

Two-scale spatial models for binary data

Cécile Hardouin¹  · Noel Cressie²

Accepted: 6 August 2017 / Published online: 17 August 2017
© Springer-Verlag GmbH Germany 2017

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-

✉ Cécile Hardouin
hardouin@parisnanterre.fr

¹ MODAL'X, Université Paris Nanterre, Nanterre, France

² NIASRA, School of Mathematics and Applied Statistics, University of Wollongong, Wollongong, Australia

logistic model, which belongs to Besag's auto-models class (Besag 1974); 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, Augustin et al. (1996), He et al. (2003) and Sanderson et al. (2005) in ecology, Gumpertz et al. (1997) in epidemiology, Koutsias (2003) in image analysis, and Jiang et al. (2015) and Moon and Russell (2008) in land-use. The auto-logistic model has been reparameterized by Caragea and Kaiser (2009), which helps with interpretation of the parameters, and this was extended recently to the spatio-temporal case by Wang and Zheng (2013).

In a hierarchical framework, when the data are noisy and missing, a Generalised Linear Model (McCullagh and Nelder 1989; McCulloch et al. 2001) can be implemented with a link appropriate for binary data. The logit link is canonical, but models involving other link functions have been developed; for example, in Marsh et al. (2000) and LeSage et al. (2011), a spatial probit model is applied to problems in agriculture and economics, and Roy et al. (2016) introduce a Bayesian spatial probit model that is more robust against extreme observations. When the spatial variable represents presence/absence of a rare event, Elkin and Calabrese (2015) suggest the quantile function of the Generalised 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; see the discussion in Sect. 6. 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 it combines a geostatistical model with a Markov random field (MRF) model in a new approach.

Section 2 is devoted to the description of the model; we display its properties and behaviours by varying different values of the parameters in Sect. 3. The results of this section show how we can identify the Grid resolution Δ . We present parameter estimation in Sect. 4, where both geostatistical and MRF parameters are involved. Section 5 contains an application to modelling the occurrence of corn borer larvae in agricultural fields in Iowa, USA. Finally, in Sect. 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 $D \subset \mathbb{R}^2$; $D = \{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n\}$ is a finite set of sites, with $\mathbf{s}_i = (s_{i1}, s_{i2})^T$ for $i = 1, \dots, n$, and we denote $P(D)$ as its perimeter (or boundary). We are thinking of D as having no missing locations, but the following definition for perimeter covers all cases:

$$P(D) = \{\mathbf{s} \in D: \text{the number of nearest neighbours of } \mathbf{s} \text{ is } < 4\}.$$

For the sake of simplicity, we assume that D is a fine regular lattice, but the situation can be generalised to irregular lattices, as is discussed in Sect. 6.

Let $\mathbf{Z} = (Z(\mathbf{s}) : \mathbf{s} \in D)^T$ be the process on D , taking its values in the state space $E = \{0, 1\}^n$. We consider two scales of spatial dependence, which occur locally at fine-scale resolution 1, and at a coarse-scale resolution Δ , where $\Delta \in \{2, 3, \dots\}$. Here we assume that Δ is 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 Sect. 5 but, in what follows, Δ is not 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 for each site $\mathbf{s} \in P(G)$, the four sites at distance Δ from \mathbf{s} belong to D , see Fig. 1.

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

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

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

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

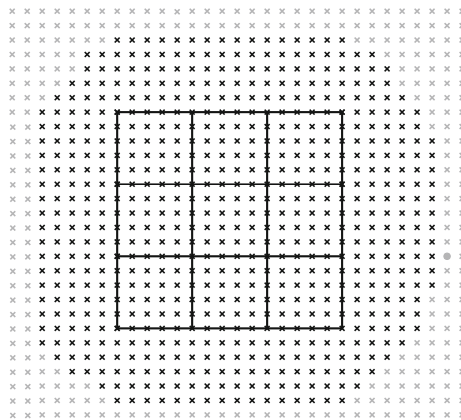


Fig. 1 The Background (*solid crosses*) and the Grid (*solid lines*)

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

We now turn to spatial-process modelling. We start with the Background in Sect. 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 Sect. 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 $\{\mathbf{s}_i : i = 1, \dots, n_B\}$, for each $\mathbf{s}_i \in B$, we write the following independent conditional distributions for $Z_B(\mathbf{s}_i)$ given $\boldsymbol{\varepsilon} = (\varepsilon(\mathbf{s}_1), \dots, \varepsilon(\mathbf{s}_n))^T$ as those given by Bernoulli random variables,

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

where

$$p(\mathbf{s}_i) = \frac{e^{\varepsilon(\mathbf{s}_i)}}{1 + e^{\varepsilon(\mathbf{s}_i)}; \quad i = 1, \dots, n.$$

The hidden process $\boldsymbol{\varepsilon}$ 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; see the discussion in Sect. 6. That is,

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

2.3 Coarse-scale process on the Grid

We define the model on the Grid conditional on the Background using a Markov random field (MRF) 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 $N_G(\mathbf{s}) = \{\mathbf{u} \in G : \mathbf{u} = \mathbf{s} \pm (\Delta, 0), \mathbf{s} \pm (0, \Delta)\}$, and let $\langle \mathbf{u}, \mathbf{v} \rangle_G$ denote the edge between \mathbf{u} and \mathbf{v} , where $\mathbf{u} \in N_G(\mathbf{v})$. Our conditional model for the Grid values is the spatial auto-logistic model:

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

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

From Besag (1974), we know that these conditional distributions are compatible and the joint distribution is,

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

where the energy U is given by

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

and C is the normalising 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 $N_B(\mathbf{s}) \equiv \{\mathbf{u} = (u_1, u_2) \in B : |u_1 - s_1| \leq \Delta, |u_2 - s_2| \leq \Delta\}$ is the set of the 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 $\alpha_B(\mathbf{s}) + \frac{\beta}{2} > 0$ favours state 1 and $\alpha_B(\mathbf{s}) + \frac{\beta}{2} < 0$ favours state 0. In building the model, we want to have the same type of equilibrium behaviour on the Grid as on the Background: We want equilibrium of states 0 and 1, or prevalence of the same state 0 or 1, on both B and G . 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 $\gamma = -\frac{\alpha + \beta}{2}$. When $\gamma = -\frac{\alpha + \beta}{2}$, we can write $\alpha_B(\mathbf{s}) + \frac{\beta}{2} = \alpha \left(\frac{1}{|N_B(\mathbf{s})|} \sum_{\mathbf{u} \in N_B(\mathbf{s})} Z_B(\mathbf{u}) - \frac{1}{2} \right)$, which 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 $0 < \frac{\sum_{\mathbf{u} \in N_B(\mathbf{s})} Z_B(\mathbf{u})}{|N_B(\mathbf{s})|} < \frac{1}{2}$. Thus, the general behaviour on the Background will be reflected in the Grid.

For the reason given above, we calibrate our model with $\gamma = -\frac{\alpha + \beta}{2}$. Hence, we can re-write (4) as:

$$\begin{aligned} \pi(\mathbf{Z}_G | \mathbf{Z}_B) = C^{-1} \exp \left[\sum_{\mathbf{s} \in G} Z_G(\mathbf{s}) \left\{ \alpha \left(\frac{1}{|N_B(\mathbf{s})|} \sum_{\mathbf{u} \in N_B(\mathbf{s})} Z_B(\mathbf{u}) - \frac{1}{2} \right) \right. \right. \\ \left. \left. + \beta \left(\frac{1}{2|N_G(\mathbf{s})|} \sum_{\mathbf{u} \in N_G(\mathbf{s})} Z_G(\mathbf{u}) - \frac{1}{2} \right) \right\} \right], \end{aligned} \quad (6)$$

where the normalising 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 equilibrium value of $\frac{1}{2}$; one is for the Background and the other is for the Grid. The sum related to the Grid is divided by $2|N_G(\mathbf{s})|$ because each edge $\langle \cdot, \cdot \rangle_G$ is added twice in this expression.

Let $[A|B]$ denote the conditional probability distribution of A given B . Then the model (1) and (2) for $[\mathbf{Z}_B]$ and the model (6) for $[\mathbf{Z}_G|\mathbf{Z}_B]$ defines the two-scale spatial model, $[\mathbf{Z}_G, \mathbf{Z}_B]$. In Sect. 4, we discuss estimation of the model's parameters Σ , α , and β . The theoretical properties that would be obtained from $[\mathbf{Z}_G|\mathbf{Z}_B]$ are largely intractable, so in the following section, we illustrate a number of them through simulation.

3 The model's properties through simulation

In this section, we show through simulations that the two-scale spatial model given by (1), (2), and (6) allows both competitive and co-operative behaviours. Further, we look for possible edge effects, and we examine measures that can better represent the spatial dependence. We introduce the *spatial odds ratio*, which accounts 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 Fig. 1. That is $G = \{(s_{i_1}, s_{i_2}) : i_1, i_2 = 5, 9, \dots, 49\}$.

Following the model's hierarchical description in Sect. 2, we simulated forward as follows: In Step 1, we simulated a Gaussian random field $\boldsymbol{\varepsilon}$ on the domain D with spatial covariance Σ , using the the R package RandomFields (Schlather et al. 2015). In Step 2, we simulated independent Bernoulli random variables on the Background. Finally, in Step 3, we used a Gibbs sampler (with 2000 runs) to simulate the auto-logistic model on the Grid.

1. The Gaussian random field $\boldsymbol{\varepsilon}$ 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 $\sigma_\varepsilon^2 = 1$.

2. Conditional on the simulated ε , we simulate independent Bernoulli random variables $Z_B(\mathbf{s})$ on the Background, with parameters $p(\mathbf{s}) = \frac{e^{\varepsilon(\mathbf{s})}}{1 + e^{\varepsilon(\mathbf{s})}}$; $\mathbf{s} \in B$.
3. For the standard auto-logistic model with constant α and β , values of β that would give weak spatial dependence, stronger spatial dependence, and very strong spatial 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 simulations for different values of parameters α and β , both possibly negative and positive. Specifically, we ran simulations for $\alpha \in \{-6, -5, \dots, -1, 1, \dots, 5, 6\}$ 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 $\mathbf{Z} = (\mathbf{Z}_B^T, \mathbf{Z}_G^T)^T$. The properties of our model for \mathbf{Z} are obtained by Monte Carlo averaging of $\{\mathbf{Z}^{(l)} : l = 1, \dots, L\}$.

3.2 Simulation 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 $\mathbf{Z} = (\mathbf{Z}_B^T, \mathbf{Z}_G^T)^T$, and Fig. 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 strengthened by the Background dependence. The proportion of 1s equals 0.5372 on D , and 0.5000 on the Grid.

In order to study edge effects, we considered the first, middle, and last columns of the lattice, and computed 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 , and we see similar behaviour whether in the middle or on the edge of the lattice. Similar results were obtained for all sets of parameters (and we would obviously obtain analogous results if we fix s_2 and look at the average values functions of s_1). 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

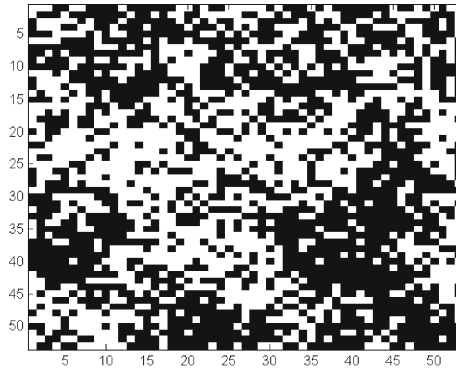


Fig. 2 One realization of \mathbf{Z} with $\sigma_\varepsilon^2 = 1$, $\theta = 5$, $\alpha = \beta = 2$

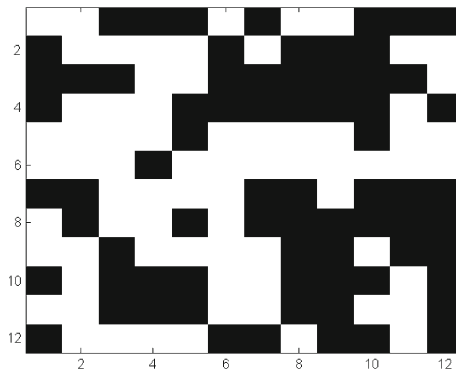


Fig. 3 Representation of \mathbf{Z}_G extracted from \mathbf{Z}

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 notions of relative risk and odds ratio to a binary spatial setting. Since we obtain very similar results for both measures, we only present here 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

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

where $p_{jk}(\mathbf{s}, h, \mathbf{e})$ is the probability of the pair $(Z(\mathbf{s}) = j, Z(\mathbf{s} + h\mathbf{e}) = k)$, and the vector \mathbf{e} defines the direction of interest. The quantity in (7) is a property of the model, which we study here via simulation. Based on the $L = 1600$ simulations, we can define Monte Carlo averages \hat{p}_{jk} that estimate p_{jk} ; these allow us to compute

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

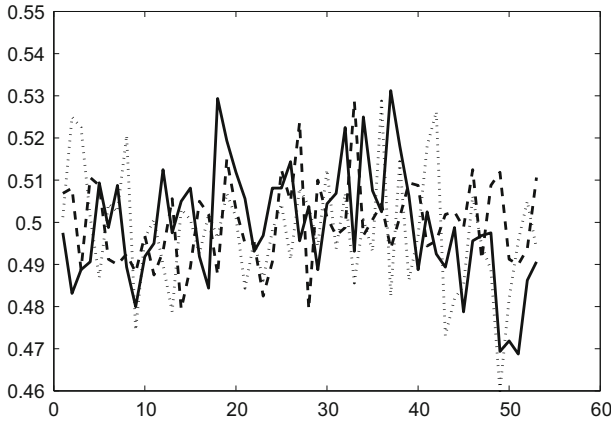


Fig. 4 Average values of sites on the first (line) middle (dotted line) and last (dashed line) columns of the lattice

where $\hat{p}_{jk}(s, h, \mathbf{e}_i) \equiv \frac{1}{L} \sum_{l=1}^L \mathbf{1}_{\{Z^{(l)}(s)=j, Z^{(l)}(s+h\mathbf{e}_i)=k\}}$, and $\mathbf{1}_A$ denotes the indicator function for the set A . 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 spatial-dependence property to the spatial covariance $\text{cov}(Z(\mathbf{s}), Z(\mathbf{s}+h\mathbf{e}_i)) \equiv C_{\mathbf{e}_i}(s, h)$, we computed it from Monte Carlo averages as follows:

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

where $\bar{Z}(s) = \frac{1}{L} \sum_{l=1}^L Z^{(l)}(s)$. Then the global covariance and global SOR are spatial averages, respectively given by

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

and

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

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

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

We plotted the values (9) versus h , $h = 1, \dots, 5\Delta = 20$, 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

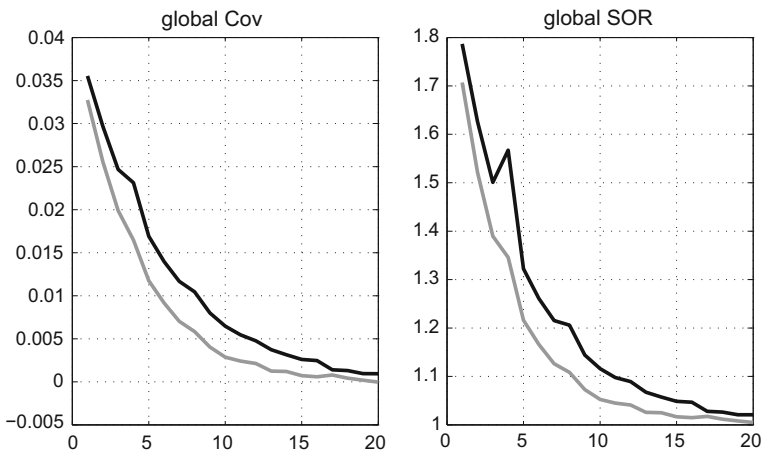


Fig. 5 Global covariance (*left panel*) and global SOR (*right panel*) in horizontal direction \mathbf{e}_1 (*black*) and diagonal direction \mathbf{e}_2 (*grey*) for $\sigma_\varepsilon^2 = 1$, $\theta = 5$, $\alpha = 4$, $\beta = 4$. Here, $\Delta = 4$

over 1600 simulations, and consequently the curves are smooth; if estimated from a single realization, the curves are much more irregular, and some features such as peaks and bumps used to determine Δ will be much harder to discern (e.g., see Sect. 5).

As expected, the covariance values as well as the SOR values increase with parameter θ ; however, more interesting to see is 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 Fig. 6. An important observation is that, in most cases, Δ is more easily detected from the SOR than from the covariance. 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 (i.e., no diagonal dependence) 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 peak 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 use SOR rather than covariance to explore the behaviour of 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 shows no apparent bumps or concavities. We observe a bump in the middle, at lag $h = \Delta = 4$, but none for higher lags. The plot on the right-hand side shows that when β is negative, we can also obtain a dip.

Finally, we observe that both the covariance and the magnitude of the possible peaks increase with $\alpha > 0$, as shown in Fig. 7, and this is more obvious for larger

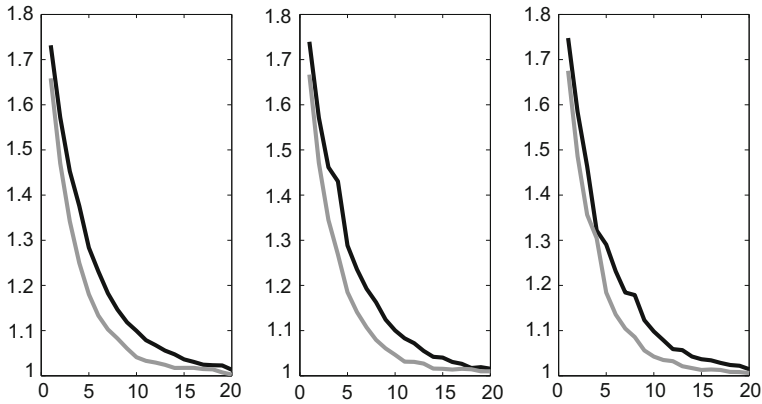


Fig. 6 SOR in horizontal direction \mathbf{e}_1 (black) and diagonal direction \mathbf{e}_2 (grey) for $\sigma_\varepsilon^2 = 1$, $\theta = 5$. Left $(\alpha, \beta) = (3, 1)$. Middle $(\alpha, \beta) = (2, 2)$. Right $(\alpha, \beta) = (10, -5)$. Here, $\Delta = 4$

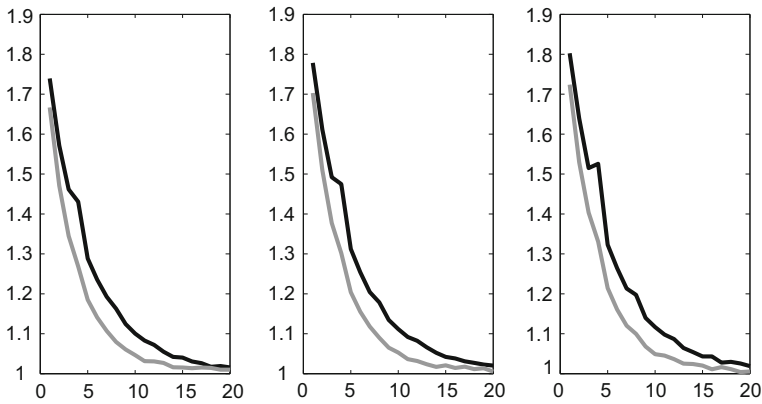


Fig. 7 SOR in horizontal direction \mathbf{e}_1 (black) and diagonal direction \mathbf{e}_2 (grey) for $\sigma_\varepsilon^2 = 1$, $\theta = 5$; $\beta = 2$. Left $\alpha = 2$. Middle $\alpha = 4$. Right $\alpha = 6$. Here, $\Delta = 4$

values of the exponential parameter θ . From Fig. 8, we see that the magnitude of the peaks can be increased by increasing β .

We conducted other simulations to study the covariance and SOR values for different sites, belonging to the Grid or to the Background, located in the domain's centre or in the border region. Our conclusion is that the results seen in Figs. 5, 6, 7 and 8 are essentially repeated for the other sites.

4 Parameter estimation

In this section, we consider the task of obtaining maximum likelihood estimates of the parameters $(\varphi_\varepsilon, \alpha, \beta)$, where $\varphi_\varepsilon \equiv (\sigma_\varepsilon^2, \theta)$ are parameters of the Background process \mathbf{Z}_B . We suppose that Δ is known; in practice, there is information about Δ in the empirical SOR (Sects. 3, 5).

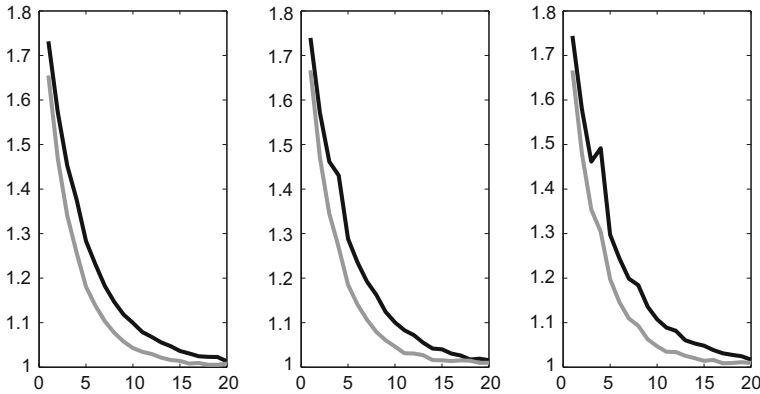


Fig. 8 SOR in horizontal direction \mathbf{e}_1 (black) and vertical direction \mathbf{e}_2 (grey) for $\sigma_\varepsilon^2 = 1$, $\theta = 5$; $\alpha = 2$. Left $\beta = 1$. Middle $\beta = 3$. Right $\beta = 4$. Here, $\Delta = 4$

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 normalising constant, the auto-model parameters α and β are estimated by maximizing the (conditional on the Background) pseudo-likelihood introduced by Besag (1977). The second term involves the latent Gaussian field’s parameters σ_ε^2 and θ ; their estimation in a hierarchical statistical model typically requires an EM algorithm; see Dempster et al. (1977) or McLachlan and Krishnan (2008). The E-step needs the expectation of the latent field $\boldsymbol{\varepsilon}$ given the observations, but we do not know in closed form the distribution over which the expectation is taken. There are several ways to overcome this issue: A common approach is to use Monte Carlo procedures; see for instance Robert and Casella (2004) and Cappé et al. (2005). Here, instead, we use Laplace approximations to approximate the intractable integrals. Recall that the notation $[A|B]$ denotes the conditional probability distribution of A given B .

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

$$\begin{aligned}
 [\mathbf{Z} | \varphi_\varepsilon, \alpha, \beta] &= \int [\boldsymbol{\varepsilon}, \mathbf{Z}_B, \mathbf{Z}_G | \varphi_\varepsilon, \alpha, \beta] d\boldsymbol{\varepsilon} \\
 &= \int [\mathbf{Z}_G | \mathbf{Z}_B, \boldsymbol{\varepsilon}, \varphi_\varepsilon, \alpha, \beta][\mathbf{Z}_B | \boldsymbol{\varepsilon}, \varphi_\varepsilon, \alpha, \beta][\boldsymbol{\varepsilon} | \varphi_\varepsilon, \alpha, \beta] d\boldsymbol{\varepsilon} \\
 &= [\mathbf{Z}_G | \mathbf{Z}_B, \alpha, \beta] \times \int [\mathbf{Z}_B | \boldsymbol{\varepsilon}][\boldsymbol{\varepsilon} | \varphi_\varepsilon] d\boldsymbol{\varepsilon}.
 \end{aligned}
 \tag{10}$$

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

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

The normalising constant $C(\alpha, \beta)$ is potentially problematic, since it depends on α and β ; 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 MRF context, another method is to replace maximizing the likelihood with maximizing the (conditional) pseudo likelihood (Besag 1974, 1977; Cressie 1993; Gaetan and Guyon 2010), which allows fast and easy computation. The pseudo likelihood here would be the product of conditional probabilities, $\prod_{s \in G} \pi_s(Z_G(s) | Z_B, \alpha, \beta)$, expressed as a function of α and β . Of course, maximum pseudo likelihood estimators are less efficient (Besag 1977).

The second term in (10) involves the Gaussian distribution of ϵ and the following conditional distribution of Z_B given ϵ ,

$$[Z_B | \epsilon] = \prod_{s \in B} p(s)^{Z(s)}(1 - p(s))^{1-Z(s)} = \prod_{s \in B} e^{\epsilon(s)Z(s)} \frac{1}{1 + e^{\epsilon(s)}}.$$

Now, the underlying spatial process ϵ is not observed, but we can define the complete likelihood based on both Z and the missing data ϵ .

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

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

where

$$A_1(\alpha, \beta) = \sum_{s \in G} \log[Z_G(s) | Z_B, \alpha, \beta],$$

and

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

where recall that $[Z_G(\cdot) | Z_B, \alpha, \beta]$ and Σ are given in (3) and (2), respectively.

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 Z_B on the Background B . The second term will be used to obtain an EM estimate of $\varphi_\epsilon = (\sigma_\epsilon^2, \theta)$.

4.1 Estimation of the Grid parameters

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

It is straightforward to obtain

$$A_1(\alpha, \beta) = \sum_{\mathbf{s} \in G} \left(\alpha(V(\mathbf{s}) - \frac{1}{2}) - \frac{\beta}{2} + \frac{\beta}{8} \sum_{\mathbf{u} \in N_G(\mathbf{s})} Z_G(\mathbf{u}) \right) Z_G(\mathbf{s}) - \sum_{\mathbf{s} \in G} \log \left(1 + \exp \left\{ \alpha(V(\mathbf{s}) - \frac{1}{2}) - \frac{\beta}{2} + \frac{\beta}{8} \sum_{\mathbf{u} \in N_G(\mathbf{s})} Z_G(\mathbf{u}) \right\} \right), \quad (12)$$

where $V(\mathbf{s}) \equiv \frac{\sum_{\mathbf{u} \in N_B(\mathbf{s})} Z_G(\mathbf{u})}{|N_B(\mathbf{s})|}$.

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 relevant component of the pseudo complete log likelihood. Recall that $\varphi_\varepsilon = (\sigma_\varepsilon^2, \theta)$ and

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

Since $\boldsymbol{\varepsilon}$ has not been observed, estimation is performed using the EM algorithm; see [Dempster et al. \(1977\)](#) and [McLachlan and Krishnan \(2008\)](#). Here we define

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

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

- the E-step computes the expectation $q(\varphi_\varepsilon, \hat{\varphi}_\varepsilon^{(k-1)})$, and
- the M-step maximizes $q(\varphi_\varepsilon, \hat{\varphi}_\varepsilon^{(k-1)})$ with respect to φ_ε ; that is, $\hat{\varphi}_\varepsilon^{(k)} = \arg \max_{\varphi_\varepsilon} q(\varphi_\varepsilon, \hat{\varphi}_\varepsilon^{(k-1)})$.

In our case, the E-step is problematic, since we do not have a closed-form expression for the conditional distribution of $\boldsymbol{\varepsilon}$ given the observations \mathbf{Z} . There are several possible approaches, one being to implement a stochastic EM (SEM) algorithm (e.g., [Robert and Casella 2004](#); [McLachlan and Krishnan 2008](#)), 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 $\boldsymbol{\varepsilon}$. Choosing the “right” proposal density (see Chib and Greenberg 1995; Roberts and Rosenthal 2001) can be problematic, and when datasets are large, computations can be very slow.

4.2.2 Laplace approximations in the EM algorithm

We now derive Laplace approximations (LA) to approximate the E-step 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 $A_2(\varphi_\varepsilon; \mathbf{Z}_B, \boldsymbol{\varepsilon})$. Let us denote $\boldsymbol{\varepsilon}_m$ as the vector maximizing $A_2(\varphi_\varepsilon; \mathbf{Z}_B, \boldsymbol{\varepsilon})$; then, a second-order Taylor series expansion for $A_2(\varphi_\varepsilon; \mathbf{Z}_B, \boldsymbol{\varepsilon})$ around $\boldsymbol{\varepsilon}_m$ yields:

$$A_2(\varphi_\varepsilon; \mathbf{Z}_B, \boldsymbol{\varepsilon}) = A_2(\varphi_\varepsilon; \mathbf{Z}_B, \boldsymbol{\varepsilon}_m) + (\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}_m)^T \frac{\partial}{\partial \boldsymbol{\varepsilon}} A_2(\varphi_\varepsilon; \mathbf{Z}_B, \boldsymbol{\varepsilon}) + \frac{1}{2} (\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}_m)^T \frac{\partial^2}{\partial \boldsymbol{\varepsilon} \partial \boldsymbol{\varepsilon}^T} A_2(\varphi_\varepsilon; \mathbf{Z}_B, \boldsymbol{\varepsilon}) \Big|_{\boldsymbol{\varepsilon}=\boldsymbol{\varepsilon}_m} (\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}_m) + \dots$$

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

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

where $H(\boldsymbol{\varepsilon}_m) = \frac{\partial^2}{\partial \boldsymbol{\varepsilon} \partial \boldsymbol{\varepsilon}^T} A_2(\varphi_\varepsilon; \mathbf{Z}_B, \boldsymbol{\varepsilon}) \Big|_{\boldsymbol{\varepsilon}=\boldsymbol{\varepsilon}_m}$.

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

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

We deduce from (15) that $E[\boldsymbol{\varepsilon} \mid \mathbf{Z}_B, \varphi_\varepsilon] \simeq \boldsymbol{\varepsilon}_m$ and $\text{var}(\boldsymbol{\varepsilon} \mid \mathbf{Z}_B, \varphi_\varepsilon) \simeq -H(\boldsymbol{\varepsilon}_m)^{-1}$.

The remaining term in (13), for which we need an approximation of its expectation, is $E[\ln(1 + e^{\boldsymbol{\varepsilon}(\mathbf{s})}) \mid \mathbf{Z}_B, \varphi_\varepsilon]$. We again use a second-order Taylor-series expansion, now of $\ln(1 + e^{\boldsymbol{\varepsilon}(\mathbf{s})})$ around $\boldsymbol{\varepsilon}_m(\mathbf{s})$:

$$\ln(1 + e^{\boldsymbol{\varepsilon}(\mathbf{s})}) = \ln(1 + e^{\boldsymbol{\varepsilon}_m(\mathbf{s})}) + (\boldsymbol{\varepsilon}(\mathbf{s}) - \boldsymbol{\varepsilon}_m(\mathbf{s})) \frac{e^{\boldsymbol{\varepsilon}_m(\mathbf{s})}}{1 + e^{\boldsymbol{\varepsilon}_m(\mathbf{s})}} + \frac{1}{2} (\boldsymbol{\varepsilon}(\mathbf{s}) - \boldsymbol{\varepsilon}_m(\mathbf{s}))^2 \frac{e^{\boldsymbol{\varepsilon}_m(\mathbf{s})}}{(1 + e^{\boldsymbol{\varepsilon}_m(\mathbf{s})})^2} + \dots$$

Taking the expectation, we obtain

$$E[\ln(1 + e^{\varepsilon(\mathbf{s})}) \mid \mathbf{Z}_B, \varphi_\varepsilon] \simeq \ln(1 + e^{\varepsilon_m(\mathbf{s})}) - \frac{1}{2} \frac{e^{\varepsilon_m(\mathbf{s})}}{(1 + e^{\varepsilon_m(\mathbf{s})})^2} (H(\boldsymbol{\varepsilon}_m)^{-1})_{\text{ss}},$$

where A_{ss} denotes the diagonal element of the square matrix A corresponding to the spatial location \mathbf{s} .

Finally, we obtain the Laplace approximation of the expectation as:

$$E[A_2(\varphi_\varepsilon; \mathbf{Z}_B, \boldsymbol{\varepsilon}) \mid \mathbf{Z}_B, \hat{\varphi}_\varepsilon^{(k)}] \simeq \tilde{q}(\varphi_\varepsilon, \hat{\varphi}_\varepsilon^{(k)}),$$

where

$$\begin{aligned} \tilde{q}(\varphi_\varepsilon, \hat{\varphi}_\varepsilon^{(k)}) = & - \sum_{\mathbf{s} \in B} \left[\ln(1 + e^{\varepsilon_m^{(k)}(\mathbf{s})}) + \frac{1}{2} \frac{e^{\varepsilon_m^{(k)}(\mathbf{s})}}{(1 + e^{\varepsilon_m^{(k)}(\mathbf{s})})^2} (H(\boldsymbol{\varepsilon}_m^{(k)})^{-1})_{\text{ss}} \right] \\ & + \sum_{\mathbf{s} \in B} \varepsilon_m^{(k)}(\mathbf{s}) Z_B(\mathbf{s}) - \frac{1}{2} \ln(\det \Sigma) - \frac{1}{2} \boldsymbol{\varepsilon}_m^{(k)T} \Sigma^{-1} \boldsymbol{\varepsilon}_m^{(k)} \\ & - \frac{1}{2} \text{tr}(\Sigma^{-1} (-H(\boldsymbol{\varepsilon}_m^{(k)})^{-1})) - \frac{|B|}{2} \log(2\pi). \end{aligned} \tag{16}$$

Finally, starting with an initialization $\hat{\varphi}_\varepsilon^{(0)}$, the k -th iteration of the EM algorithm is achieved in the following two steps: at the E-step, we compute the mode, $\boldsymbol{\varepsilon}_m^{(k-1)}$, of $A_2(\hat{\varphi}_\varepsilon^{(k-1)}; \mathbf{Z}_B, \boldsymbol{\varepsilon})$, the Hessian $-H(\boldsymbol{\varepsilon}_m^{(k-1)})$, and $\tilde{q}(\varphi_\varepsilon, \hat{\varphi}_\varepsilon^{(k-1)})$. We give the details for the computation of the mode $\boldsymbol{\varepsilon}_m$ and the matrix $H(\boldsymbol{\varepsilon}_m)$ in the Appendix.

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

$$f(\theta, \sigma_\varepsilon^2) = \frac{1}{\sigma_\varepsilon^2} \boldsymbol{\varepsilon}_m^T Q(\theta)^{-1} \boldsymbol{\varepsilon}_m + \frac{1}{\sigma_\varepsilon^2} \text{tr}(Q(\theta)^{-1} (-H(\boldsymbol{\varepsilon}_m)^{-1})) + n \ln \sigma_\varepsilon^2 + \ln(\det Q(\theta)),$$

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:

$$\sigma_\varepsilon^2(\theta) = \frac{1}{n} \left[\boldsymbol{\varepsilon}_m^T Q(\theta)^{-1} \boldsymbol{\varepsilon}_m + \text{tr}(Q(\theta)^{-1} (-H(\boldsymbol{\varepsilon}_m)^{-1})) \right]. \tag{17}$$

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

$$\begin{aligned} g(\theta, \hat{\varphi}_\varepsilon^{(k-1)}) = & n \ln \left[\frac{1}{n} \left[\boldsymbol{\varepsilon}_m^{(k-1)T} Q(\theta)^{-1} \boldsymbol{\varepsilon}_m^{(k-1)} + \text{tr}(Q(\theta)^{-1} (-H(\boldsymbol{\varepsilon}_m^{(k-1)})^{-1})) \right] \right] \\ & + \ln(\det Q(\theta)). \end{aligned}$$

Table 1 Summary of two-scale model parameter estimates

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

4.3 Simulation experiments

We ran $L = 1600$ simulations, as described in Sect. 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 Sects. 4.1 and 4.2. We present in Table 1 the means and standard deviations of the estimates, obtained from the 1600 simulations.

We observe a negative bias, especially for the Background parameters. This bias may come from the Laplace approximation and might be reduced using higher-order terms. Looking inside the EM algorithm, we observe that, at each iteration, very often the new value of the estimate $\hat{\theta}^{(k)}$ is less than $\hat{\theta}^{(k-1)}$. The initial bias for parameter $\hat{\theta}$ transfers to a bias for $\widehat{\sigma}_{\varepsilon}^2$, since the latter is obtained directly from $\hat{\theta}$ by (17).

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.

Estimation of the Grid parameters is obtained from the 144 observations in \mathbf{Z}_G , which is not a large number. 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 estimate $\hat{\theta}$. For lattice size 53×53 , the parameter $\theta = 20$ induces high spatial dependence, and a typical realization on this lattice does not show enough contrast to estimate θ accurately. Nevertheless, a model-based spatial 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 (McGuire et al. 1957). 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. (2001) 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

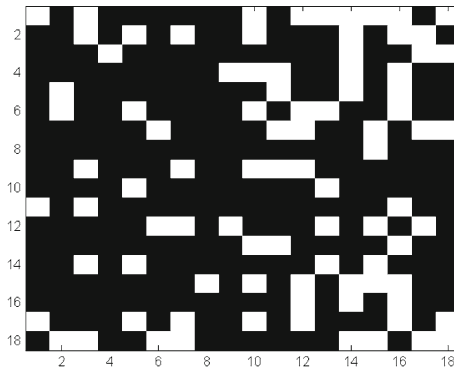


Fig. 9 Corn borers data given in Lee et al. (2001)

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 $u_i \in \{1, \dots, 18\}$ denote the E-W position and $v_i \in \{1, \dots, 18\}$ the N-S position of subplot i in a regular 18×18 lattice S ; define $\mathbf{s}_i = (u_i, v_i)$ and, for $i = 1, \dots, 324$,

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

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

In their paper, Lee et al. considered 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 sizes 3 and 4 associated with the eight-nearest-neighbour system. They show that the extra cliques contained in Model 2, due 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 neighbours, the spread of infestation among spatial subplots is higher if infested neighbours occur in all directions rather than in just one direction. Having infested neighbours in all directions implies that biological conditions are favourable for infestation throughout the whole immediate region, as opposed to a situation in which conditions are favourable 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, $\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \Sigma)$, 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

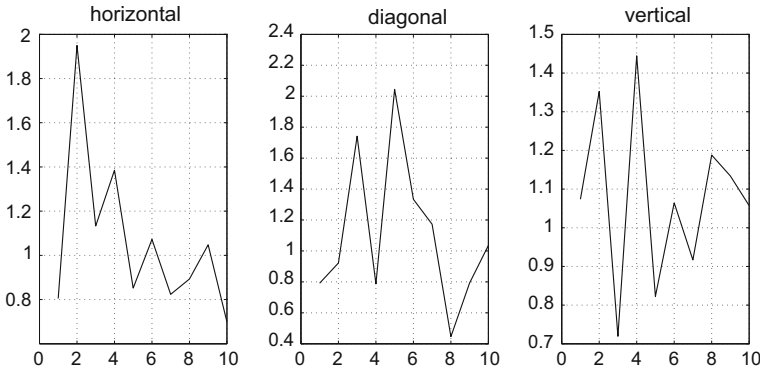


Fig. 10 Empirical SOR in directions \mathbf{e}_1 (left), \mathbf{e}_2 (middle), and \mathbf{e}_3 (right), for lags $h = 1, \dots, 10$. The units of lag in direction \mathbf{e}_2 have to be multiplied by $\sqrt{2}$ to compare to the units of lag in the directions \mathbf{e}_2 and \mathbf{e}_3

by (4) and (5), and with parameters α and β . We estimate the parameters $(\sigma_\epsilon^2, \theta, \alpha, \beta)$ using the procedure described in Sect. 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 a single dataset, we compute the empirical SOR at lag h in direction \mathbf{e} , as follows:

$$\widehat{SOR}_e(h) = \frac{\hat{p}_{00}(h, \mathbf{e}) \times \hat{p}_{11}(h, \mathbf{e})}{\hat{p}_{01}(h, \mathbf{e}) \times \hat{p}_{10}(h, \mathbf{e})}, \tag{18}$$

where $\hat{p}_{jk}(h, \mathbf{e}) = \frac{1}{|D_{\mathbf{e},h}|} \sum_{\mathbf{s} \in D_{\mathbf{e},h}} \mathbf{1}_{\{Z(\mathbf{s})=j, Z(\mathbf{s}+\mathbf{h}\mathbf{e})=k\}}$, $j, k = 0, 1$, and recall that $D_{\mathbf{e},h} = \{\mathbf{s} \in D : \mathbf{s} + \mathbf{h}\mathbf{e} \in D\}$.

In Fig. 10, we present the plots of the spatial odds ratios as a function of h in the three directions, \mathbf{e}_1 , \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 $\Delta_{N-S} = 4$. Based on this exploratory spatial data analysis, 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 involve, respectively, 64 and 32 sites.

Now we want to specify the location of the Grids G_1 and G_2 ; for that purpose, we again use the empirical SOR. We compute $\widehat{SOR}_{\mathbf{e}_1}(\Delta_{E-W})$, $\widehat{SOR}_{\mathbf{e}_3}(\Delta_{N-S})$, and

$$\widehat{SOR}_G(\Delta_{E-W}, \Delta_{N-S}) = \frac{(\hat{p}_{00}(\Delta_{E-W}, \mathbf{e}_1) + \hat{p}_{00}(\Delta_{N-S}, \mathbf{e}_3)) \times (\hat{p}_{11}(\Delta_{E-W}, \mathbf{e}_1) + \hat{p}_{11}(\Delta_{N-S}, \mathbf{e}_3))}{(\hat{p}_{01}(\Delta_{E-W}, \mathbf{e}_1) + \hat{p}_{01}(\Delta_{N-S}, \mathbf{e}_3)) \times (\hat{p}_{10}(\Delta_{E-W}, \mathbf{e}_1) + \hat{p}_{10}(\Delta_{N-S}, \mathbf{e}_3))}, \tag{19}$$

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

	$\hat{\alpha}$	$\hat{\beta}$	$A_1(\hat{\alpha}, \hat{\beta})$	$\hat{\sigma}_\varepsilon^2$	$\hat{\theta}$	$\tilde{q}(\hat{\varphi}_\varepsilon^{(k_f)}, \hat{\varphi}_\varepsilon^{(k_f)})$
Grid $G_1(2, 2)$	1.0089	4.8032	-0.3944	0.7250	37.8321	220.9151
Grid $G_2(4, 2)$	4.0538	1.7894	-0.2314	0.9432	21.6741	52.0199

for different positionings. We specify each location of the Grid by giving the coordinates (in terms of row and column numbers) of its north-west corner. For both Grid G_1 and Grid G_2 , the largest $\widehat{SOR}_{G_1}(\Delta_{E-W}, \Delta_{N-S})$ and $\widehat{SOR}_{G_2}(\Delta_{E-W}, \Delta_{N-S})$ values were obtained for positioning the Grid at (2, 2).

For each of the two Grids, we estimated the parameters using the procedure described in Sect. 4. We note that since the sizes of the Grids are small, we could use the true likelihood instead of the pseudo-likelihood for estimation of the Grid parameters, but we chose here to follow the general procedure given in Sect. 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, $A_1(\hat{\alpha}, \hat{\beta})$ and $\tilde{q}(\hat{\varphi}_\varepsilon^{(k_f)}, \hat{\varphi}_\varepsilon^{(k_f)})$, defined by (12) and (16), where for the latter, $\hat{\varphi}_\varepsilon^{(k_f)}$ is the value of $\hat{\varphi}_\varepsilon$ obtained at the final iteration k_f of the EM algorithm.

First, we can see that strong spatial dependence is estimated for both models. Indeed, the scale parameter θ of the spatial covariance is at least 21, which is quite large; and for both Grids, the interaction parameter β indicates strong spatial dependence. However, we can see that spatial dependence is much stronger in the model associated with Grid G_1 ; indeed, both spatial-interaction parameters θ (in the Background) and β (in the Grid) are greater than their values obtained for Grid G_2 . Grid G_2 has a larger estimate of α and a smaller estimate of β . Finally, the values of $\tilde{q}(\hat{\varphi}_\varepsilon^{(k_f)}, \hat{\varphi}_\varepsilon^{(k_f)})$ indicate preference for the two-scale model with Grid G_1 .

Now, since, Grid G_2 has different scale on the E-W and N-S directions, it is interesting to consider different values for parameter β . Thus, we adapt our model and consider the East-West two-nearest neighbourhood and North-South two-nearest neighbourhood, defined respectively by, for each $\mathbf{s} \in G$,

$$N_{G,E-W}(\mathbf{s}) = \{\mathbf{u} \in G : \mathbf{u} = \mathbf{s} \pm (0, \Delta_{E-W})\},$$

and

$$N_{G,N-S}(\mathbf{s}) = \{\mathbf{u} \in G : \mathbf{u} = \mathbf{s} \pm (\Delta_{N-S}, 0)\}.$$

We apply the procedure for both Grids G_1 and G_2 . The results are summarized in Table 3. Estimates of the Background parameters are not impacted. In both cases (Grids G_1 and G_2), the estimate value of α in the anisotropic model is a bit lowered compared to the isotropic model, which goes with the fact that $\hat{\beta}_{E-W} + \hat{\beta}_{N-S}$ is a bit larger than $\hat{\beta}$. The log-pseudo-likelihood values $A_1(\hat{\alpha}, \hat{\beta}_{E-W}, \hat{\beta}_{N-S})$ are quite similar to those obtained for the isotropic model. Finally, we compare $\hat{\beta}_{E-W}$ and $\hat{\beta}_{N-S}$; we have $\hat{\beta}_{E-W} < \hat{\beta}_{N-S}$ for Grid G_1 , and $\hat{\beta}_{E-W} > \hat{\beta}_{N-S}$ for Grid G_2 ; this is in accordance

Table 3 Estimates of the Grid parameters and log likelihood values for Grids G_1 and G_2 , for the non isotropic model

	$\hat{\alpha}$	$\hat{\beta}_{E-W}$	$\hat{\beta}_{N-S}$	$A_1(\hat{\alpha}, \hat{\beta}_{E-W}, \hat{\beta}_{N-S})$
Grid $G_1(2, 2)$	0.7523	2.2126	2.8772	-0.3806
Grid $G_2(4, 2)$	4.2634	1.1966	0.8540	-0.2343

with the empirical Spatial Odds Ratio $\widehat{SOR}_{e_1}(\Delta_{E-W})$ and $\widehat{SOR}_{e_3}(\Delta_{N-S})$, equaling respectively 4.75 and 5.42 for Grid G_1 and 3.17 and 3.00 for Grid G_2 .

6 Discussion and conclusions

In this paper, we have developed a hierarchical (in spatial 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 Background dependence is captured through Bernoulli random variables with success probabilities given by the logit transformation of the latent process. Then, conditional on this Background, binary data on a Grid at a coarser resolution is overlaid; 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 Sect. 4 requires computation of an inverse covariance matrix, which can be problematic for large data sets. In this case, we could consider modelling $\boldsymbol{\varepsilon}$ with a Spatial Random Effects (SRE) model, as described by [Cressie and Johannesson \(2008\)](#); see also [Kang and Cressie \(2011\)](#) and [Sengupta and Cressie \(2013a\)](#). Other reduced-rank approaches could also be used (e.g. [Wikle and Hooten 2010](#)), or indeed the inverse covariance matrix could be modelled directly (e.g., [Lindgren et al. 2011](#)).

The methodology presented here for regular lattices could be generalised to irregular lattices. The Background (representing fine-scale variation) would be irregular, but the Grid (representing coarse-scale variation) is regular and obtained by moving nearest locations on the irregular lattice to the Grid locations. In [Cressie and Kornak \(2003\)](#), it is shown that the effect on spatial variability is small.

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

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

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

where $\mathbf{X}(\cdot)$ 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 generalised linear model. The binary

Markov random field on the superimposed Grid could quite naturally be generalised to a spatial auto-model from the same member of the exponential family that is used to generate the Background.

Simple assumptions on spatial dependence on the Grid 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. (2001). This might be particularly useful within the context of image analysis, where higher-order interactions allow for improved image processing (e.g., see Descombes et al. 1995; Tjelmeland and Besag 1998).

Appendix

In Sect. 4, we use the EM algorithm where the E-step is to compute the Laplace approximation, $\tilde{q}(\varphi_\varepsilon, \hat{\varphi}_\varepsilon^{(k)})$, given by (16); its expression depends on the mode $\boldsymbol{\varepsilon}_m$ and on the Hessian $H(\boldsymbol{\varepsilon}_m)$ of $A_2(\varphi_\varepsilon; \mathbf{Z}_B, \boldsymbol{\varepsilon})$, computed at the mode. Recall that the process $\boldsymbol{\varepsilon}$ can be written as a vector, $\boldsymbol{\varepsilon}^T = (\boldsymbol{\varepsilon}_B^T, \boldsymbol{\varepsilon}_G^T)$, with covariance matrix, $\Sigma = \begin{pmatrix} \Sigma_B & \Sigma_{BG} \\ \Sigma_{GB} & \Sigma_G \end{pmatrix}$. We note $\Sigma^{-1} = \begin{pmatrix} A & C^T \\ C & B \end{pmatrix}$, 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,

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

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

$$\mathbf{Z}_B - \text{vec} \left(\frac{e^{\boldsymbol{\varepsilon}_B}}{1 + e^{\boldsymbol{\varepsilon}_B}} \right) + (C^T B^{-1} C - A) \boldsymbol{\varepsilon}_B = 0.$$

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

A simple calculation gives the Hessian $H(\boldsymbol{\varepsilon}_m)$; that is,

$$-H(\boldsymbol{\varepsilon}_m) = \begin{pmatrix} \text{diag} \left(\frac{e^{\boldsymbol{\varepsilon}_{B,m}}}{(1 + e^{\boldsymbol{\varepsilon}_{B,m}})^2} \right) + A & C^T \\ C & B \end{pmatrix}.$$

Acknowledgements 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 Project, DP 150104576.

References

- 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, pp 1–54
- Aitchison J (1982) The statistical analysis of compositional data. *J R Stat Soc Ser B* 44(2):139–177
- Albert JH, Chib S (1993) Bayesian analysis of binary and polychotomous response data. *J Am Stat Assoc* 88(422):669–679
- Augustin NH, Muggleston MA, Buckland ST (1996) An autologistic model for the spatial distribution of wildlife. *J Appl Ecol* 33(2):339–347
- Besag J (1974) Spatial interaction and the statistical analysis of lattice systems. *J R Stat Soc Ser B* 36(2):192–236
- Besag J (1977) Efficiency of pseudo likelihood estimation for simple Gaussian fields. *Biometrika* 64:616–618
- Cappé O, Moulines E, Rydén T (2005) *Inference in hidden Markov models*. Springer series in statistics. Springer, New York
- Caragea P, Kaiser M (2009) Autologistic models with interpretable parameters. *J Agric Biol Environ Stat* 14(3):281–300
- Chen M-H, Shao Q-M, Ibrahim JG (2000) *Monte Carlo methods in Bayesian computation*. Springer series in statistics. Springer, New York
- Chib S, Greenberg E (1995) Understanding the Metropolis algorithm. *Am Stat* 49(4):327–335
- Chib S, Greenberg E (1996) Markov Chain Monte Carlo simulation methods in econometrics. *Econom Theory* 12:409–431
- Cressie N (1993) *Statistics for spatial data*, rev edn. Wiley, New York
- Cressie N, Johannesson G (2008) Fixed rank kriging for very large spatial data sets. *J R Stat Soc Ser B* 70:209–226
- Cressie N, Kornak J (2003) Spatial statistics in the presence of location error with an application to remote sensing of the environment. *Stat Sci* 18(4):436–456
- Cressie N, Wikle CK (2011) *Statistics for spatio-temporal data*. Wiley, Hoboken
- 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
- Dempster AP, Laird N, Rubin D (1977) Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc Ser B* 39:1–38
- Diggle PJ, Tawn JA, Moyeed RA (1988) Model-based geostatistics. *J R Stat Soc Ser C* 47:299–350
- Elkink JA, Calabrese R (2015) Estimating binary spatial autoregressive models for rare events. In: *The annual meeting of the american political science association*, Washington DC, pp 28–31
- Gaetan C, Guyon X (2010) *Spatial statistics and modeling*. Springer series in statistics. Springer, New York
- Gumpertz ML, Graham JM, Ristaino JB (1997) Autologistic model of spatial pattern of *Phytophthora* epidemic in bell pepper: effects of soil variables on disease presence. *J Agric Biol Environ Stat* 2:131–156
- Guyon X (1995) *Random fields on a network: modeling. Statistics and applications*. Springer, New York
- He F, Zhou J, Zhu H (2003) Autologistic regression model for the distribution of vegetation. *J Agric Biol Environ Stat* 8(2):205–222
- 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. *J Geogr Sci* 25(7):836–850
- Kang EL, Cressie N (2011) Bayesian inference for the spatial random effects model. *J Am Stat Assoc* 106:972–983
- Katzfuss M, Cressie N (2012) Bayesian hierarchical spatio-temporal smoothing for very large datasets. *Environmetrics* 23:94–107
- Koutsias N (2003) An autologistic regression model for increasing the accuracy of burned surface mapping using landsat thematic mapper data. *Int J Remote Sens* 24:2199–2204

- Lee J, Kaiser MS, Cressie N (2001) Multiway dependence in exponential family conditional distributions. *J Multivar Anal* 79:171–190
- LeSage JP, Pace RK, Lam N, Campanella R, Liu X (2011) New Orleans business recovery in the aftermath of Hurricane Katrina. *J R Stat Soc Ser A* 174:1007–1027
- Lindgren F, Rue H, Lindström J (2011) An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *J R Stat Soc Ser B* 73(4):423–498
- Lindsay BG (1988) Composite likelihood methods. *Contemp Math* 80:221–239
- Liu JS (2008) Monte Carlo strategies in scientific computing. Springer series in statistics. Springer, New York
- Marsh L, Mittelhammer RC, Huffaker RG (2000) Probit with spatial correlation by field plot: Potato leafroll virus net necrosis in potatoes. *J Agric Biol Environ Stat* 5:22–36
- McCullagh P, Nelder JA (1989) Generalized linear models, 2nd edn. Chapman and Hall, London
- McCulloch CE, Searle SR, Neuhaus JM (2001) Generalized, linear, and mixed models. Wiley, New York
- McGuire J, Brindley T, Bancroft T (1957) The distribution of European corn borer larvae *Pyrausta nubi/alisi* (Hbn.) in field corn. *Biometrics* 13:65–78
- McLachlan GJ, Krishnan T (2008) The EM algorithm and extensions, 2nd edn. Wiley, New York
- Moon S, Russell GJ (2008) Predicting product purchase from inferred customer similarity: an autologistic model approach. *Manag Sci* 54(1):71–82
- Robert CP, Casella G (2004) Monte Carlo statistical methods. Springer, New York
- Roberts GO, Rosenthal JS (2001) Optimal scaling for various Metropolis-Hastings algorithms. *Stat Sci* 16(4):351–367
- Roy V, Evangelou E, Zhu Z (2016) Efficient estimation and prediction for the Bayesian binary spatial model with flexible link functions. *Biometrics* 72(1):289–298
- Rue H, Held L (2005) Gaussian Markov random fields: theory and applications. Chapman & Hall, London
- Sanderson RA, Eyre MD, Rushton SP (2005) Distribution of selected macroinvertebrates in a mosaic of temporary and permanent freshwater ponds as explained by autologistic models. *Ecography* 28(3):55–362
- Schlather M, Malinowski A, Menck PJ, Oesting M, Storkorb K (2015) Analysis, simulation and prediction of multivariate random fields with package RandomFields. *J Stat Softw* 63(8):1–25
- Sengupta A, Cressie N (2013a) Empirical hierarchical modeling for count data using the Spatial Random Effects model. *Spat Econ Anal* 8(3):389–418
- Sengupta A, Cressie N (2013b) Hierarchical statistical modeling of big spatial datasets using the exponential family of distributions. *Spat Stat* 4:14–44
- Tjelmeland H, Besag J (1998) Markov random fields with higher-order interactions. *Scand J Stat* 25:415–433
- Wang Z, Zheng Y (2013) Analysis of binary data via a centered spatial-temporal autologistic regression model. *Environ Ecol Stat* 20(1):37–57
- Wikle CK, Hooten MB (2010) A general science-based framework for spatio-temporal dynamical models. *Test* 19:417–451