Supplementary materials for this article are available at https://doi.org/10.1007/s13253-022-00512-3.

A Spatial Logistic Regression Model Based on a Valid Skew-Gaussian Latent Field

Vahid TADAYON and Mohammad Mehdi SABER

Logistic regression is commonly used to estimate the association of one (or more) independent variable(s) with a binary- dependent outcome. In many applications latent sources are both spatially dependent and non-Gaussian; thus, it is desirable to exploit both properties jointly. Spatial logistic regression is a well-established technique of including spatial dependence in logistic regression models. In this paper, we develop a spatial logistic regression model based on a valid skew-Gaussian random field. For parameter estimation, we use a Monte Carlo extension of the EM algorithm along with an approximation based on the standard logistic function. A simulation study is applied in order to determine the performance of the proposed model and also to compare the results with a recently introduced model with established efficiency. The identifiability of the parameters is investigated as well. As an illustrative purpose, an application to the Meuse heavy metals dataset is presented.

Supplementary materials accompanying this paper appear online.

Key Words: Binary spatial data; MCEM algorithm; Spatial modeling; Non-Gaussian random field.

1. INTRODUCTION

Many practical studies in public health, ecology and many other disciplines rely on binary spatial data. However, most of the conventional spatial analyses were designed to address the problem of estimation/prediction based on continuous observations. In the case of binary variables, for instance, diagnosis of groundwater pollution, there are only two possible outcomes, present (denoted as 1) or absent (denoted as 0). The logistic regression model is a well-known and well-documented methodology which is used in many contexts, specifically, in the presence of spatial dependence, see for example, Lin and Clayton (2005), Zhu et al. (2005), Xie et al. (2005), Tayyebi et al. (2010), Wu and Zhang (2013), Diggle and Giorgi (2016).

In a spatial framework, Paciorek (2007) focused on a large binary dataset and compared penalized likelihood and Bayesian models based on fit, speed and ease of implementation. He

V. Tadayon · M. M. Saber (✉) Department of Statistics, Higher Education Center of Eghlid, Eghlid, Iran
E-mail: vahidtadayon24@gmail.com; mmsaber@eghlid.ac.ir.

Journal of Agricultural, Biological, and Environmental Statistics, Volume 28, Number 1, Pages 59–73
https://doi.org/10.1007/s13253-022-00512-3

also devised an effective Markov chain Monte Carlo (MCMC) sampling scheme to address slow mixing of MCMC techniques in a generalized linear mixed model (GLMM). Zhu et al. (2008) studied logistic regression analysis of binary lattice data using a spatial–temporal autologistic regression model in a frequentist approach and used Monte Carlo maximum likelihood estimators for parameter estimation. To handle computational and inferential challenges posed by high-dimensional binary spatial data, Chang et al. (2016) presented a novel calibration method for computer models and applied a generalized principal component-based dimension reduction method. Sengupta et al. (2016) used a reduced-rank spatial random effects model to account for remote sensing datasets that can be massive in size and non-stationary in space. They estimated the parameters using an expectation–maximization (EM) algorithm. Nisa et al. (2019) focused on the estimation of propensity score as a method which is used to reduce bias due to confounding factors in the estimation of the treatment impact on observational data. They incorporated a spatial logistic regression model and used an EM algorithm to handle maximum likelihood estimation. Hardouin (2019) presented a variational method for parameter estimation in a logistic spatial regression since the expectations in the E-step of the EM algorithm were not available in closed-from expressions. Zhang et al. (2021) proposed a multivariate skew-elliptical link model for correlated binary responses, which included the multivariate probit model as a special case.

Intrinsically, the inference of a logistic regression model involves a hidden unobserved process, although in all aforementioned studies the hidden process has been treated as a user-friendly Gaussian random field. Nevertheless, in a whole range of applications, non-Gaussianity of the latent component arises explicitly from the existence of spatial/spatiotemporal heterogeneities. Thus, some active efforts to seek departures from Gaussianity called for some applicable strategies to handle some of the potential weaknesses associated with the transformation methods. A review of the most recent studies on this topic has been deemed by Tadayon and Torabi (2019) and Tadayon and Rasekh (2019). Mahmoudian (2018) discussed that most of previous skewed spatial models were ill-defined according to the consistency condition of the Kolmogorov existence theorem (Billingsley 2008) as their parametrization of the skewed distributions does not directly allow for an extension to a spatial random field model. Using the multivariate skew-normal distribution of Sahu et al. (2003) (SSN) they proposed a valid random field model with a skew structure to tackle non-Gaussian features and claimed that their random field is particularly convenient for computation. In addition, Mahmoudian (2018) expressed that the induced skewness under this family is not mixed with the spatial correlations.

To the best of our knowledge, the literature on modeling skewness in the case of binary spatial data is very scarce (Hosseini et al. 2011; Afroughi 2015). This design is very useful when our interest is to capture spatial dependence and avoid inefficient estimates by manipulating the data. In this paper, we focus on implementing the valid flexible skew-Gaussian random field introduced by Mahmoudian (2018) to address both spatial dependence and (possible) skewness through a logistic regression model. The plan of the remainder of this paper is as follows. The following section introduces our proposed spatial logistic regression model based on a valid skew-Gaussian random field and explains our methodology of estimating the model parameters. An analysis of a synthetic data is described in Sect. 3.

Section 4 analyzes the Meuse heavy metals dataset as an application of our methodology. Finally, the paper ends with some conclusions and final remarks (Sect. 5).

2. THE SPATIAL MODEL

Logistic regression is generally a kind of multiple regression model to analyze the relationship between a binary outcome and independent variables. Let $\mathbf{Z}(\mathbf{S}) = (Z(s_1), \ldots, Z(s_n))^{\mathrm{T}}$ be an observable vector of spatially dependent binary variables at locations $\mathbf{S} = (s_1, \ldots, s_n)^{\mathrm{T}}$. In a hierarchical setting, it is conventional to model $\mathbf{Z}(\mathbf{S})$ as Bernoulli variables, whose means depend on an underlying spatial process $\mathbf{Y}(\mathbf{S}) = (Y(s_1), \ldots Y(s_n))^{\mathrm{T}}$ such that $Z(s_i)$s are conditionally independent, given the hidden process $\mathbf{Y}(\mathbf{S})$. Like Tadayon and Torabi (2022), the specific hierarchical model we investigate has the following representation

$$\begin{aligned} Z(s)\,|Y(s) &\sim \mathrm{Ber}\left(p(s) = \frac{e^{Y(s)}}{1+e^{Y(s)}}\right) \\ Y(s) &= \mathbf{x}(s)^{\mathrm{T}}\boldsymbol{\beta} + \gamma W(s) + \varepsilon(s), \end{aligned} \tag{1}$$

where $\boldsymbol{\beta}$ is a vector of k unknown parameters with corresponding $\mathbf{x}(s) = (x_1(s), \ldots, x_k(s))^{\mathrm{T}}$ as a vector of known covariates that captures the large-scale spatial variation, γ is a scale parameter, $W(\cdot)$ takes account of the non-Gaussian features through a valid skewed random field in a latent mode. Finally, the white noise error $\varepsilon(\cdot) \sim \mathrm{N}\left(0, \tau^2\right)$ is considered to be independent of $W(\cdot)$. Evidently,

$$\begin{aligned} \Pr[Z(s) = z\,|Y(s) = y] &= p(s)^z[1-p(s)]^{1-z} \\ &= \frac{1}{1+\exp\{-y(2z-1)\}}. \end{aligned}$$

We consider $\mathbf{W}(\mathbf{S}) = (W(s_1), \ldots, W(s_n))^{\mathrm{T}}$ as the SSN process

$$\mathbf{W}(\cdot) \sim \mathrm{SSN}_n\left[-\sqrt{\frac{2}{\pi}}\delta\mathbf{1}_n, H, \delta I_n\right], \tag{2}$$

with the probability density function

$$f(\mathbf{w}) = 2^n\phi_n\left[\mathbf{w}; -\sqrt{\frac{2}{\pi}}\delta\mathbf{1}_n, H+\delta^2 I_n\right]\Phi_n\left[\delta\left(H+\delta^2 I_n\right)^{-1}(\mathbf{w}+\sqrt{\frac{2}{\pi}}\delta\mathbf{1}_n); \mathbf{0}, \Delta\right],$$

mean $\mathbf{0}$ and covariance matrix $H + (1-2/\pi)\delta^2 I_n$, where $\mathbf{w} \in \mathbb{R}^n$, $\Delta = I_n - \delta^2\left[H+\delta^2 I_n\right]^{-1}$, $\mathbf{1}_n$ denotes an $n \times 1$ vector of ones and I_n is the identity matrix. $\phi_n(\cdot; \mu, \Sigma)$ and $\Phi_n(\cdot; \mu, \Sigma)$ represent the normal density and the normal cumulative distribution function of $\mathrm{N}_n(\mu, \Sigma)$, respectively. The second term in the covariance matrix can be viewed as a nugget effect in geostatistics. The exponential correlation function is chosen for the entries of H such that $H_{ij} = \exp\{-\|h\|/\psi\} = \exp\left\{-\left\|s_i - s_j\right\|/\psi\right\}$, where ψ is the range

parameter. Therefore, the complete log likelihood function of $\boldsymbol{\eta} = \left(\boldsymbol{\beta}, \gamma, \tau^2, \delta, \psi\right)^{\mathrm{T}}$ is given by

$$\begin{aligned}\ell(\boldsymbol{\eta}) = &-\sum_i \ln\left(1 + e^{Y(s_i)}\right) + \sum_i Y(s_i)\, Z(s_i) \\ &-\frac{1}{2\tau^2}\boldsymbol{\varepsilon}^{\mathrm{T}}\boldsymbol{\varepsilon} - \ln\left|H + \delta^2 I_n\right| - \frac{n}{2}\ln \pi^2\tau^2 \\ &-\frac{1}{2}\left[\mathbf{W} + \sqrt{\frac{2}{\pi}}\delta\mathbf{1}_n\right]^{\mathrm{T}}\left[H + \delta^2 I_n\right]^{-1}\left[\mathbf{W} + \sqrt{\frac{2}{\pi}}\delta\mathbf{1}_n\right] \\ &+\ln \Phi_n\left[\delta\left(H + \delta^2 I_n\right)^{-1}(\mathbf{W} + \sqrt{\frac{2}{\pi}}\delta\mathbf{1}_n); \mathbf{0}, \Delta\right], \end{aligned} \tag{3}$$

where $|\cdot|$ denotes the determinant. Since the likelihood function $\ell(\boldsymbol{\eta})$ is analytically intractable, one can use a natural extension of the EM algorithm that employs Monte Carlo methods (MCEM algorithm) to estimate the model parameters $\boldsymbol{\eta}$. In order to be self-contained, we recall that the EM algorithm operates on the so-called Q-function where at the tth *E-step* is defined by

$$Q\left(\boldsymbol{\eta}\,|\boldsymbol{\eta}^t\right) = E\left[\ell\left(\boldsymbol{\eta}\,|\mathbf{Z}, \boldsymbol{\eta}^t\right)\right] = \int \ell(\boldsymbol{\eta})\, f\left(\mathbf{W}, \boldsymbol{\varepsilon}\,|\mathbf{Z}, \boldsymbol{\eta}^t\right) d\mathbf{W} d\boldsymbol{\varepsilon}. \tag{4}$$

The *M-step* is to maximize Q with respect to $\boldsymbol{\eta}$ to obtain $\boldsymbol{\eta}^{t+1} = \arg\max_{\boldsymbol{\eta}\in\Theta} Q\left(\boldsymbol{\eta}\,|\boldsymbol{\eta}^t\right)$, where Θ is the parameter space. When the integral in equation (4) is analytically intractable or very high dimensional the MCEM algorithm presents a modification of the EM algorithm where the expectation in the *E-step* is computed numerically through Monte Carlo simulation. By replacing the conditional expectations in (4) with the corresponding Monte Carlo approximations, we can write

$$Q\left(\boldsymbol{\eta}\,|\boldsymbol{\eta}^t\right) \approx \frac{1}{M}\sum_{m=1}^{M} \ell\left(\boldsymbol{\eta}^t; \mathbf{Z}, \mathbf{W}^{(m)}, \boldsymbol{\varepsilon}^{(m)}\right),$$

and employ an optimization procedure to maximize $Q\left(\boldsymbol{\eta}\,|\boldsymbol{\eta}^t\right)$ with respect to $\boldsymbol{\eta}$. These steps are repeated until convergence conditions of the MCMC were satisfied through the Gelman–Rubin convergence diagnostics (Gelman and Rubin 1992). At the tth iteration of the MCEM algorithm, we need to calculate some conditional expectations of the form of $E_i\left[\mathrm{g}\left(\mathbf{W}, \boldsymbol{\varepsilon}\right)|\mathbf{Z}\right]$, $i \in \{1, \ldots, 7\}$ for some function g of $\mathbf{W}$ and $\boldsymbol{\varepsilon}$. These conditional expectations that are shown in Equation (A1) of Appendix as an extended form of (4) may not have explicit forms and need to be substituted by their Monte Carlo approximations. We use the notation $\mathbb{E}_i^t(\cdot)$ to show the corresponding approximation of the ith conditional expectation $E_i(\cdot)$ whenever it does not have a closed form. $\mathbb{E}_i^t$ can be calculated based on samples $\left\{\mathbf{W}^{(m)}, \boldsymbol{\varepsilon}^{(m)}\right\}_{m=1}^{M}$ from the joint distribution $f_{\mathbf{W},\boldsymbol{\varepsilon}|\mathbf{Z},\boldsymbol{\eta}}$ as

$$\mathbb{E}_i^t = M^{-1}\sum\nolimits_{m=1}^{M} \mathrm{g}_i\left(\mathbf{Z}, \mathbf{W}^{(m)}, \boldsymbol{\varepsilon}^{(m)}; \boldsymbol{\eta}^t\right).$$

For details regarding the updates of the model parameters through the M-step see the Appendix in which a variational method is used to estimate the parameters. Generating samples from the joint distribution $f_{\mathbf{W},\boldsymbol{\varepsilon}|\mathbf{Z},\boldsymbol{\eta}}$ requires a MCMC algorithm. To that end, we explore the full conditional distributions as follows.

- $\boldsymbol{\varepsilon}\,|\mathbf{Z}, \mathbf{W}, \boldsymbol{\eta}$: According to the details of the variational method described in Appendix, we can write

$$f\left(\boldsymbol{\varepsilon}\,|\mathbf{Z}, \mathbf{W}, \boldsymbol{\eta}\right) \propto \exp\left\{-\frac{1}{2\tau^2}\sum\nolimits_i \varepsilon_i^2 + \sum\nolimits_i Z_i\varepsilon_i - \frac{1}{2}\sum\nolimits_i \varepsilon_i \right.$$
$$\left. - \sum\nolimits_i \lambda\left(\theta_i\right)\left(\varepsilon_i^2 + 2\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\varepsilon_i + 2\gamma W_i\varepsilon_i\right)\right\},$$

therefore, the full conditional distribution of ε_is is approximately proportional to a normal density as

$$\varepsilon_i\,|\mathbf{Z}, \mathbf{W}, \boldsymbol{\eta} \overset{d}{\simeq} \mathrm{N}\left[\frac{Z_i - 2\lambda\left(\theta_i\right)\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + \gamma W_i\right) - 0.5}{\tau^{-2} + 2\lambda\left(\theta_i\right)}, \frac{1}{\tau^{-2} + 2\lambda\left(\theta_i\right)}\right].$$

- $\mathbf{W}\,|\mathbf{Z}, \boldsymbol{\varepsilon}, \boldsymbol{\eta}$: With regard to the hierarchical representation of the SSN distribution based on a normal and a truncated normal distributions, we can rewrite $\mathbf{W}$ as

$$\mathbf{W}\,|\mathbf{V} = \mathbf{v} \sim \mathrm{N}_n\left[\boldsymbol{\mu}_{\mathbf{v}}, H\right], \qquad \boldsymbol{\mu}_{\mathbf{v}} = \delta\left(\mathbf{v} - \sqrt{\frac{2}{\pi}}\mathbf{1}_n\right),$$

where $\mathbf{V} \sim \mathrm{N}_n\left[\mathbf{0}, I_n\right]\mathbf{I}_{\{\mathbb{R}_+^n\}}\left(\mathbf{V}\right)$ and $\mathbf{I}_{\{\cdot\}}\left(\cdot\right)$ denotes the indicator function. Therefore,

$$f\left(\mathbf{W}\,|\mathbf{Z}, \boldsymbol{\varepsilon}\,, \boldsymbol{\eta}\right) \propto f\left(\mathbf{Z}\,|\mathbf{W}, \boldsymbol{\varepsilon}\,, \boldsymbol{\eta}\right) f\left(\mathbf{W}\,|\mathbf{V}, \boldsymbol{\eta}\right) f\left(\mathbf{V}\right)$$
$$\propto \exp\left\{-\frac{1}{2}\gamma\sum\nolimits_i W_i - \gamma^2\sum\nolimits_i \lambda\left(\theta_i\right) W_i^2 - 2\gamma\sum\nolimits_i \lambda\left(\theta_i\right)\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}W_i\right.$$
$$\left. - 2\gamma\sum\nolimits_i \lambda\left(\theta_i\right)\varepsilon_i W_i + \gamma\sum\nolimits_i Z_i W_i\right\}$$
$$\times \exp\left\{-\frac{1}{2}\left(\mathbf{W}^{\mathrm{T}}H^{-1}\mathbf{W} - 2\boldsymbol{\mu}_{\mathbf{v}}^{\mathrm{T}}H^{-1}\mathbf{W}\right)\right\}.$$

One can synthesize the above terms to obtain $\mathbf{W}\,|\mathbf{Z}, \mathbf{V}, \boldsymbol{\varepsilon}, \boldsymbol{\eta} \overset{d}{\simeq} \mathrm{N}_n\left[\boldsymbol{\mu}_{\mathbf{w}|\cdot}, \Sigma_{\mathbf{w}|\cdot}\right]$, where

$$\boldsymbol{\mu}_{\mathbf{w}|\cdot} = \Sigma_{\mathbf{w}|\cdot}\left(H^{-1}\boldsymbol{\mu}_{\mathbf{v}} - \frac{\mathcal{C}}{2}\right), \qquad \Sigma_{\mathbf{w}|\cdot} = \left(H^{-1} + \mathcal{D}\right)^{-1},$$

in which $\mathcal{C}$ is an $n \times 1$ vector with elements $c_i = \gamma\left(1 + 4\lambda\left(\theta_i\right)\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + 4\lambda\left(\theta_i\right)\varepsilon_i - 2Z_i\right)$ and $\mathcal{D}$ is a diagonal matrix as $\mathcal{D} = 2\gamma^2\mathrm{diag}\left(\lambda\left(\theta_1\right), \ldots, \lambda\left(\theta_n\right)\right)$.

- $\mathbf{V}\,|\mathbf{Z}, \mathbf{W}, \boldsymbol{\varepsilon}, \boldsymbol{\eta}$:

$$f\left(\mathbf{V}\,|\mathbf{Z}, \mathbf{W}, \boldsymbol{\varepsilon}\,, \boldsymbol{\eta}\right) \propto f\left(\mathbf{W}\,|\mathbf{V}, \boldsymbol{\eta}\right) f\left(\mathbf{V}\right)$$
$$\propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{\mu}_{\mathbf{v}}^{\mathrm{T}}H^{-1}\boldsymbol{\mu}_{\mathbf{v}} - 2\boldsymbol{\mu}_{\mathbf{v}}^{\mathrm{T}}H^{-1}\mathbf{W} + \mathbf{V}^{\mathrm{T}}\mathbf{V}\right)\right\}\mathbf{I}_{\{\mathbb{R}_+^n\}}\left(\mathbf{V}\right),$$

hence, $\mathbf{V}\,|\mathbf{Z}, \mathbf{W}, \boldsymbol{\varepsilon}, \boldsymbol{\eta} \sim \mathrm{N}_n \left[\Sigma_{\mathbf{v}|\cdot} \left(\delta H^{-1}\mathbf{W} - \sqrt{\frac{2}{\pi}} H^{-1}\mathbf{1}_n\right), \Sigma_{\mathbf{v}|\cdot} \right] \mathbf{I}_{\{\mathbb{R}^n_+\}}(\mathbf{V})$, where its covariance matrix can be written as $\Sigma_{\mathbf{v}|\cdot} = \left(\delta^2 H^{-1} + I_n\right)^{-1}$.

3. ANALYSIS OF A SYNTHETIC DATASET

We now assess the performance of the proposed model using a synthetic dataset along with making a comparison between our results and the one is resulted by applying the model presented in Hardouin (2019). Thus, the contribution of this section is twofold. First, the performance of the presented model is evaluated in estimating the parameters using the response variable generated from model (1) (using algorithm 1), and then, the effect of sample size (the number of spatial locations) on model performance is examined. Ultimately, the results are compared with that of its competitor. All computations were performed using the publicly available statistical software R.

To address our goals, we use algorithm 1 to generate spatially correlated binary data $Z(s_i)$ with $E[Z(s_i)] = p(s_i)$ and $\rho\left[Z(s_i), Z(s_j)\right] = H_{ij} = \exp\{-\|h\|/\psi\}$. We did three distinct simulations, each with $\mathcal{R} = 500$ generated datasets for three different sample sizes as $n = 200, 400$ and 800 observations, respectively. For all three simulations, we set $M = 100$. In each simulation study, the sites are uniformly distributed over the region $(0, 10) \times (0, 10)$. The data were simulated from the model (1) with $x_i \sim N(0, 1)$ where the true values of the model parameters has been shown in Table 1 which also summarizes the results. Notice that choosing $\psi = 3.5$ in each simulation yields the rough values 0.02 and 0.99 for $\exp\{-\|h\|/\psi\}$ as the approximations of the maximum and minimum dependencies based on the presented exponential correlation function correspond to the smallest and largest distances between the selected locations, respectively.

Table 1 specially reports the bias criterion for an arbitrary parameter, say ϑ, as

$$\mathrm{Bias}\left(\hat{\vartheta}\right) = \mathcal{R}^{-1} \sum\nolimits_{r=1}^{\mathcal{R}} \left(\hat{\vartheta}^{(r)} - \vartheta\right)$$

and also the empirical variance of each estimation as

$$\mathrm{MSE}\left(\hat{\vartheta}\right) = \mathcal{R}^{-1} \sum\nolimits_{r=1}^{\mathcal{R}} \left(\hat{\vartheta}^{(r)} - \bar{\vartheta}\right)^2 \tag{5}$$

to assess the performance of the proposed methodology, where $\bar{\vartheta} = \mathcal{R}^{-1} \sum_{r=1}^{\mathcal{R}} \hat{\vartheta}^{(r)}$. It is worthwhile mentioning that in Hardouin (2019)'s approach spatial variation is captured through the term $\varepsilon(\cdot)$ with the same exponential correlation function, i.e., $\boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, H^*)$, where $H^*_{ij} = \tau^2 \exp\{-\|h\|/\psi\}$. Eventually, the resultant Akaike information criterion (AIC) values were used to compare model performance. This benchmark, which is the most popular criterion for model assessment in the literature, is calculated as AIC $=$ 2[#model parameters $- \dot{\boldsymbol{\ell}}_{y,\lambda}$]. The AIC values corroborate better performance of the proposed model compared to its competitor. Note that Table 1 compares parameter estimates for the data generated from the proposed model with a misspecified model considered in

Table 1. Bias value and the empirical variance (Evar) of the estimated parameters for the proposed and Hardouin (2019)'s approaches based on three different simulations with $n = 200$, 400 and 800

	Parameter	True value	Proposed model		Hardouin's model	
			Bias	EVar	Bias	EVar
$n = 200$	β_0	1.5	0.025	0.097	−0.837	1.657
	β_1	0.8	0.022	0.088	1.017	1.544
	γ	0.5	−0.031	0.092	–	–
	δ	0.9	0.096	0.094	–	–
	τ^2	0.4	−0.091	0.097	1.530	1.991
	ψ	3.5	0.099	0.093	1.153	2.023
	AIC		862.4		994.1	
	MCR		13.5%		32.8%	
$n = 400$	β_0	1.5	0.019	0.085	0.840	1.651
	β_1	0.8	−0.017	0.069	0.981	1.538
	γ	0.5	0.02	0.078	–	–
	δ	0.9	−0.089	0.08	–	–
	τ^2	0.4	0.081	0.082	1.147	1.993
	ψ	3.5	−0.09	0.089	−1.148	2.004
	AIC		841.3		980.7	
	MCR		11.7%		31.7%	
$n = 800$	β_0	1.5	−0.011	0.075	−0.710	1.624
	β_1	0.8	−0.01	0.06	0.932	1.517
	γ	0.5	0.017	0.071	–	–
	δ	0.9	0.073	0.069	–	–
	τ^2	0.4	−0.074	0.072	1.135	1.985
	ψ	3.5	−0.85	0.081	1.132	1.976
	AIC		840.6		971.1	
	MCR		10.4%		28.5%	

AIC Akaike information criterion, *MCR* misclassification rate

Hardouin (2019), so it is expected that parameter estimates will be biased for the misspecified model. To address this issue, we use the mean squared prediction error (MSPE) to assess the performance of suggested strategy. In classification problems prediction error is commonly defined as the probability of an incorrect classification, also called the misclassification rate (MCR). To compute MCR, we randomly drop $n/10$ observations from each simulation; then, MCR is calculated by $\text{MCR} = (n/10)^{-1} \sum_{i=1}^{n/10} I\left(\widehat{Z}_i \neq Z_i\right)$, where $I(\cdot)$ is the indicator function that is equal to one when its input is true. The results which have been reported as percentage in Table 1 represent lower MCRs for the suggested model and also show that as the sample size increases MCRs decrease.

Algorithm 1 Generate spatially correlated skewed binary data $Z(s)$ from model (1)

I. Generate $\mathbf{V}$ from $\mathrm{N}_n[\mathbf{0}, I_n]\mathbf{I}_{\{\mathbb{R}_+^n\}}(\mathbf{V})$,

II. Generate normally distributed and spatially correlated $\mathbf{W}|\mathbf{V} = \mathbf{v} \sim \mathrm{N}_n\left[\boldsymbol{\mu}_{\mathbf{v}}, H\right]$ with $\rho\left[W(s_i)|\mathbf{V}, W\left(s_j\right)|\mathbf{V}\right] = H_{ij}$,

III. Calculate $Y(s_i) = \beta_0 + \beta_1 x_i + \gamma W_i + \varepsilon_i$ and determine $p(s_i) = e^{Y(s_i)}/(1 + e^{Y(s_i)})$,

IV. Generate $Z(s_i)$ from $\mathrm{Ber}(p(s_i))$.

Table 2. Bias value and the empirical variance (Evar) of the estimated parameters for the proposed model based on a simulation with $n = 800$ and $\psi = 0.1$. MCR = misclassification rate

	Parameter	True value	Bias	EVar
$n = 800$	β_0	1.5	0.012	0.075
	β_1	0.8	−0.013	0.062
	γ	0.5	0.018	0.070
	δ	0.9	−0.075	0.074
	τ^2	0.4	−0.077	0.066
	ψ	0.1	0.079	0.079
	MCR		10.4%	

Table 3. Identifiability of two parameters τ^2 and δ based on three different simulations with $n = 200$

δ	True value	0.7	1	1.3
	Bias value	0.092	−0.087	−0.09
	Empirical variance	0.093	0.09	0.093
τ^2	True value	0.3	0.5	0.8
	Bias value	−0.083	0.087	0.083
	Empirical variance	0.091	0.090	0.090

To evaluate the performance of the proposed model in different scale of spatial dependence, we did another simulation based on $n = 800$, however, in this case, we fixed all parameters $\beta_0, \beta_1, \gamma, \delta$ and τ^2 the same as what considered before in Table 1 and only altered ψ to $\psi = 0.1$ which allows the spatial dependency to vary from almost 0 to 0.5. We are aware that in practical issues this value should be considered according to the autocorrelation function relative to the size of the domain, however, it has been chosen to assess the performance of the proposed model in the case of low spatial dependence. The results that are presented in Table 2 could readily be compared with the corresponding part of Table 1. Patently, the results substantiate stability in the performance (bias, the empirical variance and MCR) of the proposed model in the case of low spatial dependence.

Finally, since the inference may be challenging in identifying the nugget effect components (τ^2 and δ), here, we discuss to what extent information about these parameters can be recovered from data. To assess identifiability of each of these parameters, say δ, three datasets (of size 200) were generated from the proposed model with different values of δ (and fixed values for other parameters, as described in Table 1). Then, the estimated values were obtained. The same applies for inference on τ^2. Table 3 indicates that the data allow for meaningful inference on the model's nugget effect components.

4. APPLICATION: THE MEUSE HEAVY METALS DATA

In this section, we illustrate our proposed methodology using a well-known real dataset in the literature on spatial statistics. The Meuse dataset which has been documented in detail by Rikken and Van Rijn (1993) and Burrough and McDonnell (1998) and have been studied

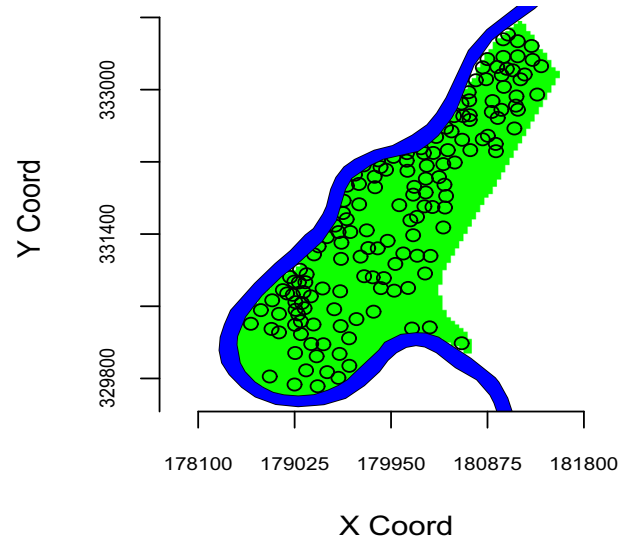

Figure 1. Study area, coordinates are in RDM, the Dutch topographical map coordinate system. The blue color shows the Meuse River .

frequently in several geostatistical researches, comprise heavy metals measurements in the topsoil in a flood plain along the Meuse River west of the municipality of Stein, Limburg, the Netherlands. The dataset is available in the R package `sp` and can be loaded with the data function as `data(meuse)`. The measures consist of 155 soil samples collected in an area of approximately $15m \times 15m$ which were analyzed for their concentration of toxic heavy metals (zinc, lead, copper, and cadmium) in ppm. Figure 1 below depicts a schematic description of the region and sampling locations.

We chose the binary variable *lime* as our response of interest Z and simultaneously in order to find the most related covariates to our response, corrected AIC (AICC) introduced by Hoeting et al. (2006) (for geostatistical model selection) was used. AICC is given by

$$\text{AICC} = 2\left[n\frac{p+k+1}{n-p-k-2} - \ell_{y,\lambda}\right],$$

where, p shows the number of regression coefficients including an intercept term, k is the number of parameters associated with the autocorrelation function and n is the number of observed sites. Considering four variables *zinc, lead, copper,* and *cadmium* as potential covariates, we investigated among all $2^4 - 1$ feasible embedded models and ultimately, an overall consideration (that are not presented here) resulted in a model with two covariates *lead* and *zinc*. Although the simulation study showed that estimation of parameters does not depend on initial values for parameters, we use estimates of ordinary GLM for initial values of regression coefficients, i.e., $\beta_0 = -3.34$, $\beta_{lead} = -0.03$ and $\beta_{zinc} = 0.01$. From $p(s) = e^{Y(s)}/\left(1 + e^{Y(s)}\right)$ and Equation (1) we can write

$$\gamma W(s) + \varepsilon(s) \simeq \ln\left(\frac{\bar{Z}}{1-\bar{Z}}\right) - \mathbf{x}(s)^{\mathrm{T}}\boldsymbol{\beta},$$

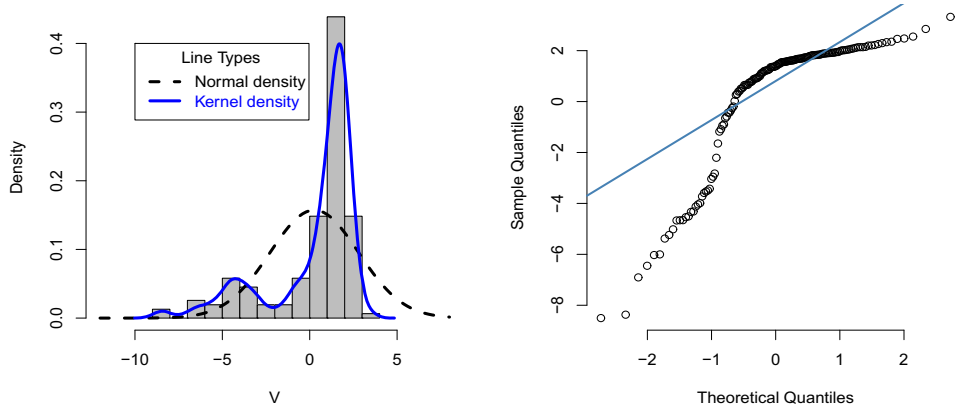

Figure 2. Left panel displays the histogram of $V(s)$ with its kernel density estimate and right panels shows its normal QQ-plot .

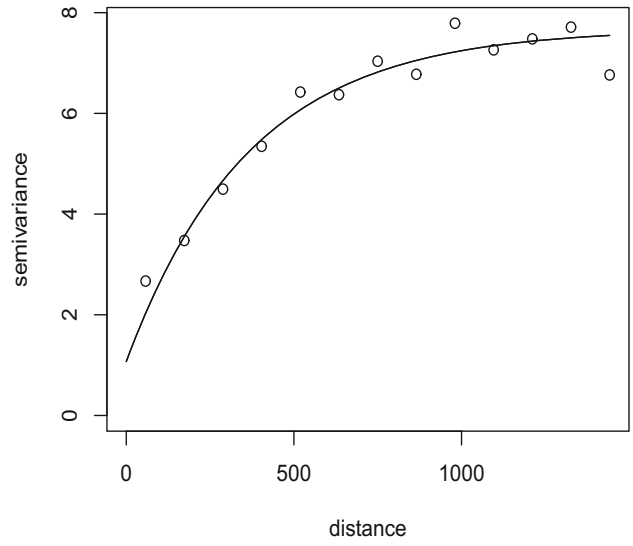

Figure 3. The empirical semi-variogram of $V(s)$.

where $\widehat{p(s)} = \bar{Z} = 0.284$. Now, we define $V(s) = \gamma W(s) + \varepsilon(s)$ which is approximately given by $V(s) = 2.422 + 0.038lead(s) - 0.017zinc(s)$. We can see that $V(s)$ is a member of the SSN family (2). The Q-Q plot and histogram of $V(s)$ are demonstrated in Fig. 2. As a result of simple exploratory data analysis, the histogram shows a non-Gaussian feature, which confirms the suitability of implementing the proposed model based on above-mentioned skew random field. The empirical semi-variogram of $V(s)$ was plotted in Fig. 3. The best model was exponential with parameters $nugget\ effect = 1.07, sill = 6.60$ and $range = 367.24$. Table 4 displays the model parameter estimates and the corresponding standard error for the proposed model and MCRs of both competitor models. The presented estimated-values/MCRs were calculated as the mean of estimated values/MCRs over $\mathcal{R} = 20$ runs of the program.

Table 4. The estimated values and standard error of parameters based on the presented model. Misclassification rate (MCR) has been presented for both competitor models

Parameter	Estimated value	Standard error
β_0	−2.11	0.032
β_{lead}	−1.14	0.034
β_{zinc}	0.86	0.031
γ	1.97	0.043
δ	−2.02	0.040
τ^2	1.36	0.041
ψ	335.17	0.52
MCR	Proposed model	13.1%
	Hardouin's model	34.11%

5. CONCLUSION

The present study concentrated on implementing a valid flexible skew-Gaussian random field based on the skew-normal family introduced by Sahu et al. (2003) to capture both spatial dependence and (possible) skewness through a logistic regression model. Declaring that directly maximizing the likelihood function of observed data is intractable, a Monte Carlo extension of EM algorithm was developed to compute the maximum likelihood estimate of model parameters. Moreover, a simulation study was conducted to assess the performances of the proposed model and also to investigate the effect of sample size on the results. Finally, a real data application regarding the presence of lime in the topsoil along the Meuse River was also analyzed in which, the concentration of toxic heavy metals zinc and lead were considered as two covariates.

Overall, the proposed model added flexibility to the class of spatial logistic regression models often considered in the literature to account for binary spatial data. It must be mentioned that, in the spatial context, the asymptotic properties of parameter estimators strongly depend on the asymptotic regime which is considered. Specifically, two regimes can be considered, first, when the spatial domain is fixed and bounded and the density of the sampling locations increases with n (the fixed/infill domain). Second, when the spatial domain of observation is unbounded and it grows in size with the sample dimension n (the increasing domain framework). Whereas under the latter regime the maximum likelihood estimators are consistent and asymptotically normal, subject to some regularity conditions (see, for example, Mardia and Marshall (1984)), under the former analogous results do not hold and model parameters could not be consistently estimated. Besides this, in the suggested approach, the latent factors are independent for each location which results in satisfaction of mixing conditions. However, in the general case, replicates are required to obtain consistent estimates even if the number of locations is large.

An astonishing extension of this work is to investigate how the variance process can be allowed to depend on covariates which opens up an opportunity to interpret tail behavior of the process as a function of known covariates. Another step forward is to let this covariance-covariate dependence change in time. On the other hand, in the last decade, with

the wide usage of mobile applications and Global Positioning System (GPS) devices as well as the advancement of remote sensors which are accompanied with cheap data storage/computational devices, many geo-referenced data are being collected. As a result, there has been a growing enthusiasm for modeling spatial big data. The third interesting extension of this work is to scale the proposed model to big data. Moreover, in this study the exponential correlation function was chosen, although this could affect the smoothness of the process. One can choose a more flexible spatial correlation structure and compare the results. We have planned to study these approaches in our future studies.

5.1. SUPPLEMENTARY MATERIAL

Supplementary materials contain R codes for simulations and real data application conducted in this paper.

ACKNOWLEDGEMENTS

We would like to thank the Associate Editor and two reviewers for the constructive comments and suggestions, which led to an improved version of this paper.

Author Contribution VT contrived the study, conceptualized the review, reviewed and revised the manuscript. The simulation study, fitting the model to the real data, and documenting the whole manuscript were performed by the first author. MMS had the majority role in the theoretical part of the modeling and also he found an appropriate real data set. Exploratory data analysis of the real data and also some parts of the R functions were provided by the second author.

[Received February 2022. Revised July 2022. Accepted July 2022. Published Online September 2022.]

APPENDIX

In what follows, we use the notation ϑ_i to show $\vartheta\,(s_i)$. Equation (4) can be written as

$$
\begin{aligned}
Q\left(\boldsymbol{\eta}\,|\boldsymbol{\eta}^t\right) = &-\sum_i E_1\left[\ln\left\{1+\exp\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}+\gamma W_i+\varepsilon_i\right)\right\}\Big|\,\mathbf{Z}\right]\\
&+\sum_i Z_i\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}+\gamma\sum_i Z_i E_2\left[W_i\,|\mathbf{Z}\right]+\sum_i Z_i E_3\left[\varepsilon_i\,|\mathbf{Z}\right]\\
&-\frac{1}{2\tau^2}E_4\left[\boldsymbol{\varepsilon}^{\mathrm{T}}\boldsymbol{\varepsilon}\,|\mathbf{Z}\right]-\ln\left|H+\delta^2 I_n\right|-\frac{n}{2}\ln\pi^2\tau^2\\
&-\frac{1}{2}\mathrm{trace}\left\{\left(H+\delta^2 I_n\right)^{-1}E_5\left(\mathbf{W}\mathbf{W}^{\mathrm{T}}\,|\mathbf{Z}\right)\right\}\\
&-\sqrt{\frac{2}{\pi}}\delta\mathbf{1}_n^{\mathrm{T}}\left[H+\delta^2 I_n\right]^{-1}E_6\left[\mathbf{W}\,|\mathbf{Z}\right]-\frac{1}{\pi}\delta^2\mathbf{1}_n^{\mathrm{T}}\left[H+\delta^2 I_n\right]^{-1}\mathbf{1}_n\\
&+E_7\left[\ln\Phi_n\left\{\delta\left(H+\delta^2 I_n\right)^{-1}(\mathbf{W}+\sqrt{\frac{2}{\pi}}\delta\mathbf{1}_n);\,\mathbf{0},\,\Delta\right\}\Big|\,\mathbf{Z}\right],
\end{aligned}
\tag{A1}
$$

in which the fourth line has been derived from the properties $x^{\mathrm{T}}Ax = \operatorname{trace}\left(x^{\mathrm{T}}Ax\right)$ and $\operatorname{trace}(AB) = \operatorname{trace}(BA)$. A closer scrutiny shows, however, that one of the main problematic terms of (A1) is $\sum_i E_1\left[\ln\left(1+e^{Y_i}\right)\middle|\mathbf{Z}\right]$. Hardouin (2019) proposed a variational method which is based on replacing this term by an initial approximation of the logistic function $\kappa(x) = e^x/(1+e^x) = 1/\left(1+e^{-x}\right)$ which had been studied by Jaakkola and Jordan (2000) as

$$\ln\kappa(x) \geq \ln\kappa(\theta) + \frac{x-\theta}{2} - \lambda(\theta)\left(x^2-\theta^2\right), \qquad \lambda(\theta) = \frac{\kappa(\theta)-1/2}{2\theta}.$$

This variational lower bound involves the model parameters and the so-called variational parameter θ. Let $\Theta = (\theta_1, \ldots, \theta_n)^{\mathrm{T}}$, we apply this lower bound to $-\sum_i \ln\left(1+e^{Y_i}\right) = \sum_i \ln\kappa(-Y_i)$ as the first term of (3). Therefore,

$$\begin{aligned} -\sum\nolimits_i \ln\left(1+e^{Y_i}\right) \geq \sum\nolimits_i &[\ln\kappa(\theta_i) - \frac{\theta_i}{2} + \theta_i^2\lambda(\theta_i)] \\ &-\frac{1}{2}\left[\sum\nolimits_i \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + \gamma W_i + \varepsilon_i\right] \\ &-\sum\nolimits_i \lambda(\theta_i)\Big[\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right)^2 + \gamma^2 W_i^2 + \varepsilon_i^2 \\ &+2\gamma\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}W_i + 2\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\varepsilon_i + 2\gamma W_i\varepsilon_i\Big]. \end{aligned} \tag{A2}$$

The monotonicity of expectation implies that getting the (conditional) expectation of (A2) (given $\mathbf{Z}$) preserves the inequality. Now, we can write $Q\left(\boldsymbol{\eta}\middle|\boldsymbol{\eta}^t\right) \geq \tilde{Q}\left(\boldsymbol{\eta}, \Theta\middle|\boldsymbol{\eta}^t, \Theta^t\right)$, where $\tilde{Q}$ has been resulted by replacing the first term of Q with the expectation of the right hand side of (A2) given $\mathbf{Z}$, which eliminates $E_1\left[\ln(1+\exp\{Y_i\})|\mathbf{Z}\right]$ and incorporates $E_8\left[W_i^2|\mathbf{Z}\right]$ and $E_9\left[\varepsilon_i^2|\mathbf{Z}\right]$ into inference. We then use a two-stage estimation procedure in the *M-step*, where the first stage consists of maximizing $\tilde{Q}\left(\boldsymbol{\eta}, \Theta\middle|\boldsymbol{\eta}^t, \Theta^t\right)$ with respect to the model parameters for fixed Θ results in $\tilde{Q}\left(\boldsymbol{\eta}^{t+1}, \Theta\middle|\boldsymbol{\eta}^t, \Theta^t\right)$, and in the second stage, updated variational parameters Θ^{t+1} is obtained by maximizing $\tilde{Q}\left(\boldsymbol{\eta}^{t+1}, \Theta\middle|\boldsymbol{\eta}^t, \Theta^t\right)$ with respect to Θ. The updates of the model parameters are as follows. $\tau^{2^{t+1}} = n^{-1}\mathbb{E}_4^t$, $\boldsymbol{\beta}^{t+1}$ can be easily obtained as a solution of the systems of linear equations

$$\sum\nolimits_i \lambda\left(\theta_i^t\right)\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}^{t+1}\right)\mathbf{x}_i^{\mathrm{T}} = -\sum\nolimits_i\left[\frac{1}{4} + \frac{Z_i}{2} + \lambda\left(\theta_i^t\right)\left(\gamma^t\mathbb{E}_2^t + \mathbb{E}_3^t\right)\right]\mathbf{x}_i^{\mathrm{T}},$$

in which the left-hand side can be rewritten as $\left[\sum_i \mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}\right]\boldsymbol{\beta}^{t+1}$, then,

$$\boldsymbol{\beta}^{t+1} = -\left[\sum\nolimits_i \lambda\left(\theta_i^t\right)\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}\right]^{-1}\sum\nolimits_i\left[\frac{1}{4} + \frac{Z_i}{2} + \lambda\left(\theta_i^t\right)\left(\gamma^t\mathbb{E}_2^t + \mathbb{E}_3^t\right)\mathbf{x}_i\right].$$

Moreover,

$$\gamma^{t+1} = \frac{\sum_i\left[Z_i - 0.5 - 2\lambda\left(\theta_i^t\right)\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}^t + \mathbb{E}_3^t\right)\right]\mathbb{E}_2^t}{2\sum_i \lambda\left(\theta_i^t\right)\mathbb{E}_8^t},$$

$$\delta^{t+1} = \arg\max_{\delta} \left\{ -\ln \left| H^t + \delta^2 I_n \right| - \frac{1}{2} tr\{(H^t + \delta^2 I_n)^{-1} \mathbb{E}_5^t\} \right.$$
$$-\sqrt{\frac{2}{\pi}} \delta \mathbf{1}_n{}^{\mathrm{T}} \left(H^t + \delta^2 I_n \right)^{-1} \mathbb{E}_6^t - \frac{1}{\pi} \delta^2 \mathbf{1}_n{}^{\mathrm{T}} \left(H^t + \delta^2 I_n \right)^{-1} \mathbf{1}_n$$
$$\left. + \mathbb{E}_7^t \{ \ln \Phi_n (\delta (H^t + \delta^2 I_n)^{-1} (\mathbf{W} + \sqrt{\frac{2}{\pi}} \delta \mathbf{1}_n); \mathbf{0}, I_n - \delta^2 [H^t + \delta^2 I_n]^{-1}) \mid \mathbf{Z} \} \right\},$$

$$\psi^{t+1} = \arg\max_{\psi} \left\{ -\ln \left| H + \delta^{t^2} I_n \right| - \frac{1}{2} tr\{(H + \delta^{t^2} I_n)^{-1} \mathbb{E}_5^t\} \right.$$
$$-\sqrt{\frac{2}{\pi}} \delta^t \mathbf{1}_n{}^{\mathrm{T}} \left(H + \delta^{t^2} I_n \right)^{-1} \mathbb{E}_6^t - \frac{1}{\pi} \delta^{t^2} \mathbf{1}_n{}^{\mathrm{T}} \left(H + \delta^{t^2} I_n \right)^{-1} \mathbf{1}_n$$
$$\left. + \mathbb{E}_7^t \{ \ln \Phi_n (\delta^t (H + \delta^{t^2} I_n)^{-1} (\mathbf{W} + \sqrt{\frac{2}{\pi}} \delta^t \mathbf{1}_n); \mathbf{0}, I_n - \delta^{t^2} (H + \delta^{t^2} I_n)^{-1}) \mid \mathbf{Z} \} \right\},$$

$$\left(\theta_i^{t+1} \right)^2 = \left(\mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}^{t+1} \right)^2 + \left(\gamma^{t+1} \right)^2 \mathbb{E}_8^{t+1} + \mathbb{E}_9^{t+1}$$
$$+ 2\gamma^{t+1} \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}^{t+1} \mathbb{E}_2^{t+1} + 2\mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}^{t+1} \mathbb{E}_3^{t+1} + 2\gamma^{t+1} \mathbb{E}_2^{t+1} \mathbb{E}_3^{t+1}.$$

REFERENCES

Afroughi S (2015) Bayesian inference of spatially correlated binary data using skew-normal latent variables with application in tooth caries analysis. Open J Stat 5:127–139

Billingsley P (2008) Probability and measure. Wiley, Hoboken

Burrough PA, McDonnell RA (1998) Principles of geographical information systems: spatial information systems and geostatistics

Chang W, Haran M, Applegate P, Pollard D (2016) Calibrating an ice sheet model using high-dimensional binary spatial data. J Am Stat Assoc 111(513):57–72

Diggle PJ, Giorgi E (2016) Model-based geostatistics for prevalence mapping in low-resource settings. J Am Stat Assoc 111(515):1096–1120

Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. Stat Sci 7(4):457–472

Hardouin C (2019) A variational method for parameter estimation in a logistic spatial regression. Spatial Stat 31(1):1–45

Hoeting JA, Davis RA, Merton AA, Thompson SE (2006) Model selection for geostatistical models. Ecol Appl 16(1):87–98

Hosseini F, Eidsvik J, Mohammadzadeh M (2011) Approximate bayesian inference in spatial glmm with skew normal latent variables. Comput Stat Data Anal 55(4):1791–1806

Jaakkola TS, Jordan MI (2000) Bayesian parameter estimation via variational methods. Stat Comput 10(1):25–37

Lin P-S, Clayton MK et al (2005) Analysis of binary spatial data by quasi-likelihood estimating equations. Ann Stat 33(2):542–555

Mahmoudian B (2018) On the existence of some skew-gaussian random field models. Stat Prob Lett 137:331–335

Mardia KV, Marshall RJ (1984) Maximum likelihood estimation of models for residual covariance in spatial regression. Biometrika 71(1):135–146

Nisa H, Mitakda MB, Astutik S, et al. (2019) Estimation of propensity score using spatial logistic regression. In: IOP conference series: materials science and engineering, volume 546, page 052048. IOP Publishing

Paciorek CJ (2007) Computational techniques for spatial logistic regression with large data sets. Comput Stat Data Anal 51(8):3631–3653

Rikken M, Van Rijn R (1993) Soil pollution with heavy metals: in inquiry into spatial variation, cost of mapping and the risk evaluation of Copper, Cadmium, Lead and Zinc in the floodplains of the Meuse West of Stein, The Netherlands: field study report. University of Utrecht, Utrecht

Sahu SK, Dey DK, Branco MD (2003) A new class of multivariate skew distributions with applications to bayesian regression models. Canadian J Stat 31(2):129–150

Sengupta A, Cressie N, Kahn BH, Frey R (2016) Predictive inference for big, spatial, non-gaussian data: modis cloud data and its change-of-support. Aust New Zealand J Stat 58(1):15–45

Tadayon V, Rasekh A (2019) Non-gaussian covariate-dependent spatial measurement error model for analyzing big spatial data. J Agric Biol Environ Stat 24(1):49–72

Tadayon V, Torabi M (2019) Spatial models for non-gaussian data with covariate measurement error. Environmetrics 30(3):e2545

Tadayon V, Torabi M (2022) Sampling strategies for proportion and rate estimation in a spatially correlated population. Spatial Stat 47:100564

Tayyebi A, Delavar MR, Yazdanpanah MJ, Pijanowski BC, Saeedi S, Tayyebi AH (2010) A spatial logistic regression model for simulating land use patterns: a case study of the shiraz metropolitan area of iran. Advances in earth observation of global change. Springer, Berlin, pp 27–42

Wu W, Zhang L (2013) Comparison of spatial and non-spatial logistic regression models for modeling the occurrence of cloud cover in north-eastern puerto rico. Appl Geogr 37:52–62

Xie C, Huang B, Claramunt C, Chandramouli C (2005) Spatial logistic regression and gis to model rural-urban land conversion. In: Proceedings of PROCESSUS Second International Colloquium on the Behavioural Foundations of Integrated Land-use and Transportation Models: frameworks, models and applications, pages 12–15. University of Toronto

Zhang Z, Arellano-Valle RB, Genton MG, Huser R (2021) Tractable bayes of skew-elliptical link models for correlated binary data. arXiv preprint arXiv:2101.02233

Zhu J, Huang H-C, Wu J (2005) Modeling spatial-temporal binary data using markov random fields. J Agric Biol Environ Stat 10(2):212

Zhu J, Zheng Y, Carroll AL, Aukema BH (2008) Autologistic regression analysis of spatial-temporal binary data via monte carlo maximum likelihood. J Agric Biol Environ Stat 13(1):84–98