THE UNIVERSITY OF TEXAS AT SAN ANTONIO, COLLEGE OF BUSINESS

Working Paper SERIES

Date October 12, 2017

WP # 0003MSS-496-2017

# GEOSTATISTICAL BINARY DATA: MODELS, PROPERTIES AND CONNECTIONS

Victor De Oliveira1

Department of Management Science and Statistics

The University of Texas at San Antonio

San Antonio, TX 78249, USA

victor.deoliveira@utsa.edu



ONE UTSA CIRCLE
SAN ANTONIO, TEXAS 78249-0631

# GEOSTATISTICAL BINARY DATA: MODELS, PROPERTIES AND CONNECTIONS

Victor De Oliveira[1]
Department of Management Science and Statistics
The University of Texas at San Antonio
San Antonio, TX 78249, USA
victor.deoliveira@utsa.edu

April 7, 2017

**Abstract**

This work considers models for geostatistical data for situations in which the region where the phenomenon of interest varies is partitioned into two disjoint subregions, which is called a binary map. The goals of this work are threefold. First, a review is provided of the classes of models that have been proposed so far for geostatistical binary data as well as a description of their main features. Second, a generalization is provided of a spatial multivariate probit model that eases regression function modeling, interpretation of the regression parameters, and establishing connections with other models. The second-order properties of this model are studied in some detail. Finally, connections between the aforementioned classes of models are established, showing that some of these are reformulations (reparametrizations) of the other models.

**Key words**: Clipped Gaussian random field; Gaussian copula model; Generalized linear mixed model; Indicator kriging; Multivariate probit model.

*JEL Classifications: C21, C31, C53*

[1]This project was funded by the University of Texas at San Antonio, Office of the Vice President for Research.

# 1 Introduction

This work considers models for spatial data for situations in which the region where the phenomenon of interest varies, say $D$, is partitioned into two disjoint subregions, $D = B \cup W$, called a binary map of $D$. This map is often unknown. Geostatistical binary data is a type of spatial data that arises when observations are taken at single sampling locations, so the measurements determine to which subregion, $B$ or $W$, each sampling location belongs to. In addition, location–dependent covariates that help explain the binary map may also be available. Examples of geostatistical binary data include the determination of disease status (presence or absence) of trees sampled from a forest, and the determination of rock type in a geological formation composed of only two rock types. Among the most common goals for the analysis of this kind of data are predicting to which, $B$ or $W$, an unsampled location belongs to (prediction/classification problem), and estimating the effects of covariates on the binary map (regression problem).

Early works on the analysis of geostatistical binary data include those of Journel (1983) and Solow (1986) who proposed a kriging variant called *indicator kriging* for spatial prediction, and Albert and McShane (1995) and Gotway and Stroup (1997) who proposed quasi–likelihood and generalized estimating equations for spatial regression (see also Lin and Clayton, 2005). These works were based on moment specifications. The main appeals of these early methods are methodological simplicity and similarity to models for other types of data (e.g., repeated measures data), but these may be conceptually unsound, as discussed in Section 5. Later, De Oliveira (1997, 2000), Nott and Wilson (1997), and Heagerty and Lele (1998) proposed the clipped Gaussian random field, a type of multivariate probit model that is conceptually sound, although is also more challenging to fit. Diggle, Tawn and Moyeed (1998) proposed a large class of hierarchical models for geostatistical data, a subclass of which has been used to describe binary (or more generally binomial) data. This class of models, which can be seen as a type of generalized linear mixed model, seems to be currently the most commonly used for the analysis of discrete geostatistical data. Finally, another class of models is that of spatial Gaussian copulas proposed by Madsen (2009), Kazianka and Pilz (2010), Kazianka (2013), Bai, Kang and Song (2014), and Han and De Oliveira (2016). The bulk of the effort in all of the aforementioned works has been focused on methods for fitting these models and their application to the analysis of spatial datasets. Comparatively, much less effort has been placed

on the study of the properties of these classes of models. But their properties are needed to assess the scope and limitation of these models for an adequate representation of real data. In addition, possible relations between these models have been hinted in the literature, but not studied in detail. This work aims at filling these gaps.

The goals of this work are threefold. First, a review is provided of the aforementioned classes of models proposed for the description of geostatistical binary data: clipped Gaussian random fields, generalized linear mixed models, Gaussian copula models and moment–based models. Second, a generalization is provided of the clipped Gaussian random field that eases regression function modeling, interpretation of the regression parameters, and the establishment of connections with other models. The second–order properties of this generalization of the clipped Gaussian random field are studied in some detail. Third, connections between the aforementioned classes of models are established, showing that some of these models are reformulations (reparametrizations) of the other models. The article closes with a discussion about the practical implications of the results for the modeling of geostatistical binary data.

## 2 Clipped Gaussian Random Fields

For a wide variety of contexts, a convenient strategy to model binary outcomes is to threshold latent continuous variables, where the latter are commonly described by Gaussian models. This leads to the so–called probit models (Collett, 2003). This strategy has been used to model binary time series by Kedem (1980), and to model spatial data by De Oliveira (1997, 2000), Nott and Wilson (1997), Heagerty and Lele (1998), Gelfand, Ravishanker and Ecker (2000), and Oman, Landsman, Carmel and Kadmon (2007), among many others. All of these can be viewed as multivariate probit models, but following De Oliveira (2000) they are called here clipped Gaussian random fields. This model is particularly appealing since many spatial binary data are actually generated—either by convenience or by limitations of the measuring device—by thresholding an underlying (unobserved) continuous random field, and Gaussian random fields are convenient models for the latter. The underlying continuous random field may either have a well defined physical interpretation or just serve as a convenient device to model spatial association among the binary data, beyond that described by the covariates. Here I describe a generalization of the clipped Gaussian random field proposed by De Oliveira (1997, 2000) and Heagerty and Lele (1998). The generalization allows the binary random field

to have an arbitrary mean function, which is convenient to ease interpretation of the regression parameters, and establish connections with other models. Later, some of its main second–order properties are described in detail to assess the model flexibility. In what follows $D$ denotes the region of interest. It is assumed for concreteness to be a subset of $\mathbb{R}^2$, but the model can also be used to describe spatial binary data in Euclidean spaces of other dimensions.

Let $\{Y(\mathbf{s}) : \mathbf{s} \in D\}$ be a binary random field, meaning that for any $\mathbf{s} \in D$, $Y(\mathbf{s})$ takes only two values, coded for convenience as 0 and 1. In addition to the binary response we may also have a set of $p$ location–dependent covariates $\mathbf{f}(\mathbf{s}) = (f_1(\mathbf{s}), \ldots, f_p(\mathbf{s}))'$, with $f_1(\mathbf{s}) \equiv 1$. A clipped Gaussian model (CGM) is obtained by clipping (thresholding) a latent Gaussian random field, so it is defined as

$$Y(\mathbf{s}) = \mathbf{1}\{Z(\mathbf{s}) > c\}, \tag{1}$$

where $\mathbf{1}\{A\}$ is the indicator function of the event $A$, $c \in \mathbb{R}$ is a threshold and $\{Z(\mathbf{s}) : \mathbf{s} \in D\}$ is an unobserved Gaussian random field with mean function $\nu(\mathbf{s})$ and covariance function $C(\mathbf{s}, \mathbf{u})$. The mean function can have *any* form, usually depending on the covariates $\mathbf{f}(\mathbf{s})$ and regression parameters $\boldsymbol{\beta} \in \mathbb{R}^p$, while the covariance function is assumed of the form

$$C(\mathbf{s}, \mathbf{u}) = \sigma^2\big((1 - \tau^2)K_\theta(\mathbf{s}, \mathbf{u}) + \tau^2\mathbf{1}\{\mathbf{s} = \mathbf{u}\}\big), \tag{2}$$

with $\sigma^2 > 0$, $\tau^2 \in [0, 1]$, and $K_\theta(\mathbf{s}, \mathbf{u})$ a correlation function parameterized by $\theta$ that is continuous everywhere. Heagerty and Lele (1998) studied model (1) with $\nu(\mathbf{s}) = \boldsymbol{\beta}'\mathbf{f}(\mathbf{s})$ and $\tau^2 > 0$ unknown, while De Oliveira (2000) studied model (1) with the same $\nu(\mathbf{s})$ and $\tau^2 = 0$. Then, the specification of the above CGM involves choosing the functions $\nu(\mathbf{s})$ and $C(\mathbf{s}, \mathbf{u})$.

It was shown in Heagerty and Lele (1998) and De Oliveira (2000) that the above model, with $\nu(\mathbf{s}) = \boldsymbol{\beta}'\mathbf{f}(\mathbf{s})$, is not identifiable from the binary data. An identifiable and interpretable model is obtained by fixing $\sigma^2$ and $c$, which without loss of generality are commonly set at $\sigma^2 = 1$ and $c = 0$, so in this case it holds that $\text{var}\{Z(\mathbf{s})\} = 1$. This choice is kept for the model (1) with arbitrary $\nu(\mathbf{s})$. In addition, any parameters in the correlation function $K_\theta(\mathbf{s}, \mathbf{u})$ that control smoothness/roughness of the random field $Z(\cdot)$ need to be fixed to avoid problems of near–non–identifiability, since the binary data contain little or no information about them; see De Oliveira (2000) for a discussion on this issue.

## 2.1 Second–Order Structure

This section describes some properties of the mean and covariance functions of clipped Gaussian random fields. The mean function of the CGM (1) can be set at *any* desired function taking values in $(0,1)$, say $\mu(\mathbf{s})$, by an appropriate choice of the mean function of the latent random field. Specifically

$$\begin{aligned} E\{Y(\mathbf{s})\} &= P\{Y(\mathbf{s})=1\} = P\{Z(\mathbf{s})-\nu(\mathbf{s}) > -\nu(\mathbf{s})\} \\ &= \Phi\big(\nu(\mathbf{s})\big), \end{aligned}$$

where $\Phi(\cdot)$ is the cdf of the standard normal distribution. So if $\nu(\mathbf{s}) := \Phi^{-1}(\mu(\mathbf{s}))$ is set (with $\Phi^{-1}(\cdot)$ the inverse function of $\Phi(\cdot)$), then it holds that $E\{Y(\mathbf{s})\} = \mu(\mathbf{s})$. The commonly used probit and logit regression functions are obtained by setting $\nu(\mathbf{s})$ equal to, respectively, $\boldsymbol{\beta}'\mathbf{f}(\mathbf{s})$ and $\Phi^{-1}\big(\big(1+\exp(-\boldsymbol{\beta}'\mathbf{f}(\mathbf{s}))\big)^{-1}\big)$. Hence, it is possible to have a CGM with the (logit) mean function $\mu(\mathbf{s}) = \big(1+\exp(-\boldsymbol{\beta}'\mathbf{f}(\mathbf{s}))\big)^{-1}$, and in this case the regression parameters have the usual interpretation as the (marginal) effects of the covariates on the log–odds. From now on I assume in the CGM (1) that $\nu(\mathbf{s}) = \Phi^{-1}(\mu(\mathbf{s}))$, where $\mu(\mathbf{s})$ is the desired mean/regression function of the binary random field. In practice the form of $\mu(\mathbf{s})$ is usually chosen from exploratory data analysis.

For any choice of mean function $\mu(\cdot)$ and any $\mathbf{s} \neq \mathbf{u}$, the correlation function of the CGM (1) is given by

$$\begin{aligned} & \operatorname{corr}\{Y(\mathbf{s}), Y(\mathbf{u})\} \\ = \; & \frac{P\{Y(\mathbf{s})=1, Y(\mathbf{u})=1\} \;-\; \mu(\mathbf{s})\mu(\mathbf{u})}{\big(\mu(\mathbf{s})(1-\mu(\mathbf{s}))\mu(\mathbf{u})(1-\mu(\mathbf{u}))\big)^{1/2}} \\ = \; & \Big(P\{Z(\mathbf{s})-\Phi^{-1}(\mu(\mathbf{s})) > -\Phi^{-1}(\mu(\mathbf{s})), Z(\mathbf{u})-\Phi^{-1}(\mu(\mathbf{u})) > -\Phi^{-1}(\mu(\mathbf{u}))\} \;-\; \mu(\mathbf{s})\mu(\mathbf{u})\Big) \\ & \qquad\qquad \times \Big(\mu(\mathbf{s})(1-\mu(\mathbf{s}))\mu(\mathbf{u})(1-\mu(\mathbf{u}))\Big)^{-1/2} \\ = \; & \frac{\Phi_2\big(\Phi^{-1}(\mu(\mathbf{s})), \Phi^{-1}(\mu(\mathbf{u})); \; (1-\tau^2)K_\theta(\mathbf{s},\mathbf{u})\big) \;-\; \mu(\mathbf{s})\mu(\mathbf{u})}{\big(\mu(\mathbf{s})(1-\mu(\mathbf{s}))\mu(\mathbf{u})(1-\mu(\mathbf{u}))\big)^{1/2}}, \end{aligned} \tag{3}$$

where $\Phi_2(t_1, t_2;\; \rho)$ is the cdf of the bivariate normal distribution with means $(0,0)$, variances $(1,1)$ and correlation $\rho$. The range of possible correlations under the CGM (1) is investigated next.

Let $(X_1, X_2)$ be an arbitrary binary random vector with respective marginals $\text{Ber}(p_1)$ and[2]

[2] $\text{Ber}(p)$ denotes the Bernoulli distribution with parameter $p$.

$\text{Ber}(p_2)$, where $p_1, p_2 \in (0,1)$. It is well known that the maximum possible correlation between $X_1$ and $X_2$ over all possible joint distributions having these marginals depends on $p_1$ and $p_2$, and is given by the Frechet–Hoeffding upper bound (Nelsen, 2006)

$$\text{corr}\{X_1, X_2\} \leq \min\Big\{\Big(\frac{p_1(1-p_2)}{p_2(1-p_1)}\Big)^{1/2},\ \Big(\frac{p_2(1-p_1)}{p_1(1-p_2)}\Big)^{1/2}\Big\}. \tag{4}$$

To find the maximum possible correlation under the CGM (1) for a given mean function $\mu(\cdot)$, we use the following properties of $\Phi_2(t_1, t_2;\ \rho)$ (see Patel and Read, 1996).

**Lemma 2.1** *For any fixed values of $t_1, t_2 \in \mathbb{R}$ it holds that:*

*(a) $\Phi_2(t_1, t_2;\ \rho)$ is a strictly increasing function of $\rho$.*

*(b) $\lim_{\rho \to 1^-} \Phi_2(t_1, t_2;\ \rho) = \Phi(\min\{t_1, t_2\})$.*

Using (3) and Lemma 2.1(a), it holds that under CGM (1) and for $\mu(\mathbf{s})$ and $\mu(\mathbf{u})$ fixed, an upper bound for $\text{corr}\{Y(\mathbf{s}), Y(\mathbf{u})\}$ over all values of $\tau^2 \in [0,1]$ and $K_\theta(\mathbf{s}, \mathbf{u}) \in [0,1]$ is given by

$$\begin{aligned}
& \lim_{\rho \to 1^-} \frac{\Phi_2\big(\Phi^{-1}(\mu(\mathbf{s})), \Phi^{-1}(\mu(\mathbf{u}));\ \rho\big) - \mu(\mathbf{s})\mu(\mathbf{u})}{\big(\mu(\mathbf{s})(1-\mu(\mathbf{s}))\mu(\mathbf{u})(1-\mu(\mathbf{u}))\big)^{1/2}} \\
= \ & \frac{\Phi\big(\min\{\Phi^{-1}(\mu(\mathbf{s})), \Phi^{-1}(\mu(\mathbf{u}))\} - \mu(\mathbf{s})\mu(\mathbf{u})}{\big(\mu(\mathbf{s})(1-\mu(\mathbf{s}))\mu(\mathbf{u})(1-\mu(\mathbf{u}))\big)^{1/2}}, \qquad \text{by Lemma 2.1(b)} \\
= \ & \frac{\min\{\mu(\mathbf{s}), \mu(\mathbf{u})\} - \mu(\mathbf{s})\mu(\mathbf{u})}{\big(\mu(\mathbf{s})(1-\mu(\mathbf{s}))\mu(\mathbf{u})(1-\mu(\mathbf{u}))\big)^{1/2}}, \qquad \text{since } \Phi(\cdot) \text{ is strictly increasing} \\
= \ & \min\Big\{\Big(\frac{\mu(\mathbf{s})(1-\mu(\mathbf{u}))}{\mu(\mathbf{u})(1-\mu(\mathbf{s}))}\Big)^{1/2},\ \Big(\frac{\mu(\mathbf{u})(1-\mu(\mathbf{s}))}{\mu(\mathbf{s})(1-\mu(\mathbf{u}))}\Big)^{1/2}\Big\}.
\end{aligned}$$

This upper bound agrees with the Frechet–Hoeffding upper bound (4). Moreover, this upper bound is tight as it is achieved when $K_\theta(\mathbf{s}, \mathbf{u}) = 1$ and $\tau^2 = 0$, so the CGM (1) allows the full correlation range that is feasible for a given mean function $\mu(\cdot)$.

Next, it is shown that the continuity or discontinuity of the correlation function (2) on the 'diagonal' $\mathbf{s} = \mathbf{u}$ is inherited by the correlation function (3) of the binary random field. Let us assume the mean function $\mu(\mathbf{s}) \in (0,1)$ is continuous for all $\mathbf{s} \in D$. If $\tau^2 = 0$, it follows from (3) that $\text{corr}\{Y(\mathbf{s}), Y(\mathbf{u})\}$ is continuous for any $(\mathbf{s}, \mathbf{u})$, since all the functions involved are continuous and the denominator does not vanish. If $\tau^2 > 0$, we have by the same argument

that $\text{corr}\{Y(\mathbf{s}), Y(\mathbf{u})\}$ is continuous for any $\mathbf{s} \neq \mathbf{u}$. But in this case

$$\begin{aligned}
\lim_{\mathbf{u}\to\mathbf{s}} \text{corr}\{Y(\mathbf{s}), Y(\mathbf{u})\} &= \frac{\Phi_2\big(\Phi^{-1}(\mu(\mathbf{s})), \Phi^{-1}(\mu(\mathbf{s})); \ (1-\tau^2)\big) - \mu^2(\mathbf{s})}{\mu(\mathbf{s})(1-\mu(\mathbf{s}))} \\
&< \lim_{\rho\to 1} \frac{\Phi_2\big(\Phi^{-1}(\mu(\mathbf{s})), \Phi^{-1}(\mu(\mathbf{s})); \ \rho\big) - \mu^2(\mathbf{s})}{\mu(\mathbf{s})(1-\mu(\mathbf{s}))} \\
&= \frac{\mu(\mathbf{s}) - \mu^2(\mathbf{s})}{\mu(\mathbf{s})(1-\mu(\mathbf{s}))} = 1 \quad (= \text{corr}\{Y(\mathbf{s}), Y(\mathbf{s})\}), \qquad (5)
\end{aligned}$$

where the inequality follows from Lemma 2.1(b) and the last equality from Lemma 2.1(a). Then when $\tau^2 > 0$, $\text{corr}\{Y(\mathbf{s}), Y(\mathbf{u})\}$ is discontinuous on the diagonal $\mathbf{s} = \mathbf{u}$ (it displays the so–called 'nugget effect'), and the Frechet–Hoeffding upper bound is not achievable. This fact is illustrated in the next section. The second–order properties of the CGM derived above are summarized in the following:

**Result 1.** The CGM (1) has a flexible second–order structure: The model can have any $\mu(\mathbf{s}) \in (0,1)$ as its mean function, its correlation function (3) may or may not include a nugget effect, and in the latter case, its range can take any value that is feasible for the assumed mean function $\mu(\mathbf{s})$.

## 2.2 Further Properties of Correlation Functions

This section investigates how the correlation function of the CGM (1) depends on the correlation function of the latent random field and mean function $\mu(\cdot)$. From (3) we have $\text{corr}\{Y(\mathbf{s}), Y(\mathbf{u})\} = G(\tau^2, K_\theta(\mathbf{s},\mathbf{u}), \mu(\mathbf{s}), \mu(\mathbf{u}))$, for a certain function $G$ say. Since a non-linear transformation of a Gaussian process has always a 'whitening effect' (Koyak, 1987), we have that $\text{corr}\{Y(\mathbf{s}), Y(\mathbf{u})\} \leq C(\mathbf{s},\mathbf{u})$. Unlike the cases of smooth one–to–one transformations explored in De Oliveira (2003), in which the correlation functions of the transformed and untransformed Gaussian random fields are close, for CGMs the whitening effect is substantial and the size of it depends on $\mu(\mathbf{s})$ and $\tau^2$ (see below). Also, from Lemma 2.1(a) follows that, everything else being equal, $\text{corr}\{Y(\mathbf{s}), Y(\mathbf{u})\}$ is an increasing function of $K_\theta(\mathbf{s},\mathbf{u})$ and a decreasing function of $\tau^2$. To graphically illustrate these and other features of the dependence of $\text{corr}\{Y(\mathbf{s}), Y(\mathbf{u})\}$ on $\tau^2$, $K_\theta(\mathbf{s},\mathbf{u})$ and $\mu(\cdot)$, let us consider first the case of constant mean, so $\mu(\mathbf{s}) = \mu(\mathbf{u}) = \mu$. In this case, an alternative expression for (3) is given by the classic formula (Kedem, 1980, p. 35)

$$\text{corr}\{Y(\mathbf{s}), Y(\mathbf{u})\} = \frac{\frac{1}{2\pi}\int_0^{(1-\tau^2)K_\theta(\mathbf{s},\mathbf{u})} (1-t^2)^{-1/2} \exp\big(-\frac{(\Phi^{-1}(\mu))^2}{1+t}\big) dt}{\mu(1-\mu)}, \qquad \mathbf{s} \neq \mathbf{u},$$

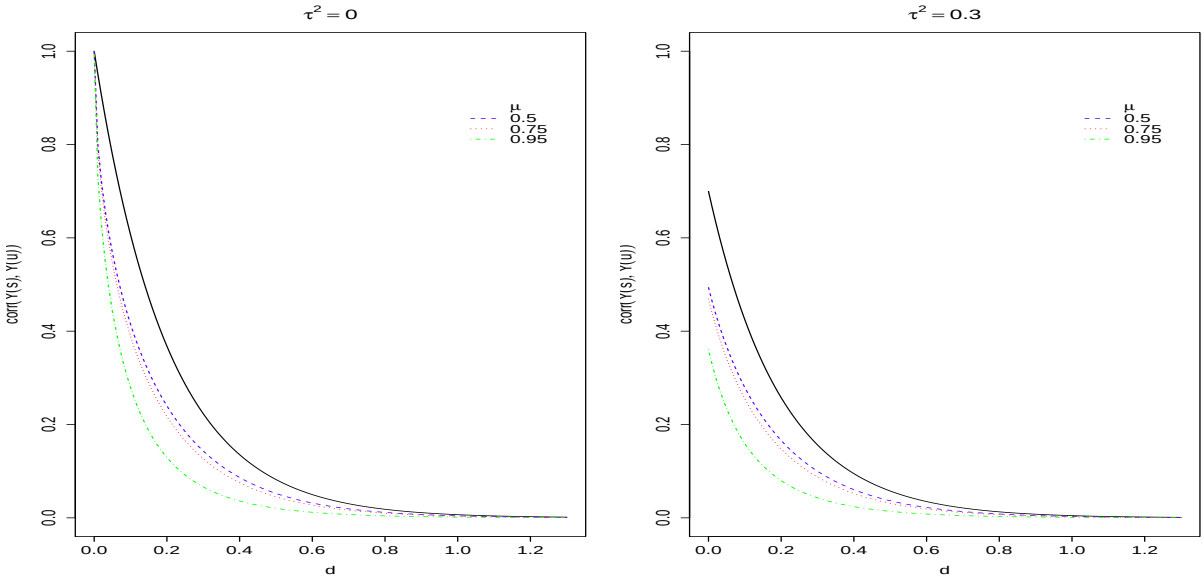


Figure 1: Plots of $\text{corr}\{Y(\mathbf{s}), Y(\mathbf{u})\}$ as a function of distance $d = ||\mathbf{s} - \mathbf{u}||$, for $K_\theta(\mathbf{s}, \mathbf{u}) = \exp(-d/0.2)$ (solid black line) and different values of $\mu$, when $\tau^2 = 0$ (left panel) and $\tau^2 = 0.3$ (right panel).

where the integral in the right hand side is easily approximated by numerical quadrature, e.g., by the `R` function `integrate`. From this follows that, as a function of $\mu$, $\text{corr}\{Y(\mathbf{s}), Y(\mathbf{u})\}$ is symmetric around $\frac{1}{2}$ since[3] it takes the same value at $\mu$ and $1 - \mu$. Figure 1 displays $\text{corr}\{Y(\mathbf{s}), Y(\mathbf{u})\}$ as a function of distance $d = ||\mathbf{s}-\mathbf{u}||$ and different values of $\mu$, when $K_\theta(\mathbf{s}, \mathbf{u}) = \exp(-d/0.2)$ (solid black line), and $\tau^2 = 0$ (left panel) and $\tau^2 = 0.3$ (right panel). The presence of a nugget effect (discontinuity at the origin) when $\tau^2 = 0.3$ is apparent as well as the whitening effect (reduced correlation), which increases as $\mu$ departs from $\frac{1}{2}$. Figure 2 displays $\text{corr}\{Y(\mathbf{s}), Y(\mathbf{u})\}$ as a function of $\mu$ for different values of $K_\theta(\mathbf{s}, \mathbf{u})$, when $\tau^2 = 0$ (left panel) and $\tau^2 = 0.3$ (right panel). These figures show that the whitening effect is substantial in most cases. Everything else being equal, the whitening effect increases with the magnitude of the nugget parameter $\tau^2$, and also increases as $|\mu - \frac{1}{2}|$ increase.

When $\mu(\cdot)$ is not constant, computing $\text{corr}\{Y(\mathbf{s}), Y(\mathbf{u})\}$ in (3) requires the computation of $\Phi_2(t_1, t_2; \rho)$, which has no closed–form. There are numerous ways to accurately approximate this function, either by explicit approximations or by numerical quadrature, many of which are

[3]For any $x \in (0, 1)$, $\Phi^{-1}(1 - x) = -\Phi^{-1}(x)$.

reviewed in Patel and Read (1996, Chapter 9) and Genz and Bretz (2009, Chapter 2). The latter approximation is implemented in the `R` package `mvtnorm`. The features described above about the dependence of $\text{corr}\{Y(\mathbf{s}), Y(\mathbf{u})\}$ on $\tau^2$, $K_\theta(\mathbf{s}, \mathbf{u})$ and $\mu(\cdot)$ are also expected to hold when $\mu(\cdot)$ is not constant, although they are more difficult to visualize.

### 2.3 Hierarchical Formulation

The specification of the binary random field in (1) is given as a many–to–one transformation of a latent Gaussian random field. Alternatively, when $\tau^2 > 0$, this model admits a hierarchical formulation (Gelfand, et al. 2000; Oman, et al. 2007). The latent Gaussian random field $Z(\cdot)$ in (1) can be decomposed as

$$Z(\mathbf{s}) \stackrel{d}{=} \Gamma(\mathbf{s}) + \epsilon(\mathbf{s}), \qquad \mathbf{s} \in D,$$

where $\Gamma(\cdot)$ is a Gaussian random field with mean function $\nu(\mathbf{s})$ and covariance function $(1 - \tau^2)K_\theta(\mathbf{s}, \mathbf{u})$, $\epsilon(\cdot)$ is a Gaussian random field with mean 0 and covariance function $\tau^2 \mathbf{1}\{\mathbf{s} = \mathbf{u}\}$, and $\Gamma(\cdot)$ and $\epsilon(\cdot)$ are independent random fields; $\stackrel{d}{=}$ denotes equality of the family of finite–dimensional distributions. From this follows that for any set of distinct locations $\mathbf{s}_1, \ldots, \mathbf{s}_n \in D$,

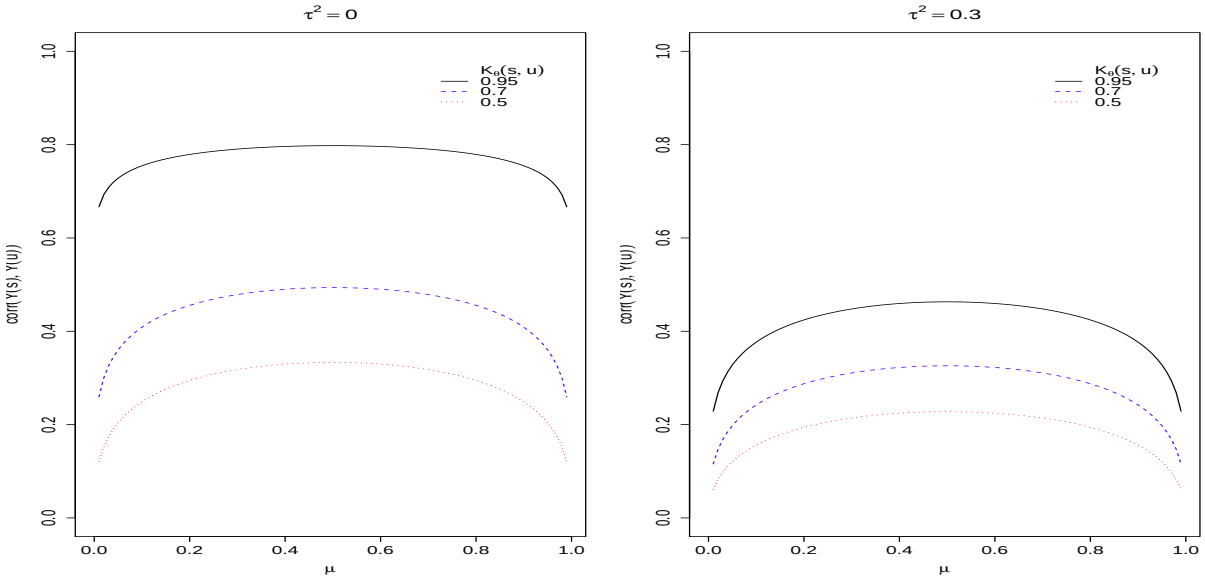


Figure 2: Plots of $\text{corr}\{Y(\mathbf{s}), Y(\mathbf{u})\}$ as a function of $\mu$ when $\tau^2 = 0$ (left panel) and $\tau^2 = 0.3$ (right panel), for different values of $K_\theta(\mathbf{s}, \mathbf{u})$.

$Y(\mathbf{s}_i)$ is a function of only $\Gamma(\mathbf{s}_i)$ and $\epsilon(\mathbf{s}_i)$, so if $\mathbf{\Gamma} = (\Gamma(\mathbf{s}_1), \ldots, \Gamma(\mathbf{s}_n))$, it holds that

$$\begin{aligned} P\{Y(\mathbf{s}_i) = 1 \mid \mathbf{\Gamma}\} &= P\{\Gamma(\mathbf{s}_i) + \epsilon(\mathbf{s}_i) > 0 \mid \Gamma(\mathbf{s}_i)\} = P\Big\{\frac{\epsilon(\mathbf{s}_i)}{\tau} > -\frac{\Gamma(\mathbf{s}_i)}{\tau} \mid \Gamma(\mathbf{s}_i)\Big\} \\ &= \Phi\Big(\frac{\Gamma(\mathbf{s}_i)}{\tau}\Big). \end{aligned}$$

Then when $\tau^2 > 0$, the CGM (1) is equivalent to the following hierarchical model:

(1A) For any set of distinct locations $\mathbf{s}_1, \ldots, \mathbf{s}_n \in D$, $Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n)$ are conditionally independent given $\mathbf{\Gamma} = (\Gamma(\mathbf{s}_1), \ldots, \Gamma(\mathbf{s}_n))'$, and

$$\begin{aligned} Y(\mathbf{s}_i) \mid \mathbf{\Gamma} &\stackrel{d}{=} Y(\mathbf{s}_i) \mid \Gamma(\mathbf{s}_i) \\ &\sim \text{Ber}\Big(\Phi\Big(\frac{\Gamma(\mathbf{s}_i)}{\tau}\Big)\Big), \qquad i = 1, \ldots, n. \end{aligned}$$

(2A) The random field $\Gamma(\cdot)$ is Gaussian with mean function $\nu(\mathbf{s})$ and covariance function $(1 - \tau^2)K_\theta(\mathbf{s}, \mathbf{u})$.

A hierarchical formulation of the CGM (1) is not possible when $\tau^2 = 0$. This hierarchical formulation of the CGM would make clear its connection to another class of hierarchical models that have been proposed for geostatistical binary data, to be described next.

# 3 Generalized Linear Mixed Models

Diggle et al. (1998) proposed a large class of models to describe geostatistical data that combines generalized linear models with Gaussian random fields, and as such it is a class of generalized linear mixed model. Although in principle the proposed models can describe a wide variety of non–Gaussian geostatistical data, they have been mostly used to describe discrete geostatistical data; see Christensen and Waagepetersen (2002), Diggle, Moyeed, Rowlingson and Thomson (2002), Zhang (2002), Diggle and Ribeiro (2007), and Jing and De Oliveira (2015).

For the case of geostatistical binary data (or more generally binomial data), the model is particularly appealing in the following context. Suppose the binary data represent occurrences of an event of interest whose probability of occurrence (or relative risk) varies spatially and may depend on observed location–dependent covariates, as well as on unobserved/unknown spatially varying variables (random effects). Such data could be described by the following hierarchical model:

(1B) For any set of distinct locations $\mathbf{s}_1, \ldots, \mathbf{s}_n \in D$, $Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n)$ are conditionally independent given $\mathbf{S} = (S(\mathbf{s}_1), \ldots, S(\mathbf{s}_n))$, and

$$\begin{aligned} Y(\mathbf{s}_i) \mid \mathbf{S} &\stackrel{d}{=} Y(\mathbf{s}_i) \mid S(\mathbf{s}_i) \\ &\sim \text{Ber}\big(h^{-1}(\boldsymbol{\beta}'\mathbf{f}(\mathbf{s}_i) + S(\mathbf{s}_i))\big), \qquad i = 1, \ldots, n, \end{aligned}$$

where $h : (0,1) \to \mathbb{R}$ is a known link function, assumed strictly increasing.

(2B) The random field $S(\cdot)$ is Gaussian with mean 0 and covariance function $\sigma^2\rho_\theta(\mathbf{s}, \mathbf{u})$, where $\sigma^2 > 0$ and $\rho_\theta(\mathbf{s}, \mathbf{u})$ is a correlation function.

In the above, $S(\cdot)$ represents unobserved sources of spatial variation affecting the spatially varying probability (relative risk), and the correlation function $\rho_\theta(\mathbf{s}, \mathbf{u})$ is often (but not always) assumed continuous everywhere. This is called a spatial generalized linear mixed model (GLMM), whose specification requires choosing the link function $h(\cdot)$ and correlation function $\rho_\theta(\mathbf{s}, \mathbf{u})$.

Berrett and Calder (2016) carried an empirical comparison, in terms of prediction performance, between the GLMM with probit link and the CGM with $\nu(\mathbf{s}) = \boldsymbol{\beta}'\mathbf{f}(\mathbf{s})$ and $\tau^2 = 0$, which are different models. But when $\tau^2 > 0$, by comparing (1A)–(2A) and (1B)–(2B) we have the following:

**Result 2.** When the probit link function is used, $h(u) = \Phi^{-1}(u)$, the above GLMM (1B)–(2B) is a reparameterization of the CGM (1A)–(2A), obtained by setting in the latter

$$\nu(\mathbf{s}) = (1+\sigma^2)^{-1/2}\boldsymbol{\beta}'\mathbf{f}(\mathbf{s}), \quad \tau^2 = (1+\sigma^2)^{-1} \quad \text{and} \quad K_\theta(\mathbf{s}, \mathbf{u}) = \rho_\theta(\mathbf{s}, \mathbf{u}).$$

Therefore in this case the GLMM and CGM are the *same* model.

Hence, the GLMM with probit link also satisfy the second–order properties described in Sections 2.1 and 2.2. When $h(\cdot) \neq \Phi^{-1}(\cdot)$, the GLMM (1B)–(2B) and CGM (1A)–(2A) are different models, and in general the former lacks closed–form expressions for the mean and covariance functions. In this case the second–order properties of the GLMM are less transparent than those of the CGM, specially in regard to the interpretation of the regression parameters (see below).

Most applications of GLMM have used the logit link function, $h(u) = \log\big(u/(1-u)\big)$. In this case, standard moment decompositions and a delta method argument show that for $\sigma$ *close*

to zero, it holds that for any $\mathbf{s}, \mathbf{u} \in D$ (Coull and Agresti, 2000)

$$\begin{aligned} E\{Y(\mathbf{s})\} &\approx \frac{\exp(\boldsymbol{\beta}'\mathbf{f}(\mathbf{s}))}{1+\exp(\boldsymbol{\beta}'\mathbf{f}(\mathbf{s}))} =: \mu^a(\mathbf{s}) \\ \operatorname{cov}\{Y(\mathbf{s}), Y(\mathbf{u})\} &\approx \sigma^2\mu^a(\mathbf{s})(1-\mu^a(\mathbf{s}))\mu^a(\mathbf{u})(1-\mu^a(\mathbf{u}))\rho_\theta(\mathbf{s},\mathbf{u}) \\ &\quad + \mu^a(\mathbf{s})(1-\mu^a(\mathbf{s}))\big(1-\sigma^2\mu^a(\mathbf{s})(1-\mu^a(\mathbf{s}))\big)\mathbf{1}\{\mathbf{s}=\mathbf{u}\}; \end{aligned}$$

The latter is guaranteed to be a positive–definite function, provided $\sigma$ is close enough to zero. When $h(\cdot)$ is the logit link but $\sigma$ is not small, the above approximations are poor and the regression parameters $\boldsymbol{\beta}$ do not have a marginal interpretation. Although they have a conditional interpretation (Diggle et al. 1998), in the spatial context this conditional interpretation is not of as much interest as marginal interpretations since repeated measures are rare in typical geostatistical data (but see below).

### 3.1 Variations

Several variations of the GLMM (1B)–(2B) have been recently proposed. In the context of temporal or spatial binary data with replications, both marginal and conditional interpretations of the regression parameters are relevant. Parzen et al. (2011) [in a longitudinal setting] and Boehm, Reich and Bandyopadhyay (2013) [in a spatial setting] proposed a modification of the GLMM with logit link, where $S(\cdot)$ in level (2B) becomes a stationary zero–mean non–Gaussian random field defined by a Gaussian copula model (see Section 4), whose family of marginal distributions is tailored to make the conditional and marginal interpretations of the regression parameters agree; these marginal distributions are the so–called 'bridge distributions' derived by Wang and Louis (2003).

Wang, Dey and Banerjee (2010) and Roy, Evangelou and Zhu (2016) proposed extending the GLMM by using a parametric family of link functions $h_\xi(\cdot)$ in level (1B), where the parameter $\xi$ is estimated from the data. The authors advocate this extended model as more capable of fitting datasets with asymmetric responses, and producing predictions for the latent process values that are robust against model misspecification. The downside of this model is that marginal interpretations of the regression parameters become problematic, if at all possible.

# 4 Gaussian Copula Models

Copulas are continuous multivariate distribution functions with unif$(0,1)$ marginals. Their usefulness and recent upsurge in interest and applications rely on the celebrated result of Abe Sklar, stating that any multivariate cdf $H$ can be obtained by evaluating a copula at the marginal cdfs of $H$ (Nelsen, 2006). Specifically, if $(X_1,\ldots,X_n)$ is a random vector with joint cdf $H$, then

$$H(x_1,\ldots,x_n) = C(F_1(x_1),\ldots,F_n(x_n)),$$

where $C(\cdot)$ is a copula in $\mathbb{R}^n$ and $F_1,\ldots,F_n$ are the marginal cdfs of $X_1,\ldots,X_n$. This allows the construction of models for dependent data by separately modeling the marginal and association structures of the random variables involved.

Madsen (2009), Kazianka and Pilz (2010), Kazianka (2013), Bai, Kang and Song (2014) and Han and De Oliveira (2016) proposed the use of Gaussian copulas to model geostatistical discrete data, where $C(u_1,\ldots,u_n) = \Phi_K(\Phi^{-1}(u_1),\ldots,\Phi^{-1}(u_n))$ and $\Phi_K(\cdot)$ is the joint cdf of the multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $K$ with diagonal entries all equal to 1. Let $\mathcal{F} = \{F_{\mathbf{s}}(\cdot) : \mathbf{s} \in D\}$ be a family of distributions, where $F_{\mathbf{s}}(\cdot)$ is the intended cdf of $Y(\mathbf{s})$. A Gaussian copula random field for $Y(\cdot)$ having $\mathcal{F}$ as its family of marginal distributions is given by (Han and De Oliveira, 2016)

$$Y(\mathbf{s}) = F_{\mathbf{s}}^{-1}\big(\Phi(\tilde{Z}(\mathbf{s}))\big), \qquad \mathbf{s} \in D, \tag{6}$$

where $\tilde{Z}(\cdot)$ is a Gaussian random field with mean 0, variance 1 and correlation function $C(\mathbf{s},\mathbf{u})$ as in (2), with $\sigma^2 = 1$, and $F_{\mathbf{s}}^{-1}(\cdot)$ is the quantile function that corresponds to $F_{\mathbf{s}}(\cdot)$; this is called a Gaussian copula model (GCM), which is specified by the family of cdfs $\mathcal{F}$ and the correlation function $C(\mathbf{s},\mathbf{u})$.

For the case of geostatistical binary data, let $\mathcal{F} = \{\text{Ber}(\mu(\mathbf{s})) : \mathbf{s} \in D\}$ be the family of proposed marginal distributions. In this case we have

$$F_{\mathbf{s}}(x) = P\{Y(\mathbf{s}) \le x\} = \begin{cases} 0 & \text{if } x < 0 \\ 1-\mu(\mathbf{s}) & \text{if } 0 \le x < 1 \\ 1 & \text{if } x \ge 1 \end{cases}, \qquad x \in \mathbb{R},$$

and

$$\begin{aligned} F_{\mathbf{s}}^{-1}(u) &= \inf\{x \in \mathbb{R} : F_{\mathbf{s}}(x) \geq u\}, \qquad u \in (0,1) \\ &= \begin{cases} 0 & \text{if } 0 < u \leq 1 - \mu(\mathbf{s}) \\ 1 & \text{if } 1 - \mu(\mathbf{s}) < u < 1 \end{cases}. \end{aligned}$$

Hence, the GCM (6) becomes

$$\begin{aligned} Y(\mathbf{s}) &= \mathbf{1}\{1 - \mu(\mathbf{s}) < \Phi(\tilde{Z}(\mathbf{s})) < 1\} = \mathbf{1}\{\tilde{Z}(\mathbf{s}) - \Phi^{-1}(1 - \mu(\mathbf{s})) > 0\} \\ &= \mathbf{1}\{Z(\mathbf{s}) > 0\}, \end{aligned}$$

where $Z(\mathbf{s}) := \tilde{Z}(\mathbf{s}) - \Phi^{-1}(1 - \mu(\mathbf{s})) = \tilde{Z}(\mathbf{s}) + \Phi^{-1}(\mu(\mathbf{s}))$ is a Gaussian random field with mean function $\Phi^{-1}(\mu(\mathbf{s}))$ and correlation function $C(\mathbf{s}, \mathbf{u})$. Therefore we have the following:

**Result 3.** The GCM (6) specified by the family of marginal distributions $\{\text{Ber}(\mu(\mathbf{s})) : \mathbf{s} \in D\}$ and the correlation function $C(\mathbf{s}, \mathbf{u})$, and the CGM (1) specified by $\nu(\mathbf{s}) = \Phi^{-1}(\mu(\mathbf{s}))$ and $C(\mathbf{s}, \mathbf{u})$ are the *same* model.

The above connection was noted by Bai et al. (2014) for the particular case when $\mu(\mathbf{s}) = \Phi(\boldsymbol{\beta}'\mathbf{f}(\mathbf{s}))$.

# 5 Moment–Based Models

A different approach for the analysis of geostatistical binary data, more in line with traditional geostatistical methods, is the use of models that only specify the mean and correlation functions of the binary random field. Examples of this include Journel (1983), Solow (1986), Albert and McShane (1995), Gotway and Stroup (1997) and Lin and Clayton (2005). In these works $E\{Y(\mathbf{s})\}$ was given by a constant or a logit regression function, and $\text{corr}\{Y(\mathbf{s}), Y(\mathbf{u})\}$ by a traditional isotropic positive define function in $\mathbb{R}^2$ that, up to a reparameterization, is given by (2) with $\sigma^2 = 1$. For $K_\theta(d)$, Albert and McShane (1995) used the exponential model while Gotway and Stroup (1997) used the spherical model. These authors proposed using quasi–likelihood and generalized estimating equations to estimate the model parameters. This approach appears attractive from a practical point of view, but it may be probabilistically flawed. Even though for any positive definite function (2) there exists a Gaussian random field having this as its correlation function, the same is not true for binary random fields. For instance, Matheron (1989) showed that there is no binary random field having $K_\theta(d) =$

$\exp(-(d/\theta)^2)$ as its correlation function. For many other positive definite functions (e.g., the spherical model used by Gotway and Stroup, 1997), it seems to be unknown whether or not they are compatible with a binary random field. A similar issue arises with the specification of the nugget parameter $\tau^2$. For CGM and GLMM the nugget (the size of the discontinuity of the correlation functions along the diagonal $\mathbf{s} = \mathbf{u}$) is a function of the mean function $\mu(\cdot)$ (see (5) and Figure 1), but Albert and McShane (1995) assumed this to be a (functionally) independent parameter. The existence of binary random fields with such a feature seems questionable. Lin and Clayton (2005) proposed a consistent model but under a very narrow specification: the sampling locations form a regular grid, covariates are binary, and the correlation function is isotropic, exponential and continuous at 0 (no nugget). De Oliveira (2000) and Moyeed and Papritz (2002) raised the issue of possible incompatibility in the use of such models for spatial prediction (indicator kriging), where the latter article even suggested that most indicator kriging applications may rely on a flawed probabilistic model. Building a solid foundation for this approach would require finding conditions that guarantee the compatibility between a positive definite function $C(\mathbf{s}, \mathbf{u})$ and a given family of Bernoulli distributions (i.e., a mean function $\mu(\mathbf{s})$); this author is unaware of any such condition.

# 6 Discussion

The findings of this article have several practical implications for the modeling of geostatistical binary data. First, the review of models proposed in the literature shows that there are essentially only two types of valid models: clipped Gaussian random fields and generalized linear mixed models with link functions different than the probit link. Of these, the second–order properties of the former seem to be more flexible and better understood than those of the latter. Second, the model equivalences stated in Results 2 and 3 imply that the understanding gained about one class of models helps understand better the other classes of models. For instance, the second–order properties discussed in Sections 2.1 and 2.2 also hold for the GLMM (1B)–(2B) with probit link function and the GCM (6). In addition, any software designed for fitting one class of models can also be used to fit the other classes, as these are just reformulations (reparametrizations) of the former. Finally, moment–based models are in general not valid, as they are probabilistically flawed, although the practical implications of using such models are unclear.

## References


Albert, P.S. and McShane, L.M. (1995), A Generalized Estimating Equations Approach for Spatially Correlated Binary Data: Applications to the Analysis of Neuroimaging Data, *Biometrics*, 51, 627-638.

Bai, Y., Kang, J. and Song, P.X.-K. (2014), Efficient Pairwise Composite Likelihood Estimation for Spatial–Clustered Data, *Biometrics*, 70, 661-670.

Berrett, C. and Calder, C.A. (2016), Bayesian Spatial Binary Classification, *Spatial Statistics*, 16, 72-102.

Boehm, L., Reich, B.J. and Bandyopadhyay, D. (2013), Bridging Conditional and Marginal Inference for Spatially Referenced Binary Data, *Biometrics*, 69, 545-554.

Collett, D. (2003), *Modelling Binary Data*, 2nd ed., Chapman & Hall/CRC.

Christensen, O.F. and Waagepetersen, R. (2002), Bayesian Prediction of Spatial Count Data Using Generalized Linear Mixed Models, *Biometrics*, 58, 280-286.

Coull, B.A. and Agresti, A. (2000), Random Effects Modeling of Multiple Binomial Responses Using the Multivariate Binomial Logit–Normal Distribution, *Biometrics*, 56, 73-80.

De Oliveira, V. (2003), A Note on the Correlation Structure of Transformed Gaussian Random Fields, *Australian and New Zealand Journal of Statistics*, 45, 353-366.

De Oliveira, V. (2000), Bayesian Prediction of Clipped Gaussian Random Fields, *Computational Statistics and Data Analysis*, 34, 299-314.

De Oliveira, V. (1997), Prediction in Some Classes of Non–Gaussian Random Fields, Unpublished Ph.D. Dissertation, University of Maryland–College Park.

Diggle, P.J and Ribeiro, P.J. (2007), *Model–Based Geostatistics*, Springer–Verlag.

Diggle, P.J., Moyeed, R.A., Rowlingson, B.S. and Thomson, M.C. (2002), Childhood Malaria in the Gambia: A Case Study in Model-based Geostatistics, *Applied Statistics*, 51, 493-506.

Diggle, P.J., Tawn, J.A. and Moyeed, R.A. (1998), Model–Based Geostatistics (with discussion), *Applied Statistics*, 47, 299-350.

Gelfand, A.E., Ravishanker, N. and Ecker, M.D. (2000), Modeling and Inference for Point–Referenced Binary Spatial Data. In: D.K. Dey, S.K. Ghosh and Mallick, B.K. (eds.), *Generalized Linear Models: A Bayesian Perspective*, pp 373-386. Marcel Dekker.

Genz, A. and Bretz, F. (2009), *Computation of Multivariate Normal and t Probabilities*, Springer.

Gotway, C.A. and Stroup, W.W. (1997), A Generalized Linear Model Approach to Spatial Data Analysis and Prediction, *Journal of Agricultural, Biological and Environmental Statistics*, 2, 157-178.

Han, Z. and De Oliveira, V. (2016), On the Correlation Structure of Gaussian Copula Models for Geostatistical Count Data, *Australian and New Zealand Journal of Statistics*, 58, 47-69.

Heagerty, P.J. and Lele, S.R. (1998), A Composite Likelihood Approach to Binary Spatial Data, *Journal of the American Statistical Association*, 93, 1099-1111.

Jing, L. and De Oliveira, V. (2015), geoCount: An R Package for the Analysis of Geostatistical Count Data, *Journal of Statistical Software*, 63 (11), 1-33.

Journel, A.G. (1983), Nonparametric Estimation of Spatial Distributions, *Mathematical Geology*, 15, 445-468.

Kazianka, H. (2013), Approximate Copula–Based Estimation and Prediction of Discrete Spatial Data, *Stochastic Environmental Research and Risk Assessment*, 27, 2015-2026.

Kazianka, H. and Pilz, J. (2010), Copula–Based Geostatistical Modeling of Continuous and Discrete Data Including Covariates, *Stochastic Environmental Research and Risk Assessment*, 24, 661-673.

Kedem, B. (1980), *Binary Time Series*, Marcel Dekker.

Koyak, R.A. (1987), On Measuring Internal Dependence in a Set of Random Variables, *Annals of Statistics*, 15, 1215-1228.

Lin, P.-S. and Clayton, M.K. (2005), Analysis of Binary Spatial Data by Quasi–Likelihood Estimating Equations, *The Annals of Statistics*, 33, 542-555.

Madsen, L. (2009), Maximum Likelihood Estimation of Regression Parameters With Spatially Dependent Discrete Data, *Journal of Agricultural, Biological, and Environmental Statistics*, 14, 375-391.

Matheron, G. (1989), The Internal Consistency of Models in Geostatistics. In: M. Armstrong (ed.), *Geostatistics, Volume 1*, pp 21-38. Kluwer Academic Publishers.

Moyeed, R.A. and Papritz, A. (2002), An Empirical Comparison of Kriging Methods for Nonlinear Spatial Point Prediction, *Mathematical Geology*, 34, 365-386.

Nott, D.J. and Wilson, R.J. (1997), Parameter Estimation for Excursion Set Texture Models, *Signal Processing*, 63, 199-210.

Nelsen, R.B. (2006), *An Introduction to Copulas*, 2nd ed., Springer-Verlag.

Oman, S.D., Landsman, V., Carmel, Y. and Kadmon, R. (2007), Analyzing Spatially Distributed Binary Data Using Independent–Block Estimating Equations, *Biometrics*, 63, 892-900.

Patel, J.K. and Read, C.B. (1996), *Handbook of the Normal Distribution*, 2nd ed., Marcel Dekker.

Parzen, M., Ghosh, S., Lipsitz, S., Sinha, D., Fitzmaurice, G.M., Mallick, B. and Ibrahim, J.G. (2011), A Generalized Linear Mixed Model for Longitudinal Binary Data With a Marginal Logit Link Function, *Annals of Applied Statistics*, 5, 449-467.

Solow, A.R. (1986), Mapping by Simple Indicator Kriging, *Mathematical Geology*, 18, 335-354.

Roy, V., Evangelou, E. and Zhu, Z. (2016), Efficient Estimation and Prediction for the Bayesian Binary Spatial Model with Flexible Link Functions, *Biometrics*, 72, 289-298.

Wang, X., Dey, D.K. and Banerjee, S. (2010), Non–Gaussian Hierarchical Generalized Linear Geostatistical Model Selection. In: M.H. Chen, D.K. Dey, P. Müller, D. Sun and K. Ye (eds.), *Frontiers of Statistical Decision Making and Bayesian Analysis—In Honor of James O. Berger*, pp 484-497. Springer.

Wang, Z. and Louis, T.A. (2003), Matching Conditional and Marginal Shapes in Binary Random Intercept Models Using a Bridge Distribution Function, *Biometrika*, 90, 765-775.

Zhang, H. (2002), On Estimation and Prediction for Spatial Generalized Linear Mixed Models, *Biometrics*, 58, 129-136.