Statistica Sinica Preprint No: SS-2023-0237

Title	Joint Mean-angle Model for Spatial Binary Data
Manuscript ID	SS-2023-0237
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202023.0237
Complete List of Authors	Cheng Peng, Renwen Luo, Yang Han and Jianxin Pan
Corresponding Authors	Jianxin Pan
E-mails	jianxinpan@uic.edu.cn

Statistica Sinica

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0237

Joint Mean-Angle Model for Spatial Binary Data

Cheng Peng[1], Renwen Luo[2], Yang Han[1], and Jianxin Pan[3,2]

[1] *University of Manchester, UK*

[2] *Guangdong Provincial Key Laboratory of Interdisciplinary Research and Application for Data Science, BNU-HKBU United International College, China*

[3] *Research Center for Mathematics, Beijing Normal University, China*

Abstract: The analysis of spatially correlated binary data has received substantial attention in geo-statistical research but is very challenging due to the intricacy of the distributional form. Two principal objectives include examining the dependence of binary response on covariates of interest and quantifying the covariances or correlations between pairs of outcomes. While the literature has sufficiently addressed the modelling issue of the mean structure of a binary response, the characterization of the covariances between pairs of binary responses in terms of covariates is not clear. In this paper, we propose methods to explain such characterizations by using a latent Gaussian copula model with alternative hypersphere decomposition of the covariance matrix. Correctly specifying the covariance matrix is crucial not only for the high efficiency of mean parameters but also for scientific interest. The key is to model the marginal mean and pairwise covariance, simultaneously, for spatial binary data. Two generalized estimating equations are proposed to estimate the parameters, and the asymptotic properties of the resulting estimators are investigated. To evaluate the performance of the methods, we conduct simulation studies and provide real data analysis for illustration.

Statistica Sinica

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0237

Key words and phrases: Alternative hypersphere decomposition, Generalized estimating equation, Joint mean-angle model, Latent Gaussian copula model

1. Introduction

Binary data widely arise in a range of scientific fields such as public health, biomedicine, economics, and agriculture. The collection of binary data is often more straightforward and cost-effective than gathering accurate continuous measurements since binary data only takes on two possible values. This paper focuses on the analysis of spatially correlated binary data, where the spatial dependence that cannot be ignored in making statistical inferences relies on the distances between sampling locations. Our research was motivated by a geostatistical study of Bovine Tuberculosis (bTB) infection of cattle herds and badger setts in Ireland. In this study, the binary outcome is whether the cattle herd or badger sett is infected with bTB or not. On the one hand, the primary objective is to describe the binary outcome as a function of the available covariates. On the other hand, since bTB is an infectious disease that can spread through the air and adjacent animal herds may be spatially associated, modelling the spatial correlation between animal herds is also of great importance.

In the literature, three main modelling frameworks have been proposed to analyse spatial binary data: conditional models, multivariate probit models and marginal models. Conditional models, also known as random effects models, as-

Statistica Sinica

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0237

sume that the binary responses are independent conditional on random effects. For example, the generalized linear geostatistical models (GLGMs) proposed by Diggle and Paulo (2007, Chap. 4) extend the generalized linear mixed model for spatial data by using the realizations of a stationary Gaussian random field as the random effects. Bayesian inference for parameter estimation can be implemented using Markov chain Monte Carlo (Diggle and Paulo, 2007, Chap. 4) or integrated nested Laplace approximation (Brown, 2015) to approximate the posterior distribution. However, deriving closed-form expressions for the marginal means and pairwise covariances of binary responses is generally not possible. As pointed out by Oliveira (2020), the mean regression parameter in GLGM only has a conditional interpretation that is of less interest than marginal interpretations when there are no repeated measures in the data (e.g., spatial data).

The multivariate probit model is another popular approach for analyzing binary data, which was first introduced by Ashford and Sowden (1970) and is also known by different names such as latent Gaussian (copula) model (Fan et al., 2017), clipped Gaussian random field (Oliveira, 2020), and GeoCopula model (Bai et al., 2014). This modelling approach assumes that binary variables are indicators of whether latent Gaussian random variables exceed certain thresholds. Oliveira (2020) demonstrated that a generalized linear geostatistical model is equivalent to a multivariate probit model when there is a non-zero nugget effect. However, due to the integration and matrix inversion, estimating the parameters

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0237

using the full likelihood can be computationally intensive, especially for large data sets. To overcome these challenges, Heagerty and Lele (1998) proposed a composite likelihood method that approximated the full log-likelihood by the sum of log-likelihoods for adjacent pairs of binary observations.

The third approach, marginal models with generalized estimating equations (GEEs, Liang and Zeger (1986)), is also a viable alternative that can circumvent the complete specification of full likelihood and the computational intensity of the estimation procedure. Oman et al. (2007) introduced estimating equations for spatially correlated binary data whose sampling locations can be divided into independent blocks and revealed that the GEE-type method may be more efficient than the aforementioned composite likelihood (Heagerty and Lele, 1998). However, the GEE-type method also has some theoretical drawbacks due to its underlying model assumptions. First, the failure to correctly specify the working correlation structure may lead to a great loss of efficiency, as highlighted by (Wang and Carey, 2003). Second, the lack of an objective function makes model selection a challenging task. Third, for binary data, the elements of the working correlation matrix in GEE may not satisfy the Fréchet-Hoeffding bounds (Nelsen, 2007, Chap. 2) imposed on the Pearson correlation coefficients, which is likely to cause misleading conclusions (Sabo and Chaganty, 2010). For example, the generalized and quasi-likelihood estimating equations proposed by Albert and McShane (1995) and Lin and Clayton (2005) ignore this constraint when mod-

Statistica Sinica

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0237

elling spatially correlated binary data.

In the aforementioned modelling techniques for spatial binary data, covariance modelling plays a crucial role, since a correctly specified covariance matrix not only enhances the efficiency of mean parameters but also avoids the infeasibility caused by violating some constraints such as Fréchet-Hoeffding bounds and positive definiteness. Moreover, the covariance matrix itself is of scientific interest. Common practice involves trying correlation structures such as Matérn family and spherical family (Diggle and Paulo, 2007, Chap. 4) and selecting the optimal structure using criteria like QIC (Pan, 2001). However, estimating correlation structure parameters must take the aforementioned constraints into account, and the structural assumption of the covariance matrix may be too restrictive since, under most circumstances, the true covariance matrix lacks a specific structure. For binary data, odds ratio (Carey et al., 1993) is an alternative measure of pairwise association as the bounds on odds ratio are less restrictive than the Fréchet bounds on Pearson correlation coefficient (Chaganty and Joe, 2006). Correlation coefficients can be explicitly expressed as a function of odds ratios, but odds ratios lack a straightforward interpretation compared to the former, as they are only defined in terms of odds rather than probabilities or marginal means. Another way of specifying a covariance matrix is to transform the constrained elements to unrestricted parameters and then model them with given covariates using linear regression. For example, Ye and Pan (2006) decomposed covariance matrix

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0237

into a generalized autoregressive matrix and an innovation variance matrix via the Modified Cholesky Decomposition method (MCD, Pourahmadi (1999)) and proposed three generalized estimating equations for the marginal means, generalized autoregressive parameters and innovation variances. Similar decomposition methods include Alternative Cholesky Decomposition (ACD, Chen and Dunson (2003)) and Hypersphere Decomposition (HPC, Rebonato and Jaeckel (2000)). However, Cholesky-type decomposition methods which are commonly used for modelling the covariance matrices of temporally correlated data cannot be applied to spatial data, because they require a known order of observations that is absent in spatial data. The Alternative Hypersphere Decomposition method (AHPC) proposed by Li and Pan (2022) addressed the order issue in HPC by redefining the angles modelled through linear regression and giving them new geometric interpretations, but it cannot guarantee the positive definiteness of the resulting covariance matrix.

In this paper, based on the latent Gaussian copula model (Fan et al., 2017) and Alternative Hypersphere Decomposition method (Li and Pan, 2022), we propose a novel joint modelling approach for the marginal mean and covariance matrix of spatial binary data. The latent Gaussian copula model assumes that binary responses are produced by thresholding a latent Gaussian random vector with the cutoff points, and so the correlation between binary responses can be expressed as a nonlinear function of the correlation between the corresponding la-

Statistica Sinica

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0237

tent Gaussian random variables. Here the nonlinear function is termed the bridge function, and we prove that the pairwise correlations of binary responses obtained through this bridge function lie within the Fréchet-Hoeffding bounds. The correlation matrix of the latent Gaussian random vector is modelled using AHPC with covariates like the standard Euclidean distance between sampling locations. Calibration approaches (Choi et al., 2019; Huang et al., 2017) can help find an appropriate surrogate when the latent correlation matrix is not positive definite. For example, Choi et al. (2019) minimizes the distance (e.g., spectral-norm and scaled Frobenius-norm) between the resulting correlation matrix and surrogate while keeping the minimum eigenvalue of the surrogate positive. Therefore, our proposed model satisfies the Fréchet-Hoeffding bounds and finds the dependence of the covariance matrix elements on given covariates. Although it does not guarantee the positive definiteness of the correlation matrix, a surrogate can be calculated and used for non-positive definite cases. We introduce two generalized estimating equations to estimate the mean and latent correlation parameters, and also prove that the parameter estimators are consistent and asymptotically normally distributed. Overall, the main contributions and novelties of our method are the ability to jointly model the marginal mean and covariance matrix of spatial binary data, the use of AHPC to model the latent correlation matrix with covariates, and the derivation of consistent and asymptotically normal parameter estimators.

Statistica Sinica

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0237

In Section 2, the latent Gaussian copula model and Alternative Hypersphere Decomposition (AHPC) are briefly introduced, followed by a description of our joint mean-angle model and its estimation procedure. Section 3 presents the asymptotic properties of parameter estimators. To demonstrate the effectiveness of the proposed method, numerical studies and practical real data analysis are shown in Section 4 and Section 5, respectively. Finally, we summarize our findings and provide some concluding remarks in Section 6. All technical proofs and additional simulation results can be found in the supplementary materials.

2. Methodology and estimation procedure

Consider a binary response variable Y_i observed at the i-th sampling location s_i, where s_i is typically represented as a 2-dimensional vector of geographical coordinates (e.g., latitude and longitude), and $i \in \{1, ..., n\}$. Let $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top$ be the $n \times 1$ vector of binary responses, and let $\boldsymbol{s} = (s_1^\top, \ldots, s_n^\top)^\top$ be the $n \times 2$ matrix of sampling locations. Suppose $\mu_i = \mathrm{E}(Y_i|X_i)$ and $\boldsymbol{\Sigma} = \mathrm{Var}(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{s})$ denote the first two moments of binary responses, where X_i is a p_β-dimensional vector of covariates and $\boldsymbol{X} = (X_1, \ldots, X_n)^\top$ is the $n \times p_\beta$ design matrix. Here, we assume that the covariance matrix $\boldsymbol{\Sigma}$ is positive definite, without loss of generality.

To provide a clear exposition of the proposed joint mean-angle model, we present the latent Gaussian copula model and Alternative Hypersphere Decomposition in the following two subsections. These subsections serve as preliminary

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0237

materials since the JMA model builds upon their concepts.

2.1 Latent Gaussian copula model for spatial binary data

For spatial binary data, the latent Gaussian copula model is defined as,

Definition 1. An n-dimensional binary random vector $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top$ follows the latent Gaussian copula model $LGC_n(\mathbf{0}, \mathbf{R}, \mathbf{c})$, if its j-th component is denoted by $Y_j = I(Z_j > c_j)$ for all $j \in \{1, \ldots, n\}$, where $I(\cdot)$ is the indicator function. The vector $\mathbf{Z} = (Z_1, \ldots, Z_n)^\top$ is an n-dimensional Gaussian random vector and is distributed as $\mathcal{N}_n(\mathbf{0}, \mathbf{R})$, where $\mathbf{R} = (R_{ij})_{1 \le i,j \le n}$ is the latent covariance matrix with diagonal elements set to 1's (i.e., latent correlation matrix). $\mathbf{c} = (c_1, \ldots, c_n)^\top$ is a vector of constant cutoff points.

The cumulative distribution function (cdf) of the standard univariate normal distribution and the standard bivariate normal distribution with correlation τ are denoted by $\Phi(\cdot)$ and $\Phi_2(\cdot, \cdot; \tau)$, respectively. Let σ_{ij} represent the (i, j)-th entry of the covariance matrix $\boldsymbol{\Sigma}$. According to Definition 1, the values of σ_{ij}, R_{ij}, c_i and c_j are required to satisfy the relationship,

$$\begin{aligned}\sigma_{ij} &= \mathrm{E}(Y_iY_j) - \mathrm{E}(Y_i)\mathrm{E}(Y_j) = \Pr(Z_i > c_i, Z_j > c_j) - \Pr(Z_i > c_i)\Pr(Z_j > c_j)\\ &= \Pr(Z_i \le c_i, Z_j \le c_j) - \Pr(Z_i \le c_i)\Pr(Z_j \le c_j) = \Phi_2(c_i, c_j; R_{ij}) - \Phi(c_i)\Phi(c_j).\end{aligned} \tag{2.1}$$

The covariance elements $|\sigma_{ij}| \le 1$ due to the probabilities in equation (2.1) varying between 0 and 1. Furthermore, for any $i, j \in \{1, \ldots, n\}$ such that $i \ne j$, the correlation coefficient $\varrho_{ij} = \sigma_{ij}/(\sigma_{ii}\sigma_{jj})^{1/2}$ between Y_i and Y_j can be shown to fall within the Fréchet-Hoeffding bounds:

$$\max\left\{-\left[\mu_i\mu_j/\{(1-\mu_i)(1-\mu_j)\}\right]^{1/2}, -\left\{(1-\mu_i)(1-\mu_j)/(\mu_i\mu_j)\right\}^{1/2}\right\} < \varrho_{ij}$$
$$< \min\left\{\left[\mu_i(1-\mu_j)/\{\mu_j(1-\mu_i)\}\right]^{1/2}, \left[\mu_j(1-\mu_i)/\{\mu_i(1-\mu_j)\}\right]^{1/2}\right\},$$

and the derivation is given in the supplementary materials. For any fixed (c_i, c_j), equation (2.1) denotes a bridge function $F(t; c_i, c_j) = \Phi_2(c_i, c_j; t) - \Phi(c_i)\Phi(c_j)$. In particular, we have $\sigma_{ij} = F(R_{ij}; c_i, c_j)$ so that $F(t; c_i, c_j)$ links the correlation R_{ij} between the latent Gaussian variables Z_i and Z_j to the covariance σ_{ij} between the binary variables Y_i and Y_j. It has been demonstrated in Fan et al. (2017) that $F(t; c_i, c_j)$ is a strictly monotonic increasing function of t on the interval $(-1, 1)$, for any fixed c_i and c_j.

Note that the values of cutoff points c_i and c_j in the bridge function $F(t; c_i, c_j)$ are often unknown and need to be estimated. A common approach is to estimate the cutoff points by $c_i = \Phi^{-1}(1-\mu_i)$ and $c_j = \Phi^{-1}(1-\mu_j)$. This is because the marginal means μ_i and μ_j are the probabilities of observing $Y_i = 1$ and $Y_j = 1$, respectively, which can be expressed as $\mu_i = \mathrm{E}(Y_i) = \Pr(Y_i = 1) = \Pr(Z_i > c_i) = 1 - \Pr(Z_i \le c_i) = 1 - \Phi(c_i)$, and similarly $\mu_j = 1 - \Phi(c_j)$.

2.2 Alternative Hypersphere Decomposition

The original Hypersphere Decomposition was proposed by Rebonato and Jaeckel (2000) with the aim of constructing a model for the correlation matrix. Rapisarda

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0237

et al. (2007) subsequently improved this decomposition method and provided a geometric interpretation for the angle parameters.

For a covariance matrix $\mathbf{\Sigma}^*$, it can be initially decomposed as,

$$\mathbf{\Sigma}^* = \mathbf{D}^*\mathbf{R}^*\mathbf{D}^*,$$

where $\mathbf{R}^* = (R^*_{ij})_{1\le i,j\le n}$ is the correlation matrix and $\mathbf{D}^*$ is a diagonal matrix with diagonal elements being the standard deviations $\sigma_{ii} = \{\mathrm{Var}(Y_i)\}^{1/2}$ of Y_i, $i \in \{1,\dots,n\}$. To further investigate the correlation matrix $\mathbf{R}^*$, the Hypersphere Decomposition method is applied, which decomposes $\mathbf{R}^*$ as,

$$\mathbf{R}^* = \mathbf{F}\mathbf{F}^\top,$$

where $\mathbf{F}$ is a lower triangular matrix containing elements defined by trigonometric functions of hyperspherical angles (Rapisarda et al., 2007). Specifically, for $1 \le j \le i \le n$, the (i,j)-th element of $\mathbf{F}$ is given by

$$f_{ij} = \begin{cases} 1, & \text{if } i=1, j=1, \\ \cos(\omega^*_{ij}), & \text{if } 2 \le i \le n, j=1 \\ \cos(\omega^*_{ij})\prod_{l=1}^{j-1}\sin(\omega^*_{il}), & \text{if } 2 \le j < i \le n \\ \prod_{l=1}^{j-1}\sin(\omega^*_{il}), & \text{if } 2 \le i \le n, j=i, \end{cases}$$

where $\cos(\omega^*_{ij})$ and $\sin(\omega^*_{ij})$ are trigonometric functions of hyperspherical angle $\omega^*_{ij} \in [0,\pi)$. Let $\mathbf{\Omega}^* = (\omega^*_{ij})_{1\le i,j\le n}$ be the hyperspherical angle matrix. Hypersphere Decomposition projects the unit row vectors of $\mathbf{F}$ into a hypersphere coordinate so that the Pearson correlation coefficient R^*_{ij} can be expressed as a

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0237

function of ω^*_{ij},

$$R^*_{ij} = \cos(\omega^*_{ij}) \prod_{l=1}^{j-1} \sin(\omega^*_{il}) \sin(\omega^*_{jl}) + \sum_{l=1}^{j-1} \left\{ \cos(\omega^*_{il}) \cos(\omega^*_{jl}) \prod_{t=1}^{l-1} \sin(\omega^*_{it}) \sin(\omega^*_{jt}) \right\}. \tag{2.2}$$

One of the primary drawbacks of the Hypersphere Decomposition (HPC) method is its order dependence. This means that the order of observations affects the values of elements ω^*_{ij} in the hyperspherical angle matrix $\mathbf{\Omega}^*$ (see supplementary materials for detailed proof). Consequently, if one constructs a regression model for correlation R^*_{ij} using ω^*_{ij} and (2.2), the order-dependence of HPC implies that the estimates of regression parameters are determined by the order of observations. However, in the case of spatial binary data which do not have a natural order, the regression model for σ_{ij} should not be influenced by the order of observations. To address this issue, we introduce the Alternative Hypersphere Decomposition (AHPC) proposed by Li and Pan (2022), in which the hyperspherical angle matrix $\mathbf{\Omega}^*$ is redefined from that used in HPC as follows,

$$\mathbf{\Omega}^* = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ \omega^*_{21} & 0 & 0 & \cdots & 0 & 0 \\ \omega^*_{31} & \omega^*_{32} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \ddots & \vdots \\ \omega^*_{n1} & \omega^*_{n2} & \omega^*_{n3} & \cdots & \omega^*_{nn-1} & 0 \end{bmatrix},$$

with the element $\omega^*_{ij} \in [0, \pi)$ being the angle between row vectors $\vec{F_i}$ and $\vec{F_j}$ of matrix $\mathbf{F}$ in HPC. Specifically, it holds that $\omega^*_{ij} =< \vec{F}_i, \vec{F}_j >$, $1 \leq j < i \leq n$.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0237

Given the above definition, we can establish a relationship between the correlation R^*_{ij} and the angle ω^*_{ij},

$$\cos(\omega^*_{ij}) = \cos < \vec{F}_i, \vec{F}_j >= \vec{F}_i^\top \vec{F}_j = R^*_{ij}, \tag{2.3}$$

As the cosine function is a strictly decreasing function on the interval $[0, \pi)$, equation (2.3) gives a one-to-one mapping between R^*_{ij} and ω^*_{ij} so that the value of ω^*_{ij} is uniquely determined by the value of R^*_{ij}. Moreover, for all $1 \leq j < i \leq n$, the values of ω^*_{ij} remain invariant to any permutation of $\{Y_i\}_{i=1}^n$, implying that AHPC is order independent. Therefore, we consider using AHPC to model the latent correlation matrix of spatial binary responses. In particular, the elements ω^*_{ij} in angle matrix $\mathbf{\Omega}^*$ are modelled by

$$\omega^*_{ij} = f(\zeta_{ij}^\top \gamma) = \arctan(\zeta_{ij}^\top \gamma) + \pi/2,$$

where ζ_{ij} is a vector of known covariates and γ is the associated parameter vector. The link function $f(\cdot) = \arctan(\cdot) + \pi/2$ is used to ensure that the angle $\omega^*_{ij} \in [0, \pi)$. In addition, when the dimension of the correlation matrix is either greater than or equal to three, the geometric interpretation of AHPC induces the constraint $\omega^*_{kj} \leq \omega^*_{ij} + \omega^*_{ki}$ for all $i < j < k$.

Note that the AHPC method cannot guarantee the positive definiteness of the estimated correlation matrix $\hat{\mathbf{R}}^*$. Therefore, if $\hat{\mathbf{R}}^*$ is not positive definite, we suggest implementing the fixed support positive-definite (FSPD) modification (Choi et al., 2019) to obtain an appropriate surrogate.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0237

2.3 Joint mean-angle model and GEE

Based on the AHPC method, we propose a joint mean-angle (JMA) model for the marginal mean and latent correlation of binary responses,

$$g(\mu_i) = X_i^\top \beta, \quad R_{ij} = \cos(\omega_{ij}) = \cos\left\{\arctan(\zeta_{ij}^\top \gamma) + (\pi/2)\right\}, \tag{2.4}$$

where X_i and ζ_{ij} are $p_\beta \times 1$ and $p_\gamma \times 1$ vectors of covariates, respectively. β and γ are the associated parameters in the marginal mean and latent correlation model. Let p_β and p_γ denote the dimensions of β and γ, respectively. R_{ij} represents the (i, j)-th entry of correlation matrix $\mathbf{R}$ in the latent Gaussian copula model $LGC_n(\mathbf{0}, \mathbf{R}, \mathbf{c})$ and ω_{ij} is the corresponding angle in AHPC method. The link function $g(\cdot)$ is assumed to be known, monotone, and differentiable. In our study, the covariates X_i and ζ_{ij} may comprise baseline covariates, polynomials in spatial distance, and their interactions. For example, if we utilize the polynomials in spatial distance to model the angles, the covariate vector ζ_{ij} can take the form:

$$\zeta_{ij} = (1, d(s_i, s_j), \ldots, d^{p_\gamma - 1}(s_i, s_j))^\top,$$

where $d(s_i, s_j)$ is the Euclidean distance between locations s_i and s_j.

The joint model proposed for spatial binary data combines the latent Gaussian copula model with the AHPC method. Compared with previous methods, it not only makes the estimated correlation coefficients naturally satisfy the Fréchet-Hoeffding bounds, but also account for the correlation coefficients through a regression model with geographical covariates.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0237

Throughout this paper, we introduce the notation $\text{vech}^*(\mathbf{A})$ to represent the vectorization operator of the lower off-diagonal elements of a symmetric matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$. When a scalar function $h(\cdot)$ acts on an m-dimensional square matrix $\mathbf{B} = (B_{ij})_{1\le i,j\le m}$, we denote the resulting matrix as $h(\mathbf{B}) = (h(B_{ij}))_{1\le i,j\le m}$. Based on these notations, we propose two generalized estimating equations to estimate the mean regression parameter β and the latent angle regression parameter γ in the JMA model (2.4),

$$S_1(\beta) = \left(\partial \boldsymbol{\mu}^\top/\partial\beta\right) \boldsymbol{\Sigma}^{-1}\{\boldsymbol{y} - \boldsymbol{\mu}(\boldsymbol{X}\beta)\} = 0, \tag{2.5}$$

$$S_2(\gamma)|_{\beta=\hat{\beta}} = \left(\partial \boldsymbol{\eta}^\top/\partial\gamma\right) \mathbf{M}^{-1}(\mathbf{H} - \boldsymbol{\eta}) = 0, \tag{2.6}$$

where

$$\mathbf{H} = \text{vech}^*(\hat{\boldsymbol{\Sigma}}) = \text{vech}^*[\{\boldsymbol{y} - \boldsymbol{\mu}(\boldsymbol{X}\hat{\beta})\}\{\boldsymbol{y} - \boldsymbol{\mu}(\boldsymbol{X}\hat{\beta})\}^\top],$$

$$\boldsymbol{\eta} = \text{vech}^*[F\{\mathbf{R}(\gamma); \mathbf{c}(\hat{\beta})\}];$$

the parameter estimator $\hat{\beta}$ in $S_2(\gamma)$ is the solution of the generalized estimating equation (2.5); $\partial\boldsymbol{\mu}^\top/\partial\beta$ is the $p_\beta \times n$ matrix with i-th column being $\partial\mu_i/\partial\beta = \dot{g}^{-1}(X_i^\top\beta)X_i$ and $\dot{g}^{-1}(\cdot)$ being the derivative of the inverse of logit link function $g(\cdot)$; $\partial\boldsymbol{\eta}^\top/\partial\gamma$ is the $p_\gamma \times \{n(n-1)/2\}$ matrix with k-th column being $\partial\eta_k/\partial\gamma = \dot{F}(R_{ij}; c_i, c_j)(\partial R_{ij}/\partial\gamma)$. Here $\partial R_{ij}/\partial\gamma = -\sin\{\arctan(\zeta_{ij}^\top\gamma) + (\pi/2)\}\zeta_{ij}/\{1 + (\zeta_{ij}^\top\gamma)^2\}$, $\dot{F}(R_{ij}; c_i, c_j)$ is the derivative of the bridge function F with respect to R_{ij} and $k = i - j(j+1)/2 + (j-1)n$, $1 \le j \le n$, $j+1 \le i \le n$. In addition, $F\{\mathbf{R}(\gamma); \mathbf{c}\} = (F\{R_{ij}(\gamma); c_i, c_j\})_{1\le i,j\le n}$ and $\mathbf{M} = \text{cov}(\mathbf{H})$ is the covariance

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0237

matrix of $\mathbf{H}$. We simultaneously solve the generalized estimating equations (2.5) and (2.6) to obtain the parameter estimators $\hat{\beta}$ and $\hat{\gamma}$, which also implies that our approach assigns equal importance to the marginal mean and latent correlation when modelling spatial binary data.

The requirement for $S_2(\gamma)$ to be conditional on $\hat{\beta}$ is motivated by the fact that the true value of β is typically unknown. If β were known, we could replace $\hat{\beta}$ in $S_2(\gamma)$ with the true value and then solve $S_2(\gamma) = 0$ to estimate γ, as $\mathrm{E}(S_2(\gamma)) = 0$ in this case. However, since β is unknown, we use a consistent estimator, $\hat{\beta}$, instead. By plugging $\hat{\beta}$ into $S_2(\gamma)$ and requiring it to be conditional on $\hat{\beta}$, we construct an estimating equation for γ. The equation (2.6) is essentially an asymptotic generalized estimating equation since $\mathrm{E}(S_2(\gamma)) \neq 0$ and $\mathrm{E}(S_2(\gamma)) \to 0$ as $n \to \infty$. Note that although $S_2(\gamma)$ is conditional on the parameter estimator $\hat{\beta}$, the resulting parameter estimator $\hat{\gamma}$ is still consistent, which is shown in Section 3.

It is imperative to specify the covariance matrix $\mathbf{M} = \mathrm{cov}(\mathbf{H})$ in $S_2(\gamma)$ before solving the equation (2.6). While the expressions for diagonal elements in $\mathbf{M}$ are clear, the off-diagonal elements are usually mathematically intractable. To overcome this, we approximate the off-diagonal elements in $\mathbf{M}$ using a working structure. In line with Ye and Pan (2006), we approximate $\mathbf{M}$ by a sandwich working covariance $\mathbf{D}^{1/2}\mathbf{\Gamma}(\delta)\mathbf{D}^{1/2}$. Here, $\mathbf{D} = \mathrm{diag}\{\mathrm{Var}(r_1 r_2), \mathrm{Var}(r_1 r_3), \ldots, \mathrm{Var}(r_{n-1} r_n)\}$ ($r_i = y_i - \mu_i$, $1 \leq i \leq n$), $\mathbf{\Gamma}(\delta)$ is a working correlation matrix used to approximate the correlation between H_i and H_j $(i \neq j)$ and δ is the working parameter. Since y_i can only be 0 or 1, $y_i^2 = y_i$ and $r_i^2 = (y_i - \mu_i)^2 = (1 - 2\mu_i) y_i + \mu_i^2$. Then

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0237

for all $1 \leq i < j \leq n$, the diagonal elements of matrix $\mathbf{D}$ can be calculated by

$$\begin{aligned}\mathrm{Var}(r_i r_j) &= \mathrm{E}(r_i^2 r_j^2) - \{\mathrm{E}(r_i r_j)\}^2 = \mathrm{E}[\{(1-2\mu_i)y_i + \mu_i^2\}\{(1-2\mu_j)y_j + \mu_j^2\}] - \sigma_{ij}^2 \\ &= \{(1-2\mu_i)(1-2\mu_j)\mathrm{E}(y_i y_j)\} + (1-2\mu_i)\mu_i\mu_j^2 + (1-2\mu_j)\mu_j\mu_i^2 + \mu_i^2\mu_j^2 - \sigma_{ij}^2 \\ &= (1-2\mu_i)(1-2\mu_j)(\sigma_{ij} + \mu_i\mu_j) + (1-2\mu_i)\mu_i\mu_j^2 + (1-2\mu_j)\mu_j\mu_i^2 + \mu_i^2\mu_j^2 - \sigma_{ij}^2.\end{aligned}$$

In the context of spatial data, the working correlation matrix $\mathbf{\Gamma}(\delta)$ often takes on structures like independence structure and compound symmetry structure, among others. As is the case with the conventional generalized estimating equation for the marginal mean (Liang and Zeger, 1986), the choice of working parameter δ has no impact on the consistency of the latent angle parameter estimator $\hat{\gamma}$, but it does affect the efficiency. In support of this claim, the simulation studies and real data analysis presented in Section 4 and Section 5 provide compelling evidence.

2.4 The main algorithm

As outlined in Subsection 2.3, the estimators of parameters β and γ are the solutions of the generalized estimating equations (2.5) and (2.6). The form of $S_2(\gamma)|_{\beta=\hat{\beta}}$ inspires us to update one parameter when holding the other fixed. We use the quasi-Fisher scoring algorithm to obtain the numerical solutions for parameters β and γ. In particular, given $\mathbf{\Sigma}$, we update the estimator of the mean parameter β by

$$\beta^{(k+1)} = \beta^{(k)} + \left\{\left(\partial\boldsymbol{\mu}^\top/\partial\beta\right)\mathbf{\Sigma}^{-1}\left(\partial\boldsymbol{\mu}^\top/\partial\beta\right)^\top\right\}^{-1}\left[\left(\partial\boldsymbol{\mu}^\top/\partial\beta\right)\mathbf{\Sigma}^{-1}\left\{\mathbf{y} - \boldsymbol{\mu}(\boldsymbol{X}\beta)\right\}\right]\Bigg|_{\beta=\beta^{(k)}}. \tag{2.7}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0237

On the other hand, given $\beta = \hat{\beta}$ and $\mathbf{M}$, the value for the latent angle parameter γ is calculated through

$$\gamma^{(k+1)} = \gamma^{(k)} + \left\{ \left(\partial \boldsymbol{\eta}^{\top} / \partial \gamma \right) \mathbf{M}^{-1} \left(\partial \boldsymbol{\eta}^{\top} / \partial \gamma \right)^{\top} \right\}^{-1} \left\{ \left(\partial \boldsymbol{\eta}^{\top} / \partial \gamma \right) \mathbf{M}^{-1} \left(\mathbf{H} - \boldsymbol{\eta} \right) \right\} \Bigg|_{\gamma = \gamma^{(k)}, \beta = \hat{\beta}} . \tag{2.8}$$

The equations (2.7) and (2.8) imply that the parameters can be estimated iteratively by weighted generalized least squares. To sum up, we use the following algorithm to obtain the parameter estimates in our model:

Step 1. Use GLM algorithm to obtain a starting value $\beta^{(0)}$ of mean parameter β. Give the latent angle parameter γ a starting value $\gamma^{(0)}$ (e.g., $\gamma^{(0)} = (0, \ldots, 0)^{\top}$, which leads the latent correlation matrix $\mathbf{R}^{(0)}$ to be an identity matrix $I_n = \mathrm{diag}(1, \ldots, 1)$) and set $k = 0$.

Step 2. Compute the estimates of mean values $\hat{\boldsymbol{\mu}}^{(k)} = g^{-1}(\boldsymbol{X}\beta^{(k)})$ and the estimated covariance matrix $\hat{\boldsymbol{\Sigma}}^{(k)} = (\mathbf{y} - \hat{\boldsymbol{\mu}}^{(k)})^{\top}(\mathbf{y} - \hat{\boldsymbol{\mu}}^{(k)})$.

Step 3. Calculate the latent correlation matrix $\mathbf{R}^{(k)} = \left(R_{jj'}^{(k)} \right)_{1 \le j, j' \le n}$ with

$$R_{jj'}^{(k)} = \cos\{\arctan(\zeta_{jj'}^{\top} \gamma^{(k)}) + (\pi/2)\},$$

and estimate the cutoff points $\mathbf{c}^{(k)} = \Phi^{-1}(\mathbf{1} - \hat{\boldsymbol{\mu}}^{(k)})$. Next, use equation (2.1) to obtain the covariance matrix $\boldsymbol{\Sigma}^{(k)}$. Employ the quasi-Fisher scoring algorithm and the weighted generalized least squares estimator (2.8) to calculate $\gamma^{(k+1)}$. Replace $\gamma^{(k)}$ with $\gamma^{(k+1)}$, and update the latent correlation matrix $\mathbf{R}^{(k)}$ and covariance matrix $\boldsymbol{\Sigma}^{(k)}$ accordingly.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0237

Step 4. Similarly, based on the covariance matrix $\mathbf{\Sigma}^{(k)}$, obtain the parameter estimate $\beta^{(k+1)}$ through (2.7). Then replace $\beta^{(k)}$ with $\beta^{(k+1)}$.

Step 5. Repeat Steps 2-4 until convergence is achieved for the parameter estimators $\hat{\beta}$ and $\hat{\gamma}$.

Note that the Fisher scoring algorithm is a form of Newton's method which replaces the Hessian matrix with the expected Fisher information matrix, and the quasi-Fisher scoring algorithm substitutes the score function in the Fisher scoring algorithm with the quasi-score function obtained from quasi-likelihood. The rate of convergence of the Fisher scoring method and advantages over Newton Raphson method have been discussed in the literature (Osborne, 1992; Demidenko, 2013). In subsection S2.3 of the supplementary materials, a series of simulation studies are also conducted to study the convergence of our proposed algorithm. For example, the effect of the convergence criterion and starting parameter value on the number of iterations needed for the convergence and parameter estimates.

3. Asymptotic properties

To study the rates of convergence for parameter estimators $\hat{\beta}$ and $\hat{\gamma}$, we first give a set of regularity conditions:

(A1) The dimensions of X_i and ζ_{ij}, denoted by p_β and p_γ, respectively, are fixed. The first four moments of the binary responses exist, and the inverse link function $g^{-1}(\cdot)$ has a bounded second derivative.

Statistica Sinica

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0237

(A2) The parametric spaces Θ_1 and Θ_2 are compact subsets of $\mathbb{R}^{p_\beta}$ and $\mathbb{R}^{p_\gamma}$, respectively. The true parameter values β and γ lie in the interiors of their corresponding parameter spaces.

(A3) All covariates X_i and ζ_{ij}, as well as every element of the vector

$$\big(\exp(X_1^\top\beta)/\{1+\exp(X_1^\top\beta)\}^2,\ldots,\exp(X_n^\top\beta)/\{1+\exp(X_n^\top\beta)\}^2\big)^\top$$

and the inverse matrix M^{-1} are bounded.

Under (A1), the existence of the first four moments of binary responses can guarantee the consistent estimation of the parameters in our joint model. Condition (A2) is conventionally made in linear models, and condition (A3) is normally satisfied. Let $N = n(n-1)/2$. Then we have the following two theorems,

Theorem 1. Assume that the generalized estimating equations (2.5) and (2.6) have only one root respectively. If the conditions (A1)–(A3) hold, the parameter estimators $\hat\beta$ and $\hat\gamma$ are $\sqrt{n}$-consistent. That is, $\|\hat\beta-\beta\|_2 = O_p(n^{-1/2})$ and $\|\hat\gamma - \gamma\|_2 = O_p(n^{-1/2})$.

Under Conditions (A1)–(A3), the following necessary conditions for asymptotic normality are valid.

(A4) $n^{-1/2}\left(\partial\boldsymbol{\mu}^\top/\partial\beta\right)\boldsymbol{\Sigma}^{-1}\{\mathbf{y}-\boldsymbol{\mu}(\boldsymbol{X}\beta)\} \xrightarrow{D} \mathcal{N}_{p_\beta}(\mathbf{0},\boldsymbol{\Lambda}_\beta)$, where the variance matrix $\boldsymbol{\Lambda}_\beta = n^{-1}\mathrm{Var}\{S_1(\beta)\}$.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0237

(A5) $N^{-1/2}\left(\partial\boldsymbol{\eta}^\top/\partial\gamma\right)\mathbf{M}^{-1}\text{vech}^*\{\mathbf{r}\mathbf{r}^\top - \text{E}(\mathbf{r}\mathbf{r}^\top)\} \xrightarrow{D} \mathcal{N}_{p_\gamma}(\mathbf{0}, \boldsymbol{\Lambda}_\gamma)$, where $\mathbf{r} = \mathbf{y} - \boldsymbol{\mu}$ and the variance matrix $\boldsymbol{\Lambda}_\gamma = n^{-1}\text{Var}\{S_2(\gamma)\}$.

Based on Conditions (A1)–(A3), we can easily derive (A4) and (A5) from the central limit theorem. Note that the consistent estimations of variance matrices $\boldsymbol{\Lambda}_\beta$ and $\boldsymbol{\Lambda}_\gamma$ rely on the availability of independent replications of observations (Albert and McShane, 1995; Heagerty and Lele, 1998; Heagerty and Lumley, 2000). However, the spatial binary data considered in the article only have one single observation for each sampling location. Based on the idea of Feng et al. (2014) and Varin et al. (2011), we can generate independent draws $\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_K$ from the fitted model and then estimate $\boldsymbol{\Lambda}_\beta$ and $\boldsymbol{\Lambda}_\gamma$ by sample covariance matrices of $\{S_1(\hat{\beta}; \tilde{\mathbf{y}}_k), k = 1, \ldots, K\}$ and $\{S_2(\hat{\gamma}; \tilde{\mathbf{y}}_k), k = 1, \ldots, K\}$, respectively.

The following theorem presents the asymptotic normality of parameter estimators $\hat{\beta}$ and $\hat{\gamma}$ separately, with all notations explained in the supplementary materials.

Theorem 2. Under conditions (A1)–(A3), the generalized estimating equation estimators $\hat{\beta}$ and $\hat{\gamma}$ are asymptotically normal, respectively. That is,

$$n^{1/2}(\hat{\beta} - \beta) \to \mathcal{N}_{p_\beta}(\mathbf{0}, \mathbf{V}_\beta^{-1}\boldsymbol{\Lambda}_\beta\mathbf{V}_\beta^{-1}),$$

$$N^{1/2}(\hat{\gamma} - \gamma)|_{\hat{\beta}} \to \mathcal{N}_{p_\gamma}(\mathbf{V}_\gamma^{-1}\mathbf{J}^*, \mathbf{V}_\gamma^{-1}\boldsymbol{\Lambda}_\gamma\mathbf{V}_\gamma^{-1}),$$

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0237

in distribution as $n, N \to \infty$, where

$$\mathbf{V}_\beta = \lim_{n\to\infty} n^{-1} \left(\partial \boldsymbol{\mu}^\top / \partial\beta\right) \boldsymbol{\Sigma}^{-1} \left(\partial \boldsymbol{\mu}^\top / \partial\beta\right)^\top,$$

$$\mathbf{V}_\gamma = \lim_{N\to\infty} N^{-1} \left(\partial \boldsymbol{\eta}^\top / \partial\gamma\right) \mathbf{M}^{-1} \left(\partial \boldsymbol{\eta}^\top / \partial\gamma\right)^\top,$$

$$\mathbf{J}^* = \lim_{N\to\infty} N^{-1/2} \left(\partial \boldsymbol{\eta}^\top / \partial\gamma\right) \mathbf{M}^{-1} \left\{\partial \mathrm{vech}^*(\mathbf{r}\mathbf{r}^\top)^\top / \partial\beta + \partial\boldsymbol{\eta}^\top / \partial\beta\right\}^\top (\hat{\beta} - \beta).$$

It should be noted that $\mathbf{V}_\gamma^{-1}\mathbf{J}^*$ represents the asymptotically conditional mean, whose value is determined by $\|\hat{\beta} - \beta\|_2$. Since $N^{-1/2}\mathbf{V}_\gamma^{-1}\mathbf{J}^* = O_p(n^{-1/2})$, the estimator $\hat{\gamma}$ is asymptotically unbiased for the parameter γ.

4. Simulation studies

To evaluate the numerical performance of our proposed joint mean-angle model and GEE estimation procedure, we conduct simulation studies using 10 × 10 regular lattice data. The regular lattice design involves $n = 100$ observations distributed uniformly on a grid square within the range of $[0, 1] \times [0, 1]$. Let $s_i = (locx_i, locy_i)$ denote the spatial location of i-th observation, where $locx_i$ and $locy_i$ are the x- and y-coordinates, respectively. To generate the binary responses, we consider the joint mean-angle model that determines the marginal means and latent correlation matrix,

$$g(\mu_i) = x_{1i}\beta_1 + x_{2i}\beta_2 + x_{3i}\beta_3 + x_{4i}\beta_4,$$

$$\omega_{ij} = \arctan(\zeta_{1ij}\gamma_1 + \zeta_{2ij}\gamma_2 + \zeta_{3ij}\gamma_3) + \pi/2,$$

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0237

where $1 \leq i, j \leq n$, $g(\cdot)$ is the logit link function, mean parameter $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)^\top = (1, -0.5, 0.3, -0.7)^\top$ and latent angle parameter $\gamma = (\gamma_1, \gamma_2, \gamma_3)^\top = (-0.2, 0.4, -0.2)^\top$; $(x_{1i}, x_{2i}, x_{3i})^\top$ is sampled from a three-dimensional Gaussian distribution $\mathbb{N}_3(\mathbf{0}, 0.5^2\tilde{\mathbf{R}})$, where $\tilde{\mathbf{R}}$ is of the AR(1) correlation structure with a parameter value 0.2; x_{4i} is independent of x_{ki}, $k = 1, 2, 3$, and follows a uniform distribution $\mathbb{U}(-1/2, 1/2)$; $\zeta_{1ij} = 1, \zeta_{2ij} = ||s_i - s_j||_2 = \{(locx_i - locx_j)^2 + (locy_i - locy_j)^2\}^{1/2}$ and $\zeta_{3ij} = ||s_i - s_j||_2^2$. The simulation setup for angles $\{\omega_{ij}, 1 \leq i, j \leq n\}$ is verified to satisfy the constraints of angles in AHPC. Latent random variables $\{z_i, i = 1, \ldots, n\}$ are generated from a Gaussian distribution $\mathbb{N}_n(\mathbf{0}, \mathbf{R})$, where $\mathbf{R} = (R_{ij})_{1\leq i,j\leq n} = (\cos(\omega_{ij}))_{1\leq i,j\leq n}$. Finally, the simulated binary response of i-th observation is obtained by $y_i = I(z_i > c_i)$, where $I(\cdot)$ is an indicator function and the cutoff point $c_i = \Phi^{-1}(1 - \mu_i)$. Note that both the latent correlation matrix $\mathbf{R}$ and true covariance matrix $\mathbf{\Sigma}_0$ are positive definite.

In this paper, we use four different approaches to model and analyze the simulated spatial binary data: (1) the joint mean-angle model (JMA) fitted by two generalized estimating equations (GEEs); (2) the generalized linear geostatistical model (GLGM) fitted by integrated nested Laplace approximation (INLA); (3) the independence generalized linear model (GLM) fitted by maximum likelihood (ML); (4) conventional generalized estimating equation (GEE) for marginal mean. In our proposed GEE method, the entries of estimated covariance matrix $\hat{\mathbf{\Sigma}} = \hat{\mathbf{r}}\hat{\mathbf{r}}^\top$ are taken as the responses in the generalized estimating equation (2.6).

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0237

The sandwich working covariance $\mathbf{D}^{1/2}\mathbf{\Gamma}(\delta)\mathbf{D}^{1/2}$ is used with exchangeable correlation structure (i.e., compound symmetry) for $\mathbf{\Gamma}(\delta)$. The effect of misspecification of $\mathbf{\Gamma}(\delta)$ on estimating β and γ is studied by setting the working correlation parameter $\delta \in \{0, 0.2, 0.4, 0.6, 0.8\}$ in our simulations. For the estimations of variance matrices $\mathbf{\Lambda}_\beta$ and $\mathbf{\Lambda}_\gamma$, we generate $K = 10$ independent draws of the response $\mathbf{y}$ from the fitted model. The R package "geostatsp" (Brown, 2015) can be utilized to fit the generalized linear geostatistical model (GLGM) with Bayesian INLA, and the Matérn correlation with shape parameter $\kappa = 1$ (one default value in the package) is chosen as the correlation structure for the underlying spatial process. As for the conventional GEE (Liang and Zeger, 1986), the working correlation structure assumed for spatially correlated binary data is exchangeable and the R package "gee" is applied to the estimation of mean structure parameters. The alternative methods presented in the simulation study section are applied to each outcome separately.

The simulation results reported in Table 1 and Table 2 are averaged over 500 independent replications. Table 1 displays the estimation outcomes of all the parameters in four distinct modelling approaches. It is clear that our proposed method can provide precise estimates for the parameters β and γ in the joint mean-angel model. In addition, the resulting estimators of parameters β and γ are robust against misspecification of working correlation matrix $\mathbf{\Gamma}(\delta)$. Table 2 presents the sample deviation (SD) of 500 estimates, the average (SE) of 500

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0237

estimated standard errors, and the coverage probability (CoPr) of the confidence interval for each parameter. It is evident that the gaps between the SDs and SEs are small, which indicates that the standard error formulas shown in Section 3 can be used to estimate the variability of the parameter estimators. From these two tables, we note that our proposed method performs slightly better than the other three modelling approaches, in terms of the efficiency of mean parameter estimator $\hat{\beta}$. Furthermore, the angle parameter estimator $\hat{\gamma}$ can substitute for the parameter γ in our joint mean-angle model to calculate the estimates of latent correlation matrix $\mathbf{R}$ and covariance matrix $\mathbf{\Sigma}$, whereas the generalized linear geostatistical model (GLGM) cannot provide a closed-form expression for the covariance matrix $\mathbf{\Sigma}$ of binary responses $\boldsymbol{Y}$.

We also conduct simulation experiments to compare all the methods under varying correlation strengths (e.g., weak, moderate, and strong). It can be observed that when the latent correlation strength is moderate or strong, the proposed joint modelling method is an improvement over the existing marginal model methodologies, in terms of the standard deviations (SDs) of mean parameter estimator $\hat{\beta}$. Besides, the improvements are more obvious than that of the weak association case. Meanwhile, we augment the sample size (e.g., $n \in \{144, 225, 400\}$) and then find that the biases and standard deviations of parameter estimators in our proposed method decline as the sample size increases. Due to the length limit, all the details are deferred to the supplementary materials.

Statistica Sinica

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0237

Table 1. Simulation results on spatial binary data over 500 replications. Average of biases of 500 estimates for each parameter.

	True	JMA($\delta = 0$)	JMA($\delta = 0.2$)	JMA($\delta = 0.4$)	JMA($\delta = 0.6$)	JMA($\delta = 0.8$)	GLGM	GEE	ML
β_1	1	0.0527	0.0721	0.0740	0.0739	0.0756	0.1040	0.0580	0.0551
β_2	-0.5	-0.0471	-0.0562	-0.0528	-0.0508	-0.0490	-0.0599	-0.0485	-0.0534
β_3	0.3	0.0028	0.0109	0.0112	0.0108	0.0087	0.0085	0.0063	0.0093
β_4	-0.7	-0.0793	-0.0966	-0.1007	-0.1007	-0.0968	-0.0915	-0.0751	-0.0662
γ_1	-0.2	-0.0117	-0.0478	-0.0484	-0.0450	-0.0475			
γ_2	0.4	0.0327	0.0323	0.0420	0.0349	0.0403			
γ_3	-0.2	-0.0160	-0.0191	-0.0272	-0.0228	-0.0264			

5. Data Analysis

We now turn to the example of bovine tuberculosis (bTB) infection (Kelly, 2013; Griffin et al., 2005) and apply the foregoing techniques to the analysis of data. The data we analyze here were collected from the Four Area Project (FAP) which was designed to study badger removal in four counties in Ireland. The study comprises 417 cattle herds and 251 badger setts, whose spatial locations were recorded using GIS coordinates. The binary outcome variable indicates whether at least one individual in a cattle herd or badger sett tested positive for bTB or not. Figure 1 shows the data on bTB infection, with black and white circles representing the bTB positive and negative cases, respectively. The covariates for cattle herds consist of the logarithm of herd size ($logsize$) and a binary variable indicating prior infection (ph). For badger setts, the size of sett

Statistica Sinica

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0237

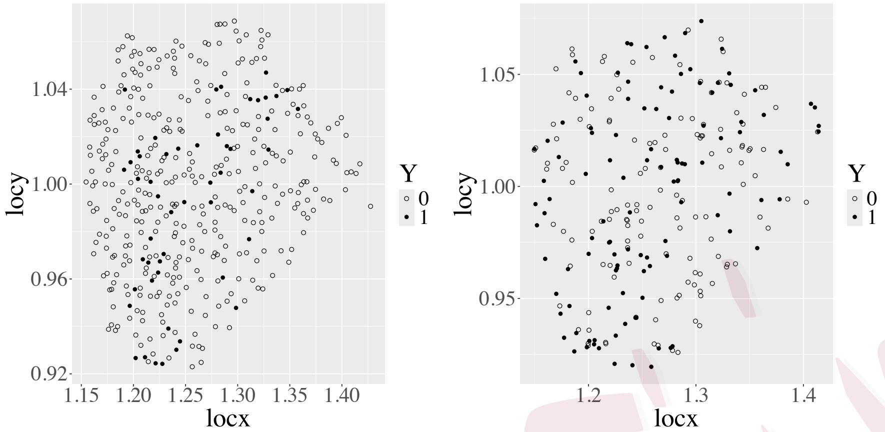
Table 2. The standard deviations (SD), averages (SE) of estimated standard errors, and the coverage probabilities (CoPr) of the confidence intervals for parameter estimators $\hat{\beta}$ and $\hat{\gamma}$. The confidence level is 0.95.

Parameter		JMA($\delta = 0$)	JMA($\delta = 0.2$)	JMA($\delta = 0.4$)	JMA($\delta = 0.6$)	JMA($\delta = 0.8$)	GLGM	GEE	ML
β_1	SD	0.4790	0.4964	0.4962	0.4986	0.4981	0.4962	0.4898	0.4920
	SE	0.4736	0.4848	0.4845	0.4848	0.4851	0.4848	0.4721	0.4713
	CoPr	0.9440	0.9420	0.9420	0.9400	0.9400	0.9500	0.9440	0.9440
β_2	SD	0.4661	0.4848	0.4839	0.4842	0.4839	0.4810	0.4758	0.4797
	SE	0.4586	0.4594	0.4597	0.4596	0.4593	0.4733	0.4615	0.4613
	CoPr	0.9440	0.9380	0.9400	0.9400	0.9400	0.9440	0.9460	0.9480
β_3	SD	0.4517	0.4828	0.4813	0.4805	0.4771	0.4824	0.4500	0.4526
	SE	0.4457	0.4474	0.4470	0.4473	0.4470	0.4626	0.4490	0.4499
	CoPr	0.9460	0.9400	0.9380	0.9380	0.9400	0.9480	0.9400	0.9460
β_4	SD	0.8132	0.8213	0.8196	0.8227	0.8181	0.8098	0.8176	0.8188
	SE	0.7546	0.7558	0.7563	0.7569	0.7568	0.7811	0.7577	0.7588
	CoPr	0.9540	0.9480	0.9500	0.9480	0.9520	0.9500	0.9520	0.9520
γ_1	SD	0.2280	0.3752	0.3775	0.3773	0.3804			
	SE	0.2333	0.4493	0.4659	0.4616	0.4755			
	CoPr	0.9540	0.9500	0.9520	0.9520	0.9520			
γ_2	SD	0.6780	0.7372	0.7418	0.7405	0.7432			
	SE	0.6674	0.8063	0.8186	0.8310	0.8279			
	CoPr	0.9540	0.9440	0.9440	0.9440	0.9460			
γ_3	SD	0.5194	0.5652	0.5672	0.5663	0.5680			
	SE	0.5110	0.6122	0.6238	0.6341	0.6280			
	CoPr	0.9480	0.9460	0.9480	0.9500	0.9480			

Statistica Sinica

($size$) is the only covariate. Our scientific interest is two-fold: first, to estimate the association between bTB infection and measured covariates; and second, to account for the spatial correlations by a regression model with some covariates

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0237



(a) bTB infection for cattle herds (b) bTB infection for badger setts

Figure 1. The bTB infection plots for cattle herds and badger setts, where locx and locy are the GIS x- and y-coordinates.

like standard Euclidean distances between sampling locations.

In accordance with the joint mean-angle model introduced in Section 2, we adopt a logistic function of measured covariates to model the marginal mean of the infection status y_i of i-th cattle herd or badger sett,

- Cattle herds: $\text{logit}(\mu_i) = \ln\{\mu_i/(1-\mu_i)\} = \beta_1 + \beta_2 ph_i + \beta_3 logsize_i$,

- Badger setts: $\text{logit}(\mu_i) = \ln\{\mu_i/(1-\mu_i)\} = \beta_1 + \beta_2 size_i$,

where $\mu_i = \text{E}(y_i)$, and using the following two angle models to specify the latent correlation matrix $\mathbf{R} = (R_{ij})_{1\le i,j\le n}$, for $1 \le i, j \le n$,

$$\text{(i)} \quad R_{ij} = \cos(\omega_{ij}) = \cos\{\arctan(\zeta_{1ij}\gamma_1 + \zeta_{2ij}\gamma_2) + (\pi/2)\},$$

$$\text{(ii)} \quad R_{ij} = \cos(\omega_{ij}) = \cos\{\arctan(\zeta_{1ij}\gamma_1 + \zeta_{2ij}\gamma_2 + \zeta_{3ij}\gamma_3) + (\pi/2)\},$$

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0237

where the covariates $\zeta_{1ij} = 1, \zeta_{2ij} = ||s_i - s_j||_2 = \{(locx_i - locx_j)^2 + (locy_i - locy_j)^2\}^{1/2}$, and $\zeta_{3ij} = ||s_i - s_j||_2^2$. We also compare our proposed method with three other modelling approaches. Table 3 and Table 4 summarize the estimates of mean parameter β and angle parameter γ, and their corresponding standard errors for the bTB infection of cattle herds and badger setts, respectively. Note that we generate $K = 10$ independent draws of the response $\mathbf{y}$ from the fitted model to calculate the standard errors of parameter estimators. The results in Table 3 imply that the infection status of a cattle herd is positively related to its size and previous infection. Similarly, it can be observed from Table 4 that a badger sett of larger size is associated with a higher likelihood of bTB infection.

In Table 3 and Table 4, we also observe that the estimates of the mean parameter β and angle parameter γ, along with their corresponding standard errors, are relatively stable across different values of the working parameter δ when using our proposed method. Although the difference between our proposed method and the other three modelling approaches is minor in terms of the mean parameter estimates and their standard errors, the other three approaches fail to provide an appropriate specification of the marginal pairwise correlation between each pair of binary responses.

The QIC and RJ presented in Table 3 and Table 4 are quasi-likelihood under the independence model criterion for the mean parameter estimator $\hat{\beta}$ and Rotnitzky-Jewell criterion for the estimated covariance matrix $\hat{\mathbf{\Sigma}}$. Both criteria are used to select the optimal angle model and the optimal value of the working correlation parameter δ. The QIC proposed by Pan (2001) for the mean parameter estimator $\hat{\beta}$ is defined as

$$\mathrm{QIC} = -2\sum_{i=1}^{n}\left[y_i \ln\left\{\hat{\mu}_i/(1-\hat{\mu}_i)\right\} + \ln(1-\hat{\mu}_i)\right] + 2\mathrm{trace}(\boldsymbol{\Upsilon}^{-1}\hat{\boldsymbol{V}}_{\boldsymbol{\Sigma}}),$$

where $\hat{\mu}_i = \mu_i(\hat{\beta})$ is an estimate of μ_i; $\boldsymbol{\Upsilon}$ and $\hat{\boldsymbol{V}}_{\boldsymbol{\Sigma}}$ are the covariance matrices of mean parameter estimator $\hat{\beta}$ obtained by using $\mathrm{diag}(\mu_1(1-\mu_1),\ldots,\mu_n(1-\mu_n))$ and $F(\mathbf{R},\mathbf{c}) = (F(R_{ij},c_i,c_j))_{1\le i,j\le n}$ as the covariance matrix $\boldsymbol{\Sigma}$ in the generalized estimating equation (2.5), respectively. According to Hin et al. (2007), the Rotnitzky-Jewell criterion (RJ) has the following form,

$$\mathrm{RJ} = \left[\left\{1 - \mathrm{trace}(Q)/p_Q\right\}^2 + \left\{1 - \mathrm{trace}(Q^2)/p_Q\right\}^2\right]^{1/2},$$

where

$$Q = \left\{\left(\partial\boldsymbol{\mu}^\top/\partial\beta\right)\boldsymbol{\Sigma}^{-1}\left(\partial\boldsymbol{\mu}^\top/\partial\beta\right)^\top\right\}^{-1}\left\{\left(\partial\boldsymbol{\mu}^\top/\partial\beta\right)\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})(\mathbf{y}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}\left(\partial\boldsymbol{\mu}^\top/\partial\beta\right)^\top\right\}$$

is a p_Q-dimensional square matrix. If the estimate of $\boldsymbol{\Sigma}$ approximates the true covariance matrix $\boldsymbol{\Sigma}_0$, Q will be close to an identity matrix and RJ will also tend to zero. That is, a smaller value of RJ indicates a better fit for the correlation structure. For the cattle herd infection data, both QIC and RJ suggest that the JMA1 angle model with a working independence structure (i.e., $\delta = 0$) is the optimal choice. It implies that the covariance between two data points decreases as their spatial distance increases since the slope γ_2 is estimated to be positive. In the case of badger sett infection, we observe that the QIC and RJ

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0237

Table 3. The estimates of mean and angle parameters for bTB cattle herd data. Standard errors are in parenthesis. The joint mean-angle models which use the angle models (i) and (ii) to characterize the latent correlation matrix **R** are referred to as JMA1 and JMA2, respectively.

Method	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	QIC	RJ
ML	-2.5490(0.2262)	0.8396(0.3238)	0.9378(0.2250)					
GEE	-2.5844(0.1142)	0.8462(0.3139)	0.9479(0.2132)					
GLGM	-2.5398(0.2263)	0.8273(0.3241)	0.9417(0.2248)					
JMA1(δ=0)	-2.5907(0.2012)	0.8425(0.3167)	0.9552(0.2224)	-0.0328(0.0374)	0.0140(0.0135)		281.9777	0.3427
JMA1(δ=0.2)	-2.6128(0.2060)	0.8439(0.3209)	0.9636(0.2266)	-0.0384(0.0525)	0.0073(0.0066)		281.9309	0.3026
JMA1(δ=0.4)	-2.6128(0.2060)	0.8439(0.3209)	0.9636(0.2266)	-0.0384(0.0491)	0.0074(0.0072)		281.9309	0.3026
JMA1(δ=0.6)	-2.6128(0.2060)	0.8439(0.3209)	0.9636(0.2266)	-0.0384(0.0488)	0.0074(0.0075)		281.9309	0.3026
JMA1(δ=0.8)	-2.6128(0.2060)	0.8439(0.3209)	0.9636(0.2266)	-0.0384(0.0335)	0.0074(0.0075)		281.9309	0.3026
JMA2(δ=0)	-2.6459(0.2162)	0.8260(0.3288)	0.9871(0.2366)	-0.1504(0.4576)	0.0308(0.0970)	-0.0012(0.0039)	282.1111	0.3799
JMA2(δ=0.2)	-2.6692(0.2228)	0.8238(0.3337)	0.9958(0.2416)	-0.1584(0.5556)	0.0314(0.1040)	-0.0012(0.0037)	282.2403	0.4124
JMA2(δ=0.4)	-2.6692(0.2228)	0.8238(0.3337)	0.9958(0.2416)	-0.1584(0.5491)	0.0314(0.1041)	-0.0012(0.0038)	282.2403	0.4124
JMA2(δ=0.6)	-2.6692(0.2228)	0.8238(0.3337)	0.9958(0.2416)	-0.1584(0.5419)	0.0314(0.1076)	-0.0012(0.0043)	282.2403	0.4124
JMA2(δ=0.8)	-2.6692(0.2228)	0.8238(0.3337)	0.9958(0.2416)	-0.1584(0.5482)	0.0314(0.0995)	-0.0012(0.0037)	282.2403	0.4124

values are similar across all scenarios, which makes it difficult to determine the optimal model. However, our analysis shows that the off-diagonal elements of the estimated covariance matrix $\mathbf{\Sigma}$ for both two angle models are almost zero, indicating negligible spatial correlation between each pair of badger setts. This result implies that modelling the correlation between badger setts may not be necessary for this particular data set, and a simpler model can be sufficient for inference.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0237

Table 4. The estimates of mean and angle parameters for bTB badger sett data. Standard errors are in parenthesis. The joint mean-angle models which use the angle models (i) and (ii) to characterize the latent correlation matrix $\mathbf{R}$ are referred to as JMA1 and JMA2, respectively.

Method	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	QIC	RJ
ML	-1.2231(0.2559)	0.4636(0.1015)					
GEE	-1.2271(0.2234)	0.4635(0.1015)					
GLGM	-1.2231(0.2561)	0.4635(0.1015)					
JMA1(δ=0)	-1.2091(0.2355)	0.4630(0.1017)	0.0042(0.0110)	0.0025(0.0024)		326.0247	2.5204
JMA1(δ=0.2)	-1.2194(0.2352)	0.4634(0.1016)	0.0317(0.0437)	0.0013(0.0019)		326.0813	2.5685
JMA1(δ=0.4)	-1.2194(0.2352)	0.4634(0.1016)	0.0319(0.0670)	0.0013(0.0018)		326.0813	2.5686
JMA1(δ=0.6)	-1.2195(0.2352)	0.4634(0.1016)	0.0319(0.0382)	0.0013(0.0020)		326.0814	2.5686
JMA1(δ=0.8)	-1.2195(0.2352)	0.4634(0.1016)	0.0319(0.0621)	0.0013(0.0022)		326.0814	2.5686
JMA2(δ=0)	-1.3023(0.2547)	0.4707(0.1014)	-0.1060(0.2294)	0.0266(0.0556)	-0.0013(0.0026)	326.0537	2.4510
JMA2(δ=0.2)	-1.2712(0.2396)	0.4672(0.1016)	-0.0870(0.2192)	0.0260(0.0560)	-0.0012(0.0025)	325.9661	2.4597
JMA2(δ=0.4)	-1.2710(0.2395)	0.4672(0.1016)	-0.0869(0.2384)	0.0260(0.0556)	-0.0012(0.0025)	325.9661	2.4597
JMA2(δ=0.6)	-1.2711(0.2395)	0.4672(0.1016)	-0.0869(0.2216)	0.0260(0.0519)	-0.0012(0.0024)	325.9661	2.4597
JMA2(δ=0.8)	-1.2711(0.2395)	0.4672(0.1016)	-0.0869(0.2648)	0.0260(0.0538)	-0.0012(0.0025)	325.9661	2.4597

6. Discussion

This paper introduces a novel joint mean-angle model for analyzing spatially correlated binary data, with the goal of quantifying the dependence of binary responses on covariates and characterizing the spatial covariances between pairs of responses. The formulation of the pairwise covariance model builds on the latent Gaussian coupla model and Alternative Hypersphere Decomposition (AHPC).

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0237

The former assumes that the binary responses are generated by dichotomizing a latent Gaussian random vector with a set of cutoff points and ensures that the estimated correlation coefficients fall within the Fréchet-Hoeffding bounds. The latter enables the transformed correlation coefficients (i.e., the angles) to be related to measured covariates through a regression model. However, the resulting covariance matrix for spatial binary data may not always be positive definite. In this case, the FSPD estimator suggested by Choi et al. (2019) is a simple positive definite surrogate with some theoretical advantages. Together with a logistic regression model for the marginal means, our proposed method provides a flexible modelling framework for spatially correlated binary data.

In this paper, considering the consistency of mean parameter estimator $\hat{\beta}$ and the computational cost involved in the estimation procedure, we use two separate estimating equations rather than the joint form of GEE2 (Zhao and Prentice, 1990; Liang et al., 1992) to estimate the mean structure parameter β and the latent angle parameter γ. We adopt compound symmetry as the working correlation structure in our second-order generalized estimating equation (2.6) due to its closed-form inverse. The theoretical properties such as the consistency and asymptotic normality of the parameter estimators $\hat{\beta}$ and $\hat{\gamma}$ are established. We assess the performance of the proposed method through numerical simulations and data analysis. The results imply that our proposed method performs better than the aforementioned marginal models in terms of the efficiency of the mean

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0237

parameter estimator if the correlation strength is moderate or strong. Even though the method may perform similarly to the marginal models when the correlations between binary responses are weak, it can still be used to estimate the correlation matrix of binary responses and interpret how the correlation varies as the distance between sampling locations changes, which cannot be achieved by using the other aforementioned methods.

We also identify some directions which may need more work. First, extending the joint model to other types of spatial categorical data, such as binomial and count data, may require the development of a new bridge function that links the latent correlations to the covariances of categorical responses. Second, since the AHPC method allows us to model the correlation coefficients with covariates through a parametric regression model, additional flexibility can be introduced to the proposed method by modelling both the marginal means and pairwise correlations in a non-parametric or semi-parametric manner. Finally, the AHPC method cannot guarantee the positive semi-definiteness of the estimated covariance matrix. Therefore, further research is required to investigate a covariance modelling method that can ensure the positive definiteness of the estimated covariance matrix while also having geometric interpretations and addressing the order dependence issue like the AHPC method.

Statistica Sinica

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0237

Supplementary Material

The supplementary material contains the R-code used in our research and a PDF file that includes technical proofs and some further simulation studies.

Acknowledgments

The authors gratefully acknowledge the financial support of the China Scholarship Council scholarship under Grant 202006240040.

References

Albert, P. S. and L. M. McShane (1995). A generalized estimating equation approach for spatially correlated binary data: applications to the analysis of neuroimaging data. *Biometrics 51*(2), 627–638.

Ashford, J. R. and R. R. Sowden (1970). Multi-variate probit analysis. *Biometrics 26*(3), 535–546.

Bai, Y., J. Kang, and P. Song (2014). Efficient pairwise composite likelihood estimation for spatial-clustered data. *Biometrics 70*(3), 661–670.

Brown, P. E. (2015). Model-based geostatistics the easy way. *Journal of Statistical Software 63*(12), 1–24.

Carey, V., S. L. Zeger, and P. J. Diggle (1993). Modelling multivariate binary data with alternating logistic regressions. *Biometrika 80*(3), 517–526.

Chaganty, N. R. and H. Joe (2006). Range of correlation matrices for dependent bernoulli random variables. *Biometrika 93*(1), 197–206.

Statistica Sinica

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0237

Chen, Z. and D. B. Dunson (2003). Random effects selection in linear mixed models. *Biometrics 59*(4), 762–769.

Choi, Y.-G., J. Lim, A. Roy, and J. Park (2019). Positive-definite modification of covariance matrix estimators via linear shrinkage. *Journal of Multivariate Analysis 124*, 234–249.

Demidenko, E. (2013). *Mixed models: Theory and applications with R, 2nd ed.* Wiley.

Diggle, P. J. and J. R. Paulo (2007). *Model-based geostatistics*. Springer New York.

Fan, J., H. Liu, Y. Ning, and H. Zou (2017). High dimensional semiparametric latent graphical model for mixed data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 79*(2), 405–421.

Feng, X., J. Zhu, P. Lin, and M. Steen-Adams (2014, 12). Composite likelihood estimation for models of spatial ordinal data and spatial proportional data with zero/one values. *Environmetrics 25*, 571–583.

Griffin, J. M., D. H. Williams, G. E. Kelly, T. A. Clegg, I. O'Boyle, J. D. Collins, and S. J. More (2005). The impact of badger removal on the control of tuberculosis in cattle herds in ireland. *Preventive Veterinary Medicine 67*(4), 237–266.

Heagerty, P. J. and S. R. Lele (1998). A composite likelihood approach to binary spatial data. *Journal of the American Statistical Association 93*(443), 1099–1111.

Heagerty, P. J. and T. Lumley (2000). Window subsampling of estimating functions with application to regression models. *Journal of the American Statistical Association 95*(449), 197–211.

Hin, L.-Y., V. J. Carey, and Y.-G. Wang (2007). Criteria for working-correlation-structure selection in gee: assessment via simulation. *The American Statistician 61*(4), 360–364.

Statistica Sinica

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0237

Huang, C., D. Farewell, and J. Pan (2017). A calibration method for non-positive definite covariance matrix in multivariate data analysis. *Journal of Multivariate Analysis 157*, 45–52.

Kelly, G. E. (2013). Joint spatio-temporal modeling of mycobacterium bovis infections in badgers and cattle - results from the irish four area project. *Statistical Communications in Infectious Diseases 5*(1), 1–16.

Li, Q. and J. Pan (2022). Permutation variation and alternative hyper-sphere decomposition. *Mathematics 10*(4), 562.

Liang, K.-Y. and S. L. Zeger (1986). Longitudinal data analysis using generalized linear models. *Biometrika 73*(1), 13–22.

Liang, K.-Y., S. L. Zeger, and B. Qaqish (1992). Multivariate regression analyses for categorical data (with discussion). *Journal of the Royal Statistical Society: Series B (Methodological) 54*(1), 3–40.

Lin, P.-S. and M. K. Clayton (2005). Analysis of binary spatial data by quasi-likelihood estimating equations. *The Annals of Statistics 33*(2), 542–555.

Nelsen, R. B. (2007). *An introduction to copulas.* Springer-Verlag.

Oliveira, V. D. (2020). Models for geostatistical binary data: properties and connections. *The American Statistician 74*(1), 72–79.

Oman, S. D., V. Landsman, Y. Carmel, and R. Kadmon (2007). Analyzing spatially distributed binary data using independent-block estimating equations. *Biometrics 63*(3), 892–900.

Osborne, M. R. (1992). Fisher's method of scoring. *International Statistical Review / Revue Internationale de Statistique 60*(1), 99–117.

Pan, W. (2001). Akaike's information criteria in generalized estimating equations. *Biometrics 57*(1),

Statistica Sinica

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0237

120–125.

Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika 86*(3), 677–690.

Rapisarda, F., D. Brigo, and F. Mercurio (2007). Parameterizing correlations: a geometric interpretation. *IMA Journal of Management Mathematics 18*(1), 55–73.

Rebonato, R. and P. Jaeckel (2000). The most general methodology for creating a valid correlation matrix for risk management and option pricing purposes. *Journal of Risk 2*(2), 17–27.

Sabo, R. T. and N. R. Chaganty (2010). What can go wrong when ignoring correlation bounds in the use of generalized estimating equations. *Statistics in Medicine 29*(24), 2501–2507.

Varin, C., N. Reid, and D. Firth (2011). An overview of composite likelihood methods. *Statistica Sinica 21*(1), 5–42.

Wang, Y. G. and V. Carey (2003). Working correlation structure misspecification, estimation and covariate design: implications for generalised estimating equations performance. *Biometrika 90*(1), 29–41.

Ye, H. and J. Pan (2006). Modelling covariance structures in generalized estimating equations for longitudinal data. *Biometrika 93*(4), 927–941.

Zhao, L. P. and R. L. Prentice (1990). Correlated binary regression using a quadratic exponential model. *Biometrika 77*(3), 642–648.

Cheng Peng, Department of Mathematics, University of Manchester, Oxford Road, Manchester, UK.

E-mail: (cheng.peng-4@postgrad.manchester.ac.uk)

Statistica Sinica

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0237

Renwen Luo, Division of Science and Technology, United International College (BNU-HKBU), 2000 Jintong Road, Zhuhai, Guangdong Province, China.

E-mail: (luorenwen@uic.edu.cn)

Yang Han, Department of Mathematics, University of Manchester, Oxford Road, Manchester, UK.

E-mail: (yang.han@manchester.ac.uk)

Jianxin Pan, Division of Science and Technology, United International College (BNU-HKBU), 2000 Jintong Road, Tangjiawan, Zhuhai, Guangdong Province, China.

E-mail: (jianxinpan@uic.edu.cn)

Statistica Sinica