ELSEVIER

Computational Statistics & Data Analysis 34 (2000) 299–314

COMPUTATIONAL STATISTICS & DATA ANALYSIS

www.elsevier.com/locate/csda

Bayesian prediction of clipped Gaussian random fields

Victor De Oliveira

Departamento de Cómputo Científico y Estadística, Universidad Simón Bolívar, Apartado 89000, Caracas 1080-A, Venezuela

Received 1 March 1999; received in revised form 1 November 1999

Abstract

This work provides a framework to perform prediction in some types of binary random fields. It is assumed the binary random field is obtained by clipping a Gaussian random field at a fixed level. The model, following a Bayesian approach, is used to map a binary outcome over a bounded region D of the plane: For each location $\boldsymbol{s}_0 \in D$, we compute the optimal predictor of $Z(\boldsymbol{s}_0)$, 0 or 1, given the binary data from a realization of the random field, and provide measures of prediction uncertainty amenable for binary outcomes. The optimal predictor and the measure of prediction uncertainty are computed through data augmentation using Markov Chain Monte Carlo methods; a less computationally demanding plug-in approach is also described. A brief description of a geostatistical method called indicator kriging is given as well as some of its shortcomings. The prediction ability of the model is illustrated with two simulated binary maps, obtaining satisfactory results, and comparisons between the Bayesian, plug-in, and indicator kriging approaches are given.

Keywords: Binary map; Binary random field; Data augmentation; Indicator kriging; Semivariogram; Spatial prediction

1. Introduction

This work considers the mapping of a region composed of only two features, based on observations at a finite number of locations. More specifically, consider a region $D \subset \mathbb{R}^d$ that is partitioned in *two* disjoint subregions, $D = B \cup W$, which we call a *binary map* of D. Examples of binary maps include a contaminated region where B is formed by the locations where the contaminant concentration surpasses a safety

E-mail address: vdo@cesma.usb.ve (V. De Oliveira).

0167-9473/00/$ - see front matter
PII: S 0167-9473(99)00103-6

level, and a geologic formation in a region composed of only two rock types. The binary map is observed at a finite number of sampling locations, where it is known to which, B or W, each of them belong. The goal is to estimate the subregion B.

A general approach to this problem is to assume that the binary map is a realization of a *binary random field* $\{Z(\boldsymbol{s}),\ \boldsymbol{s} \in D\}$. The problem can then be seen as one of spatial prediction and the binary map is estimated using a predictor $\hat{Z}(.)$ for the binary random field:

$$\hat{B} = \{\boldsymbol{s} \in D\colon \hat{Z}(\boldsymbol{s}) = 1\},\ \hat{W} = D - \hat{B}.$$

The main hurdle for prediction in binary random fields is the scarcity of suitable and flexible models. By this, we mean families of finite-dimensional distributions defined in terms of interpretable components, as is the case for Gaussian random fields, which make likelihood-based inference possible.

In this article, we assume that B is the *excursion set* obtained by 'clipping' an underlying Gaussian random field at a fixed level. The resulting *clipped Gaussian* random field provides a mathematical framework to perform likelihood-based inference for binary maps, having also some attractive features. First, some binary maps are *actually* generated — either by convenience or by the characteristics of the recording device — by thresholding an underlying (unobserved) continuous process, as is the case for the contaminated region example mentioned above. Second, even when there is no apparent physical mechanism generating the binary map by thresholding, the model can still be used as a modeling device applicable in situations where high degree of smoothness is expected in the binary map, as is the case in the geologic formation example mentioned above.

The idea of thresholding a continuous latent variable to model binary outcomes was used by Finney (1947) for univariate binary data, Ashford and Sowden (1970) for bivariate binary data, and Ochi and Prentice (1984) for equicorrelated multivariate binary data. The same idea has been used for modeling binary time series by Carlin and Polson (1992) and Kedem (1980), and for modeling binary spatial processes by De Oliveira (1997), Heagerty and Lele (1998), and Nott and Wilson (1997). Recently, Albert and Chib (1993) proposed the explicit use of the latent variables, using Markov chain Monte Carlo (MCMC) methods, to perform Bayesian analysis on this kind of models for univariate binary and ordinal data. Chib and Greenberg (1998) extended these methods to analyze multivariate binary data.

We study how to perform inference and prediction for the clipped Gaussian random field using the Bayesian approach. For an unsampled location $\boldsymbol{s}_0 \in D$, we compute $P\{Z(\boldsymbol{s}_0) = 1|data\}$ by appealing to data augmentation and MCMC methods. These probabilities are used to estimate the binary map and to compute a measure of prediction uncertainty amenable for binary data. A less computationally intensive plug-in approach is also described.

Also, we describe an alternative geostatistical approach called *indicator kriging* and point some of its drawbacks. This method was initially proposed by Switzer (1977) for estimation of the distribution function of a stationary random field, and later used by Journel (1983) as a non-parametric method for prediction in random fields, and by Solow (1986) for prediction in binary random fields. For its implementation,

indicator kriging only requires the second-order specification of the binary random field.

The paper is organized as follows. Section 2 describes the clipped Gaussian random field and discusses parameterization issues to assure model identifiability. Section 3 describes the Markov chain Monte Carlo algorithm for generating a sample from the joint posterior distribution of model parameters and augmented data. This sample is used for estimating the model parameters as well as estimating the binary map and computing measures of prediction uncertainty. Section 4 describes the indicator kriging approach. Section 5 illustrates the proposed model with two simulated binary maps, and discusses the empirical findings. Conclusions are given in Section 6.

2. The model

Let $\{Z(\boldsymbol{s}),\ \boldsymbol{s}\in D\}$, $D\subset\mathbb{R}^d$, $d\geq 1$, be a *binary random field*, meaning a random field where for every location $\boldsymbol{s}$, $Z(\boldsymbol{s})$ takes only the value 0 or 1. Suppose we have n observations $\boldsymbol{Z}=(Z(\boldsymbol{s}_1),\ldots,Z(\boldsymbol{s}_n))'$ from a single realization of this random field, $\boldsymbol{s}_1,\ldots,\boldsymbol{s}_n$ being known sampling locations in D. Based on $\boldsymbol{Z}$ and our prior knowledge about the random field, we want to predict the unobserved random vector $\boldsymbol{Z}_0=(Z(\boldsymbol{s}_{01}),\ldots,Z(\boldsymbol{s}_{0k}))'$, where $\boldsymbol{s}_{01},\ldots,\boldsymbol{s}_{0k}$ are also in D.

Let $\{Y(\boldsymbol{s}),\ \boldsymbol{s}\in D\}$ be an underlying Gaussian random field, with $E\{Y(\boldsymbol{s})\}=\boldsymbol{\beta}'\underline{f}(\boldsymbol{s})$; $\boldsymbol{\beta}=(\beta_1,\ldots,\beta_p)'\in R^p$ are unknown regression parameters, $\underline{f}(\boldsymbol{s})=(f_1(\boldsymbol{s}),\ldots,f_p(\boldsymbol{s}))'$ are known location-dependent covariates, with $f_1(\boldsymbol{s})\equiv 1$, and $Cov\{Y(\boldsymbol{s}),Y(\boldsymbol{u})\}=\sigma^2K_{\vartheta}(\boldsymbol{s},\boldsymbol{u})$; $\sigma^2>0$ is the variance and $\boldsymbol{\vartheta}=(\theta_1,\ldots,\theta_q)'\in\Theta\subset\mathbb{R}^q$ are correlation parameters of the random field $Y(.)$. The random field $Y(.)$ may describe the spatial variation of some real quantity that we are unable to observe, or it may just be considered a modeling device.

We assume the binary random field is obtained by *clipping* the Gaussian random field at the threshold ζ, meaning

$$Z(\boldsymbol{s})=\mathbf{1}_{\{Y(\boldsymbol{s})>\zeta\}},\quad \boldsymbol{s}\in D, \tag{1}$$

where $\mathbf{1}_A$ denotes the indicator function of the event A. From (1) and the Gaussian assumption about $Y(.)$, the family of finite-dimensional distributions for $Z(.)$ can be derived, making likelihood-based inference possible.

Let $\boldsymbol{z}=(z_1,\ldots,z_n)'$ be the observed binary data, $z_i=Z(\boldsymbol{s}_i)$, $i=1,\ldots,n$. The likelihood of the parameters $(\boldsymbol{\beta},\sigma^2,\boldsymbol{\vartheta},\zeta)'$ based on $\boldsymbol{z}$ is given by

$$\int_{A(z_1)}\cdots\int_{A(z_n)}\left(\frac{1}{2\pi\sigma^2}\right)^{n/2}|\Sigma_{\vartheta}|^{-1/2}\exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y}-X\boldsymbol{\beta})'\Sigma_{\vartheta}^{-1}(\boldsymbol{y}-X\boldsymbol{\beta})\right\}\mathrm{d}\boldsymbol{y}, \tag{2}$$

$$A(z_i)=\begin{cases}(-\infty,\zeta] & \text{if } z_i=0\\ (\zeta,\infty) & \text{if } z_i=1\end{cases},\quad i=1,\ldots,n,$$

where X is a known $n\times p$ design matrix defined by $X_{ij}=f_j(\boldsymbol{s}_i)$, assumed to have full rank, and Σ_{ϑ} is a $n\times n$ correlation matrix defined by $\Sigma_{\vartheta,ij}=K_{\vartheta}(\boldsymbol{s}_i,\boldsymbol{s}_j)$, assumed positive definite $\forall\,\boldsymbol{\vartheta}\in\Theta$.

It is apparent from (2) that a direct Bayesian analysis of this model is intractable, even when using MCMC methods, so we appeal to what is called data augmentation. This strategy was used by Albert and Chib (1993) and Chib and Greenberg (1998) for regression analysis in probit models, and by Carlin and Polson (1992) for analysis of categorical time series. We describe the approach in Section 3 and show how it is used for estimating the binary map. In the remainder of this section we discuss parameterization issues and the specification of the prior distribution for the model parameters.

Note from (2) that under this full model, the parameters are not (likelihood) identifiable. It can be shown that for any constants $a > 0$, $c \in R$, and $\boldsymbol{e}_1 = (1, 0, \ldots, 0)'$, the parameter vectors $(a\boldsymbol{\beta} - c\boldsymbol{e}_1, a^2\sigma^2, \vartheta, a\zeta - c)'$ have the same likelihood. To avoid this problem, we fix σ^2 and ζ at (arbitrary) known values σ_0^2 and ζ_0 (usually $\sigma_0^2 = 1$ and $\zeta_0 = 0$).

In principle, the model could be implemented using an arbitrary (parametric) correlation function, but some care must be exercised in its specification to avoid near-non-identifiability of its parameters and convergence problems of the MCMC algorithm. To illustrate this, consider using the isotropic correlation function $K_\vartheta(l) = \theta_1^{l^{\theta_2}}$, where l is Euclidean distance. The parameter $\theta_1 \in (0, 1)$ controls the *range of correlation*, while $\theta_2 \in (0, 2]$ controls the *roughness* of the realizations of the random field $Y(.)$. After clipping $Y(.)$, the binary data contain no information about θ_2, so there is no hope to tell different values of θ_2 apart, even if the complete binary realization were observed; similar identifiability problems were reported by Diggle et al. (1998). Therefore, correlation parameters controlling the smoothness/roughness of the random field $Y(.)$ should not be left free, but fixed at appropriate values. For the remainder of the paper, we assume that the correlation function $K_\vartheta(.)$ depends only on one parameter controlling the range of correlation (e.g., exponential, squared exponential, spherical), so our unknown model parameters are $\boldsymbol{\eta} = (\boldsymbol{\beta}, \vartheta)'$, being now (likelihood) identifiable.

Regarding the prior distribution, $\boldsymbol{\beta}$ and ϑ are assumed independent a priori with $\boldsymbol{\beta} \sim \mathrm{N_p}(\boldsymbol{m}_0, V_0^{-1})$, $\boldsymbol{m}_0$ and V_0^{-1} chosen to reflect our prior information about $P\{Z(\boldsymbol{s}) = 1\}$; see De Oliveira (1997) for details on how the second-order structure of the clipped Gaussian random field depends on the model parameters. A convenient 'non-informative' prior distribution is obtained by setting $\boldsymbol{m}_0 = \zeta_0\boldsymbol{e}_1$ and $V_0 = (t_0/\sigma_0^2)I_p$, $t_0 > 0$ small and I_p the $p \times p$ identity matrix. This choice implies that, a priori, $P\{Z(\boldsymbol{s}) = 1\} = \frac{1}{2}\ \forall\, \boldsymbol{s} \in D$, and assures that inferences are mainly driven by the likelihood. The prior $p(\vartheta)$ would depend on the selected correlation function $K_\vartheta(\boldsymbol{s}, \boldsymbol{u})$, but in general we use a vague proper prior for it. Then, our prior distribution for the model parameters is given by

$$p(\boldsymbol{\eta}) \propto \exp\{-\tfrac{1}{2}(\boldsymbol{\beta} - \boldsymbol{m}_0)' V_0 (\boldsymbol{\beta} - \boldsymbol{m}_0)\}\, p(\vartheta). \tag{3}$$

Remark. The arbitrary choices of σ^2 and ζ have no influence neither on inference about $\boldsymbol{\eta}$ nor on inference about the binary map, as described in the next section. If $(\sigma_1^2, \zeta_1)'$ and $(\sigma_2^2, \zeta_2)'$ are two possible choices, then there exist constants $a > 0$, $c \in R$ such that $\sigma_2^2 = a^2\sigma_1^2$ and $\zeta_2 = a\zeta_1 - c$. The corresponding latent model parameters

$(\boldsymbol{\beta}_1, \vartheta_1)'$ and $(\boldsymbol{\beta}_2, \vartheta_2)'$, being related as $\boldsymbol{\beta}_2 = a\boldsymbol{\beta}_1 - c\boldsymbol{e}_1$ and $\vartheta_2 = \vartheta_1$, have the same likelihood (based on $\boldsymbol{z}$) as well as the same prior distribution (when the non-informative prior is used), and hence $p(\boldsymbol{y}, \boldsymbol{\eta}|\boldsymbol{z})$ does not depend on which choice is used.

3. Inference for the binary random field

Let $\boldsymbol{y} = (y_1, \ldots, y_n)'$ and $\boldsymbol{y}_0 = (y_{01}, \ldots, y_{0k})'$ be the latent vectors corresponding respectively to $\boldsymbol{z}$ and $\boldsymbol{z}_0$, where $\boldsymbol{z}_0 = (z_{01}, \ldots, z_{0k})'$ is the binary vector we want to predict, $y_i = Y(\boldsymbol{s}_i)$, $y_{0j} = Y(\boldsymbol{s}_{0j})$, and $z_{0j} = Z(\boldsymbol{s}_{0j})$; $i = 1, \ldots, n$, $j = 1, \ldots, k$. We work with the *extended model* $p(\boldsymbol{y}_0, \boldsymbol{y}, \boldsymbol{\eta}, \boldsymbol{z})$ that explicitly includes the latent vectors $\boldsymbol{y}$ and $\boldsymbol{y}_0$ as unknown 'parameters', and make inferences using MCMC methods (Chib and Greenberg, 1995; Gilks et al. 1996).

3.1. Sampling from the posterior distribution

We briefly describe a MCMC algorithm that parallels the one proposed by Chib and Greenberg (1998) to obtain a sample from $p(\boldsymbol{y}, \boldsymbol{\eta}|\boldsymbol{z}) \propto p(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{\eta}) p(\boldsymbol{y}|\boldsymbol{\eta}) p(\boldsymbol{\eta})$.

The full conditional distribution of each component of the latent vector $\boldsymbol{y}$ is given by

$$\begin{aligned} p(y_i|\boldsymbol{y}_{(i)}, \boldsymbol{\eta}, \boldsymbol{z}) &= p(y_i|\boldsymbol{y}_{(i)}, \boldsymbol{\eta}, z_i) \\ &\propto N(\boldsymbol{\beta}' \underline{f}(\boldsymbol{s}_i) + \boldsymbol{v}'_{\vartheta,i} \Sigma^{-1}_{\vartheta,(i)} (\boldsymbol{y}_{(i)} - X_{(i)}\boldsymbol{\beta}), 1 - \boldsymbol{v}'_{\vartheta,i} \Sigma^{-1}_{\vartheta,(i)} \boldsymbol{v}_{\vartheta,i}) \mathbf{1}_{\{y_i \in A(z_i)\}}, \end{aligned} \tag{4}$$

where $\boldsymbol{y}_{(i)} = \{y_j, j \neq i\}$ so this is a normal distribution truncated to $A(z_i)$, which is generated using standard algorithms (Devroye, 1986; Geweke, 1991). The matrix $\Sigma_{\vartheta,(i)}$ is Σ_ϑ with the ith column and ith row removed, $\boldsymbol{v}'_{\vartheta,i}$ is the ith row of Σ_ϑ with its ith element removed, and the matrix $X_{(i)}$ is X with the ith row removed, $i = 1, \ldots, n$. Note that the matrices $\Sigma_{\vartheta,(i)}$ need not be directly inverted since all matrices $\Sigma^{-1}_{\vartheta,(i)}$ can be computed at once from Σ^{-1}_ϑ (Christensen et al., 1992).

The full conditional distribution of $\boldsymbol{\beta}$ is given by

$$p(\boldsymbol{\beta}|\vartheta, \boldsymbol{y}, \boldsymbol{z}) = p(\boldsymbol{\beta}|\vartheta, \boldsymbol{y}) = \mathrm{N_p}(\tilde{\boldsymbol{m}}_0, \tilde{V}_0^{-1}), \tag{5}$$

where $\tilde{\boldsymbol{m}}_0 = \tilde{V}_0^{-1}(V_0\boldsymbol{m}_0 + X'\Sigma^{-1}_\vartheta \boldsymbol{y})$ and $\tilde{V}_0 = V_0 + X'\Sigma^{-1}_\vartheta X$, making this full conditional distribution straightforward to simulate.

Lastly, the full conditional distribution of ϑ is given by

$$\begin{aligned} p(\vartheta|\boldsymbol{\beta}, \boldsymbol{y}, \boldsymbol{z}) &= p(\vartheta|\boldsymbol{\beta}, \boldsymbol{y}) \\ &\propto |\Sigma_\vartheta|^{-1/2} \exp\{-\tfrac{1}{2}(\boldsymbol{y} - X\boldsymbol{\beta})' \Sigma^{-1}_\vartheta (\boldsymbol{y} - X\boldsymbol{\beta})\} p(\vartheta), \end{aligned} \tag{6}$$

which is of non-standard form and difficult to directly sample from, so we perform a Metropolis-Hastings step. Given current values $\boldsymbol{\beta}, \vartheta$ and $\boldsymbol{y}$, a candidate value ϑ' for the next iteration is generated from a proposal distribution $q(.|\vartheta)$, and this candidate is accepted as the next iterate with probability

$$\alpha(\vartheta, \vartheta') = \min\left\{ \frac{p(\vartheta'|\boldsymbol{\beta}, \boldsymbol{y}) q(\vartheta|\vartheta')}{p(\vartheta|\boldsymbol{\beta}, \boldsymbol{y}) q(\vartheta'|\vartheta)}, 1 \right\}. \tag{7}$$

If not accepted, the next iterate is equal to ϑ. The choice of proposal distribution depends on the selected correlation function $K_\vartheta(\boldsymbol{s},\boldsymbol{u})$, so it is problem specific. A particular proposal distribution is given in Example 1.

By successively sampling from (4), $i=1,\ldots,n$, (5), and the Metropolis-Hastings step described above, and discarding an initial section of the generated sequence of say b iterations, we obtain a sample $\{(\boldsymbol{y}^{(t)},\boldsymbol{\eta}^{(t)})',\ t=b+1,\ldots,m\}$ (approximately) distributed as $p(\boldsymbol{y},\boldsymbol{\eta}|\boldsymbol{z})$. From this sample we can make inference about the model parameters $\boldsymbol{\eta}$ as well as inference about the binary map, as described in the next section.

3.2. Prediction of $\boldsymbol{Z}_0$

The problem of predicting $\boldsymbol{Z}_0$ is formulated as a decision problem, following the terminology given in Berger (1985), and based on the *Bayesian predictive distribution* $p(\boldsymbol{z}_0|\boldsymbol{z})$. Let $\tilde{\boldsymbol{Z}}_0=(\tilde{Z}_{01},\ldots,\tilde{Z}_{0k})'$, $\tilde{Z}_{0j}=\tilde{Z}(\boldsymbol{s}_{0j})$, be an arbitrary predictor of $\boldsymbol{Z}_0$. We restrict attention to additive loss functions

$$L(\boldsymbol{Z}_0,\tilde{\boldsymbol{Z}}_0)=\frac{1}{k}\sum_{j=1}^{k}[l_0\mathbf{1}_{\{\tilde{Z}_{0j}=1,Z_{0j}=0\}}+l_1\mathbf{1}_{\{\tilde{Z}_{0j}=0,Z_{0j}=1\}}], \tag{8}$$

where $l_i>0$ is the loss of mis-predicting $Z(\boldsymbol{s}_{0j})$ by $1-i$, $i=0,1$ (l_i might be location-dependent); a correct prediction incurs in no loss. The *optimal Bayes* predictor of $\boldsymbol{Z}_0$, defined as the predictor that minimizes the *Bayesian expected loss*, is given by

$$\hat{Z}_{0j}=\begin{cases}1 \text{ if } P\{Z_{0j}=1|\boldsymbol{z}\}>\frac{l_0}{l_0+l_1}\\ 0 \text{ otherwise}\end{cases},\quad j=1,\ldots,k. \tag{9}$$

This predictor was suggested by Solow (1986) who proposed estimating $P\{Z_{0j}=1|\boldsymbol{z}\}$ using indicator kriging (see Section 4), assuming the mean and covariance functions of the binary random field were known. In the special case when $l_0=l_1=1$, (8) becomes the mis-prediction rate (i.e., $(1/k)\sum_{j=1}^{k}\mathbf{1}_{\{\tilde{Z}_{0j}\neq Z_{0j}\}}$) and $\hat{Z}_{0j}=\mathbf{1}_{\{P(Z_{0j}=1|\boldsymbol{z})>0.5\}}$.

The Bayesian expected loss of the optimal Bayes predictor (9) is given by

$$E\{L(\boldsymbol{Z}_0,\hat{\boldsymbol{Z}}_0)|\boldsymbol{z}\}=\frac{1}{k}\sum_{j=1}^{k}l_{1-\hat{Z}_{0j}}|\hat{Z}_{0j}-P\{Z_{0j}=1|\boldsymbol{z}\}|, \tag{10}$$

which is a *global* measure of uncertainty about the estimated binary map providing information about its quality.

To compute (9) note that from (1), $P\{Z_{0j}=1|\boldsymbol{z}\}=P\{Y_{0j}>0|\boldsymbol{z}\}$, $j=1,\ldots,k$, so these conditional probabilities can be computed using a sample from $p(\boldsymbol{y}_0|\boldsymbol{z})$. For each posterior draw $(\boldsymbol{y}^{(t)},\boldsymbol{\eta}^{(t)})'$, a draw $\boldsymbol{y}_0^{(t)}$ from $p(\boldsymbol{y}_0|\boldsymbol{y}^{(t)},\boldsymbol{\eta}^{(t)})$ is readily generated since $p(\boldsymbol{y}_0|\boldsymbol{y},\boldsymbol{\eta})=N_k(X_0\boldsymbol{\beta}+B_\vartheta\Sigma_\vartheta^{-1}(\boldsymbol{y}-X\boldsymbol{\beta}),E_\vartheta-B_\vartheta\Sigma_\vartheta^{-1}B_\vartheta')$, where X_0 is a known $k\times p$ matrix, $X_{0,ij}=f_j(\boldsymbol{s}_{0i})$, and E_ϑ, B_ϑ are respectively $k\times k$ and $k\times n$ correlation matrices: $E_{\vartheta,ij}=K_\vartheta(\boldsymbol{s}_{0i},\boldsymbol{s}_{0j})$, $B_{\vartheta,ij}=K_\vartheta(\boldsymbol{s}_{0i},\boldsymbol{s}_j)$. The sequence $\{\boldsymbol{y}_0^{(t)};\ t=b+1,\ldots,m\}$ is a

sample (approximately) from $p(\boldsymbol{y}_0|\boldsymbol{z})$. Then, the Bayesian estimates of the conditional probabilities are given by

$$\hat{P}^B\{Z_{0j}=1|\boldsymbol{z}\} \approx \frac{1}{m-b}\sum_{t=b+1}^{m} \mathbf{1}_{\{y_{0j}^{(t)}>0\}}, \quad j=1,\ldots,k. \tag{11}$$

A measure of *local* prediction uncertainty is given by

$$\sigma(\boldsymbol{s}_{0j}) = l_{1-\hat{Z}_{0j}}|\hat{Z}_{0j} - P\{Z_{0j}=1|\boldsymbol{z}\}|, \quad j=1,\ldots,k. \tag{12}$$

For instance, if $l_0 = l_1 = 1$, then $\sigma(\boldsymbol{s}_{0j}) \in [0, \frac{1}{2}]$. Values close to zero correspond to locations with small predictive uncertainty, while values close to a half correspond to locations with large predictive uncertainty. Complete inference about the binary map consists of the estimated binary map obtained using (9) and (11) in place of the conditional probabilities, and the map of local prediction uncertainty (12).

Estimating the binary map using the previously described *Bayesian approach* is computationally very demanding in the cases of large datasets (n large) or large maps (k large). A way to reduce the computational burden when k is large is to use the following *plug-in* approach: From the sample of $p(\boldsymbol{y}, \boldsymbol{\eta}|\boldsymbol{z})$, estimates for $\boldsymbol{y}$ and $\boldsymbol{\eta}$ are first obtained, say $\hat{\boldsymbol{y}}$ and $\hat{\boldsymbol{\eta}} = (\hat{\boldsymbol{\beta}}, \hat{\vartheta})'$. Treating these estimates as known, we have that the plug-in estimates of the conditional probabilities are given by

$$\begin{aligned}\hat{P}^P\{Z_{0j}=1|\boldsymbol{z}\} &= P\{Y_{0j}>0|\hat{\boldsymbol{y}}, \hat{\boldsymbol{\eta}}\} \\ &= \Phi\left(\frac{\boldsymbol{x}_{0,j}'\hat{\boldsymbol{\beta}} + \boldsymbol{b}_{\hat{\vartheta},j}'\Sigma_{\hat{\vartheta}}^{-1}(\hat{\boldsymbol{y}} - X\hat{\boldsymbol{\beta}})}{(1 - \boldsymbol{b}_{\hat{\vartheta},j}'\Sigma_{\hat{\vartheta}}^{-1}\boldsymbol{b}_{\hat{\vartheta},j})^{0.5}}\right), \quad j=1,\ldots,k,\end{aligned} \tag{13}$$

where $\boldsymbol{x}_{0,j}'$ is the jth row of X_0, $\boldsymbol{b}_{\vartheta,j}'$ is the jth row of B_ϑ, and $\Phi(.)$ is the cdf of the standard normal distribution; the binary map is estimated using (9) with (13) in place of the conditional probabilities. This approach ignores the uncertainty about the model parameters and latent data in the estimation of $P\{Z_{0j}=1|\boldsymbol{z}\}$, which may or may not produce accurate estimates.

4. Indicator kriging

Let $\{Z(\boldsymbol{s}),\ \boldsymbol{s}\in D\}$ be a stationary binary random field, with $P\{Z(\boldsymbol{s})=1\}=\omega$, and isotropic semivariogram function $\gamma(l) = \frac{1}{2}\mathrm{var}\{Z(\boldsymbol{s}) - Z(\boldsymbol{u})\}$, $l = ||\boldsymbol{s}-\boldsymbol{u}||$. Indicator kriging is a simple method proposed by Solow (1986) for prediction in binary random fields. It consists of estimating $P\{Z(\boldsymbol{s}_0)=1|\boldsymbol{z}\}$ by its best linear unbiased estimator based on $\boldsymbol{z}$, which is given by

$$\hat{P}^{\mathrm{K}}\{Z(\boldsymbol{s}_0)=1|\boldsymbol{z}\} = \left(\boldsymbol{\gamma} + \frac{(1 - \mathbf{1}'\boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma})}{\mathbf{1}'\boldsymbol{\Gamma}^{-1}\mathbf{1}}\mathbf{1}\right)'\boldsymbol{\Gamma}^{-1}\boldsymbol{z}, \tag{14}$$

where $\boldsymbol{\Gamma}_{ij} = \gamma(||\boldsymbol{s}_i - \boldsymbol{s}_j||)$, $\boldsymbol{\gamma}_i = \gamma(||\boldsymbol{s}_0 - \boldsymbol{s}_i||)$, and $\mathbf{1} = (1,\ldots,1)'$; $i,j = 1,\ldots,n$ (Cressie, 1993, p. 281); the binary map is estimated using (9) with (14) in place of the conditional probabilities (see Goovaerts et al. (1997) for an application of indicator kriging). Besides ignoring the uncertainty about correlation parameters, (14) is, in general, an

inferior estimator for the conditional probabilities than (11), as the former is optimal only among the class of linear unbiased estimators, while the latter is optimal among the class of all estimators.

The indicator kriging approach has some caveats for its implementation. First, there is no guarantee that $\hat{P}^{\mathrm{K}}\{Z(\boldsymbol{s}_0)=1|\boldsymbol{z}\}$ lie in $[0,1]$. Practitioners argue that inadmissible estimates occur rarely, in which case estimates are set to the closest admissible value, 0 or 1. Second, unlike Gaussian random fields, not every conditionally negative definite function $\gamma(.)$ (Cressie, 1993, p. 60) is compatible with some binary random field. As shown by Matheron (1989), for some models $\gamma(.)$ (e.g. the squared exponential model), there exist no binary random field having $\gamma(.)$ as their semivariogram, and for other models $\gamma(.)$ of common use in practice (e.g. the spherical model), it is not known whether or not they are compatible with some binary random field. Papritz and Moyeed (1997) note that "it appears that most indicator kriging applications rely on a flawed probabilistic model". The effects of using such flawed model on inference about the binary map are unclear.

A way to avoid this potential incompatibility is to work with the semivariogram function of a clipped Gaussian random field, with $p=1$. In this case, $\omega=\Phi(\beta)$, and using the formula in Kedem (1980, p. 35), the correlation function of $Z(.)$ is given by

$$R_{\boldsymbol{\eta}}(l)=\frac{\frac{1}{2\pi}\int_0^{K_\vartheta(l)}\frac{\exp\{-\frac{\beta^2}{1+t}\}}{(1-t^2)^{0.5}}\,\mathrm{d}t}{\omega(1-\omega)}. \tag{15}$$

Hence, $\gamma_{\boldsymbol{\eta}}(l)=\omega(1-\omega)(1-R_{\boldsymbol{\eta}}(l))$ is by construction compatible with a binary random field, and choosing $\gamma(.)$ translates into choosing a correlation function $K(.)$ for the random field $Y(.)$. The computation of $R_{\boldsymbol{\eta}}(l)$ has to be done in general by numerical quadrature, but note that $R_{\boldsymbol{\eta}}(l)$ is insensitive to values of β close to zero, say $\beta\in[0,0.6]$, in which case $R_{\boldsymbol{\eta}}(l)\approx R_{(0,\vartheta)}(l)=(2/\pi)sin^{-1}(K_\vartheta(l))$ (Switzer, 1977).

5. Examples

We now illustrate the proposed methodology using two binary maps of the region $D=[1,20]\times[1,20]\subset R^2$, where D was discretized in a 20×20 regular lattice. Both binary maps were obtained by clipping, at level $\zeta_0=0$, a realization of a Gaussian random field with mean $\beta=0.5$ (so $\omega=0.691$), variance $\sigma_0^2=1$, but with different isotropic correlation functions. We also compare the prediction performance of the Bayesian, plug-in, and indicator kriging approaches using two sampling schemes.

5.1. Example 1

Fig. 1(a) displays a realization of a clipped Gaussian random field with $K(l)=0.8^l$. This binary map will be estimated at $k=364$ 'unobserved' locations of the regular lattice, using the regular sampling scheme data shown in Fig. 2(a) where $n=36$; 0 represents a white pixel and 1 a black pixel. We use the sampling model described

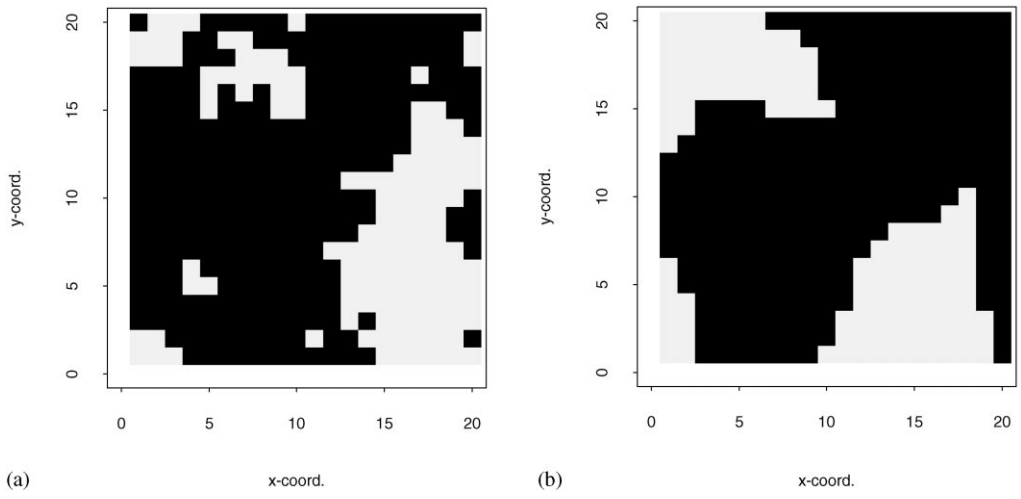

Fig. 1. Realizations on $D = [1,20] \times [1,20]$ of two clipped Gaussian random fields with mean 0.5, variance 1, and correlation function (a) $K(l) = 0.8^l$, (b) $K(l) = 0.92^{l^{1.9}}$.

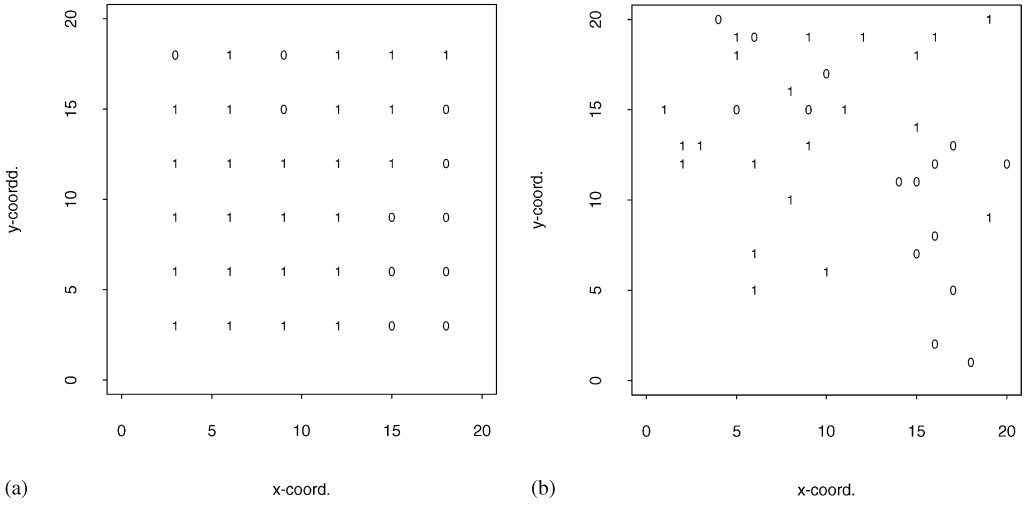

Fig. 2. Datasets: (a) regular sampling scheme, (b) irregular sampling scheme.

in Section 2, with constant mean β and isotropic correlation function $K_\vartheta(l) = \vartheta^l$. The prior distribution for $\boldsymbol{\eta} = (\beta, \vartheta)' \in R \times (0,1)$ is as in (3), with $m_0 = 0$, $t_0 = 0.05$, and $\vartheta \sim \text{unif}(0,1)$.

First, we use the algorithm described in Section 3.1 to obtain a sample from the posterior distribution of $(\boldsymbol{y}, \boldsymbol{\eta})'$. To perform the random walk Metropolis–Hastings step, we transform the correlation parameter ϑ to $\xi = \text{logit}(\vartheta)$, generate ξ' from the proposal density $q_\xi(.|\xi) = N(\xi, \psi^2)$, and accept the value $\vartheta' = \text{logit}^{-1}(\xi')$ with probability (7), where from the change of variable formula we have

$$\frac{q(\vartheta|\vartheta')}{q(\vartheta'|\vartheta)} = \frac{\vartheta'(1-\vartheta')}{\vartheta(1-\vartheta)}.$$

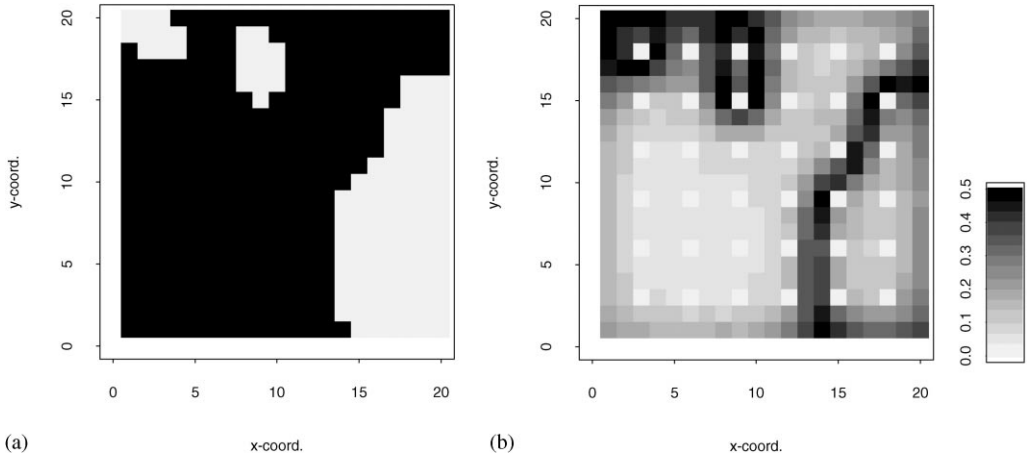

Fig. 3. Inference about the binary map in Fig. 1(a) based on the regular sampling scheme dataset: (a) estimated binary map, (b) map of local prediction uncertainty.

The tuning parameter ψ^2 was chosen following the rule-of-thumb used by many practitioners of selecting, by experimentation, a value producing an empirical rate of acceptance in the range 0.4-0.5 (see Gelman et al. (1996) for a theoretical justification). We used $\psi^2 = 0.64$, for which the empirical rate of acceptance was 0.46.

To diagnose convergence, we generated three runs of the algorithm, each of $m = 3000$ iterations, with different sets of starting values. The first $b = 1000$ iterations were discarded, and the subsequent 2000 iterations were used for graphical and statistical analyses. Trace plots (not shown) of β, ϑ, and each latent datum y_i appeared (visually) stationary for each run, and displayed similar behaviors: the 5%, 50% and 95% sample quantiles of each parameter and latent datum were close across runs. All these suggest convergence of the algorithm. The posterior medians of ω and ϑ are respectively 0.636 and 0.787, while their 95% credible intervals are $(0.219, 0.887)$ and $(0.331, 0.941)$.

Second, we generated a sample from $p(\boldsymbol{y}_0 \mid \boldsymbol{z})$ as described in Section 3.2, and computed the optimal predictor (9) with estimated conditional probabilities (11), where $l_0 = l_1 = 1$. Fig. 3(a) shows the estimated binary map, having a mis-prediction rate $MPR = \frac{1}{364}\sum_{j=1}^{364} \mathbf{1}_{\{\hat{Z}_{0j} \neq Z_{0j}\}} = 0.159$. Note that, as for continuous data, the estimated binary map is smoother than the true binary map. Fig. 3(b) shows the map of local prediction uncertainty (12). This map identifies subregions with high prediction uncertainty, corresponding to 'boundaries' where prediction is expected to be difficult. The *global* measure of uncertainty (10) is 0.205, which is a useful quantity for assessing the quality of the estimated binary map.

To illustrate the above point, we repeated the previous analysis to estimate the binary map in Fig. 1(a), but now based on the irregular sampling scheme data displayed in Fig. 2(b) (same sample size); the locations were chosen at random. The estimated binary map is shown in Fig. 4(a), having a $MPR = 0.187$. It is clear that this estimated binary map is of poorer quality than the one in Fig. 3(a), but note that the *MPR* is not available from the data. On the other hand, the deterioration

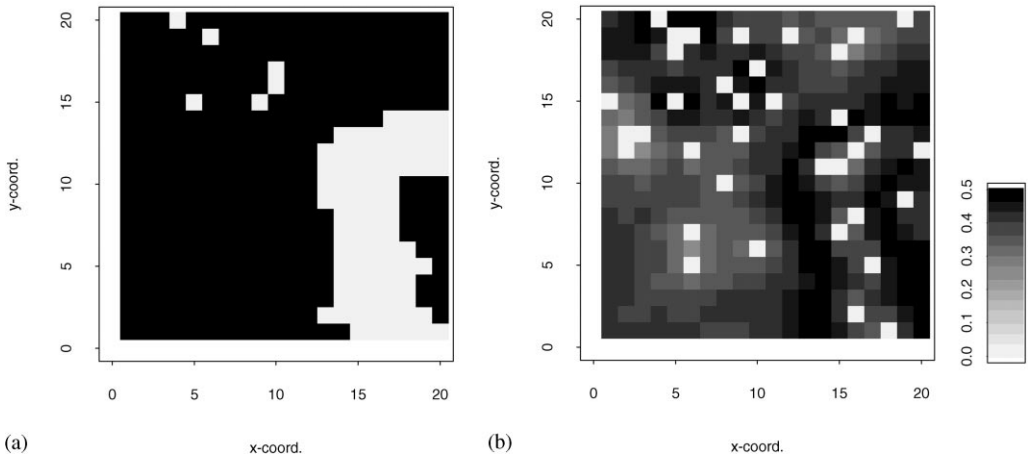

Fig. 4. Inference about the binary map in Fig. 1(a) based on the irregular sampling scheme dataset: (a) estimated binary map, (b) map of local prediction uncertainty.

in the quality of the estimated binary map based on the irregular sampling scheme data is clearly seen in the map of local prediction uncertainty (12) displayed in Fig. 4(b). This map has almost everywhere larger values than the one in Fig. 3(b), and the global measure of uncertainty (10) is 0.406, about double of the one based on the regular sampling scheme data.

We applied the plug-in approach described at the end of Section 3.2, where each component of $(\boldsymbol{y}, \boldsymbol{\eta})'$ was estimated by the sample median of the corresponding component in the MCMC output. For the regular sampling scheme data, the estimated binary map has a $MPR = 0.157$, and differs from the map in Fig. 3(a) in just 5 locations. The mean absolute difference between the maps of prediction uncertainty for the Bayesian and plug-in approaches is 0.017, with a maximum absolute difference of 0.058. On the other hand, for the irregular sampling scheme data, the estimated binary map has a $MPR = 0.217$, and differs from the map in Fig. 4(a) in 31 locations. The mean absolute difference between the maps of prediction uncertainty for the Bayesian and plug-in approaches is 0.024, with a maximum absolute difference of 0.098. Hence, the resulting inferences using the plug-in approach were close to the ones using the Bayesian approach for the regular sampling scheme data, but not so for the irregular sampling scheme data.

We also applied the indicator kriging approach described in Section 4, using the semivariogram function obtained from (15) with $K_\vartheta(l) = \vartheta^l$. The model parameters were estimated by weighted-least squares (WLS), as is commonly done for continuous geostatistical data (Cressie, 1993). For the regular sampling scheme data, the estimates were $\hat{\beta} = 0$ and $\hat{\vartheta} = 0.867$; Fig. 5(a) shows the empirical and fitted semivariograms. The poor estimate $\hat{\beta}$ is due to the insensitivity of $R_{\boldsymbol{\eta}}(l)$ to values of β close to zero, which makes WLS inappropriate for the estimation of β. Fig. 6(a) shows the estimated binary map obtained by indicator kriging, having a $MPR = 0.162$, and differing from the one in Fig. 3(a) in 9 locations (in 18 locations $\hat{P}^{\mathrm{K}}\{Z(\boldsymbol{s}_0) = 1 \mid \boldsymbol{z}\}$ lied outside $[0, 1]$, so these were set to 0 or 1). For the irregular sampling scheme

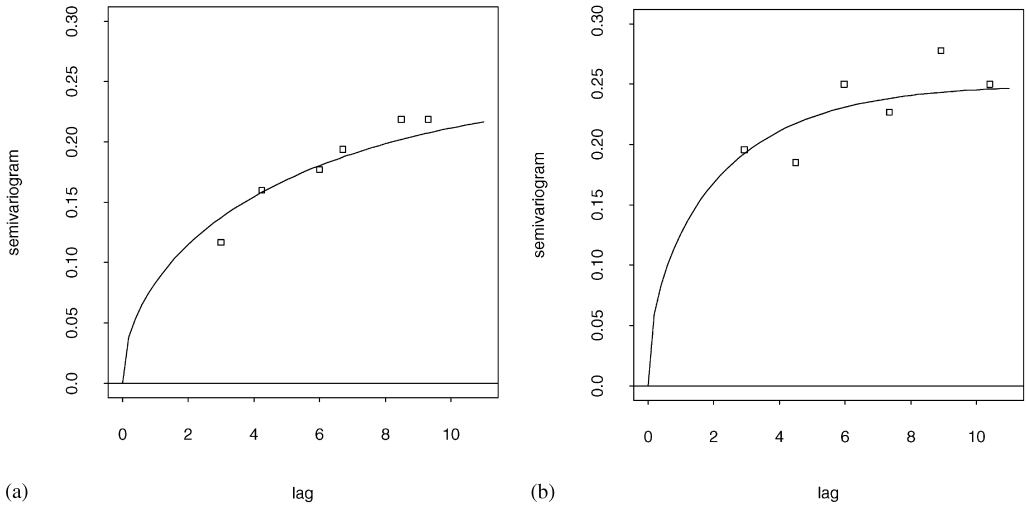

Fig. 5. Empirical (□) and fitted (—) semivariograms of the clipped Gaussian random field in Example 1: (a) from the regular sampling scheme dataset, (b) from the irregular sampling scheme dataset.

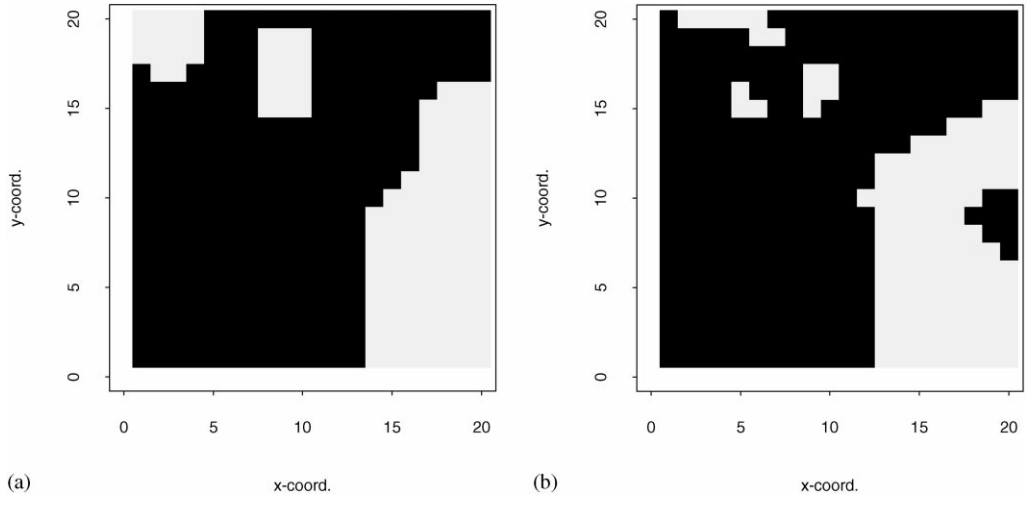

Fig. 6. Estimation of the binary map in Fig. 1(a) using indicator kriging: (a) from the regular sampling scheme dataset, (b) from the irregular sampling scheme dataset.

data, the estimates were $\hat{\beta} = 0$ and $\hat{\vartheta} = 0.7021$; Fig. 5(b) shows the empirical and fitted semivariograms. Fig. 6(b) shows the estimated binary map obtained by indicator kriging, having a $MPR = 0.151$, and differing from the one in Fig. 4(a) in 33 locations (in all locations $\hat{P}^{\mathrm{K}}\{Z(\boldsymbol{s}_0) = 1|\boldsymbol{z}\}$ lied inside $[0, 1]$). Hence, the resulting inference is slightly inferior to the ones obtained using either the Bayesian or plug-in approaches for the regular sampling scheme data, but the opposite holds for the irregular sampling scheme data.

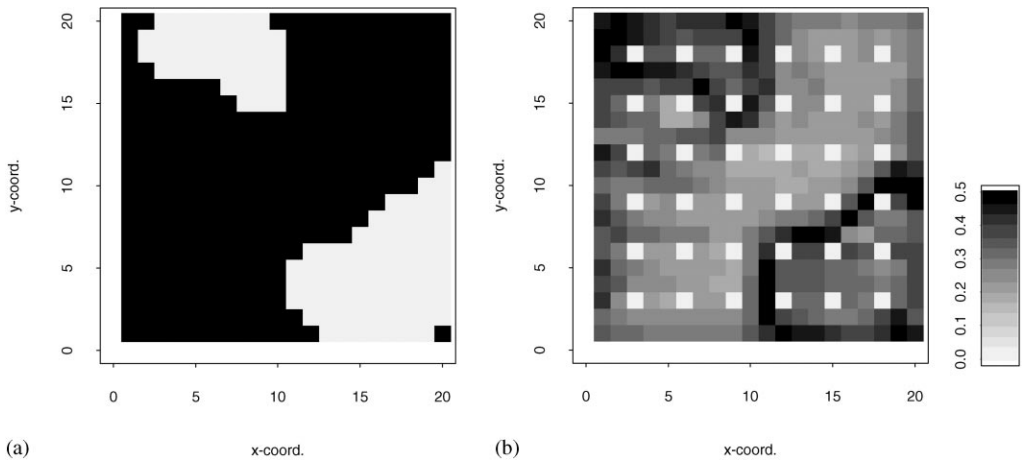

Fig. 7. Inference about the binary map in Fig. 1(b) based on the regular sampling scheme dataset: (a) estimated binary map, (b) map of local prediction uncertainty.

5.2. Example 2

Fig. 1(b) displays a realization of a clipped Gaussian random field with $K(l) = 0.92^{l^{1.9}}$. This binary map has proportions of 'black' and 'white' similar to those in Fig. 1(a), but is smoother than the latter; the value $\vartheta = 0.92$ was chosen to make the correlation at distance $l = 3$ (the minimum distance between two locations) the same as in the model used in Example 1. This binary map will also be estimated at $k = 364$ locations of the regular lattice using the regular sampling scheme data. We use the sampling model and prior distribution for $\boldsymbol{\eta} = (\beta, \vartheta)' \in R \times (0,1)$ as in Example 1, but now with isotropic correlation function $K_\vartheta(l) = \vartheta^{l^{1.9}}$.

Repeating the MCMC algorithm and diagnostics described in Example 1, we obtained a sample from $p(\boldsymbol{y}, \boldsymbol{\eta}|\boldsymbol{z})$. The posterior medians of ω and ϑ are respectively 0.679 and 0.899, while their 95% credible intervals are $(0.399, 0.870)$ and $(0.363, 0.957)$. Again, we generated a sample from $p(\boldsymbol{y}_0 \mid \boldsymbol{z})$, and computed the optimal predictor (9) using estimated conditional probabilities (11). Fig. 7(a) shows the estimated binary map, having a $MPR = 0.175$, and Fig. 7(b) shows the map of local prediction uncertainty (12), where the global measure of uncertainty (10) is 0.312. The inference about this binary map is not as good as the one about the binary map in Fig. 1(a).

5.3. Discussion

The clipped Gaussian random field showed good prediction performance. The main findings from the empirical comparisons in the above examples are:

(a) None of the three approaches had uniformly superior predictive performance. These approaches use, in essence, the same prediction rule (9), but differ on the way the conditional probabilities $P\{Z(\boldsymbol{s}_0) = 1|\boldsymbol{z}\}$ are estimated. Even though the Bayesian estimator of these conditional probabilities is expected to be better than the plug-in

and indicator kriging estimators, this does not imply the binary map obtained by the Bayesian approach will have better predictive performance than either the plug-in or indicator kriging approaches. This somewhat counterintuitive observation might be explained by the results in Friedman (1997), for the closely related problem of classification of binary outcomes. He showed that the bias-variance trade-off for classification error is very different than the one corresponding to estimation error of the conditional probabilities, in a way that, more accurate conditional probability estimates do not necessarily lead to better classification performance, and sometimes can make it worse. This certainly requires further research.

(b) The inference about the binary map depends heavily on the correlation structure of the underlying Gaussian random field, an observation also reported by Heagerty and Lele (1998). This makes the selection of the covariance function for this model a point of extreme importance, perhaps even more than in the case of continuous outcomes. The development of model selection criteria amenable for binary random fields will be explored elsewhere.

(c) The inference about the binary map is very dependent on the sampling scheme that is used, which is also the case for continuous outcomes.

6. Conclusions

Most of the work on prediction in random fields based on geostatistical data has been intended for continuous variables, and the case of binary variables has barely been studied. Binary data seem to be more difficult to deal with than continuous data, because of the scarcity of flexible and adequate models for binary random fields.

The clipped Gaussian random field has been proposed by De Oliveira (1997) and Heagerty and Lele (1998) to model binary spatial data, which can be used in situations where high degree of smoothness is expected in the binary map. This work describes the Bayesian approach for inference and prediction in clipped Gaussian random fields, which is feasible by the use of data augmentation and Markov chain Monte Carlo methods. An attractive feature of the Bayesian approach is that it provides a unified framework for inference and prediction, and uncertainty about model parameters is naturally transferred to predictive inference. In particular, this approach provides a map of prediction uncertainty, and therefore a way to assess the quality of the estimated binary map. On the negative side, this approach is computationally quite demanding, making it impractical for use with large datasets.

The approach proposed by Heagerty and Lele (1998) for inference and prediction in clipped Gaussian random fields is computationally less demanding than the Bayesian approach, so more practical for use with large datasets. On the negative side, it does not provide a unified framework for inference and prediction: parameters are estimated by penalized composite likelihood, standard errors by re-sampling, and prediction is done by local iterative conditional modes. This blend makes difficult to quantify how uncertainty about model parameters is transferred to predictive inference, and the main justification for some of these procedures rely on asymptotic grounds. Also, Diggle et al. (1998) have proposed a class of generalized linear mixed

models for prediction in non-Gaussian random fields, which could also be used for prediction in binary random fields.

The clipped Gaussian random field can be generalized to model *categorical* random fields taking values on $\{0,1,\ldots,J-1\}$, $J \geq 2$. Let $\{Y(\boldsymbol{s}),\ \boldsymbol{s} \in D\}$ be an underlying Gaussian random field as in Section 2, and let $-\infty = \zeta_0 < \zeta_1 < \cdots < \zeta_{J-1} < \zeta_J = \infty$ be a set of $J+1$ ordered thresholds. A model for a categorical random field $\{Z(\boldsymbol{s}),\ \boldsymbol{s} \in D\}$ is given by

$$Z(\boldsymbol{s}) = \sum_{j=0}^{J-1} j \mathbf{1}_{\{\zeta_j < Y(s) \leq \zeta_{j+1}\}}, \quad \boldsymbol{s} \in D.$$

Most of the properties of the clipped Gaussian random field can be extended to this more general case. In particular, the optimal predictor of $Z(\boldsymbol{s}_0)$ (with respect to a loss function similar to (8)), is $\hat{Z}(\boldsymbol{s}_0) = \arg\max_{j \in \{0,1,\ldots,J-1\}} [P\{Z(\boldsymbol{s}_0) = j \mid \boldsymbol{z}\}]$. As before, these conditional probabilities can be computed using data augmentation, although the MCMC algorithm might become more complex because of the presence of $J-2$ unknown thresholds.

Acknowledgements

I thank Benjamin Kedem and José Miguel Pérez for reading a previous version of this manuscript, and two anonymous referees for providing useful suggestions leading to an improved paper. This work was partially supported by National Aeronautics and Space Administration grant NAG52783, National Science Foundation grant DMS-9208758 to National Institute of Statistical Sciences, and benefited from computer facilities of Centro de Estadística y Software Matemático (CESMa), Universidad Simón Bolívar.

References

Albert, J.H., Chib, S., 1993. Bayesian analysis of binary and polychotomous response data. J. Amer. Statist. Assoc. 88, 669–679.

Ashford, J.R., Sowden, R.R., 1970. Multivariate probit analysis. Biometrics 26, 535–546.

Berger, J.O., 1985. Statistical Decision Theory and Bayesian Analysis. Springer, New York.

Carlin, B.P., Polson, N.G., 1992. Monte Carlo bayesian methods for discrete regression models and categorical time series. Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics 4. Oxford University Press, Oxford, pp. 577–586.

Chib, S., Greenberg, E., 1998. Analysis of multivariate probit models. Biometrika 85, 347–361.

Chib, S., Greenberg, E., 1995. Understanding the Metropolis-Hastings algorithm. Amer. Statistician 49, 327–335.

Christensen, R., Johnson, W., Pearson, L.M., 1992. Prediction diagnostics for spatial linear models. Biometrika 79, 583–591.

Cressie, N.A.C., 1993. Statistics for Spatial Data, Revised Edition. Wiley, New York.

De Oliveira, V., 1997. Prediction in some classes of non-Gaussian random fields. Unpublished Ph.D. Thesis, University of Maryland-College Park.

Devroye, L., 1986. Non-uniform Random Variate Generation. Springer, New York.

Diggle, P.J., Tawn, J.A., Moyeed, R.A., 1998. Model-based geostatistics (with discussion). Appl. Statist. 47 (3), 299–326.

Finney, D.J., 1947. The estimation from individual records of the relationship between dose and quantal response. Biometrika 34, 320–334.

Friedman, J.H., 1997. On bias, variance, 0/1-loss, and the curse-of-dimensionality. Data Mining Knowledge Discovery 1, 1997.

Gelman, A., Roberts, G.O., Gilks, W.R., 1996. Efficient Metropolis jumping rules. Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics 5. Oxford University Press, Oxford, pp. 599–607.

Geweke, J., 1991. Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints. Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface. American Statistical Association, Alexandria, VA, pp. 571–578.

Gilks, W.R., Richardson, S., Spiegelhalter, D.J., 1996. Markov chain Monte Carlo in Practice. Chapman & Hall, London.

Goovaerts, P., Webster, R., Dubois, J.-P., 1997. Assessing the risk of soil contamination in the Swiss jura using indicator geostatistics. Environ. Ecolog. Statist. 4, 31–48.

Heagerty, P.J., Lele, S.R., 1998. A composite likelihood approach to binary spatial data. J. Amer. Statist. Assoc. 93, 1099–1111.

Journel, A.G., 1983. Nonparametric estimation of spatial distributions. Math. Geology 15, 445–468.

Kedem, B., 1980. Binary Time Series. Marcel Dekker, New York.

Matheron, G., 1989. The internal consistency of models in geostatistics. Armstrong, M. (Ed.), Geostatistics, Vol. 1. Kluwer Academic Publishers, Dordrecht, pp. 21–38.

Nott, D.J., Wilson, R.J., 1997. Parameter estimation for excursion set texture models. Signal Process. 63, 199–543.

Ochi, Y., Prentice, R.L., 1984. Likelihood inference in a correlated probit regression model. Biometrika 71, 531–543.

Papritz, A., Moyeed, R.A., 1997. Empirical validation of linear and nonlinear methods for spatial point prediction. Technical Report.

Solow, A.R., 1986. Mapping by simple indicator kriging. Math. Geology 18, 335–354.

Switzer, P., 1977. Estimation of spatial distributions from point sources with application to air pollution measurement. Bull. Internat. Statist. Inst. 47, 123–137.