2016

Two-scale spatial models for binary data

Cecile Hardouin
Universite Paris Ouest Nanterre La Defense

Noel Cressie
University of Wollongong

Two-scale spatial models for binary data

Abstract

A spatial lattice model for binary data is constructed from two spatial scales linked through conditional probabilities. A coarse grid of lattice locations is specified and all remaining locations (which we call the background) capture fine-scale spatial dependence. Binary data on the coarse grid are modelled with an autologistic distribution, conditional on the binary process on the back- ground. The background behaviour is captured through a hidden Gaussian process after a logit transformation on its Bernoulli success probabilities. The likelihood is then the product of the (conditional) autologistic probability distribution and the hidden Gaussian-Bernoulli process. The parameters of the new model come from both spatial scales. A series of simulations illustrates the spatial-dependence properties of the model and likelihood-based methods are used to estimate its parameters. Presence-absence data of corn borers in the roots of corn plants are used to illustrate how the model is fitted.

National Institute for Applied Statistics Research Australia

University of Wollongong, Australia

Working Paper

07-16

Two-Scale Spatial Models for Binary Data

Cécile Hardouin and Noel Cressie

National Institute for Applied Statistics Research Australia, University of Wollongong,
Wollongong NSW 2522, Australia Phone +61 2 4221 5435, Fax +61 2 4221 4998.
Email: karink@uow.edu.au

Two-Scale Spatial Models for Binary Data

Cécile Hardouin

MODAL'X, Université Paris Ouest Nanterre La Défense, France

Noel Cressie

NIASRA, University of Wollongong, Australia

Abstract

A spatial lattice model for binary data is constructed from two spatial scales linked through conditional probabilities. A coarse grid of lattice locations is specified and all remaining locations (which we call the background) capture fine-scale spatial dependence. Binary data on the coarse grid are modelled with an autologistic distribution, conditional on the binary process on the background. The background behaviour is captured through a hidden Gaussian process after a logit transformation on its Bernoulli success probabilities. The likelihood is then the product of the (conditional) autologistic probability distribution and the hidden Gaussian–Bernoulli process. The parameters of the new model come from both spatial scales. A series of simulations illustrates the spatial-dependence properties of the model and likelihood-based methods are used to estimate its parameters. Presence-absence data of corn borers in the roots of corn plants are used to illustrate how the model is fitted.

Keywords: auto-logistic model, EM algorithm, Gaussian process, hierarchical statistical model, Laplace approximation, spatial odds-ratio

1. Introduction

Binary spatial data are involved in various domains such as economics, social sciences, ecology, image analysis, and epidemiology; see below for references. Considering the spatial framework, one common model for regularly spaced binary data is the auto-logistic model, which belongs to Besag's auto-models class ([5]); it is a particular case of a Markov Random Field, analogous to a classical logistic model, except that the explanatory variables are replaced by neighbouring values of the process. The auto-logistic model has seen a lot of use in the last 40 years and in various contexts; see, for example, [4], [22] and [42] in ecology, [20] in epidemiology, [26] in image analysis, and [23], [37] in land-use. The auto-logistic model has been reparameterized recently by Caragea and Kaiser [8], which helps with interpretation of the parameters, and this was extended to the spatio-temporal case by Wang and Zhen [47].

In a hierarchical framework, when the data are noisy and missing, a Generalized Linear Model ([33], [34]) can be implemented with a link appropriate for binary data. The logit link is canonical and a natural choice for the hidden process is auto-Gaussian (SAR model). Models involving other link functions have been developed more recently; for example, in Marsh et al. [32] and

Lesage [28], a spatial probit model is applied to problems in agriculture and economics, and Roy et al. [40] introduce a Bayesian spatial robit model that is more robust against extreme observations. When the spatial variable represents presence/absence of a rare event, Elkink and Calabrese [18] suggest the quantile function of the Generalized Extreme Value (GEV) distribution as a link function.

In this paper, we focus on binary data on a spatial lattice with two spatial scales. For the sake of simplicity, we assume that the process is observed on a regular lattice, but the model can be extended to irregular lattices; then we specify a coarse regular grid of sites, say at resolution $\Delta > 1$, where the fine-scale lattice is at resolution 1. The locations on the coarse grid define what we call the *Grid*, and all remaining locations on the underlying lattice define what we call the *Background*.

The models on the Grid and the Background account respectively for large-scale and fine-scale spatial variation. We start with the Background model, which consists of a classical hierarchical logistic model, linked to a hidden Gaussian field ε; clearly, the local spatial dependence relies on the covariance structure of this hidden field. Then, conditional on the Background observations, we consider an auto-logistic model on the Grid, where the large-scale spatial dependence is expressed via the parameters of the auto-logistic model. Thus, the final model is non-stationary, which allows us to capture spatial dependence at different scales, and combines a geostatistical model with a Markov random field (MRF) model in a new and interesting manner.

Section 2 is devoted to the description of the model; we display its properties and behaviours by varying different values of the parameters in Section 3. The results of this section show how we can identify the Grid resolution Δ. We present parameter estimation in Section 4, where both geostatistical and MRF parameters are involved. Section 5 contains an application to modeling the occurrence of corn borer larvae in agricultural fields in Iowa, USA. Finally, in Section 6 we discuss further aspects of the construction of the model and give our conclusions, followed by a technical appendix.

2. Two-scale spatial modelling

2.1. The Background and the Grid

Let us consider a two-dimensional domain of spatial-process locations $D \subset \mathbb{R}^2$; D is a finite set of n sites with the notation $D = \{\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_n\}$, with $\mathbf{s}_i = (s_{i1}, s_{i2})$ for $i = 1, ..., n$, and we denote $P(D)$ as its perimeter (or boundary). We are thinking of D as having no holes, but the following definition for perimeter covers all cases:

$$P(D) = \{\mathbf{s} \in D\text{: no. nearest neighbours of } \mathbf{s} < 4\}.$$

For the sake of simplicity, we assume that D is a fine regular lattice, but the situation can be generalized to irregular lattices.

Let $\mathbf{Z} = (Z(\mathbf{s}) : \mathbf{s} \in D)^{\mathrm{T}}$ be the process on D, taking its values in the state space $E = \{0, 1\}^D$. We consider two scales of spatial dependence, which occur locally at fine-scale resolution 1, and at a coarse-scale resolution $\Delta > 1$. This distance Δ is assumed known; in practice, it may be obtained from a preliminary

exploratory analysis of the data, or by subject matter experts. We shall come back to this point in Section 5 but, in what follows, Δ is not considered to be a parameter of the model.

Let G be a Grid such that the nodes are equally spaced at distance Δ; in order to avoid edge effects, we position the Grid in the domain, such that the edges of the Grid are at least at distance Δ from the edge of D, see Figure 1.

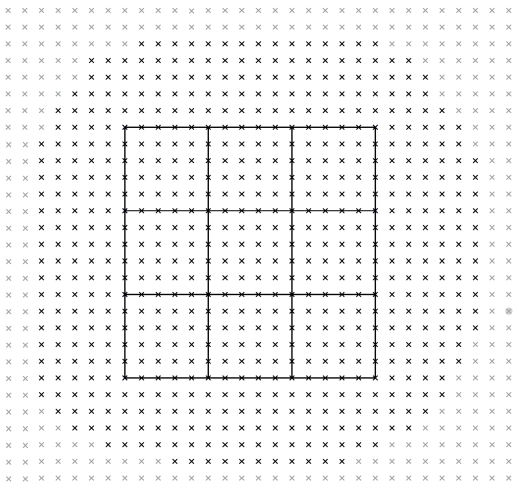

Figure 1: The Background (solid crosses) and the Grid (solid lines)

More precisely, define $D^0 \equiv \{\mathbf{s} :\| \mathbf{s} - \mathbf{u} \| \geq \Delta \ ; \ \mathbf{u} \in P(D)\} \cap D$; then we define the Grid

$$G(\Delta) = \{(k\Delta, l\Delta) : \ k = ..., -1, 0, 1, ..., \ l = ..., -1, 0, 1, ...\} \cap D^0.$$

All remaining locations outside the Grid define the so-called Background,

$$B(\Delta) = D \setminus G(\Delta).$$

For the sake of simplicity, we replace $B(\Delta)$ and $G(\Delta)$ with B and G, respectively, in all that follows. Thus, the fine-scale variation happens at the scale of the Background, while the large-scale variation happens at the scale of the Grid. Moreover, if $n_A = |A|$ is the cardinality of set A, we have $n = n_D = n_G + n_B$.

With these notations established, we write $\mathbf{Z} = (\mathbf{Z}_G^{\mathrm{T}}, \mathbf{Z}_B^{\mathrm{T}})$, where $\mathbf{Z}_G = (Z_G(\mathbf{s}) : \ \mathbf{s} \in G)^{\mathrm{T}}$, $\mathbf{Z}_B = (Z_B(\mathbf{s}) : \ \mathbf{s} \in B)^{\mathrm{T}}$, and $\mathbf{Z}_A$ denotes a spatial process on a set $A \subset D$, such that $\mathbf{Z}_A = (Z_A(\mathbf{s}) : \ \mathbf{s} \in A)^{\mathrm{T}}$.

We now turn to spatial-process modelling. We start with the Background in Section 2.2, which involves a conditional logistic model for $\mathbf{Z}_B$; then, conditional on $\mathbf{Z}_B$, we define $\mathbf{Z}_G$ on the Grid in Section 2.3.

2.2. Fine-scale process on the Background

We consider a conditional model for binary spatial data on the Background. We model the binary variables using a Bernoulli distribution, where the mean depends on an underlying (and unobserved) spatial process ε. Moreover, we assume conditional independence of the Bernoulli random variables given the hidden process.

Thus, denoting the Background locations as $\{\mathbf{s}_i : \ i = 1, ..., n_B\}$, for each $\mathbf{s}_i \in B$, we write the following independent conditional distributions for $Z_B(\mathbf{s}_i)$ given $\varepsilon = (\varepsilon(\mathbf{s}_1), ..., \varepsilon(s_n))^{\mathrm{T}}$ as those given by Bernoulli random variables,

$$Z_B(\mathbf{s}_i) \mid \varepsilon(\mathbf{s}_i) \sim Ber(p(\mathbf{s}_i)), \tag{1}$$

where

$$p(\mathbf{s}_i) = \frac{e^{\varepsilon(\mathbf{s}_i)}}{1 + e^{\varepsilon(\mathbf{s}_i)}}.$$

The hidden process ε is assumed to be Gaussian with mean $\mathbf{0}$ and spatial covariance matrix Σ. It is possible to incorporate explanatory variables in the mean but we choose not to do so initially. That is,

$$\varepsilon \sim N_n(\mathbf{0}, \Sigma). \tag{2}$$

2.3. Coarse-scale process on the Grid

We define the model on the Grid conditional on the Background using a Markov Random Field model with a neighborhood graph on the Grid, which recall has resolution Δ. For the sake of simplicity we consider here the four nearest neighbours, but the model can be modified easily to account for extra spatial dependence.

For each site $\mathbf{s} \in G$, we define the four-nearest-neighbourhood set $N_G(\mathbf{s}) = \{\mathbf{u} \in G : \ \mathbf{u} = \mathbf{s} \pm (\Delta, 0), \mathbf{s} \pm (0, \Delta)\}$. Our conditional model for the Grid values is:

$$\pi_{\mathbf{s}}(Z_G(\mathbf{s}) \mid \mathbf{Z}_B, \mathbf{Z}_{N_G(\mathbf{s})}) = \frac{\exp\left\{\alpha_B(\mathbf{s})Z_G(\mathbf{s}) + \frac{\beta}{4}\sum\limits_{\mathbf{u}\in N_G(\mathbf{s})} Z_G(\mathbf{s})Z_G(\mathbf{u})\right\}}{1 + \exp\left\{\frac{\beta}{4}\sum\limits_{\mathbf{u}\in N_G(\mathbf{s})} Z_G(\mathbf{s})Z_G(\mathbf{u})\right\}}, \tag{3}$$

where dependence on $\mathbf{Z}_B$ is captured in $\alpha_B(\mathbf{s})$; see below.

From [5], we know that these conditional distributions are compatible and the joint distribution is,

$$\pi(\mathbf{Z}_G \mid \mathbf{Z}_B) = C^{-1} \exp U(\mathbf{Z}_G; \mathbf{Z}_B) \ , \tag{4}$$

where the energy U is given by

$$U(\mathbf{Z}_G; \mathbf{Z}_B) = \sum_{\mathbf{s}\in G} \left\{ \alpha_B(\mathbf{s})Z_G(\mathbf{s}) + \frac{\beta}{4}\sum_{\mathbf{u}\in N_G(\mathbf{s})} Z_G(\mathbf{s})Z_G(\mathbf{u}) \right\}, \tag{5}$$

and C is the normalization constant, $\sum_{\mathbf{Z}_G\in\{0,1\}^{n_G}} \exp U(\mathbf{Z}_G; \mathbf{Z}_B)$. Here, β is the spatial interaction parameter, which we assume to be constant over the

Grid; $\beta < 0$ implies competitive behaviour, while $\beta > 0$ implies co-operative behaviour, and $\beta = 0$ corresponds to spatial conditional independence.

In the next paragraphs, we emphasize the role of $\mathbf{Z}_B$ in (3) and (4). Now, $\alpha_B(\mathbf{s})$ accounts for the underlying behaviour of the binary process on the Background "around" $\mathbf{s}$. We propose the following model:

$$\alpha_B(\mathbf{s}) = \gamma + \alpha \times \frac{\sum_{\mathbf{u}\in N_B(\mathbf{s})} Z_B(\mathbf{u})}{|N_B(\mathbf{s})|} \ ,$$

where $N_B(\mathbf{s}) = \{\mathbf{u} = (u_1, u_2) \in B : \ |u_1 - s_1| \leq \Delta, \ |u_2 - s_2| \leq \Delta\}$ is the set of four-nearest-neighbour Background locations. Notice that $N_B(\mathbf{s}) \cap G = \emptyset$ for all $\mathbf{s} \in G$, since $B \cap G = \emptyset$.

To understand the role of the parameters, consider the following calibration. States 0 and 1 are equiprobable if $\alpha_B(\mathbf{s}) + \frac{\beta}{2} = 0$, whereas $\alpha_B(\mathbf{s}) + \frac{\beta}{2} > 0$ favors state 1 and $\alpha_B(\mathbf{s}) + \frac{\beta}{2} < 0$ favors state 0. Now, for equal numbers of 0 and 1 in $N_B(\mathbf{s})$, $\frac{1}{|N_B(\mathbf{s})|}\sum_{\mathbf{u}\in N_B(\mathbf{s})} Z_B(\mathbf{u}) = \frac{1}{2}$, in which case $\alpha_B(\mathbf{s}) + \frac{\beta}{2} = \gamma + \frac{\alpha}{2} + \frac{\beta}{2}$, and hence we will have equilibrium on the Grid if $\gamma = -\frac{\alpha+\beta}{2}$. In this case, $\gamma = -\frac{\alpha+\beta}{2}$, we can write again $\alpha_B(\mathbf{s}) + \frac{\beta}{2} = \alpha(\frac{1}{|N_B(\mathbf{s})|}\sum_{\mathbf{u}\in N_B(\mathbf{s})} Z_B(\mathbf{u}) - \frac{1}{2})$; then the model depends only on α. If $\alpha > 0$ and if we have a predominance of 1s in $N_B(\mathbf{s})$, that is $\frac{1}{2} < \frac{\sum_{\mathbf{u}\in N_B(\mathbf{s})} Z_B(\mathbf{u})}{|N_B(\mathbf{s})|} < 1$, we obtain $\alpha_B(\mathbf{s}) + \frac{\beta}{2} > 0$, which reinforces state 1 on the Grid. The opposite happens if $\alpha > 0$ and $0 < \frac{\sum_{\mathbf{u}\in N_B(\mathbf{s})} Z_B(\mathbf{u})}{|N_B(\mathbf{s})|} < \frac{1}{2}$.

The model we ultimately choose has equilibrium on the Grid (i.e., $\gamma = -\frac{\alpha+\beta}{2}$), and hence we can re-write (4) as:

$$\pi(\mathbf{Z}_G \mid \mathbf{Z}_B) = C^{-1} \prod_{\mathbf{s}\in G} \exp Z(\mathbf{s}) \left\{ \alpha \left(\frac{1}{|N_B(\mathbf{s})|} \sum_{\mathbf{u}\in N_B(\mathbf{s})} Z_B(\mathbf{u}) - \frac{1}{2} \right) \right.$$

$$\left. + \beta \left(\frac{1}{|N_G(\mathbf{s})|} \sum_{\mathbf{u}\in N_G(\mathbf{s})} Z_G(\mathbf{u}) - \frac{1}{2} \right) \right\} \ , \qquad (6)$$

where the normalizing constant C depends on α and β. In (6), there are two terms that express the departure of the average in the neighbourhood from the true theoretical equilibrium value of $\frac{1}{2}$; one is for the Background and the other is for the Grid.

3. Simulation experiments

In this section, we show through simulations that the two-scale spatial model given by (1), (2) and (6) allows both competitive or co-operative behaviours as well. Further, we look for edge effects and we examine measures that can better represent the spatial dependence. We introduce the spatial odds ratio and the spatial relative risk, which account for dependence better than the spatial correlation when the data are binary.

3.1. Parameter settings

Let D be a square lattice of size 53×53; we fix $\Delta = 4$, and we overlay a 12×12 Grid G onto D, with an edge region of width Δ, analogous to the scheme shown in Figure 1.

Following the model's hierachical description in Section 2, we simulated forward as follows: In Step 1, we simulated a Gaussian random field ε on the domain D with spatial covariance Σ, using the the R package RandomFields ([43]). In Step 2, we simulated independent Bernoulli random variables. Finally, in Step 3, we used a Gibbs sampler (with 2000 runs) to simulate the auto-logistic model on the Grid.

Step 1. The Gaussian random field ε is simulated on D with distribution $N_n(\mathbf{0}, \Sigma)$; we choose the exponential covariance function to characterize the spatial covariance matrix Σ; that is $\Sigma = (\Sigma_{ij})$ with $\Sigma_{ij} = C(\mathbf{s}_i - \mathbf{s}_j)$ and $C(\mathbf{h}) = \sigma_\varepsilon^2 e^{-||\mathbf{h}||/\theta}$, for $\mathbf{h} \in \mathbb{R}^2$.
In order to obtain reasonable spatial dependence, we choose $\theta = 5$, and then $\theta = 20$, the latter value ensuring stronger spatial dependence. We set $\sigma_\varepsilon^2 = 1$.

Step 2. Conditional on the simulated ε, we simulate independent Bernoulli random variables $Z_B(\mathbf{s})$ on the Background, with parameters $p(\mathbf{s}) = \dfrac{e^{\varepsilon(\mathbf{s})}}{1 + e^{\varepsilon(\mathbf{s})}}$; $\mathbf{s} \in D$.

Step 3. For the standard auto-logistic model with constant α and β, values of β that would give weak spatial dependence, stronger dependence, and strong dependence would be around 3, 8, and 16, respectively. But for our two-scale model, we note a reinforcement of spatial interaction due to the Background effect, through $\alpha_B(\mathbf{s})$; then, a choice of $\beta = 2$ is large enough to obtain strong positive spatial dependence. Similarly, a strong competitive behaviour can be obtained with $\beta = -2$. We ran several simulation schemes for different values of parameters α and β, both possibly negative and positive. Specifically, we ran simulations for $\alpha \in \{-6, -5, ..., -1, 1, ..., 5, 6\}$ and $\beta \in \{-4, -3, -2, -1, 1, 2, 3, 4\}$.

Each model was simulated $L = 1600$ times, and we denote $\mathbf{Z}^{(l)}$ as the $l-$th realization of $\mathbf{Z} = (\mathbf{Z}_B^{\mathrm{T}}, \mathbf{Z}_G^{\mathrm{T}})^{\mathrm{T}}$. The statistical properties of $\mathbf{Z}$ are obtained by Monte Carlo averaging of $\{\mathbf{Z}^{(l)} : l = 1, .., L\}$.

3.2. Experimental results

3.2.1. Visualization and edge effects

Figure 2 shows one realization of the process with parameters $\sigma_\varepsilon^2 = 1,\ \theta = 5,\ \alpha = 2,\ \beta = 2$. The whole process gives $\mathbf{Z} = (\mathbf{Z}_B^{\mathrm{T}}, \mathbf{Z}_G^{\mathrm{T}})^{\mathrm{T}}$, and Figure 3 shows the corresponding $\mathbf{Z}_G$ used in the simulation of $\mathbf{Z}$. Looking at the Grid only, it is obvious that there is strong positive spatial dependence, despite the small value of $\beta = 2$. Clearly, the Grid dependence is strenghtened by the Background dependence. The proportion of $1s$ equals 0.5372 on D, and 0.5000 on the Grid.

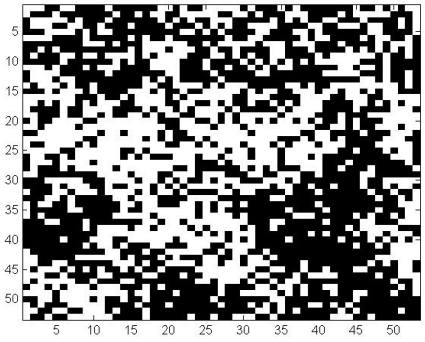

Figure 2: One realization of $\mathbf{Z}$ with $\sigma_\varepsilon^2 = 1,\ \theta = 5,\ \alpha = \beta = 2$

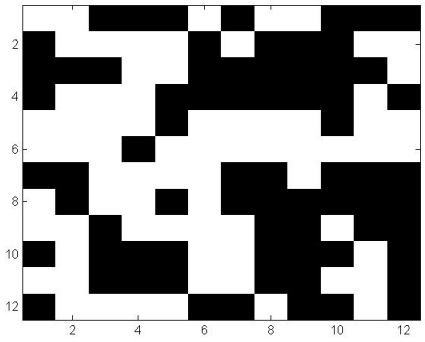

Figure 3: Representation of $\mathbf{Z}_G$ extracted from $\mathbf{Z}$

In order to study edge effects, we consider the first, middle, and last columns of the lattice, and compute the average values (taken over the $L = 1600$ simulations), $\frac{1}{L}\sum_{l=1}^{L} Z^{(l)}(s_1, s_2)$, for $s_2 = 1,\ 25,\ 53$. Figure 4 plots these average values against s_1. Similar results are obtained for all sets of parameters (and we obviously obtain analogous results if we fix s_1 and look at functions of s_2). Our conclusion is that there is no striking edge effect.

3.2.2. Covariance and spatial odds ratio.

The spatial covariance or spatial correlation are the usual measures used to quantify spatial dependence; however, when the state space is $\{0, 1\}$, there is some doubt about the usefulness of the empirical covariance. In fact, when we consider pair-values (z_i, z_j), only $(1, 1)$ contributes to the covariance. For this reason, we introduce a different characterization of spatial dependence that is more appropriate in the binary 0-1 context and incorporates all pair-values. The idea is to adapt the relative risk and odds ratio to a binary spatial setting. Since we obtain very similar results for both spatial measures, we present here

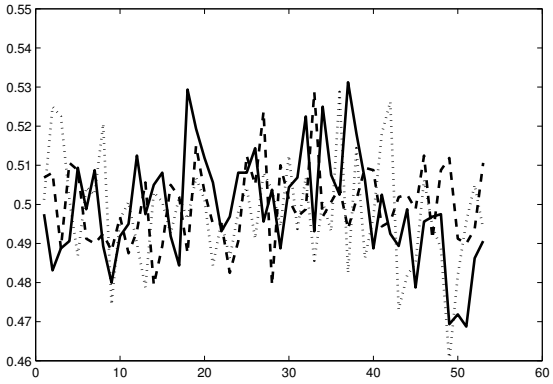

Figure 4: Average values of sites on the first (–) middle (...) and last (- -) columns of the lattice.

only the spatial odds-ratio. It is easy to derive and study the spatial relative risk in an analogous way.

For $\mathbf{s} \in D$, the spatial odds ratio (SOR), at location $\mathbf{s}$, in the direction $\mathbf{e}$ and at distance h is defined by

$$SOR_{\mathbf{e}}(\mathbf{s}, h) = \frac{p_{00}(\mathbf{s}, h, \mathbf{e}) \times p_{11}(\mathbf{s}, h, \mathbf{e})}{p_{01}(\mathbf{s}, h, \mathbf{e}) \times p_{10}(\mathbf{s}, h, \mathbf{e})}, \tag{7}$$

where $p_{jk}(\mathbf{s}, h, \mathbf{e})$ is the probability of pairs $(Z(\mathbf{s}) = j, Z(\mathbf{s} + h\mathbf{e}) = k)$, and the vector $\mathbf{e}$ defines the direction of interest. The quantity in (7) is a property of the model which can be studied via simulation. Based on the $L = 1600$ simulations, we can define Monte carlo averages that allow us to compute

$$\widehat{SOR}_{\mathbf{e}_i}(\mathbf{s}, h) = \frac{\hat{p}_{00}(\mathbf{s}, h, \mathbf{e}_i) \times \hat{p}_{11}(\mathbf{s}, h, \mathbf{e}_i)}{\hat{p}_{01}(\mathbf{s}, h, \mathbf{e}_i) \times \hat{p}_{10}(\mathbf{s}, h, \mathbf{e}_i)}, \tag{8}$$

where $\hat{p}_{jk}(\mathbf{s}, h, \mathbf{e}_i) = \dfrac{1}{L}\displaystyle\sum_{l=1}^{L} \mathbf{1}_{\left\{Z^{(l)}(\mathbf{s})=j,\ Z^{(l)}(\mathbf{s}+h\mathbf{e}_i)=k\right\}}$. Here, we shall study (8) in two directions, horizontal and diagonal, meaning that we set directional vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ with coordinates $\mathbf{e}_1 = (1, 0)$ and $\mathbf{e}_2 = (1, 1)$.
To compare this new property to the spatial covariance, we computed $cov(Z(\mathbf{s}), Z(\mathbf{s}+h\mathbf{e}_i)) = C_{\mathbf{e}_i}(\mathbf{s}, h)$ as follows:

$$C_{\mathbf{e}_i}(\mathbf{s}, h) = \frac{1}{L}\sum_{l=1}^{L}\left(Z^{(l)}(\mathbf{s}) - \overline{Z}(\mathbf{s})\right)\left(Z^{(l)}(\mathbf{s}+h\mathbf{e}_i) - \overline{Z}(\mathbf{s}+h\mathbf{e}_i)\right),$$

where $\overline{Z}(\mathbf{s}) = \dfrac{1}{L}\sum_{l=1}^{L} Z^{(l)}(\mathbf{s})$. Then the global covariance and global SOR are respectively given by

$$C_{\mathbf{e}_i}(h) = \frac{1}{|D_{\mathbf{e}_i,h}|}\sum_{\mathbf{s}\in D_{\mathbf{e}_i,h}} C_{\mathbf{e}_i}(\mathbf{s}, h),$$

and

$$SOR_{\mathbf{e}_i}(h) = \frac{1}{|D_{\mathbf{e}_i,h}|} \sum_{\mathbf{s} \in D_{\mathbf{e}_i,h}} SOR_{\mathbf{e}_i}(\mathbf{s}, h), \tag{9}$$

where $D_{\mathbf{e}_i,h} = \{\mathbf{s} \in D : \mathbf{s} + \mathbf{e}_i h \in D\}$, for $i = 1, 2$.

In the case of a continuous spatial index, the global covariance, $C_{\mathbf{e}_i}(h)$, can be written as $\int_{D_{\mathbf{e}_i,h}} cov(Z(\mathbf{s}), Z(\mathbf{s}+h\mathbf{e}_i))d\mathbf{s}$ / $\int_{D_{\mathbf{e}_i,h}} d\mathbf{s}$, and the global SOR, $SOR_{\mathbf{e}_i}(h)$, can be written as $\int_{D_{\mathbf{e}_i,h}} SOR_{\mathbf{e}_i}(\mathbf{s},h)d\mathbf{s}$ / $\int_{D_{\mathbf{e}_i,h}} d\mathbf{s}$.

We plot the values (9) versus h, $h = 1$ to $5\Delta = 20$, for different sets of parameters (θ, α, β). We summarize the main results in the text below, and present some illustrative plots for selective parameter values. These plots represent averages over 1600 simulations, and consequently the curves are smooth; an empirical version from a single realization is much more irregular and some features such as peaks and bumps used to determine Δ will be much harder to discern.

As expected, the covariance values as well as the SOR values increase with parameter θ; however, more interesting is to see the influence of the Grid. Its effect can be visually detected by the presence of a peak when $\beta > 0$ or a dip when $\beta < 0$. In intermediate cases, it is hardly observable, or not at all; see for instance the left plot of Figure 6. An important observation is that in most cases Δ is more easily detected from the SOR. Figure 5 displays the plots of the global covariance and SOR for $\sigma_\varepsilon^2 = 1$, $\theta = 5$, $\alpha = 4$, $\beta = 4$. We can see that both the covariance and SOR are decreasing with h. We can observe peaks in the horizontal direction $\mathbf{e}_1$, and almost nothing in the diagonal direction $\mathbf{e}_2$; this is expected, because our model is simulated using the four nearest neighbours and $\beta > 0$.

Furthermore, for the global SOR, we clearly observe a peak at lag $h = 4$, which is the size of Δ, and a smaller one at lag 8; there is fainter evidence of peaks at lags 12 and 16. This periodic pattern is also present in the global covariance but it is much less obvious. This underlines our recommendation that one uses SOR rather than covariance for binary data.

Figure 5 demonstrates that peaks can be seen at lags Δ, 2Δ, 3Δ, and maybe further if α and β take large positive values. In fact, the presence and magnitude of the peaks is very sensitive to both Grid parameters α and β. Figure 6 gives three plots of SOR for different values of the Markov-random-field parameters, where we hold θ fixed. The plot on the left-hand side presents no bumps or concavities. We observe a bump in the midlle, at lag $h = \Delta = 4$, but none for higher lags. The plot on the right-hand side shows that we also can obtain a dip when β is negative.

Finally, we observe that both the covariance and the magnitude of the possible peaks increase with $\alpha > 0$, as shown in Figure 7, and this is more obvious for larger values of the exponential parameter θ. From Figure 8, we see that the magnitude of the peaks can be increased by increasing β.

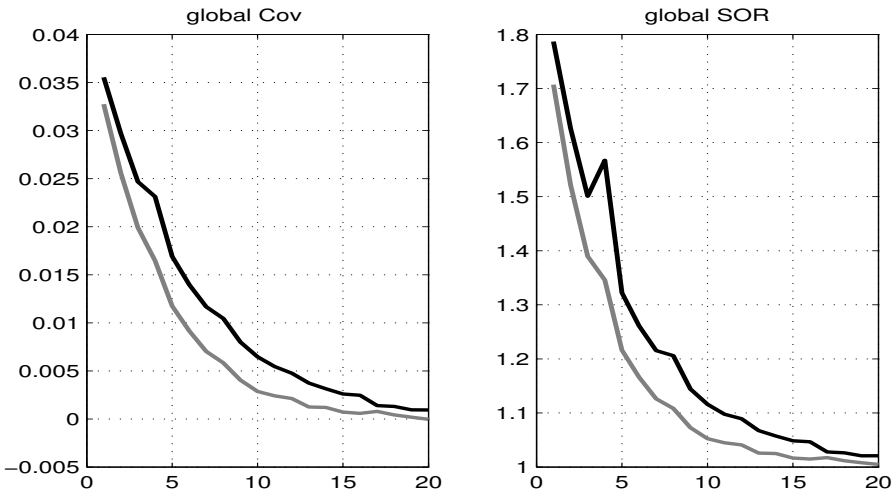

Figure 5: Global covariance (left panel) and global SOR (right panel) in horizontal direction $\mathbf{e}_1$ (black) and diagonal direction $\mathbf{e}_2$ (grey) for $\sigma_\varepsilon^2 = 1,\ \theta = 5,\ \alpha = 4,\ \beta = 4$.

We conducted other simulations to study the covariance and SOR values for different sites, belonging to the Grid or the Background, located in the domain's center or in the border region. Our conclusion is that results like those seen in Figures 5-8 are essentially the same.

4. Estimation

In this section, we consider the task of obtaining maximum likelihood estimates of the parameters $(\sigma_\varepsilon^2, \theta, \alpha, \beta)$. We suppose that Δ is known; in practice, there is information about Δ in the empirical SOR (Section 5).

The parameters come from two structures, the Grid and the Background. Because of the conditional manner in which the model is defined, the joint distribution of the process is the product of two terms, corresponding to the distribution on the Grid given the Background, times the distribution on the Background. Taking the logarithm of the likelihood, we see that estimation of the Grid parameters can be obtained separately from estimation of the Background parameters, which is an advantageous feature of our model. Due to the intractable normalizing constant, the auto-model parameters are estimated by maximizing the pseudo-likelihood introduced by Besag [6]. The second term involves the latent Gaussian field's parameters; their estimation in a hierarchical statistical model typically requires an EM algorithm; see [16] or [36]. The E-step needs the expectation of the latent field ε given the observations, but we do not know the integrated distribution. There are several ways to overcome this issue: A common approach is to use Monte Carlo procedures; see for instance [38], [7]. Here instead we use Laplace approximations to approximate the intractable integrals.

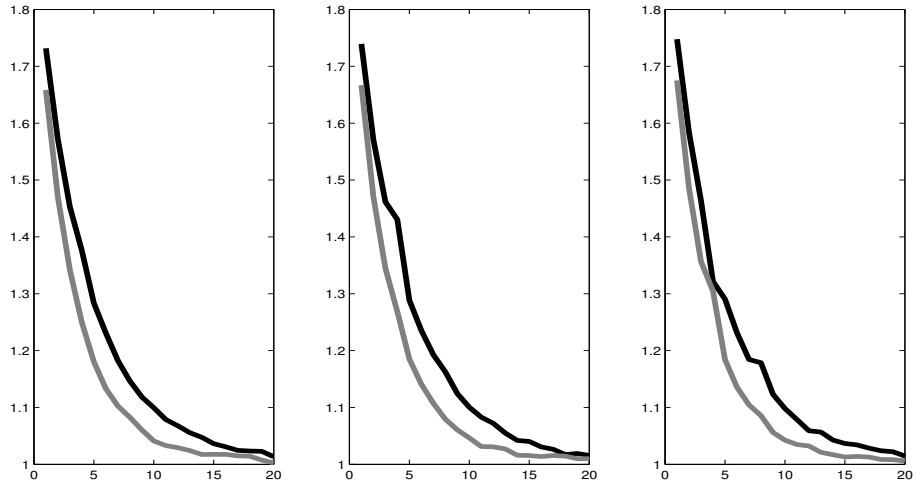

Figure 6: SOR in horizontal direction $\mathbf{e}_1$ (black) and diagonal direction $\mathbf{e}_2$ (grey) for $\sigma_\varepsilon^2 = 1,\ \theta = 5$. Left: $(\alpha, \beta) = (3, 1)$. Middle: $(\alpha, \beta) = (2, 2)$. Right: $(\alpha, \beta) = (10, -5)$.

The likelihood is given by the distribution of $\mathbf{Z} = (\mathbf{Z}_B^{\mathrm{T}}, \mathbf{Z}_G^{\mathrm{T}})^{\mathrm{T}}$:

$$\begin{aligned}
[\mathbf{Z} \mid \varphi_\varepsilon, \alpha, \beta] &= \int [\varepsilon, \mathbf{Z}_B, \mathbf{Z}_G \mid \varphi_\varepsilon, \alpha, \beta] d\varepsilon \\
&= \int [\mathbf{Z}_G \mid \mathbf{Z}_B, \varepsilon, \varphi_\varepsilon, \alpha, \beta][\mathbf{Z}_B \mid \varepsilon, \varphi_\varepsilon, \alpha, \beta][\varepsilon \mid \varphi_\varepsilon, \alpha, \beta] d\varepsilon \\
&= [\mathbf{Z}_G \mid \mathbf{Z}_B, \alpha, \beta] \times \int [\mathbf{Z}_B \mid \varepsilon][\varepsilon \mid \varphi_\varepsilon] d\varepsilon. \qquad (10)
\end{aligned}$$

The first term in (10) is explicit, given by (6):

$$[\mathbf{Z}_G \mid \mathbf{Z}_B, \alpha, \beta] = C(\alpha, \beta)^{-1} \prod_{\mathbf{s} \in G} \exp Z(\mathbf{s}) \left\{ \alpha \left(\frac{1}{|N_B(\mathbf{s})|} \sum_{\mathbf{u} \in N_B(\mathbf{s})} Z_B(\mathbf{u}) - \frac{1}{2} \right) + \beta \left(\frac{1}{|N_G(\mathbf{s})|} \sum_{\mathbf{u} \in N_G(\mathbf{s})} Z_G(\mathbf{u}) - \frac{1}{2} \right) \right\}.$$

The normalizing constant is potentially problematic; for example, for the model we considered in the previous section on a 12×12 Grid, $C(\alpha, \beta)$ is given by the summation of 2^{144} terms. Some bypass the problem by approximating C using efficient Monte Carlo methods. In the Markov-random-field context, another method is to replace maximizing the likelihood with maximizing the conditional pseudo likelihood ([5], [6], [12], [19]), which allows fast and easy computation. The pseudo likelihood here is the product of conditional probabilities, $\prod_{\mathbf{s} \in G} \pi_{\mathbf{s}}(Z_G(\mathbf{s}) \mid \mathbf{Z}_B, \alpha, \beta)$, expressed as a function of α and β.

The second term in (10) involves the Gaussian distribution of ε and the

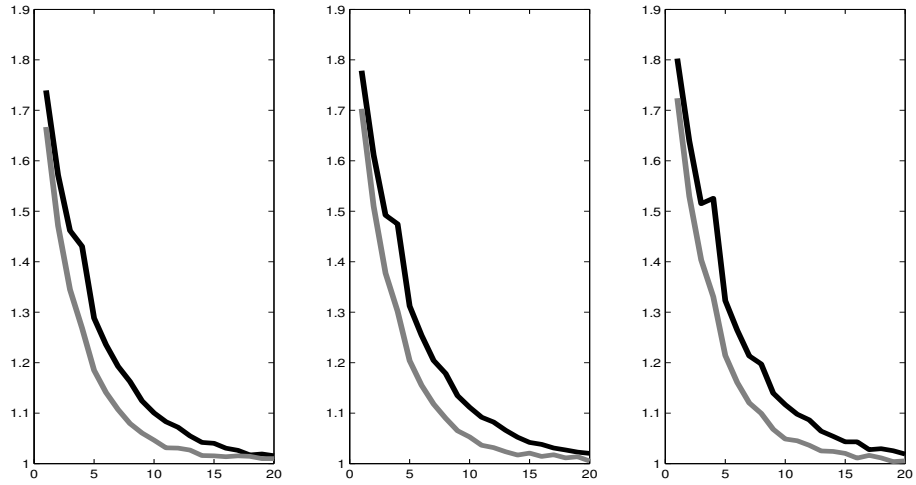

Figure 7: SOR in horizontal direction $\mathbf{e}_1$ (black) and diagonal direction $\mathbf{e}_2$ (grey) for $\sigma_\varepsilon^2 = 1,\ \theta = 5; \beta = 2$. Left: $\alpha = 2$. Middle: $\alpha = 4$. Right: $\alpha = 6$.

following conditional distribution of $\mathbf{Z}_B$ given ε,

$$[\mathbf{Z}_B \mid \varepsilon] = \prod_{\mathbf{s}\in B} p(\mathbf{s})^{Z(\mathbf{s})}(1-p(\mathbf{s}))^{1-Z(\mathbf{s})} = \prod_{\mathbf{s}\in B} e^{\varepsilon(\mathbf{s})Z(\mathbf{s})}\frac{1}{1+e^{\varepsilon(\mathbf{s})}}.$$

Now, the underlying spatial process ε is not observed, but we can define the complete likelihood of $\mathbf{Z}$ and the missing data ε.

Finally then, a quantity we call the pseudo complete log likelihood is given by

$$pl_c(\mathbf{Z}, \varepsilon; \varphi_\varepsilon, \alpha, \beta) = A_1(\alpha, \beta) + A_2(\varphi_\varepsilon), \tag{11}$$

where $\varphi_\varepsilon = (\sigma_\varepsilon^2, \theta)$,

$$A_1(\alpha, \beta) = \sum_{\mathbf{s}\in G} \log[Z_G(\mathbf{s}) \mid \mathbf{Z}_B, \alpha, \beta],$$

and

$$A_2(\varphi_\varepsilon) = -\sum_{\mathbf{s}\in B} \ln(1+e^{\varepsilon(\mathbf{s})}) + \sum_{\mathbf{s}\in B} \varepsilon(\mathbf{s})Z(\mathbf{s}) - \frac{1}{2}\ln(\det\Sigma) - \frac{1}{2}\varepsilon^{\mathrm{T}}\Sigma^{-1}\varepsilon - \frac{|B|}{2}\log(2\pi).$$

The first term, A_1, concerns the estimation of the Grid parameters, while the second term, A_2, is devoted to the estimation of the hidden Gaussian field via the observations $\mathbf{Z}_B$ on the Background B. The second term will be used to obtain an EM estimate of φ_ε.

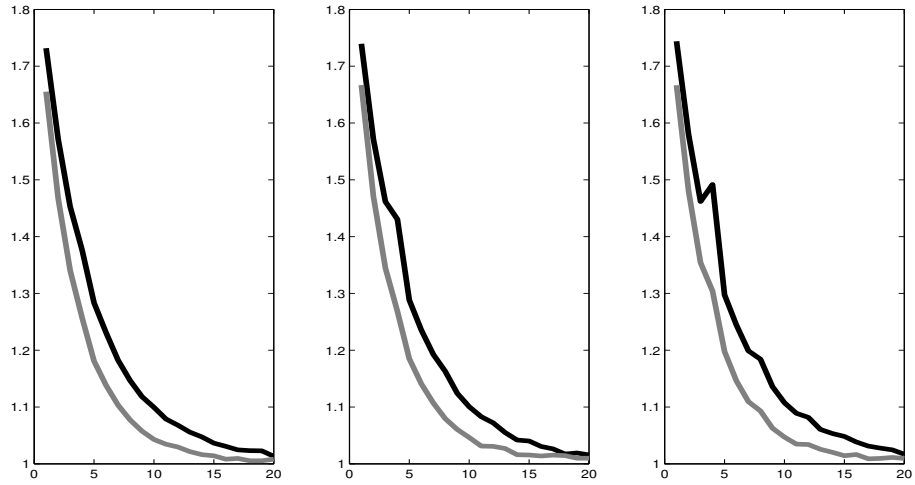

Figure 8: SOR in horizontal direction $\mathbf{e}_1$ (black) and vertical direction $\mathbf{e}_2$ (grey) for $\sigma_\varepsilon^2 = 1,\ \theta = 5; \alpha = 2$. left: $\beta = 1$. Middle: $\beta = 3$. Right: $\beta = 4$.

4.1. Estimation of the Grid parameters

The goal is to obtain $\widehat{\alpha}$ and $\widehat{\beta}$ that maximize $A_1(\alpha, \beta) = \sum_{\mathbf{s}\in G} \log[Z_G(\mathbf{s}) \mid \mathbf{Z}_B, \alpha, \beta]$. Hence,

$$A_1(\alpha,\beta) = \sum_{\mathbf{s}\in G} \left(\alpha(V(\mathbf{s}) - \frac{1}{2}) - \frac{\beta}{2} + \frac{\beta}{4} \sum_{\mathbf{u}\in N_G(\mathbf{s})} Z(\mathbf{u}) \right) Z(\mathbf{s}) - \log\left(1 + \exp\left\{ \alpha(V(\mathbf{s}) - \frac{1}{2}) - \frac{\beta}{2} + \frac{\beta}{4} \sum_{\mathbf{u}\in N_G(\mathbf{s})} Z(\mathbf{u}) \right\} \right), \quad (12)$$

where $V(\mathbf{s}) = \dfrac{\sum_{\mathbf{u}\in N_B(\mathbf{s})} Z(\mathbf{u})}{|N_B(\mathbf{s})|}$.

The maximization of this pseudo likelihood is achieved via a standard optimization algorithm and is straightforward to implement.

4.2. Estimation of the Background parameters

4.2.1. The EM algorithm

We want to obtain estimates $\widehat{\sigma_\varepsilon^2}$ and $\widehat{\theta}$ from the complete log likelihood:

$$A_2(\varphi_\varepsilon) = -\sum_{\mathbf{s}\in B} \ln(1+e^{\varepsilon(\mathbf{s})}) + \sum_{\mathbf{s}\in B} \varepsilon(\mathbf{s}) Z_B(\mathbf{s}) - \frac{1}{2}\ln(\det\Sigma) - \frac{1}{2}\varepsilon^{\mathrm{T}}\Sigma^{-1}\varepsilon - \frac{|B|}{2}\log(2\pi). \quad (13)$$

Since ε has not been observed, estimation is performed using the EM algorithm; see [16], [36].

Let us define

$$q(\varphi_\varepsilon, \hat{\varphi}_\varepsilon^{(l)}) = E\left[A_2(\varphi_\varepsilon) \mid \mathbf{Z}_B, \hat{\varphi}_\varepsilon^{(l)} \right]. \quad (14)$$

Starting with an initialization $\hat{\varphi}_\varepsilon^{(0)}$, the l-th iteration of the algorithm is achieved in two steps. For $l = 1, 2, ...$

The E-step is to compute the expectation $q(\varphi_\varepsilon, \hat{\varphi}_\varepsilon^{(l-1)})$.

The M-step is to maximize $q(\varphi_\varepsilon, \hat{\varphi}_\varepsilon^{(l-1)})$ with respect to φ_ε; that is, $\hat{\varphi}_\varepsilon^{(l)} = \arg\max_{\varphi_\varepsilon} q(\varphi_\varepsilon, \hat{\varphi}_\varepsilon^{(l-1)})$.

In our case, the E-step is problematic, since we do not have a closed-form expression for the conditional distribution of ε given the observations $\mathbf{Z}$. There are several possible approaches, one being to implement a stochastic EM (SEM) algorithm (e.g., [38], [36]), where the expectations are evaluated using Monte Carlo integration. The problem with this approach lies in the simulation, where a Metropolis algorithm is typically used to simulate the ε. Choosing the "right" proposal density (see [10], [39]) can be problematic, and when datasets are large, computations can be very slow.

Another classical remedy is to apply self-normalized importance sampling (see [38], Section 3.3). The first stage is to simulate M samples of ε according to the importance distribution $h(\varepsilon \mid \hat{\varphi}^{(l-1)})$; then the expectation $E\left[g(\varepsilon(\mathbf{s})) \mid \mathbf{Z}_B, \hat{\varphi}^{(l-1)}\right]$ is approximated by

$$\sum_{m=1}^{M} g(\varepsilon_m^{(l-1)}(\mathbf{s}))w_m^{(l-1)},$$

where $w_m^{(l-1)}$ are the normalized weights, $w_m^{(l-1)} = \frac{v_m^{(l-1)}}{\sum_{m=1}^{M} v_m^{(l-1)}}$, with $v_m = \frac{[\mathbf{Z}_B \mid \varepsilon_m, \hat{\varphi}^{(l-1)}] \times [\varepsilon_m \mid \hat{\varphi}^{(l-1)}]}{h(\varepsilon_m \mid \hat{\varphi}^{(l-1)})}$. Choosing the "right" importance distribution can be problematic; added to this, we observe a degeneracy of the weights, leading to poor estimates.

4.2.2. Laplace approximations in the EM algorithm

We now derive Laplace approximations (LA) to approximate the expectations involved in (14), which are based on second-order Taylor-series expansions of the logarithm of the integrands around their respective modes. This approach gives us a stable estimation procedure.

Write $A_2(\varphi_\varepsilon)$ more completely as $A_2(\varphi_\varepsilon; \mathbf{Z}_B, \varepsilon)$. Let us denote ε_m as the field maximizing $A_2(\varphi_\varepsilon; \mathbf{Z}_B, \varepsilon)$; then, a second-order Taylor series expansion for $A_2(\varphi_\varepsilon; \mathbf{Z}_B, \varepsilon)$ around ε_m yields:

$$\begin{aligned} A_2(\varphi_\varepsilon; \mathbf{Z}_B, \varepsilon) = A_2(\varphi_\varepsilon; \mathbf{Z}_B, \varepsilon_m) &+ (\varepsilon - \varepsilon_m)^{\mathrm{T}} \frac{\partial}{\partial \varepsilon} A_2(\varphi_\varepsilon; \mathbf{Z}_B, \varepsilon) \\ &+ \frac{1}{2}(\varepsilon - \varepsilon_m)^{\mathrm{T}} \frac{\partial^2}{\partial \varepsilon \partial \varepsilon^{\mathrm{T}}} A_2(\varphi_\varepsilon; \mathbf{Z}_B, \varepsilon_m)(\varepsilon - \varepsilon_m) + \ ... \end{aligned}$$

Now, looking at the right-hand side, the second term is zero, so we have the following approximation:

$$A_2(\varphi_\varepsilon; \mathbf{Z}_B, \varepsilon) \simeq A_2(\varphi_\varepsilon; \mathbf{Z}_B, \varepsilon_m) - \frac{1}{2}(\varepsilon - \varepsilon_m)^{\mathrm{T}}(-H(\varepsilon_m))(\varepsilon - \varepsilon_m),$$

where $H(\varepsilon_m) = \frac{\partial^2}{\partial \varepsilon \partial \varepsilon^{\mathrm{T}}} A_2(\varphi_\varepsilon; \mathbf{Z}_B, \varepsilon_m)$.

Hence, the probability density function $[\varepsilon \mid \mathbf{Z}_B, \varphi_\varepsilon]$ is approximately proportional to $\exp A_2(\varphi_\varepsilon; \mathbf{Z}_B, \varepsilon_m) \times \exp\left[-\frac{1}{2}(\varepsilon-\varepsilon_m)^{\mathrm{T}}(-H(\varepsilon_m))(\varepsilon-\varepsilon_m)\right]$; that is, $[\varepsilon \mid \mathbf{Z}_B, \varphi_\varepsilon]$ is proportional to a Gaussian density. Moreover, one can evaluate the constant that ensures a probability density, resulting in:

$$[\varepsilon \mid \mathbf{Z}_B, \varphi] \simeq \frac{1}{(2\pi)^{\frac{n}{2}} \mid -H(\varepsilon_m)\mid^{-\frac{1}{2}}} \exp\left[-\frac{1}{2}(\varepsilon-\varepsilon_m)^{\mathrm{T}}(-H(\varepsilon_m))(\varepsilon-\varepsilon_m)\right]. \tag{15}$$

We deduce from (15) that $E\left[\varepsilon \mid \mathbf{Z}_B, \varphi_\varepsilon\right] \simeq \varepsilon_m$ and $E(\varepsilon\varepsilon^{\mathrm{T}} \mid \mathbf{Z}_B, \varphi_\varepsilon) = -H(\varepsilon_m)^{-1}$, and hence we obtain the Laplace approximation of the expectation as:

$$\begin{aligned} E[A_2(\varphi_\varepsilon; \mathbf{Z}_B, \varepsilon) \mid \mathbf{Z}_B, \varphi_\varepsilon] \simeq -E[\sum_{\mathbf{s}\in B} \ln(1+e^{\varepsilon(\mathbf{s})})] + \sum_{\mathbf{s}\in B} \varepsilon_m(\mathbf{s}) Z(\mathbf{s}) - \frac{1}{2}\ln(\det\Sigma) \\ - \frac{1}{2}\varepsilon_m'\Sigma^{-1}\varepsilon_m - \frac{1}{2} trace(\Sigma^{-1}(-H(\varepsilon_m)^{-1}) - \frac{|B|}{2}\log(2\pi), \end{aligned} \tag{16}$$

where the expectation in the first term is with respect to $N(\varepsilon_m, -H(\varepsilon_m))$ distribution. That term does not involve φ_ε, and hence we are only interested in the following expression to be minimized:

$$\widetilde{A_2}(\varphi_\varepsilon) \equiv \varepsilon_m^{\mathrm{T}}\Sigma^{-1}\varepsilon_m + trace(\Sigma^{-1}(-H(\varepsilon_m)^{-1}) + \ln(\det\Sigma). \tag{17}$$

In the Appendix we give the details for the computation of the mode ε_m and the matrix $H(\varepsilon_m)$.

Finally, starting with an initialization $\hat{\varphi}_\varepsilon^{(0)}$, the l-th iteration of the EM algorithm is achieved in the following two steps: at the E-step, we compute the mode, $\varepsilon_m^{(l-1)}$, of $A_2(\hat{\varphi}_\varepsilon^{(l-1)}; \mathbf{Z}_B, \varepsilon)$, the Hessian $-H(\varepsilon_m^{(l-1)})$, and $\widetilde{A_2}(\varphi_\varepsilon, \hat{\varphi}_\varepsilon^{(l-1)})$.

At the M-step we minimize $\widetilde{A_2}(\varphi_\varepsilon, \hat{\varphi}_\varepsilon^{(l-1)})$. This is achieved by a simple minimization of a single variable function. Writing the covariance matrix as $\Sigma(\varphi_\varepsilon) = \sigma_\varepsilon^2 Q(\theta)$, we want to minimize:

$$f(\theta, \sigma_\varepsilon^2) = \frac{1}{\sigma_\varepsilon^2}\varepsilon_m^{\mathrm{T}} Q(\theta)^{-1}\varepsilon_m + \frac{1}{\sigma_\varepsilon^2} trace(Q^{-1}(-H(\varepsilon_m)^{-1}) + n\ln\sigma_\varepsilon^2 + \ln(\det Q(\theta)).$$

with respect to σ_ε^2 and θ. The derivative with respect to σ_ε^2 for a fixed θ gives the following explicit solution:

$$\sigma_\varepsilon^2(\theta) = \frac{1}{n}\left[\varepsilon_m^{\mathrm{T}} Q(\theta)^{-1}\varepsilon_m + trace(Q^{-1}(-H(\varepsilon_m)^{-1})\right]. \tag{18}$$

Then the M-step is to minimize, with respect to θ,

$$\begin{aligned} g(\theta, \hat{\varphi}_\varepsilon^{(l-1)}) =& n\ln\left[\frac{1}{n}\left[\varepsilon_m^{(l-1)\ \mathrm{T}} Q(\theta)^{-1}\varepsilon_m^{(l-1)} + trace(Q^{-1}(-H(\varepsilon_m^{(l-1)})^{-1})\right]\right] \\ &+ \ln(\det Q(\theta)). \end{aligned}$$

4.3. Simulation experiments

We ran $L = 1600$ simulations, as described in Section 3, with the values of the parameters given by $\sigma_\varepsilon^2 = 1,\ \theta = 5,\ \alpha = 2,\ \beta = 2$. Then estimation is performed on each simulation based on the procedures outlined in Sections

	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\theta}$	$\widehat{\sigma_\varepsilon^2}$
Target	2	2	5	1
Mean	1.9815	1.9297	4.6528	0.8719
Std. Dev.	1.0630	0.9223	0.7678	0.1334

Table 1: Summary of two-scale model parameter estimates

4.1 and 4.2. We present in Table 1 the mean and standard deviations of the estimates, obtained from the 1600 simulations.

We observe a negative bias, especially for the Background parameters. We think this biais comes from the Laplace approximation and might be reduced using higher order terms. Looking inside the EM-algorithm, we observe that very often, at each iteration, the new value of the estimate $\hat{\theta}^{(l)}$ is less than $\hat{\theta}^{(l-1)}$. The initial bias for parameter $\hat{\theta}$ transfers to a bias for $\hat{\sigma}_\varepsilon$ since the latter is directly computed from $\hat{\theta}$ by (18).

It is worth noticing that the EM procedure is not sensitive to the choice of starting values, and the number of iterations is often less than or equal to 8; when it is larger, we obtain a $\hat{\theta}$ that is typically highly biased.

The estimation of the Grid parameters is obtained from 144 values in $\mathbf{Z}_G$, which is not a large number of observations. In spite of this, it is encouraging that our estimates are close to the true values.

We conducted other experiments with the two-scale-model parameters given by $\sigma_\varepsilon^2 = 1,\ \theta = 20,\ \alpha = 2,\ \beta = 2$. The results we obtained were similar, but with higher bias for the parameter $\hat{\theta}$. For lattice size 53×53, the parameter $\theta = 20$ induces a large spatial dependence, and a typical realization on this lattice does not show enough contrast to estimate θ accurately. Nevertheless, a model-based prediction using this estimate can still be good.

5. Application to Corn Borers dataset

An extensive entomological field study of European corn borer larvae was conducted in northwest Iowa ([35]). The original data are available in a 1954 technical report from the Iowa State Statistical Laboratory ("Uniformity Data from European Corn Borer, Pyrausta nubilalis (Hbn.)"). Lee et al. ([27]) selected one dataset from this study, which they published, to examine whether the occurrence of corn borer larvae exhibited spatial-dependence structure. The data come from a (1/3)-acre square plot containing 36 rows, in which seeds had been planted in 36 equally spaced "hills" in each row, at an average rate of 3 seeds per hill. The area was divided into 324 regular subplots, each containing 4 hills. The response variables analyzed were defined as binary variables for the subplots, where the value 0 was obtained if corn borer larvae were absent, and the value 1 was obtained if one or more larvae were present. Now, let $u_i \in \{1, ..., 18\}$ denote the E-W position and $v_i \in \{1, ..., 18\}$ the N-S position of subplot i in a regular 18×18 lattice S; define $\mathbf{s}_i = (u_i, v_i)$ and, for $i = 1, ..., 324$,

$$Z(\mathbf{s}_i) = \begin{cases} 0 & \text{if no larvae observed in subplot } i \\ 1 & \text{if one or more larvae observed in subplot } i \end{cases}.$$

A figure showing the regular spatial lattice of subplots and observed values $\{Z(\mathbf{s}_i) : i = 1, ..., 324\}$ is presented in Fig. 9, where black illustrates the value 1, and white illustrates the value 0.

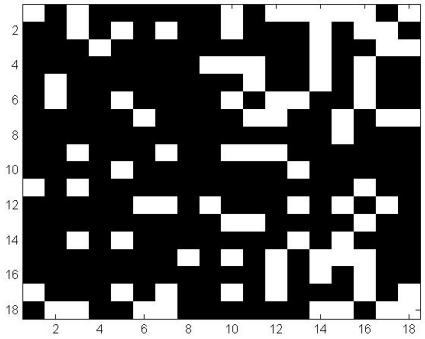

Figure 9: Corn borers data

In their work, Lee et al. consider four Bernoulli conditional models with possible multi-way dependence. The first and second models are classic auto-logistic models with pairwise-only dependence, respectively associated with the four-nearest-neighbour system and the eight-nearest-neighbour system; the third and fourth models involve cliques of size 3 and 4 associated with the eight-nearest-neighbour system. They show that the extra cliques contained in Model 2, due to to the diagonally adjacent pairs of sites, does not bring much additional information to Model 1. Models 3 and 4 allow multi-way dependence and incorporate relative positional information in observed patterns of data. Lee et al. suggest that, for a given number of infested neighbors, the spread of infestation among spatial subplots is higher if infested neighbors occur in all directions rather than in just one direction. Having infested neighbors in all directions implies that biological conditions are favorable for infestation throughout the whole immediate region, as opposed to a situation in which conditions are favorable for infestation in one direction, but not in others.

We apply our model to this dataset; more precisely, we consider a centered Gaussian spatial field, $\varepsilon \sim N_n(\mathbf{0}, \Sigma)$, with $\Sigma_{ij} = \sigma_\varepsilon^2 \exp(-||i - j||/\theta)$, and conditionally independent Bernoulli variables on the Background as defined in (1). Then, for a given Δ, we superimpose the Grid (cell size Δ) on S; we consider a Markov random field $\mathbf{Z}_G$ on the Grid with a four-nearest-neighbour system (based on Lee et al.), defined by (4) and (5), and with parameters α and β. We estimate the parameters $(\sigma_\varepsilon^2, \theta, \alpha, \beta)$ following the procedure described Section 4.

The resolution Δ of the Grid is chosen according to a preliminary exploratory step, by inspecting the spatial odds ratio at different lags. A first exploratory approach is to plot the SOR defined in (7) for different lags $h = 1$ to 10, in the directions $\mathbf{e}_1$ (E-W), $\mathbf{e}_2$ (SW-NE) and $\mathbf{e}_3$ (N-S) with coordinates $\mathbf{e}_1 = (1, 0)$, $\mathbf{e}_2 = (1, 1)$, and $\mathbf{e}_3 = (0, 1)$.

Since we have one dataset, we compute the empirical SOR at lag h in direction $\mathbf{e}$, as follows:

$$\widehat{SOR}_{\mathbf{e}}(h) = \frac{\hat{p}_{00}(h,\mathbf{e}) \times \hat{p}_{11}(h,\mathbf{e})}{\hat{p}_{01}(h,\mathbf{e}) \times \hat{p}_{10}(h,\mathbf{e})}, \tag{19}$$

where $\hat{p}_{jk}(h,\mathbf{e}) = \frac{1}{|D_{\mathbf{e},h}|} \sum_{\mathbf{s}\in D_{\mathbf{e},h}} \mathbf{1}_{\{Z(\mathbf{s})=j,\ Z(\mathbf{s}+h\mathbf{e})=k\}}$, and recall that $D_{\mathbf{e},h} = \{\mathbf{s} \in D : \mathbf{s} + \mathbf{e}h \in D\}$.

In Figure 10, we present the plots of the spatial odds ratio with respect to h in the three directions $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$.

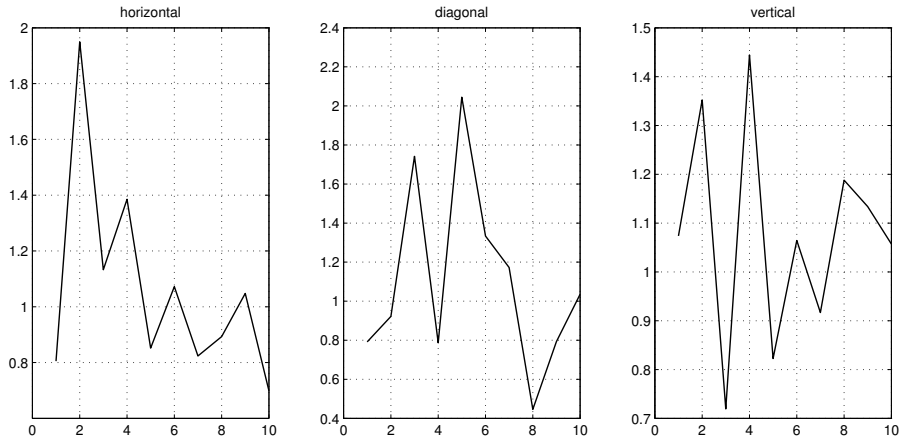

Figure 10: Empirical SOR in directions $\mathbf{e}_1$ (left), $\mathbf{e}_2$ (middle) and $\mathbf{e}_3$ (right), for lags $h = 1$ to 20. The units of lag in direction $\mathbf{e}_2$ have to be multiplied by $\sqrt{2}$ to compare to the units of lag in the directions $\mathbf{e}_2$ and $\mathbf{e}_3$

It is apparent that $\Delta_{E-W} = 2$ is a good selection for the E-W direction, while we have possible choices of $\Delta_{N-S} = 2$ and $\Delta_{N-S} = 4$ in the N-S direction; we also notice a peak in the diagonal direction which likely comes from $\Delta_{E-W} = 2$ and $\Delta_{N-S} = 4$. Finally, we consider two Grids, G_1 with lags $\Delta_{E-W} = 2$ and $\Delta_{N-S} = 2$, and G_2 with lags $\Delta_{E-W} = 2$ and $\Delta_{N-S} = 4$. Grids G_1 and G_2 both "start" at the site $\mathbf{s} = (3, 3)$ and involve, respectively, 64 and 32 sites.

For each of the two Grids, we estimate the parameters, using the procedure described in Section 4. The results are summarized in Table 2. For each Grid, we give the estimates as well as the value of the log likelihood terms $A_1(\widehat{\alpha}, \widehat{\beta})$ and $-\widetilde{A_2}(\widehat{\varphi_\varepsilon})$, defined by (12) and (17). We prefer larger values of these terms.

First, we can see that strong spatial dependence is estimated for both models; indeed, the scale parameter θ of the spatial covariance is about 20, which is quite large; for both Grids, the interaction parameter β is almost 3, which also denotes strong spatial dependence. With respect to the Background, there is no evidence of a big difference between the two models. The estimated values of θ are quite similar, as are the resulting log-likelihood values. For the Grid parameters, the interaction parameter β is on the same order (almost 3), but

	$\widehat{\alpha}$	$\widehat{\beta}$	$A_1(\widehat{\alpha},\widehat{\beta})$	$\widehat{\sigma^2_\varepsilon}$	$\widehat{\theta}$	$-\widetilde{A_2}(\widehat{\varphi_\varepsilon})$
Grid $G_1(2,2)$	1.8695	2.9452	-0.5216	0.8825	20.4916	-905.7831
Grid $G_2(4,2)$	2.5126	2.8618	-0.2613	0.9437	21.6418	-910.0022

Table 2: Estimates of the parameters and log likelihood values for Grids G_1 and G_2.

the main difference between the two models is observed for the Grid parameter α. The comparison of the values of the pseudo likelihood A_1 indicates strong preference for the two-scale model with Grid G_2.

6. Discussion and conclusions

In this paper, we have developed a hierarchical (in scale) spatial statistical model for binary data. The process model is spatially dependent and is obtained from a hidden Gaussian spatial process. This fine-scale dependence is captured through Bernoulli random variables with success probabilities given by the logit transformation of the latent process. Then, conditional on this Background level, a Grid at a coarser resolution is superimposed; the coarse-scale dependence is captured through a binary Markov random field.

Some extensions of our research could be considered. The estimation of parameters given in Section 4 requires computation of an inverse covariance matrix, which can be problematic for large data sets. In this case, we could consider modelling ε with a Spatial Random Effects (SRE) model, as described by Cressie and Johannesson ([13]; see also [24], [44]). Other reduced rank approaches could also be used (e.g. [48]), or indeed the inverse covariance matrix could be modelled directly (e.g.; Lindgren et al. [29]).

Furthermore, covariates could be incorporated in the modelling through the latent Gaussian process as follows: In (1) write,

$$p(\mathbf{s}_i) = \frac{e^{L(\mathbf{s}_i)}}{1+e^{L(\mathbf{s}_i)}},$$

$$L(\mathbf{s}_i) = \mathbf{X}(\mathbf{s}_i)'\mathbf{b} + \varepsilon(\mathbf{s}_i), \tag{20}$$

where $\mathbf{X}(.)$ denotes a p-dimensional vector of known covariates, and $\mathbf{b}$ is a p-dimensional vector of regression coefficients. This model and the estimation of its parameters is currently under investigation.

In this paper, we considered binary data, although the approach is clearly generalisable to count data or other data arising from a generalized linear model. The binary Markov random field on the superimposed Grid could quite naturally be generalized to a spatial auto-model from the same member of the exponential family that is used to generate the Background.

Simple assumptions on the spatial dependence on the Gird were made, namely the four-nearest-neighbour system and pairwise-only dependence. These assumptions can be easily extended to more neighbours and indeed the pairwise-only assumption can be relaxed in a manner similar to that given in Lee et al. [27]. This might be particularly useful within the context of image analysis,

where higher-order interactions allow for improved image processing (see for instance [15] and [46]).

Acknowledgment

Hardouin's research was conducted as part of the project Labex MME-DII (ANR11-LBX-0023-01). Cressie's research was supported by a 2015-2017 ARC Discovery Grant.

Appendix

In Section 4, we use the EM algorithm where the E-step is to compute the Laplace approximation, $\widetilde{A_2}(\varphi_\varepsilon)$, given by (17); its expression depends on the mode ε_m and on the Hessian $H(\varepsilon_m)$ of $A_2(\varphi_\varepsilon; \mathbf{Z}_B, \varepsilon)$, computed at the mode. Recall that the process ε can be written as a vector, $\varepsilon^{\mathrm{T}} = (\varepsilon_B^{\mathrm{T}}, \varepsilon_G^{\mathrm{T}})$, with covariance matrix, $\Sigma = \begin{pmatrix} \Sigma_B & \Sigma_{BG} \\ \Sigma_{GB} & \Sigma_G \end{pmatrix}$. We note $\Sigma^{-1} = \begin{pmatrix} A & C^{\mathrm{T}} \\ C & B \end{pmatrix}$, with $A = (\Sigma_B - \Sigma_{BG}\Sigma_G^{-1}\Sigma_{GB})^{-1}$, $B = (\Sigma_G - \Sigma_{GB}\Sigma_B^{-1}\Sigma_{BG})^{-1}$, and $C = -\Sigma_G^{-1}\Sigma_{GB}A = -B\Sigma_{GB}{\Sigma_B}^{-1}$. Then, with respect to this decomposition,

$$A_2(\varphi_\varepsilon; \mathbf{Z}_B, \varepsilon) = \sum_{\mathbf{s}\in B} \Big(\varepsilon(\mathbf{s})Z(\mathbf{s}) - \ln(1 + e^{\varepsilon(\mathbf{s})})\Big) - \frac{1}{2}\varepsilon_B^{\mathrm{T}} A \varepsilon_B - \frac{1}{2}\varepsilon_G^{\mathrm{T}} B \varepsilon_G - \varepsilon_G^{\mathrm{T}} C \varepsilon_B - \ln|\Sigma| - \frac{n}{2}\ln(2\pi).$$

The gradient of this function is given by $\frac{\partial}{\partial\varepsilon} A_2(\varphi_\varepsilon; \mathbf{Z}_B, \varepsilon) = \begin{pmatrix} V(\varepsilon_B) \\ V(\varepsilon_G) \end{pmatrix}$, with $V(\varepsilon_B) = \mathbf{Z}_B - vec(\frac{e^{\varepsilon_B}}{1+e^{\varepsilon_B}}) - C^{\mathrm{T}}\varepsilon_G - A\varepsilon_B$, and $V(\varepsilon_G) = -C\varepsilon_B - B\varepsilon_G$. Since $V(\varepsilon_G) = 0$, we have $\varepsilon_G = -B^{-1}C\varepsilon_B$, which results in the equation,

$$\mathbf{Z}_B - vec(\frac{e^{\varepsilon_B}}{1+e^{\varepsilon_B}}) + (C^{\mathrm{T}}B^{-1}C - A)\varepsilon_B = 0.$$

This equation can be solved by a Newton-Raphson algorithm; from it, we obtain $\varepsilon_{B,m}$, and then the mode is $\varepsilon_m^{\mathrm{T}} = (\varepsilon_{B,m}^{\mathrm{T}}, \varepsilon_{G,m}^{\mathrm{T}})$, where $\varepsilon_{G,m} = -B^{-1}C\varepsilon_{B,m}$.

A simple calculation gives the Hessian $H(\varepsilon_m)$; that is,

$$-H(\varepsilon_m) = \begin{pmatrix} diag(\frac{e^{\varepsilon_{B,m}}}{(1+e^{\varepsilon_{B,m}})^2}) + A & C^{\mathrm{T}} \\ C & B \end{pmatrix}.$$

References

[1] Aldworth J., Cressie N., 1999. Sampling designs and prediction methods for Gaussian spatial processes. In: Ghosh, S. (Ed.), Multivariate Analysis, Designs of Experiments, and Survey Sampling. Marckel Dekker, Inc., New York, NY, pp. 1-54.

[2] Aitchison J., 1982. The statistical analysis of compositional data. Journal of the Royal Statistical Society, Series B (Methodological), Vol. 44, No. 2. pp.139-177.

[3] Albert J. H., Chib S.,1993. Bayesian analysis of binary and polychotomous response data. Journal of the American Statistical Association, Vol. 88, No. 422, pp. 669-679.

[4] Augustin, N. H., Mugglestone, M. A., and Buckland, S. T., 1996. An autologistic model for the spatial distribution of wildlife. Journal of Applied Ecology, 33(2), pp. 339-347.

[5] Besag J., 1974. Spatial interaction and the statistical analysis of lattice systems. Journal of the Royal Statistical Society, Series B (Methodological), Vol. 36, No. 2., pp.192-236.

[6] Besag J., 1977. Efficiency of pseudo likelihood estimation for simple Gaussian fields. Biometrika 64, pp. 616–618.

[7] Cappé O., Moulines E., Rydén T., 2005. Inference in Hidden Markov Models. Springer Series in Statistics.

[8] Caragea P., Kaiser M., 2009. Autologistic models with interpretable parameters. Journal of Agricultural, Biological, and Environmental Statistics, 14(3), pp. 281-300.

[9] Chen M-H., Shao Q-M, Ibrahim J.G., 2000. Monte Carlo methods in Bayesian computation. Springer Series in Statistics, Springer, New York, NY.

[10] Chib S. , Greenberg E., 1995. Understanding the Metropolis algorithm. The American statistician Vol 49 n4, pp. 327-335.

[11] Chib S., Greenberg E., 1996. Markov Chain Monte Carlo simulation methods in econometrics . Econometric Theory, 12, pp. 409-431.

[12] Cressie N., 1993. Statistics for Spatial Data, rev. ed. Wiley, New York, NY.

[13] Cressie N., Johannesson G., 2008. Fixed Rank Kriging for very large spatial data sets. Journal of the Royal Statistical Society, Series B 70, pp. 209–226.

[14] Cressie N., Wikle C.K., 2011. Statistics for Spatio-Temporal Data. Wiley, Hoboken, NJ.

[15] Descombes X., Mangin J-F., Pechersky E., Sigelle M., 1995. Fine structures preserving Markov model for image processing, in Proceedings of the 9th Scandinavian Conference on Image Analysis, Uppsala, Sweden, pp. 349-356.

[16] Dempster A.P., Laird N., Rubin D., 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Series B 39, pp. 1-38.

[17] Diggle P.J., Tawn J.A., Moyeed R.A., 1988. Model-based geostatistics. Journal of the Royal Statistical Society, Series C 47, pp. 299–350.

[18] Elkink J.A., Calabrese R., 2015. Estimating binary spatial autoregressive models for rare events, the Annual Meeting of the American Political Science Association, Washington DC, pp. 28-31.

[19] Gaetan C., Guyon X., 2010. Spatial statistics and modeling. Springer series in Statistics. Springer, New York, NY.

[20] Gumpertz M. L., Graham J. M., Ristaino J. B., 1997. Autologistic model of spatial pattern of Phytophthora epidemic in bell pepper: effects of soil variables on disease presence. Journal of Agricultural, Biological, and Environmental Statistics, 2, pp. 131-156.

[21] Guyon X., 1995. Random fields on a network: modeling, statistics and applications, Springer, New York, NY.

[22] He F., Zhou J., Zhu, H., 2003. Autologistic regression model for the distribution of vegetation. Journal of Agricultural, Biological, and Environmental Statistics, 8(2), pp. 205-222.

[23] Jiang W., Chen Z., Lei X., Jia K., Wu Y., 2015. Simulating urban land use change by incorporating an autologistic regression model into a CLUE-S model. Journal of Geographical Science, 25(7), pp. 836-850.

[24] Kang E.L., Cressie N., 2011. Bayesian inference for the spatial random effects model. Journal of the American Statistical Association 106, pp. 972–983.

[25] Katzfuss M., Cressie N., 2012. Bayesian hierarchical spatio-temporal smoothing for very large datasets. Environmetrics, 23, pp. 94-107.

[26] Koutsias N., 2003. An autologistic regression model for increasing the accuracy of burned surface mapping using Landsat Thematic Mapper data. International Journal of Remote Sensing, 24, pp. 2199-2204.

[27] Lee J., Kaiser M.S., Cressie N., 2001. Multiway dependence in exponential family conditional distributions. Journal of Multivariate Analysis 79, pp. 171-190.

[28] LeSage J.P., Pace R.K., Lam N., Campanella R., Liu X., 2011. New Orleans business recovery in the aftermath of hurricane Katrina. Journal of the Royal Statistical Society: Series A (Statistics in Society), 174, pp. 1007–1027.

[29] Lindgren F., Rue H., Lindström J., 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. Journal of the Royal Statistical Society: Series B (Statistical Methodology) Volume 73, Issue 4, pp. 423–498.

[30] Lindsay, B.G., 1988. Composite likelihood methods. Contemporary Mathematics 80, pp. 221–239.

[31] Liu J.S., 2008. Monte Carlo strategies in scientific computing. Springer series in Statistics. Springer, New York, NY.

[32] Marsh L., Mittelhammer R.C., Huffaker R.G., 2000. Probit with spatial correlation by field plot:Potato leafroll virus net necrosis in potatoes. Journal of Agricultural, Biological, and Environmental Statistics, 5, pp. 22–36.

[33] McCullagh P., Nelder J.A., 1989. Generalized Linear Models, second ed. Chapman and Hall, London, UK.

[34] McCulloch C.E., Searle S.R., Neuhaus J.M., 2001. Generalized, Linear, and Mixed Models. Wiley, New York, NY.

[35] McGuire J., Brindley T., Bancroft T., 1957. The distribution of European corn borer larvae Pyrausta nubi/alis (Hbn.) in field corn, Biometrics 13, pp. 65-78.

[36] McLachlan G.J., Krishnan T., 2008. The EM Algorithm and Extensions, second ed. Wiley-Interscience, New York, NY.

[37] Moon S., Russell G. J., 2008. Predicting product purchase from inferred customer similarity: An autologistic model approach. Management Science, 54(1), pp. 71-82.

[38] Robert C.P., Casella G., 2004. Monte Carlo Statistical Methods. Springer, New York, NY.

[39] Roberts G.O., Rosenthal J.S., 2001. Optimal scaling for various Metropolis-Hastings algorithms. Statistical science, Vol. 16, No. 4, pp. 351-367

[40] Roy V., Evangelou E., Zhu Z., 2016. Efficient estimation and prediction for the Bayesian binary spatial model with flexible link functions. Biometrics Volume 72, Issue 1, pp. 289–298.

[41] Rue H., Held L., 2005. Gaussian Markov Random Fields: Theory and Applications. Chapman & Hall/CRC, London, UK.

[42] Sanderson R. A., Eyre M. D., Rushton S. P., 2005. Distribution of selected macroinvertebrates in a mosaic of temporary and permanent freshwater ponds as explained by autologistic models. Ecography, 28(3), pp. 355-362.

[43] Schlather M., Malinowski A., Menck P.J., Oesting M., Strokorb K., 2015. Analysis, Simulation and Prediction of Multivariate Random Fields with Package RandomFields. Journal of Statistical Software, 63, issue 8, pp. 1-25.

[44] Sengupta A., Cressie N., 2013. Empirical hierarchical modeling for count data using the Spatial Random Effects model, Spatial Economic Analysis, Vol. 8, No. 3, pp. 389–418.

[45] Sengupta A., Cressie N., 2013. Hierarchical statistical modeling of big spatial datasets using the exponential family of distributions, Spatial Statistics, 4, pp. 14-44.

[46] Tjelmeland H., Besag J., 1998. Markov random fields with higher-order interactions, Scandinavian Journal of Statistics 25, pp. 415-433.

[47] Wang Z., Zheng Y., 2013. Analysis of binary data via a centered spatial-temporal autologistic regression model Environmental and Ecological Statistics March 2013, Volume 20, Issue 1, pp 37-57.

[48] Wikle C.K., Hooten M.B., 2010. A general science-based framework for spatio-temporal dynamical models. Invited discussion paper for Test. 19, pp. 417-451