\hat{Z}_k^{-i}$ denote the data vector $\hat{Z}_k$, where location i is removed.
4. Find the predictor $\hat{Z}(h, \hat{\theta}_k)$ for all i such that $C_i = 1$ using expression (12) and $\hat{Z}_k^{-i}$, where h denotes the distance between coordinates (isotropy).
5. Set $\hat{Z}_{k+1} = (\mathbf{Z}_{\text{obs}}^T, \hat{\mathbf{v}}_{k+1}^T)$ where $\hat{\mathbf{v}}_{k+1} = (\max(0, \min(\hat{Z}(h, \hat{\theta}_k), LOD)) : C_i = 1)$.
6. Update θ_k from the new data $\hat{Z}_{k+1}$.
7. For iteration $k + 1$, repeat steps 4–6 until the convergence criterion is satisfied for some constants c_1 , c_2 , and c_3 chosen by the user.
8. Once convergence is attained, evaluate the estimate $\hat{\theta}$ in expression (12) to obtain the predicted values.

Note, this algorithm only works for the left-censored case. The constants c_1 , c_2 , and c_3 are chosen, such that the following three conditions are satisfied

$$\left| \frac{\widehat{\sigma}^2(\hat{\mathbf{Z}}_{k+1}) - \widehat{\sigma}^2(\hat{\mathbf{Z}}_k)}{\widehat{\sigma}^2(\hat{\mathbf{Z}}_k)} \right| \leq c_1, \quad \widehat{\sigma}^2(\hat{\mathbf{Z}}_{k+1}) \leq c_2 \widehat{\sigma}^2(\mathbf{Z}_{\text{obs}}) \quad \text{and} \quad \hat{\xi}(\hat{\mathbf{Z}}_{k+1}) > c_3 \hat{\xi}(\mathbf{Z}_{\text{obs}}), \quad (13)$$

where $\xi(X)$ represents the skewness of X .

c. SAEM algorithm

Here, we proceed as follows

1. Obtain the estimates of the mean and covariance structure by the SAEM procedure.
2. Impute the censored observations by the approximate first moments $\hat{\mathbf{Z}}^{(k)}$, obtained from the SAEM procedure.
3. Evaluate the above estimates in expression (12) to obtain predicted values for the unobserved locations. In this case, we use $\mathbf{Z}_{\text{obs}*}$ to denote the observed data and to estimate the censored observations via SAEM. This way, we can easily differentiate it from the fully observed response $\mathbf{Z}_{\text{obs}}$.

4. Application: Arsenic Contamination

In this section, we illustrate our SAEM-based estimation and prediction procedure by application to an arsenic contamination dataset (Kim et al., 2002; Goovaerts et al., 2005). We also compare and contrast the proposed SAEM method (that efficiently integrates observed and censored responses) to the available naive approaches. Concentrations of arsenic in drinking water was identified in groundwater supplies of 11 counties in southeastern Michigan: Genesee, Huron, Ingham, Jackson, Lapeer, Livingston, Oakland, Sanilac, Shiawassee, Tuscola, and Washtenaw. These data were collected at private wells sampled between 1993 and 2002. For illustration, we consider 1500 sampled points from this dataset, of which 1108 are observed and 392 (26.1%) are left-censored (i.e., falling below a LOD that varies between 0.3 and 2 $\mu\text{g/L}$).

In order to specify the mean and covariance components of the SCL model, we use some descriptive graphics. Figure 1 depicts the dataset, where panel (a) plots the sample locations, panel (b) presents the sample variogram of the observed data, and panel (c) plots the logarithm of the arsenic concentrations as a function of each coordinate pair. From this figure, it is plausible to assume a constant trend, given that panel (c) does not indicate any deterministic behavior of the arsenic concentrations. The sample variogram seems to justify an exponential covariance structure for the stochastic component. As an additional exercise,

we compared the Matérn covariance structure with various values of κ to the exponential structure. Although those fits were similar, our choice of the exponential shape was mostly driven by its simplicity and easy of interpretation.

4.1. Model specification and preliminary analysis

In this subsection, we fit the model

$$\log(Z_i) = \beta_0 + \varepsilon_i, \quad (14)$$

to the arsenic data, assuming exponential covariance for the stochastic errors ε_i . Figure 1 indicates a plausible nugget effect (variogram does not start at zero, the y-axis). We consider the estimation of this variation τ^2 in our setup. As initial values, we consider $\sigma_0^2 = 0.35$, $\phi_0 = 100$ and $\sigma_0^2 = 1.6$. For estimation, we choose a Monte Carlo sample size of $m = 20$, the maximum number of iterations $W = 200$, and a cutoff point $c = 0.2$. For the Seminaive algorithm, we choose the tuning parameters for the convergence criterion as $a_1 = 0.1$, $a_2 = 2$, $a_3 = 5$. The computational procedures were implemented using the R software package `CensSpatial` developed by the authors (Ordoñez et al., 2017).

4.2. Estimation and Prediction

Table 1 presents parameter estimates, associated log-likelihood, and model comparison (AIC/BIC) values, obtained after fitting the model (with exponential covariance structure) to the dataset via the four competing methods. Among the methods, all three measures (AIC/BIC/log-likelihood) significantly favored the SAEM adjustment (higher log-likelihood, lower AIC/BIC).

Figure 2 presents plots of the observed and fitted values from the dataset corresponding to the four methods, with the $x = y$ line overlaid to evaluate the fit. Although the performance of the SAEM, Naive 1 and Seminaive algorithms appear similar with some obvious outlying observations, Naive 2 does not seem to agree with the straight line.

Using the estimated parameters, we now proceed to compare their predictive performances using 5-fold cross-validation. To obtain out-sample prediction error, we use the mean square prediction error (MSPE) considering the squared distances between the real and predicted

values. This is defined as (see, Fridley and Dixon, 2007) $MSPE = \frac{\sum_{i=1}^n (Z_i - \hat{Z}_i)^2}{n}$, where Z_i is the observed value, $\hat{Z}_i$ is the predicted value and n is the number of samples to be predicted. The best predictive model is determined by the lowest MSPE. The MSPE of SAEM is 0.992, followed by 1.1, 1.153 and 1.203 for the Seminaive, Naive 1 and Naive 2 methods, respectively. The four methods do not vary much in terms of MSPE for this data application, with the SAEM appearing marginally better than the Seminaive, and others. This performance can certainly change with a higher proportion of censored values.

5. Simulation Study

In this section, we conduct simulation studies to evaluate the finite sample performance of the SAEM method in terms of estimation (in-sample) and prediction (out-of-sample), and compare it to the Naive 1, Naive 2 and Seminaive algorithms. In practice, we compare the bias and MSPE of the estimated parameters β , σ^2 and ϕ , and also out-of-sample prediction via 5- and 10-fold cross-validation.

For data generation, we consider a left-censored SLM with Gaussian errors, defined as

$$Z_i = \beta_0 + \beta_1 \mathbf{X}_{1i} + \beta_2 \mathbf{X}_{2i} + \varepsilon_i, \quad \varepsilon \sim N(\mathbf{0}, \sigma^2 \Psi), \quad (15)$$

where the covariates $\mathbf{X}_1$ and $\mathbf{X}_2$ are generated from the normal and Bernoulli distribution, respectively, i.e., $\mathbf{X}_1 \sim \text{Normal}(0, 1)$, $\mathbf{X}_2 \sim \text{Bernoulli}(0.5)$, with the parameters set at $\beta = (\beta_0, \beta_1, \beta_2) = (1, 0.5, 0.9)$, $\sigma^2 = 0.3$, $\phi = 3$, $\tau^2 = 0$, and Ψ as defined as in Section 2.2. For the correlation $\mathbf{R}(\phi)$, we considered the exponential function. The censoring levels were maintained at $c_1 = 25\%$ and $c_2 = 45\%$. We considered two sample sizes, $n_1 = 50$ and $n_2 = 200$. For the Seminaive algorithm, we chose the tuning parameters for the convergence criterion as suggested in Schelin and Sjöstedt-de Luna (2014), i.e., $a_1 = 0.1$, $a_2 = 2$, $a_3 = 5$.

We generated $T = 500$ Monte Carlo samples of size $n_i + n_{pred}$, $i = 1, 2$, where the first n_i observations were used to obtain parameter estimates and predictions, and the remaining $n_{pred} = 30$ were used to compare the performance of each of the competing prediction methods. We used $\sigma_0^2 = 0.2$, $\phi_0 = 0.1$ and the ordinary least squares estimator of β as initial values for each algorithm. For the Naive 1 and Naive 2 methods, we use the routine `likfit()` available in the `geOR` package (Ribeiro and Diggle, 2016) in R to obtain the estimates and predictions. For the Seminaive, we used the method described in Schelin and Sjöstedt-de Luna (2014).

Table 2 compares the bias and aggregated MSPE (combining all parameters) of model parameters, obtained after fitting the four competing methods to simulated data, varying with sample sizes and levels of censoring. Across both sample sizes and censoring levels, the SAEM method seems to uniformly produce the lowest bias and MSPE for all parameters. Intuitively, the bias of β_2 increases conspicuously for the Seminaive and Naive 1 algorithms with the increase in the censoring proportion for both sample sizes, but not so much for the Naive 2 and the SAEM. As expected, the standard deviation (SD) of the estimated β_0 (the intercept) is uniformly higher than β_1 , and σ^2 . On the overall, we observe improvement (lower SDs) with increase in sample sizes. The bias of ϕ is the largest among all parameters across all sample sizes and censoring patterns. Results from this simulation study indicate that the underlying method used to tackle censoring has a profound effect on the inference of fixed effect parameters, and prediction.

Next, for the 5 and 10-fold cross-validation (CV) study, we also generated $T = 500$ Monte Carlo samples. We use $\sigma_0^2 = 0.2$, $\phi_0 = 0.1$ and the ordinary least squares estimator of β as initial values for each method. Figures 3 and 4 present box-plots of the cross-validated

MSPEs obtained by the four competing methods, for various combinations of sample sizes and censoring levels. For the 5-fold CV, we observe from Figure 3 that the predictions are uniformly better from the SAEM method across all sample sizes and censoring levels, with the performance improving with increasing sample size. For the remaining algorithms, the Seminaive method presents better predictions than the Naive 1 and Naive 2 methods when $c = 25\%$. However, for the higher censoring level of $c = 45\%$, the Naive 1 algorithm presents the best predictions. For the 10-fold procedure, although the variability appears to be relatively higher, the results are analogous.

6. Conclusions

Motivated by a geostatistical dataset on arsenic contamination with censored responses, this paper develops a SAEM estimation procedure for ML inference and prediction to overcome some limitations of available ad hoc methods used in this context in the statistical literature (Rathbun, 2006; Schelin and Sjöstedt-de Luna, 2014). The developed method generalizes the traditional geostatistical Gaussian process regression, can be readily applied to other datasets, and easily implemented using the free R package `CensSpatial`.

We observe interesting differences in out-of-sample prediction results generated from the real data application and simulation studies. Specifically, for our data application with a relatively higher number of geostatistical points coupled with a moderate censoring (about 26%), the prediction errors from the four competing methods appear to be comparable. On the other hand, finite sample comparisons from the simulation studies (using 2 different sample sizes of 50 and 200, both of which are significantly lower than 1500) reveal a uniformly superior performance of the SAEM method over the others. From this, we infer that the effect of various ad hoc considerations to censored observations on out-of-sample predictions is negligible for relatively large geostatistical datasets with moderate censoring. However, this is certainly not the case for moderately-sized datasets with low to moderate censoring, and ad hoc methods may be avoided there.

For the present theoretical setup, we consider the exponential family based on the shape of the initial variogram plot of the dataset. Our method can be easily adapted to other types of spatial structures. Future extensions of this framework can include scale mixtures of normal distributions to accommodate heavy-tailed features in the censored responses as in De Bastiani et al. (2014). Furthermore, one may also consider extending the generalized linear mixed model framework (Diggle et al., 1998) to include censoring and measurement error in spatial models as in Li et al. (2009). All these are plausible avenues for future research, and will be considered elsewhere.

The available version of our R package `CensSpatial` conveniently handles 1500 points – the size of our illustrative dataset. However, as pointed out by a reviewer, it is not scalable to a very large number of observations. This issue of handling relatively large databases is very common in other R packages, such as `geoR`, that are popularly used in geostatistical analysis. We seek to revisit this scalability issue in a future communication. It is a non-trivial work, involving modifying our algorithms in light of research in the ‘big data’ domain.

Acknowledgments

We would like to thank the Editor, Associate Editor and two referees for their constructive comments, which led to a significantly improved version of this manuscript. This paper was written while Celso R. B. Cabral was a visiting professor in the Department of Statistics at the University of Campinas, Brazil. The research of Jose A. Ordoñez was supported by CAPES. Bandyopadhyay acknowledges partial support from grants R03DE023372, R01DE024984 and P30CA016059 (VCU Massey Cancer Center Support Grant) from the US National Institutes of Health. Celso R. B. Cabral acknowledges the support from FAPESP (Grant 2015/20922-5) and CNPq-Brazil (Grants 167731/2013-0 and 447964/2014-3). We also thank Jaymie Meliker for providing access to the arsenic contamination dataset.

References

- Akaike H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*. 1974; 19(6):716–723.
- Allasonnière S, Kuhn E, Trouvé A, et al. Construction of Bayesian deformable models via a stochastic approximation algorithm: a convergence study. *Bernoulli*. 2010; 16(3):641–678.
- Arismendi J. Multivariate Truncated Moments. *Journal of Multivariate Analysis*. 2013; 117(1):41–75.
- De Bastiani F, de Aquino Cysneiros AHM, Uribe-Opazo MA, Galea M. Influence diagnostics in elliptical spatial linear models. *TEST*. 2014; 24(2):322–340.
- De Oliveira V. Bayesian inference and prediction of Gaussian random fields based on censored data. *Journal of Computational and Graphical Statistics*. 2005; 14(1):95–115.
- Delyon B, Lavielle M, Moulines E. Convergence of a stochastic approximation version of the EM algorithm. *Annals of Statistics*. 1999; 27(1):94–128.
- Dempster A, Laird N, Rubin D. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B (Methodological)*. 1977; 39(1):1–38.
- Diggle P, Ribeiro P. *Model-Based Geostatistics*. Springer Series in Statistics. 2007
- Diggle PJ, Tawn J, Moyeed R. Model-based geostatistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*. 1998; 47(3):299–350.
- Fridley BL, Dixon P. Data augmentation for a Bayesian spatial model involving censored observations. *Environmetrics*. 2007; 18:107–123.
- Genz, A., Bretz, F., Hothorn, T., Miwa, T., Mi, X., Leisch, F., Scheipl, F. mvtnorm: Multivariate normal and t distribution. R package version 1.0–5. 2016. URL <http://CRAN.R-project.org/package=mvtnorm>
- Goovaerts P, Avruskin G, Meliker J, Slotnick M, Jacquez G, Nriagu J. Geostatistical modeling of the spatial variability of arsenic in groundwater of southeast Michigan. *Water Resources Research*. 2005; 41(7):1–19.
- Gradshteyn, IS., Ryzhik, IM. *Table of Integrals, Series and Products*. Academic Press; 1965.
- Jacqmin-Gadda H, Thiebaut R, Chene G, Commenges D. Analysis of left-censored longitudinal data with application to viral load in HIV infection. *Biostatistics*. 2000; 1:355–368. [PubMed: 12933561]
- Jank W. Implementing and diagnosing the stochastic approximation EM algorithm. *Journal of Computational and Graphical Statistics*. 2006; 15(4):803–829.
- Kim M-J, Nriagu J, Haack S. Arsenic species and chemistry in groundwater of southeast Michigan. *Environmental Pollution*. 2002; 120(2):379–390. [PubMed: 12395852]
- Kuhn E, Lavielle M. Coupling a stochastic approximation version of EM with an MCMC procedure. *European Series in Applied and Industrial Mathematics: Probability and Statistics*. 2005a; 8:115–131.
- Kuhn E, Lavielle M. Maximum likelihood estimation in nonlinear mixed effects models. *Computational Statistics & Data Analysis*. 2005b; 49(4):1020–1038.
- Li Y, Tang H, Lin X. Spatial linear mixed models with covariate measurement errors. *Statistica Sinica*. 2009; 19(3):1077–1093. [PubMed: 20046975]
- Mardia KV, Marshall R. Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika*. 1984; 71(1):135–146.

- Militino AF, Ugarte MD. Analyzing censored spatial data. *Mathematical Geology*. 1999; 31(5):551–561.
- Ordoñez, A., Galarza, CE., Lachos, VH. CensSpatial: Censored Spatial Models. R package version 1.3. 2017. URL <https://CRAN.R-project.org/package=CensSpatial>
- Panhard X, Samson A. Extension of the SAEM algorithm for nonlinear mixed models with 2 levels of random effects. *Biostatistics*. 2008; 10(1):121–135. [PubMed: 18583352]
- Rathbun SL. Spatial prediction with left-censored observations. *Journal of Agricultural, Biological, and Environmental Statistics*. 2006; 11(3):317–336.
- Ribeiro, PJ., Diggle, PJ. geoR: Analysis of Geostatistical Data. R package version 1.7–5.2. 2016. URL <https://CRAN.R-project.org/package=geoR>
- Robbins H, Monro S. A Stochastic Approximation Method. *The Annals of Mathematical Statistics*. 1951; 22(3):400–407.
- Schelin L, Sjöstedt-de Luna S. Spatial prediction in the presence of left-censoring. *Computational Statistics and Data Analysis*. 2014; 74:125–141.
- Schwarz G. Estimating the dimension of a model. *Annals of Statistics*. 1978; 6(2):461–464.
- Toscas PJ. Spatial modelling of left censored water quality data. *Environmetrics*. 2010; 21(6):632–644.
- Vaida F. Parameter convergence for EM and MM algorithms. *Statistica Sinica*. 2005; 15(3):831–840.
- Vaida F, Liu L. Fast implementation for normal mixed effects models with censored response. *Journal of Computational and Graphical Statistics*. 2009; 18:797–817. [PubMed: 25829836]
- Wei GC, Tanner MA. A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association*. 1990; 85(411):699–704.
- Wu CJ. On the convergence properties of the EM algorithm. *The Annals of Statistics*. 1983; 11(1):95–103.
- Zhou H, Lange KL. On the bumpy road to the dominant mode. *Scandinavian Journal of Statistics*. 2010; 37(4):612–631.
- Zhu H, Gu M, Peterson B. Maximum likelihood from spatial random effects models via the stochastic approximation expectation maximization algorithm. *Statistics and Computing*. 2007; 17(2):163–177.

Figure 1.

Arsenic contamination dataset. Panel (a) denotes study locations, where ‘standard’ circles represent observed values, and ‘boldface’ circles represent censored values. Panel (b) plots the sample variogram of the data, with the exponential variogram as a theoretical candidate. Panel (c) presents the logarithm of the arsenic concentrations as function of each coordinate pair.

(a)

(b)

(c)

(d)

Figure 2.

Arsenic contamination dataset. Plots of observed $\log(\text{arsenic concentration})$ and fitted values, obtained from fitting the real dataset via the four methods

(a)

(b)

(c)

(d)

Figure 3. Simulation study. Box-plots of out-of-sample MSPEs from the 5-fold cross-validation, varying with sample size and censoring levels.

(a)

(b)

(c)

(d)

Figure 4. Simulation study. Box-plots of out-of-sample MSPE's from the 10-fold cross-validation, varying with sample size and censoring levels.

Arsenic contamination dataset. Parameter estimates obtained from fitting the model with a constant trend and an exponential correlation function via the four competing methods.

Table 1

	β_0	σ^2	ϕ	τ^2	Loglik	AIC	BIC
Seminaive	-5.36	0.51	9100.00	1.19	-1294.50	2596.99	2615.74
Naive1	-5.41	0.46	9000.00	1.28	-1291.88	2591.76	2610.50
Naive2	-4.35	0.10	8999.99	1.06	-1178.45	2364.90	2383.64
SAEM	-5.27	0.85	8700.50	1.25	-1145.49	2296.98	2311.03

Table 2

Simulation study. Estimated bias and MSPE (combined across all parameters), obtained from fitting the four competing methods, varying with sample sizes and level of censoring. The numbers in parentheses are the respective standard deviations of the estimates.

n	Censoring	Method	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}^2$	$\hat{\phi}$	MSPE
50	25%	Seminaive	0.15 (0.15)	-0.09 (0.06)	-0.13 (0.12)	-0.09 (0.06)	-0.73 (1.02)	0.79 (0.16)
		Naive 1	0.16 (0.16)	-0.1 (0.05)	-0.14 (0.11)	-0.1 (0.04)	-1.92 (0.42)	0.74 (0.15)
		Naive 2	0.07 (0.13)	-0.05 (0.06)	-0.05 (0.12)	-0.06 (0.06)	-1.84 (0.41)	0.78 (0.15)
		SAEM	0.01 (0.16)	-0.01 (0.07)	-0.01 (0.12)	-0.03 (0.08)	-0.19 (1.25)	0.23 (0.17)
	45%	Seminaive	0.37 (0.16)	-0.2 (0.06)	-0.32 (0.12)	-0.14 (0.05)	-1.2 (1.13)	0.81 (0.17)
		Naive 1	0.37 (0.17)	-0.2 (0.04)	-0.32 (0.12)	-0.15 (0.04)	-2.14 (0.42)	0.75 (0.16)
		Naive 2	0.07 (0.12)	-0.07 (0.06)	-0.09 (0.13)	-0.05 (0.06)	-1.99 (0.39)	0.76 (0.15)
		SAEM	0.03 (0.18)	-0.02 (0.08)	-0.02 (0.15)	-0.04 (0.08)	-0.2 (1.45)	0.25 (0.08)
200	25%	Seminaive	0.11 (0.08)	-0.02 (0.03)	-0.06 (0.07)	-0.08 (0.06)	-1.05 (0.44)	0.57 (0.08)
		Naive 1	0.23 (0.13)	-0.14 (0.03)	-0.23 (0.05)	-0.09 (0.03)	-2.37 (0.11)	0.6 (0.07)
		Naive 2	0.06 (0.1)	-0.06 (0.03)	-0.09 (0.06)	-0.03 (0.04)	-2.28 (0.12)	0.6 (0.07)
		SAEM	0.01 (0.12)	-0.01 (0.03)	-0.01 (0.05)	-0.03 (0.05)	-0.19 (0.68)	0.13 (0.02)
	45%	Seminaive	0.37 (0.06)	-0.05 (0.05)	-0.17 (0.1)	-0.3 (0.09)	-1.73 (0.33)	0.58 (0.08)
		Naive 1	0.49 (0.13)	-0.24 (0.02)	-0.4 (0.05)	-0.14 (0.03)	-2.55 (0.11)	0.68 (0.09)
		Naive 2	-0.06 (0.09)	-0.09 (0.03)	-0.12 (0.07)	-0.01 (0.04)	2.5 (0.13)	0.59 (0.08)
		SAEM	0.07 (0.13)	-0.02 (0.03)	-0.04 (0.06)	-0.05 (0.05)	-0.37 (0.69)	0.14 (0.03)