Efficient Pairwise Composite Likelihood Estimation for Spatial-Clustered Data

Yun Bai,[1] Jian Kang,[2] and Peter X.-K. Song[1],*

[1]Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, U.S.A.
[2]Department of Biostatistics and Bioinformatics, Emory University, Atlanta, Georgia, U.S.A.
*email: pxsong@umich.edu

SUMMARY. Spatial-clustered data refer to high-dimensional correlated measurements collected from units or subjects that are spatially clustered. Such data arise frequently from studies in social and health sciences. We propose a unified modeling framework, termed as GeoCopula, to characterize both large-scale variation, and small-scale variation for various data types, including continuous data, binary data, and count data as special cases. To overcome challenges in the estimation and inference for the model parameters, we propose an efficient composite likelihood approach in that the estimation efficiency is resulted from a construction of over-identified joint composite estimating equations. Consequently, the statistical theory for the proposed estimation is developed by extending the classical theory of the generalized method of moments. A clear advantage of the proposed estimation method is the computation feasibility. We conduct several simulation studies to assess the performance of the proposed models and estimation methods for both Gaussian and binary spatial-clustered data. Results show a clear improvement on estimation efficiency over the conventional composite likelihood method. An illustrative data example is included to motivate and demonstrate the proposed method.

KEY WORDS: Gaussian copula; Generalized method of moments; Geographical cluster; Matérn class; Regression.

1. Introduction

In social and health sciences, research studies usually involve subjects that are randomly selected within a large number of geographical units. For example, among the studies of place effects on health, Chaix, Merlo, and Chauvin (2005) investigated individual and contextual factors that determined the health care utilization in France, where 10,955 people were randomly surveyed within 4421 municipals in France. To study the association of neighborhood environmental risk factors with cardiovascular diseases, Mujahid et al. (2007) used a sample of 5988 subjects selected from 576 census tracts from three states in USA. Grady (2010) assessed the impact of racial residential segregation on low birth weight from a pool of 10,277 cases nested in 1092 census tracts in Michigan. In civil engineering studies, Sener, Pendyala, and Bhat (2011) analyzed the physical activity participation levels of individuals in a family unit based on data drawn from the 2000 San Francisco Bay Area Household Travel Survey, in which individual and household socio-demographic as well as activity and travel episodes were recorded for subjects in 15,000 households.

These examples are just a glimpse of a growing number of research projects that collect data in spatial dimensions, thus necessitate the eminent need to generalize the multilevel data analysis to incorporate the spatial dependencies among the clustering units. In classic multilevel models, data across clusters are assumed to be independent, and the focus dwells on appropriately accounting for within-cluster correlations when making statistical inferences. However, when clusters are spa-

tially correlated, such as neighborhoods or census tracts, subjects from clusters are likely to be correlated due to location proximity, hence, the between-cluster independence assumption is no longer valid. Statistical analysis ignoring the spatial effect can lead to wrong standard errors of the regression coefficient estimates, which in turn biases hypothesis testing (Anselin and Griffith, 1988). As a result, in order to draw valid statistical inference, it is of critical importance to account for the between-cluster spatial correlation as well as the within-cluster correlation.

In the current literature, there are two popular modeling frameworks for analyzing spatially correlated data. One approach is based on random effects models, where mean models are specified conditional on cluster-specific random effects (e.g., Diggle et al., 2002). The spatial structures are accounted for by allowing random effects to distribute as a spatial stochastic process. For non-Gaussian data, regression parameters in such hierarchical specification only have conditional or cluster-specific interpretations, which may not be desirable when population characteristics are of interest. The other approach is the generalized estimating equation (GEE, Liang and Zeger, 1986), which specifies the mean model and covariance separately. In the covariance model, the spatial dependence is incorporated via a spatially structured working correlation matrix (e.g., Albert and McShane, 1995; Gotway and Stroup, 1997). GEE is suitable when the mean model is of central interest, since it treats spatial dependencies as nuisance components. As a result, GEE is not appropriate for spatial interpolation, which however is an important task in

many practical studies, such as disease mapping (Diggle et al., 2002).

In this article, we adopt a flexible modeling framework that models both mean and covariance structures of spatial-clustered data, termed as GeoCopula regression model. In this model, univariate margins are specified by generalized linear models, while the spatial and cluster dependencies are modeled through the multivariate Gaussian copula. The proposed framework allows us to analyze a large variety of multivariate discrete and continuous spatial-clustered data, including normal, binary, and count data as special cases. Since the mean and the dependence structure are separately formulated, regression parameters have marginal interpretations, and at the same time, spatial dependence is explicitly modeled by the copula and is not constrained by the mean model.

It is worth mentioning that Bárdossy (2006) and Bárdossy and Li (2008) proposed to use bivariate copulas as an alternative to variograms and covariance functions to describe spatial variability. They showed that the copula-based approach is more flexible in accounting for asymmetrical dependence and is superior in terms of prediction when the normality assumption is violated. Moreover, Kazianka and Pilz (2010) proposed a similar regression model in which the exponential dispersion distribution family (Jorgensen, 1997) is used as the marginal distributions and a multivariate copula is applied to model the spatial dependence. More recently, Masarotto and Varin (2012) provided a comprehensive methodological overview on the topic of Gaussian copula marginal regression models, in which they proposed an importance sampling procedure to carry out maximum likelihood estimation. Our work in this article extends the above mentioned methods to analyze more challenging multi-level spatial-clustered data, and attempts to provide a richer statistical presentation (e.g., large sample properties) of the multivariate copula regression model. Most importantly, our joint composite estimating function is new and computationally efficient in such complex data structures.

A key obstacle of preventing the wide spread of spatial analysis in contextual research is mostly due to computational issues. Almost all existing models require either high-dimensional matrix manipulations such as in GEE, or high-dimensional integrations such as in random effects models. Numerical calculations quickly become intractable for datasets with a large number of spatial units, as in the previous examples. Similar computational problems are faced by Bayesian approaches as well.

To reduce computational burdens, composite likelihood methods (Varin, 2008; Varin, Reid, and Firth, 2011) have been often used at the cost of efficiency loss. Bai, Song, and Raghunathan (2012) proposed a generalized estimation equation based approach, called Joint Composite Estimation Function or JCEF, to recover some efficiency through the weight matrix in spatio-temporal setting. In this article, we extend their idea to spatial-clustered data.

The rest of the article is structured as follows. In Section 2, the GeoCopula model is proposed and detailed for multivariate Gaussian and binary data. Section 3 proposes a joint composite estimating function approach to estimating parameters in the GeoCopula model. Large-sample properties of the proposed estimator are presented in Section 4 with all technical details available in the supplementary material. Simulation experiments are conducted in Section 5. A real-data example is illustrated in Section 6, followed by some discussions in Section 7. Section 8 outlines the web supplementary material for further details concerning Matérn covariance functions, multivariate probit model and R package `GeoCopula` as well as some technical details for readers to understand the article.

2. Model

Let Y_{si} be the outcome of the ith subject nested in geographic cluster s, where $i \in \mathcal{I}_s$, the index set of subjects in cluster s, and $s \in \mathcal{S} \subset \mathcal{R}^2$, with $\mathcal{S}$ being a collection of spatial clusters. Denote the number of subjects in cluster s as n_s, and the total number of subjects is $n = n_1 + \cdots + n_S$. Suppose each outcome Y_{si} follows a generalized linear model (McCullagh and Nelder, 1989), whose mean (or systematic component) μ_{si} is specified as a function of p covariates with an intercept, $\mathbf{x}_{si} = (1, x_{1,si}, \ldots, x_{p,si})^{\mathrm{T}}$ via a known link function h; that is, $h(\mu_{si}) = \eta(\mathbf{x}_{si}) = \mathbf{x}_{si}^{\mathrm{T}}\boldsymbol{\beta} = \beta_0 + \beta_1 x_{1,si} + \cdots + \beta_p x_{p,si}$, where $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^{\mathrm{T}}$ is a vector of regression coefficients. The cumulative distribution function (CDF) of Y_{si} is $F_{si}(y_{si}; \mu_{si}, \psi_{si})$, where ψ_{si} is the dispersion parameter. For simplicity, write the univariate CDF by $F_{si}(y_{si})$, and the corresponding density function $f_{si}(y_{si})$.

To specify a fully parametric model for all Y_{si}'s, we invoke a Gaussian copula dependence model (Song, 2000) to characterize both spatial and within-cluster correlations. The multivariate Gaussian copula is chosen for four reasons. First, Gaussian copula is analytically and theoretically well studied. Second, it encompasses some existing popular models as special cases. When margins are normal linear models, the proposed GeoCopula model becomes the spatial multivariate Gaussian distribution (Cressie, 1993), the most widely used model for spatial continuous data. When a probit link is used for binary data, the GeoCopula model results in a multivariate probit model (Heagerty and Lele, 1998). Third, the correlation matrix in the Gaussian copula enables us to model a dependence map across the entire spatial region under study. It can accommodate full dependence with correlations approaching one, and full independence with zero correlation coefficients. It allows for positive and negative correlations. Other copulas such as Archimedean copulas are not as flexible as Gaussian copula (Bárdossy, 2006; Kazianka and Pilz, 2010). Last, the correlation pattern can be formulated as functions of spatial coordinates and covariates, which can then be estimated for spatial interpolation.

Given previously specified marginal CDFs, the GeoCopula model is formulated as:

$$F(\mathbf{y}) = \Phi_n \left\{ \Phi^{-1}(F_{11}(y_{11})), \ldots, \Phi^{-1}(F_{Sn_S}(y_{Sn_S})) | \boldsymbol{\Sigma} \right\}, \quad (1)$$

where $\boldsymbol{\Sigma}$ accommodates desired spatial dependencies; see (2)–(4) for the detail. Now, we discuss two special cases derived from the GeoCopula model.

EXAMPLE 1. (**GeoCopula Special Case I: Multivariate Gaussian Regression Model**). *Assume* $Y_{si} \sim N(\mathbf{x}_{si}^{\mathrm{T}}\boldsymbol{\beta}, \sigma_{si}^2)$, *and denote its CDF as* $\Phi\left(\frac{y_{si} - \mathbf{x}_{si}^{\mathrm{T}}\boldsymbol{\beta}}{\sigma_{si}}\right)$.

Equation (1) becomes

$$F(\mathbf{y}) = \Phi_n \left(\frac{y_{11} - \mathbf{x}_{11}^{\mathrm{T}}\boldsymbol{\beta}}{\sigma_{11}}, \ldots, \frac{y_{Sn_S} - \mathbf{x}_{Sn_S}^{\mathrm{T}}\boldsymbol{\beta}}{\sigma_{Sn_S}} \,\middle|\, \boldsymbol{\Sigma} \right).$$

That is, $\mathbf{y} \sim N_n\left(\mathbf{X}\boldsymbol{\beta}, \Xi\boldsymbol{\Sigma}\Xi\right),$ *where* $\mathbf{X} = (\mathbf{x}_{11}^{\mathrm{T}}, \ldots, \mathbf{x}_{Sn_S}^{\mathrm{T}})^{\mathrm{T}}$ *and* $\Xi = \mathrm{diag}\{\sigma_{11}, \ldots, \sigma_{Sn_S}\}.$

EXAMPLE 2. **(GeoCopula Special Case II: Multivariate Probit Model).** *Assume marginally* $Y_{si} \sim \mathrm{Bernoulli}(\pi_{si}).$ *Then the CDF of* Y_{si} *is* $F_{si}(y_{si}) = (1 - \pi_{si})I(0 \le y_{si} < 1) + I(y_{si} \ge 1),$ *where* $I(\mathcal{A}) = 1$ *if event* $\mathcal{A}$ *occurs, and* $I(\mathcal{A}) = 0,$ *otherwise. Consider a probit regression model* $\pi_{si} = \Phi(\mathbf{x}_{si}^{\mathrm{T}}\boldsymbol{\beta}).$ *Plug* $F_{si}(y_{si})$ *into equation (1), we obtain a multivariate distribution for n-variate binary data which, as shown in Song (2000), has the same probability mass function as the multivariate probit model studied by Heagerty and Lele (1998).*

Consider latent normal variable $Z_{si} = \mathbf{x}_{si}^{\mathrm{T}}\boldsymbol{\beta} + \epsilon_{si},$ and $\boldsymbol{\epsilon} = (\epsilon_{11}, \ldots, \epsilon_{Sn_S})^{\mathrm{T}} \sim N(\mathbf{0}, \boldsymbol{\Sigma}).$ Define a dichotomous procedure as: $Y_{si} = I(Z_{si} > 0).$ Then $(Y_{11}, \ldots, Y_{Sn_S})^{\mathrm{T}}$ has the same probability mass function as the random vector in Example 2. That is, the multivariate probit model is a special case of the GeoCopula regression model.

Besides the versatile specifications of the marginal distributions, the correlation matrix specified in the Gaussian copula also allows a wide range of spatial correlation patterns. For example, if we assume a compound symmetry (i.e., exchangeable) structure for within-cluster correlation, then the within-cluster correlation matrix for cluster i is

$$\Sigma_{ii} = (1 - \rho)\mathbf{I}_{n_i} + \rho\mathbf{J}_{n_i}, \quad i = 1, \ldots, S \qquad (2)$$

where ρ is the correlation among individuals within the same cluster, and $\mathbf{I}_{n_i}$ is an $n_i \times n_i$ identity matrix, and $\mathbf{J}_{n_i}$ an $n_i \times n_i$ matrix with all entries being 1.

Furthermore, if we model the spatial correlation by the Matérn class across clusters, the spatial correlation matrix between observations in clusters s and t is

$$\Sigma_{st} = \frac{1}{\Gamma(\nu)2^{\nu-1}} \left(\frac{\sqrt{2\nu}d_{st}}{\alpha} \right)^{\nu} K_{\nu} \left(\frac{\sqrt{2\nu}d_{st}}{\alpha} \right) \mathbf{J}_{n_s \times n_t}, \qquad (3)$$

where d_{st} is the distance between cluster s and t, and $\mathbf{J}_{n_s \times n_t}$ is an $n_s \times n_t$ matrix with all entries being 1. That is, subjects in cluster s are equally correlated with subjects in cluster t. The strength of the correlation is a function of the distance between two clusters.

It follows that the overall correlation matrix $\boldsymbol{\Sigma}$ is a block matrix of the form

$$\boldsymbol{\Sigma} = [\Sigma_{ij}]_{S \times S}, \quad i, j = 1, \ldots, S, \qquad (4)$$

where the block-diagonal Σ_{ii} is given in (2) and the off block-diagonal Σ_{ij} is given in (3).
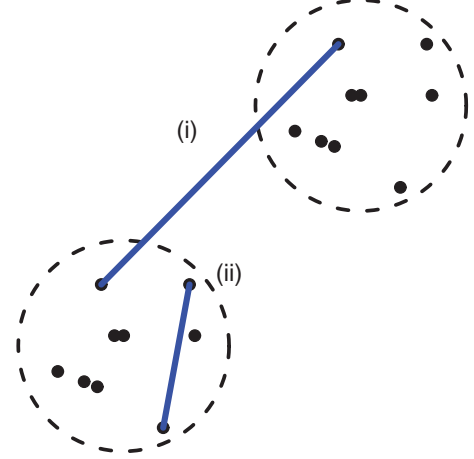

Figure 1. Configurations of spatial-clustered data with two clusters. (i) Between-cluster pair, (ii) within-cluster pair.

3. Estimation

General Theory. For a large-scale dataset, computing the distribution function of the GeoCopula models in equation (1) either requires high-dimensional integration or large matrix inversion, hence is not numerically feasible. Following Besag (1974), we consider a pseudo-likelihood approach to performing parameter estimation and inference for the GeoCopula models. Estimation functions are formulated from pairwise marginal composite likelihoods (Lindsay, 1988; Varin et al., 2011). Bai et al. (2012) proposed the joint composite estimating function (JCEF) approach to further improving the estimation efficiency. The idea is to form a quadratic objective function from different types of pairwise estimating functions. A natural grouping scheme for spatial-clustered data is to partition pairs into within-cluster and between-cluster groups (e.g., villages), as shown in Figure 1. The former contains pairs of observations from a common cluster, which are more relevant to subject-level effects and within-cluster correlations. The latter contains pairs of observations from different clusters, which capture more information relevant to cluster-level covariate effects and between-cluster spatial correlations.

To develop JCEF, the first step is to marginalize the high-dimensional CDF function in equation (1) into 2-dimensional margins. This is valid because copula models are marginally closed. Collect the parameters of interest into $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\psi}, \rho, \alpha)$, including the mean regression coefficients $\boldsymbol{\beta}$, the dispersion parameters $\boldsymbol{\psi}$, and the variance and covariance parameters ρ and α in $\boldsymbol{\Sigma}$, provided that a Matérn class spatial correlation function is used with a known parameter ν. Assume the length of $\boldsymbol{\theta}$ is q. By the property of marginal closure of the Gaussian copulas, a 2-dimensional marginal CDF is,

$$F(y_{si}, y_{tj}; \boldsymbol{\theta}) = \Phi_2\{\Phi^{-1}(F_{si}(y_{si}; \boldsymbol{\beta}, \psi_{si})), \Phi^{-1}(F_{tj}(y_{tj}; \boldsymbol{\beta}, \psi_{tj}))$$
$$| \, \boldsymbol{\Sigma}_{si,tj}(\rho, \alpha)\},$$

where $\boldsymbol{\Sigma}_{si,tj}$ is the 2×2 corresponding sub correlation matrix.

Let $f(y_{si}, y_{tj}; \boldsymbol{\theta})$ be the density with respect to $F(y_{si}, y_{tj}; \boldsymbol{\theta})$, whose explicit expression form can be adapted from stan-

dard copula density functions (Song, 2007, Section 6.3.2). Let $\boldsymbol{U}(y_{si}, y_{tj}; \boldsymbol{\theta})$ be the marginal score function associated with $f(y_{si}, y_{tj}; \boldsymbol{\theta})$, which is called the component score function (CSF). According to Varin et al. (2011), the conventional composite likelihood estimating functions sums all such CSFs within a certain distance lag d: $\boldsymbol{S}(\boldsymbol{\theta}, d) = \sum_{||s-t||<d} \boldsymbol{U}(y_{si}, y_{tj}; \boldsymbol{\theta}, d)$, where $|| \bullet ||$ is the Euclidean distance in space $\boldsymbol{R}^2$. We call this method weighted composite likelihood (WCL) approach (Bevilacqua et al., 2012). The weight is 0 or 1, depending on the distance between two clusters.

The optimal d can be determined by evaluating the Godambe information matrix (i.e., asymptotic covariance of the estimates) of the corresponding estimating equations (Bevilacqua et al., 2012). A value of d leading to the most informative set of estimating equations is then used. However, when the information matrix is computationally costly, empirical guidelines can be used. For example, from the empirical spatial variogram, one can learn the spatial dependence patterns, and choose a value for d within which pairwise correlations are fairly high. Numerous simulation experiments (Davis and Yau, 2011; Varin et al., 2011; Bai et al., 2012) have shown that including pairs within shorter distances usually results in more efficiency. This is desirable, since a substantial number of pairs can be eliminated from estimation, which greatly accelerates the computation speed.

To construct JCEF, we partition pairs into between-cluster and within-cluster groups. Label the two sets as $\mathcal{D}_{W,n}$ and $\mathcal{D}_{B,n}$, respectively. They are given by $\mathcal{D}_{W,n} = \{(s, i, s, j) : s \in \mathcal{S}, \text{ and } i \neq j, i, j \in \mathcal{I}_s\}$, and $\mathcal{D}_{B,n} = \{(s, i, t, j) : 0 < ||s-t|| \leq d; s, t \in \mathcal{S}, \text{ and } i \in \mathcal{I}_s, j \in \mathcal{I}_t\}$. Then $\mathcal{D}_n = \mathcal{D}_{W,n} \cup \mathcal{D}_{B,n} \subset \mathcal{R}^2 \times \mathcal{N}^+ \times \mathcal{R}^2 \times \mathcal{N}^+$ is the set containing all pairs used in estimation. For convenience, we write $\{\mathbf{y}(k) = (y_{si}, y_{tj})^\mathrm{T}, k = (s, i, t, j) \in \mathcal{D}_n\}$ as the paired random process.

The between-cluster CSF is constructed as $\boldsymbol{\Psi}_{B,n}(\boldsymbol{\theta}, d) = \frac{1}{|\mathcal{D}_{B,n}|} \sum_{(s,i,t,j) \in \mathcal{D}_{B,n}} \boldsymbol{U}(y_{si}, y_{tj}; \boldsymbol{\theta}, d)$. And the within-cluster CSF is $\boldsymbol{\Psi}_{W,n}(\boldsymbol{\theta}) = \frac{1}{|\mathcal{D}_{W,n}|} \sum_{(r,l,r,m) \in \mathcal{D}_{W,n}} \boldsymbol{U}(y_{rl}, y_{rm}; \boldsymbol{\theta})$, where $|A|$ is the cardinality of set A. Instead of summing two types of CSFs, we stack them into an extended CSF: $\boldsymbol{\Gamma}_n(\boldsymbol{\theta}, d) = \left(\boldsymbol{\Psi}_{B,n}^\mathrm{T}(\boldsymbol{\theta}, d), \boldsymbol{\Psi}_{W,n}^\mathrm{T}(\boldsymbol{\theta})\right)^\mathrm{T}$. Note that the dimension of $\boldsymbol{\Gamma}_n$ is larger than that of $\boldsymbol{\theta}$, leading to the so-called over-identification scenario (Hansen, 1982). To obtain an estimate of $\boldsymbol{\theta}$, following Hansen (1982) and Qu, Lindsay, and Li (2000), we form a quadratic objective function of the following form: $Q_n(\boldsymbol{\theta}, d) = \boldsymbol{\Gamma}_n^\mathrm{T}(\boldsymbol{\theta}, d)\boldsymbol{W}^{-1}\boldsymbol{\Gamma}_n(\boldsymbol{\theta}, d)$, where $\boldsymbol{W}$ is a $2p \times 2p$ positive-definite weight matrix. A JCEF estimator is defined as

$$\widehat{\boldsymbol{\theta}}_n(d) = \arg\min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} Q_n(\boldsymbol{\theta}, d). \quad (5)$$

According to Hansen (1982), the optimal weight matrix is $\boldsymbol{W}^* = \mathrm{Var}(\boldsymbol{\Gamma}_n(\boldsymbol{\theta}, d))$, in the sense that the resulting estimator has the maximum efficiency.

JCEF in multivariate Probit model. It is relatively easy to derive JCEF in the multivariate Gaussian model following the procedure outlined above. Here, we illustrate the derivation of JCEF in the GeoCopula regression model for binary data. We refer to Heagerty and Lele (1998) that considered multivariate probit model for spatial binary data.

First, the probability mass function for (Y_{si}, Y_{tj}) in the general canonical form is:

$$\log \Pr(Y_{si} = y_{si}, Y_{tj} = y_{tj}) = \alpha_0(si, tj) + \alpha_1(si, tj)y_{si}$$
$$+ \alpha_2(si, tj)y_{tj} + \alpha_3(si, tj)y_{si}y_{tj}.$$

Second, according to Zhao and Prentice (1990), the score function may be expressed as $\boldsymbol{U}_{si,tj}(\boldsymbol{\theta}) = \boldsymbol{\Delta}_{si,tj}^\mathrm{T} \boldsymbol{V}_{si,tj}^{-1} \boldsymbol{R}_{si,tj}$, with $\boldsymbol{\Delta}_{si,tj} = \frac{\partial}{\partial \boldsymbol{\theta}}(\mu_{si}, \mu_{tj}, \sigma_{si,tj})^\mathrm{T}$, $\boldsymbol{V}_{si,tj} = \mathrm{Var}(\boldsymbol{R}_{si,tj})$ and $\boldsymbol{R}_{si,tj}(\boldsymbol{\theta}) = (y_{si} - \mu_{si}, y_{tj} - \mu_{tj}, (y_{si} - \mu_{si})(y_{tj} - \mu_{tj}) - \sigma_{si,tj})^\mathrm{T}$. The detailed expression of $\boldsymbol{V}_{si,tj}$ can be found in Section 2 in the web supplementary material. Third, the GeoCopula model has $\mu_{si} = \mathrm{E}(y_{si}) = \Phi(\boldsymbol{x}_{si}^\mathrm{T}\boldsymbol{\beta})$, and $\sigma_{si,tj}^2 = \Phi_2\left(\boldsymbol{x}_{si}^\mathrm{T}\boldsymbol{\beta}, \boldsymbol{x}_{tj}^\mathrm{T}\boldsymbol{\beta} | \Sigma_{si,tj}\right) - \Phi(\boldsymbol{x}_{si}^\mathrm{T}\boldsymbol{\beta})\Phi(\boldsymbol{x}_{tj}^\mathrm{T}\boldsymbol{\beta})$. Finally, the group-based composite score functions are

$$\boldsymbol{\Psi}_{B,n}(\boldsymbol{\theta}, d) = \frac{1}{|\mathcal{D}_{B,n}|} \sum_{(s,i,t,j) \in \mathcal{D}_{B,n}} \boldsymbol{U}_{si,tj}(\boldsymbol{\theta}, d) \text{ and}$$

$$\boldsymbol{\Psi}_{W,n}(\boldsymbol{\theta}) = \frac{1}{|\mathcal{D}_{W,n}|} \sum_{(r,l,r,m) \in \mathcal{D}_{W,n}} \boldsymbol{U}_{rl,rm}(\boldsymbol{\theta}).$$

Estimation of the weight matrix. Although the optimal weight matrix $\boldsymbol{W}^* = \mathrm{Var}\{\boldsymbol{\Gamma}_n(\boldsymbol{\theta}, d)\}$ can be derived analytically using multivariate Gaussian quadrant probabilities, given the large number of possible pairs, analytic computation is not practically feasible. To mitigate this problem, we propose two approaches to estimating $\boldsymbol{W}^*$ in practice.

Parametric bootstrap. A straightforward solution to this problem is the bootstrap method (Efron, 1982; Varin et al., 2011) in that sampling data from the fully parametric model (1) is computationally efficient. To be specific, let $\widehat{\boldsymbol{\theta}}$ be a parameter estimate for the GeoCopula model (1) given observed data $\mathbf{y}$. For $m = 1, \ldots, M$, we simulate data $\mathbf{y}^{(m)}$ from the fitted model (1) with parameter $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$, and then we fit the model using $\mathbf{y}^{(m)}$ to obtain estimate $\widehat{\boldsymbol{\theta}}^{(m)}$. The parametric bootstrap estimate of the weight matrix is given by $\widehat{\boldsymbol{W}}_p(d) = \frac{1}{M} \sum_{m=1}^{M} \boldsymbol{\Gamma}_n(\widehat{\boldsymbol{\theta}}^{(m)}, d)\boldsymbol{\Gamma}_n^\mathrm{T}(\widehat{\boldsymbol{\theta}}^{(m)}, d)$. In the article, we refer to this approach as "JCEF.p", that is, the JCEF approach with the weight matrix estimated by parametric bootstrap.

Subgroup sampling. Alternatively, in spatial data analysis, estimation of the $\boldsymbol{W}^*$ matrix is mostly achieved by subsampling (Sherman and Carlestein, 1994). Specifically, let the sampling region be $\mathcal{A}_n = \mathcal{S} \times \mathcal{T}$, where $|\mathcal{A}_n| = n$. Under the assumption that, asymptotically, $|\mathcal{A}_n|\mathrm{E}\left\{\boldsymbol{\Gamma}_n(\boldsymbol{\theta})\boldsymbol{\Gamma}_n^\mathrm{T}(\boldsymbol{\theta})\right\} \to \boldsymbol{W}^*$, as $n \to \infty$, we can estimate matrix $\boldsymbol{W}^*$ using sample covariance matrix of statistics computed on either overlapping or non-overlapping subshapes of the sampling region $\mathcal{A}_n$. That is, $\widehat{\boldsymbol{W}}_s(\boldsymbol{\theta}, d) = \frac{1}{K_n} \sum_{k=1}^{K_n} |\mathcal{A}_{s(n)}^{(k)}| \left\{\boldsymbol{\Gamma}_n^{(k)}(\boldsymbol{\theta}, d) - \overline{\boldsymbol{\Gamma}}(\boldsymbol{\theta}, d)\right\} \left\{\boldsymbol{\Gamma}_n^{(k)}(\boldsymbol{\theta}, d) - \overline{\boldsymbol{\Gamma}}(\boldsymbol{\theta}, d)\right\}^\mathrm{T}$, with $\overline{\boldsymbol{\Gamma}}(\boldsymbol{\theta}, d) = \sum_{k=1}^{K_n} \boldsymbol{\Gamma}_n^{(k)}(\boldsymbol{\theta}, d)/K_n$, where $\boldsymbol{\Gamma}_n^{(k)}(\boldsymbol{\theta}, d)$ is vector $\boldsymbol{\Gamma}_n(\boldsymbol{\theta}, d)$ evaluated in $\mathcal{A}_{s(n)}^{(k)}$, $k = 1, \ldots, K_n$, a collection of (non)overlapping sub-shapes of $\mathcal{A}_n$, and K_n denotes the number of sub-shapes. This technique is useful to estimate matrix $\boldsymbol{W}^*$ and standard errors of parameter estimates. Politis and Romano (1994) showed that the optimal subsample size K_n

is proportional to $Cn^{a/(a+2)}$, where a is the dimension of the spatial domain and C is a tuning constant. There are two approaches suggested in the literature to determine constant C. One is demonstrated by Heagerty and Lumley (2000) that examine the effect of different choices of C, and the other is given by Sherman (1996) that utilizes some empirical evidence about the range of correlation for the selection of C by the following rule of thumb: If the correlation decays fast, small subsamples can be used; otherwise, large subsamples should be considered. Here, we refer to this approach as "JCEF.s", that is, the JCEF approach with the weight matrix being estimated by subgroup sampling.

Determining the smoothness parameter in Matérn kernel. The aforementioned estimation method assumes the smoothness parameter ν in the Matérn kernel (3) is known. In practice it needs to be estimated. Following the suggestions by Diggle and Ribeiro Jr (2007) and Bai et al. (2012), we propose to estimate ν using the profile quadratic objective function: $P_n(\nu, d) = Q_n(\widehat{\boldsymbol{\theta}}(\nu), d)$, where $\widehat{\boldsymbol{\theta}}(\nu)$ is a JCEF estimate in (5) with a given ν, and $Q_n(\cdot)$ is the corresponding quadratic objective function. Given a sequence of values $\nu_1, \ldots, \nu_K$, parameter ν can be estimated by minimizing $P_n(\nu, d)$, that is, $\widehat{\nu}(d) = \arg\min_k P_n(\nu_k, d)$. This $\widehat{\nu}$ along with the associated estimates $\widehat{\boldsymbol{\theta}}(\widehat{\nu})$ are the model parameter estimates.

Determining initial values. A simple and fast procedure is used to generate consistent initial estimates of $\boldsymbol{\theta}$ under independence working correlation, including (i) fit a GLM to obtain estimates of regression coefficients, $\widehat{\boldsymbol{\beta}}$; (ii) given $\widehat{\boldsymbol{\beta}}$, obtain $\widehat{\rho}$ using only the within-cluster CSF $\boldsymbol{\Psi}_{W,n}$, $\widehat{\rho} = \arg\min_\rho \boldsymbol{\Psi}_{W,n}^{\mathrm{T}}(\widehat{\boldsymbol{\beta}}, \rho, d)\boldsymbol{\Psi}_{W,n}(\widehat{\boldsymbol{\beta}}, \rho, d)$; (iii) given $\widehat{\boldsymbol{\beta}}, \widehat{\rho}$, obtain $\widehat{\alpha}$ using only the between-cluster CSF $\boldsymbol{\Psi}_{B,n}$, $\widehat{\alpha} = \arg\min_\alpha \boldsymbol{\Psi}_{B,n}^{\mathrm{T}}(\widehat{\boldsymbol{\beta}}, \widehat{\rho}, \alpha, d)\boldsymbol{\Psi}_{B,n}(\widehat{\boldsymbol{\beta}}, \widehat{\rho}, \alpha, d)$.

4. Large Sample Properties

We establish large-sample properties of the JCEF estimator $\widehat{\boldsymbol{\theta}}_n$ with the given optimal weighting matrix $\boldsymbol{W}^* = \mathrm{Var}\{\boldsymbol{\Gamma}_n(\boldsymbol{\theta}_0, d)\}$, under the increasing domain context (Mardia and Marshall, 1984). That is, the sample size increase is achieved by expanding the spatial domain. We present the two main theorems below. Details on assumptions and analytic arguments can be found in Section 3 of the supplementary material. We establish the consistency of the JCEF estimator in Theorem 1 and asymptotic normality in Theorem 2.

THEOREM 1. *Under the regularity conditions stated in Lemma 1 in the web supplementary material, if the true parameter value $\boldsymbol{\theta}_0$ is the unique minimizer of $EQ_n(\boldsymbol{\theta})$ in equation (5), and $\hat{\boldsymbol{\theta}}_n$ minimizes $Q_n(\boldsymbol{\theta})$, then $\hat{\boldsymbol{\theta}}_n \xrightarrow{p} \boldsymbol{\theta}_0$, as $n \to \infty$.*

THEOREM 2. *Under the increasing domain framework, given Assumptions 1–6 and mixing conditions (a)–(c) in the supplementary material, we have $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{L} N(\boldsymbol{0}, \boldsymbol{\Omega}(\boldsymbol{\theta}_0) \boldsymbol{\Lambda}(\boldsymbol{\theta}_0) \boldsymbol{\Omega}^{\mathrm{T}}(\boldsymbol{\theta}_0))$, $n \to \infty$, where $\boldsymbol{\Omega}(\boldsymbol{\theta}_0) = -\{\boldsymbol{I}^{\mathrm{T}}(\boldsymbol{\theta}_0)\boldsymbol{\Lambda}^{-1}(\boldsymbol{\theta}_0) \boldsymbol{I}(\boldsymbol{\theta}_0)\}^{-1}\boldsymbol{I}^{\mathrm{T}}(\boldsymbol{\theta}_0)\boldsymbol{\Lambda}^{-1}(\boldsymbol{\theta}_0)$, $\boldsymbol{\Lambda}(\boldsymbol{\theta}_0) = \lim_n n\mathrm{Var}\{\boldsymbol{\Gamma}_n(\boldsymbol{\theta}_0, d)\}$ and $\boldsymbol{I}(\boldsymbol{\theta}_0) = \lim_n \nabla_{\boldsymbol{\theta}} \boldsymbol{\Gamma}_n(\boldsymbol{\theta}_0, d)$.*

5. Simulation Experiments

We perform simulation studies to evaluate the performance of the proposed methods compared with some existing methods. In particular, four estimation methods are compared, the maximum likelihood estimation (MLE), the weighted composite likelihood estimation (WCL), the JCEF approach with the weight matrix estimated by parametric bootstrap (JCEF.p), and the JCEF approach with the weight matrix estimated by subsampling (JCEF.s). For the binary data, following Chan and Kuk (1997), we implement MLE using the EM algorithm, treating latent continuous variables Z_{si} as missing and applying Gibbs sampler to generate Monte–Carlo samples from truncated multivariate normal distributions. The subgroups in subsampling are chosen by Sherman (1996)'s method as overlapping circle subregions with a radius 4 and the number of subregions $K_n = 20$.

Set Up. We begin by conducting two simulation experiments, one based on clustered Gaussian data (Example 1) and the other based on multivariate probit model for clustered binary data (Example 2). For convenience, the number of subjects within a cluster is fixed at four. We randomly select 100 clusters from a 20×20 spatial grid with two coordinates spanning from 1 to 20. The total number of observations in one simulation run is 400.

For both experiments, the marginal mean model is specified the same with two covariates: $h(\mu_{si}) = \beta_0 + \beta_1 x_s^1 + \beta_2 x_{si}^2$, where x_s^1 is a cluster-level covariate, and x_{si}^2 is a subject-level covariate, both generated from the uniform distribution in $(0, 1)$. The correlation matrix $\boldsymbol{\Sigma}$ consists of diagonal blocks of 4×4 exchangeable correlation $\boldsymbol{\Sigma}_w$ in (2), and off-diagonal blocks of 4×4 Matérn correlation $\boldsymbol{\Sigma}_{st}$ in (3), where $\rho \in (-1, 1)$ is the within-cluster correlation, and $\alpha > 0$ is a spatial scaling parameter for between-cluster spatial correlation. The vector of parameters is $\boldsymbol{\theta} = (\beta_0, \beta_1, \beta_2, \rho, \alpha)$. For both cases, we set a common $\boldsymbol{\theta} = (1, 1, -1, 0.8, 2)$, $\psi = 1$ and the smoothness parameter $\nu = 1.5$. We generate 200 simulated datasets for each case to draw conclusions.

Estimating $\boldsymbol{\theta}$. We first evaluate the performance of estimating the parameter $\boldsymbol{\theta}$, with ν being fixed at 1.5. The profile estimation of ν will be considered in a separate simulation. We present estimation bias and root mean squared error (RMSE) in Table 1. Individual RMSE is scaled by the absolute value of the corresponding parameter, and summed up to obtain a measure of overall efficiency, called total scaled RMSE in Table 1.

As shown in Table 1, for Gaussian data, MLE, JCEF.p, and JCEF.s are comparable in terms of bias, while WCL tends to have a slightly larger bias, especially for β_2, ρ, and α. In general, MLE has the smallest RMSE, followed by JCEF.p and JCEF.s, whereas WCL has the largest RMSE. These results confirm that when the model assumption is satisfied, MLE achieves the highest efficiency, and JCEF.p is more accurate than JCEF.s. WCL appears to be the least efficient among the four methods. Clearly, the results in the Gaussian case show that both JCEF approaches improve estimation efficiency over WCL, and the JCEF.p performs comparably to MLE for some of the parameters. For the binary data case, these four methods yield similar bias and WCL appears to have large bias in estimating α. Once again, both JCEF

Table 1
Parameter estimation bias, root mean squared error (RMSE), and converge probability of 95% confidence interval of four methods (MLE, JCEF.p, JCEF.s, and WCL) using spatial-clustered Gaussian data and binary data. The true parameter is set as $\theta = (\beta_0, \beta_1, \beta_2, \rho, \alpha) = (1, 1, -1, 0.8, 2)$ and the smoothness parameter v is assumed to be known and fixed as 1.5. The number of clusters is 100 and the number of subjects per cluster is 4. The total sample size is 400. The total scaled RMSE is the summation of RMSEs scaled by the absolute true value of the corresponding parameters. Results are summarized based on 200 simulated datasets. All values are reported in unite of 10^{-2}.

		Gaussian				Binary			
$\times 10^{-2}$		MLE	JCEF.p	JCEF.s	WCL	MLE	JCEF.p	JCEF.s	WCL
	β_0	-0.25	0.32	0.09	-0.06	-2.19	-4.38	-0.91	-4.44
	β_1	0.07	-0.58	0.10	-0.78	3.82	-7.20	7.57	3.66
Bias	β_2	-0.04	0.71	-0.20	1.02	1.31	2.31	-1.38	8.35
	ρ	1.77	-1.65	2.96	-4.57	5.47	-4.73	-4.45	6.27
	α	0.23	-0.72	0.49	11.55	-16.29	7.41	-27.6	48.51
	β_0	1.87	1.91	2.00	6.37	7.35	7.71	8.16	7.84
	β_1	1.65	1.81	2.23	3.26	7.02	7.34	8.41	18.22
RMSE	β_2	1.29	1.52	1.56	3.92	5.98	6.08	7.25	8.79
	ρ	0.19	0.27	0.74	0.94	1.07	1.10	2.75	11.18
	α	9.54	9.94	12.02	12.72	15.67	15.71	17.77	36.46
Total scaled RMSE		9.83	10.55	12.73	21.08	29.52	30.37	36.13	67.05
	β_0	93.5	97.5	96.5	98.5	96.0	96.5	92.5	99.5
	β_1	95.0	98.5	96.5	96.0	94.0	97.5	96.5	100.0
Coverage	β_2	95.0	95.5	96.0	97.5	92.5	97.5	97.0	99.0
	ρ	96.5	95.0	94.5	98.0	94.5	97.0	93.5	100.0
	α	97.0	94.5	95.0	98.0	96.5	94.5	95.5	98.5

approaches achieve lower RMSE than WCL, judged by the total scaled RMSE. JCEF.p repeatedly outperforms JCEF.s with smaller RMSE.

Results of the confidence interval coverage are also reported in Table 1. It is shown that the WCL approach generally provides wider confidence intervals, and the resulting empirical coverage probabilities appear too near 100% to be reasonable. In contrast, the proposed JCEF.p and JCEF.s methods yield reasonable coverage rates. In other words, the WCL approach has lower power to detect signals than the proposed methods.

In summary, the proposed JCEF improves the estimation efficiency over the existing WCL for both Gaussian and binary data. The extent of the improvement depends on how accurately the weight matrix $\boldsymbol{W}^*$ is estimated. It seems that the weight matrix plays a more significant role in Gaussian data than in binary data for efficiency improvement. The efficiency gain also increases as the spatial dependence across clusters diminishes according to our other simulation studies (not shown here).

Determining v. We conduct a simulation study to evaluate performance of the profile QIF method for determining v discussed in Section 3. We focus on the case of spatial-clustered binary data as it is more challenging than the Gaussian case. We consider a set of v values in $(0.5, 1.0, 1.5, 2.0, 2.5, 3.0)$, and at each v value we apply the JCEF with subsampling approach (i.e., JCEF.s). We find that over 92% of 200 simulation rounds, the profile QIF method determines $\hat{v} = 1.5$. Figure 2 displays a boxplot of objective function $Q(\hat{v})$ at different v values over 200 simulated datasets. This indicates that minimizing the profile QIF objective function Q with respect to v works well.

Cluster size and sample size. To evaluate the estimation accuracy under different cluster and sample sizes, we conduct another simulation study with varying sample sizes. Multivariate probit model is chosen to mimic both real-data and model structure considered in Section 6. We randomly select S locations from 65 villages in Gambia (see Section 6 for data details) and generate at each location n binary outcomes y_{si} from a multivariate probit model with nine coefficients $\beta_j = 0.5 \times (-1)^j$ for $j = 1, \ldots, 9$ with covariance parameters $\rho = 0.5, \alpha = 2.5$, and $v = 1.5$. The number of locations S varies among 35, 45, 55, and 65, and the cluster size n_s ranges over 4, 6, 8, and 10, which results in a total of 16 cases. In Table 2, we summarize the total scaled RMSE and the max absolute bias of the nine parameter estimates over 200 simulation replicates. It is clear that when the sample size is 650 ($S = 65$, $n_s = 10$), the max absolute bias drops by 25% and the total scaled RMSE reduces by 17% compared to the case of sample size being 140 ($S = 35$, $n_s = 4$).

6. Data Example

In this section, we illustrate an application of the JCEF to a real-world data. Diggle et al. (2002) investigated the spatial variation in the prevalence of malaria among village resident children in Gambia. They developed a spatial generalized linear mixed model to account for the spatial correlation among the residuals at the village level and implemented it in a Bayesian MCMC framework. Thomson et al. (1999) used GEE to obtain regression estimates and accounted for the extra-binomial variation by a working correlation matrix with an exponential spatial correlation function. We now re-analyze this dataset using the proposed GeoCopula model.
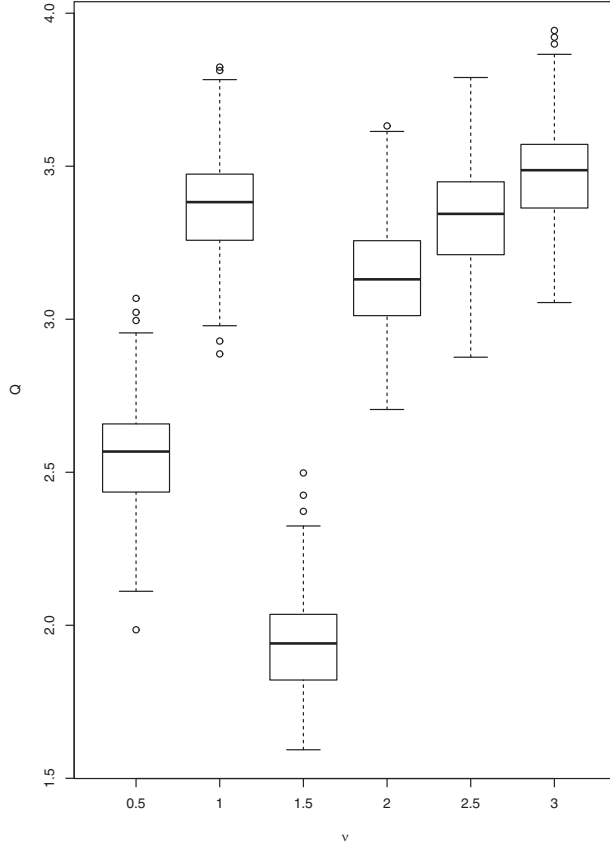
Figure 2. Boxplot that summarizes estimates of smoothness parameter ν by the profile quadratic inference function over 200 simulated datasets.

Figure 3. Spatial configuration of the sampled villages.

Two thousand thirty-five children were randomly sampled from 65 villages along the Gambia river. A graphical representation of the spatial configuration of the sampled villages is given in Figure 3. Villages scatter into four distinct regions on the map and are labeled from Area 1 to Area 5. The pairwise distances between two villages range from 0.95 to 273.3 km. The response from each child is a binary indicator of the presence of malarial parasites in the blood sample. Covariates include child level variables: age, bed net use
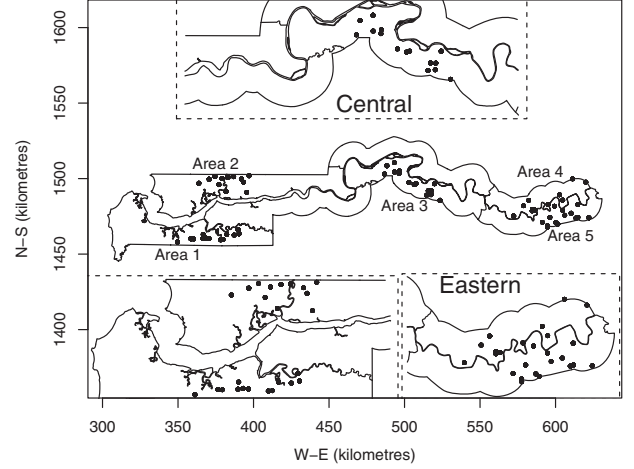
(NetUse) and whether the bed net is treated (Treated); and the village level variables: inclusion or exclusion from the primary health care (PHC) system and greenness of surrounding vegetation as derived from satellite information (Green). In the final model suggested by Diggle et al. (2002), the five-level area dummy variables (Area) are also included to adjust for the regional effects. However, information about the partition of Areas 4 and 5 is not available in the data given in R package `geoRglm` (Christensen and Ribeiro, 2002), nor can it be inferred from the map. As a result, we have to combine Areas 4 and 5 into one region in our analysis.

For child i in village s, let the binary random variable Y_{si} denote the presence of malaria (1 for yes; 0 for no). Let $\pi_{si} = E(Y_{si}|\boldsymbol{x}_{si})$ be the probability of the malaria infection. Then the probit model is:

$$
\begin{aligned}
\pi_{si} = \Phi(\boldsymbol{x}_{si}^{\mathrm{T}}\boldsymbol{\beta}) = \Phi(&\beta_0 + \beta_1 \mathrm{Age}_i + \beta_2 \mathrm{NetUse}_i + \beta_3 \mathrm{Treated}_i \\
&+ \beta_4 \mathrm{Green}_s + \beta_5 \mathrm{PHC}_s + \beta_6 \mathrm{Area}_{si}),
\end{aligned}
$$

where $\boldsymbol{\beta} = (\beta_0, \dots, \beta_6)^{\mathrm{T}}$ is the vector of corresponding regression coefficients.

The correlation matrix $\boldsymbol{\Sigma}$ is specified similar to equation (4). That is, the within-village correlation is specified as com-

Table 2

Estimation accuracy of the JCEF parameter estimation in the multivariate probit model for different cluster size (35, 45, 55, 65) and different number of observations per cluster (4, 6, 8, 10). The maximum absolute bias (MAB) across over all the parameter estimations and the total scaled RMSE (TRMSE) are summarized based on 200 replicates of simulation. All values are reported in unit of 10^{-2}.

$\times 10^{-2}$	Cluster size							
Observations	35		45		55		65	
per cluster	MAB	TRMSE	MAB	TRMSE	MAB	TRMSE	MAB	TRMSE
4	74.17	79.33	55.93	33.34	46.65	37.59	39.31	26.23
6	47.21	47.14	39.37	56.88	29.49	24.95	26.76	31.94
8	34.18	28.32	25.62	64.64	28.17	19.71	19.69	25.24
10	29.32	21.89	20.22	18.94	21.28	21.71	12.43	19.41

Table 3

Parameter estimates and 95% confidence intervals for the malaria data analysis, obtained from JCEF, WCL, and MLE at smoothness parameter $\hat{\nu} = 0.5$. We combine Areas 4 & 5 because of insufficient location information available in the data source.

	JCEF			WCL			MLE		
	Estimate	95% C.I.		Estimate	95% C.I.		Estimate	95% C.I.	
Int	-0.0221	-0.2590	0.2148	-0.1138	-0.8224	0.5949	-0.0241	-0.2383	0.1941
Age (year)	0.3723	0.0750	0.6696	0.5079	-0.1488	1.1646	0.2825	0.0903	0.6543
NetUse	-0.2156	-0.5042	0.0729	-0.1994	-0.7192	0.3205	-0.2138	-0.4800	0.0488
Treated	-0.1690	-0.3967	0.0587	-0.0450	-0.2679	0.1778	-0.1474	-0.3794	0.0414
Green	-0.0557	-0.2672	0.1558	0.0037	-0.2882	0.2956	-0.0645	-0.2477	0.1363
PHC	-0.2727	-0.7106	0.1651	-0.5945	-0.8438	-0.3452	-0.2937	-0.6877	0.1423
Area 2	-0.1948	-0.4206	0.0310	0.1087	-0.1385	0.3559	-0.2133	-0.4006	0.0110
Area 3	-0.4320	-0.6328	-0.2312	-0.4325	-0.7108	-0.1541	-0.3698	-0.6176	-0.2464
Areas 4 &5	0.5145	0.2467	0.7823	0.4057	0.0499	0.7614	0.5126	0.2639	0.7651
ρ	0.5401	0.4318	0.6447	0.5444	0.4288	0.6554	0.3918	0.4382	0.6420
α	2.4570	2.1052	2.8088	3.8955	3.1387	4.6523	1.9742	2.1353	2.7787

pound symmetry, and the between-village correlation is given by a Matérn kernel function of distance between two villages, as in equation (4). As discussed in Section 3, we minimize the profile quadratic inference function on a set of smoothness parameters $\{0.1, 0.5, 1.0, 1.5, 2.5, 3.0\}$. We find that the smoothness parameter $\hat{\nu} = 0.5$, and in this case the Matérn kernel function corresponds to the exponential decay function. This finding is in agreement with that reported in Diggle et al. (2002).

The choice of the distance lag d is based on the level of the empirical spatial correlation. Diggle et al. (2002) show that the spatial dependence decays at a fairly fast rate, so pairs of villages within 5 km are used to construct the pairwise CSF. To create subsamples for the weight matrix estimation and the standard error estimation, overlapping subregions within radius of 10 km and $K_n = 20$ are determined by Sherman's method as the sub-blocks. Given the fact that the villages scatter into four major regions, the subsampling is carried out in each region and then combined to form the overall subsample. In this way, spatial dependence patterns from different regions are all represented in the subsample.

Statistical results are summarized in Table 3, including JCEF estimates and their corresponding 95% confidence intervals obtained by the parametric bootstrap. JCEF finds that age (in years) is positively associated with malaria prevalence, and the bed net use and the treatment of the bed net tend to reduce the risk, although being only marginally significant at 0.1 significance level. Prevalence in the eastern region is significantly higher than the rest of regions. The 95% confidence interval for ρ is $(0.4318, 0.6447)$. The confidence interval for the spatial scaling parameter α is $(2.1052, 2.8088)$, corresponding to an approximately 65% decrease in dependence with 1 km increase in distance. This means that the spatial variation operates on a relatively small-scale. We find that both the JCEF estimates and the confidence intervals are close to those obtained by the MLE. In contrast, the WCL approach yields larger confidence intervals than our JCEF and the MLE for most of the model parameters, hence fails to identify some significant covariate effects (e.g., age). Also, the

findings from our JCEF approach are consistent with those in Diggle et al. (2002); for example, in the final model proposed by Diggle et al., age and being in Area 5 are positively associated with the risk of malaria. Nevertheless, it is important to note that results in this article, although similar, are not directly comparable to those in Diggle et al. (2002). The GeoCopula model provides population-level effect estimates, while the spatial linear mixed model used in Diggle et al. (2002) is a cluster-specific model. In addition, we do not have the specific boundary information for defining the same five regions as done in their analysis. Also for identification, correlation ρ instead of the variance is estimated in GeoCopula model, and our spatial scaling parameter α corresponds to the inverse of their scaling parameter.

7. Discussion

Till now, Bayesian methods are predominantly used in the analysis of spatial/temporal data, owing to the numerical limitations of the traditional likelihood methods. The proposed GeoCopula model and JCEF method in this article offer a competitive alternative package for analyzing spatial data from a frequentist perspective. The GeoCopula model furnishes both population-level regression parameter estimates and flexible within- and between-cluster spatial dependence structures. The new JCEF procedure is used to improve the estimation efficiency over the conventional composite likelihood methods (i.e., WCL). As shown in various simulation studies, the JCEF method gains significantly higher efficiency over the WCL approach for both Gaussian and binary spatial data, and is very comparable to MLE for Gaussian data. Our numerical experiences suggest that the parametric bootstrap works well albeit being slightly computationally costly, whereas the subsampling approach is faster but possibly provides underestimated standard errors when the sample size is small. Thus, the choice between parametric bootstrap and subsampling method may be made in light of the sample size.

We focus on the GeoCopula model built upon the multivariate Gaussian copula. Some disadvantages include the tail

independence and symmetrical correlation structures in the lower and upper tails. It is not clear if any modifications to account for tail-dependence are necessary to reflect spatial dependencies in practice. Although we particularly discuss the Matérn covariance function in this article, the proposed Geo-Copula model is so flexible that it may incorporate nonstationarity in both the mean and the spatial dependence models. This framework can be extended to more complex spatial dependence structures, such as those of inherent nonstationarity and anisotropy. In addition, as pointed out in Bai et al. (2012), the quadratic objective function also provides a way for a goodness-of-fit test of the mean-zero model assumption, $H_0 : E\mathbf{\Gamma}_n(\boldsymbol{\theta}) = \mathbf{0}$. This is because by Hansen (1982) $Q_n(\hat{\boldsymbol{\theta}})$ follows χ^2 distribution with degrees of freedom equal to the number of estimating functions minus the number of parameters. These are potential future research directions.

8. Supplementary Material

Web supplementary Sections 1–3 referenced in Sections 2, 3, 4 are available with this article at the *Biometrics* Website on Wiley Online Library. A user-friendly R package `GeoCopula` supplies both MAC OS X and DOS Window OS versions available on the following webpage: `web1.sph.emory.edu/users/jkang30/software/GeoCopula.html`. It provides respective R functions for estimation and inference in GeoCopula models for continuous and binary data. More details may be found in the supplementary material.

Acknowledgements

References

Albert, P. and McShane, L. (1995). A generalized estimating equations approach for spatially correlated binary data: With an application to the analysis of neuroimaging data. *Biometrics* **51**, 627–638.

Anselin, L. and Griffith, D. A. (1988). Do spatial effects really matter in regression analysis? *Papers, Regional Science Association* **65**, 11–34.

Bai, Y., Song, P. X.-K., and Raghunathan, T. (2012). Joint composite estimating functions in spatiotemporal models. *Journal of the Royal Statistical Society, Series B* **74**, 799–824.

Bárdossy, A. (2006). Copula-based geostatistical models for groundwater. *Water Resources Research* **42**, W11416.

Bárdossy, A. and Li, J. (2008). Geostatistical interpolation using copulas. *Water Resources Research* **44**, W07412.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B* **36**, 192–236.

Bevilacqua, M., Gaetan, C., Mateu, J., and Porcu, E. (2012). Estimating space and space-time covariance functions for large data sets: A weighted composite likelihood approach. *Journal of the American Statistical Association* **107**, 268–280.

Chaix, B., Merlo, J., and Chauvin, P. (2005). Comparison of a spatial approach with the multilevel approach for investigating place effects on health: The example of healthcare utilisation in France. *Journal of Epidemiology and Community Health* **59**, 517–526.

Chan, J. S. and Kuk, A. Y. (1997). Maximum likelihood estimation for probit-linear mixed models with correlated random effects. *Biometrics* **53**, 86–97.

Christensen, O. F. and Ribeiro Jr, P. J. (2002). georglm–A package for generalised linear spatial models. *R-NEWS* **2**, 26–28. ISSN 1609-3631.

Cressie, N. A. (1993). *Statistics for Spatial Data*, revised edition. New York: Wiley.

Davis, R. A. and Yau, C.-Y. (2011). Comments on pairwise likelihood in time series models. *Statistica Sinica* **21**, 255–277.

Diggle, P., Moyeed, R., Rowlingson, B., and Thomson, M. (2002). Childhood malaria in the Gambia: A case-study in model-based geostatistics. *Applied Statistics* **51**, 493–506.

Diggle, P. J. and Ribeiro Jr, P. J. (2007). *Model-Based Geostatistics*. New York: Springer.

Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*, volume 38. Montepelier, Vermont: SIAM.

Gotway, C. and Stroup, W. (1997). A generalized linear model approach to spatial data analysis. *Journal of Agricultural, Biological, and Environmental Statistics* **2**, 157–178.

Grady, S. C. (2010). Racial residential segregation impacts on low birth weight using improved neighborhood boundary definitions. *Spatial and Spatio-Temporal Epidemiology* **1**, 239–249.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* **50**, 1029–1054.

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

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

Jorgensen, B. (1997). *The Theory of Dispersion Models*. London: Chapman & Hall.

Kazianka, H. and Pilz, J. (2010). Copula-based geostatistical modeling of continuous and discrete data including covariates. *Stochastic Environmental Research and Risk Assessment* **24**, 661–673.

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

Lindsay, B. G. (1988). Composite likelihood methods. *Contemporary Mathematics* **80**, 221–239.

Mardia, K. and Marshall, R. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* **71**, 135–146.

Masarotto, G. and Varin, C. (2012). Gaussian copula marginal regression. *Electronic Journal of Statistics* **6**, 1517–1549.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, 2 edition. London: CRC Press.

Mujahid, M. S., Diez Roux, A. V., Morenoff, J. D., and Raghunathan, T. (2007). Assessing the measurement properties of neighborhood scales: From psychometrics to ecometrics. *American Journal of Epidemiology* **165**, 858–867.

Politis, D. and Romano, J. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *The Annals of Statistics* **22**, 2031–2050.

Qu, A., Lindsay, B., and Li, B. (2000). Improving generalized estimating equations using quadratic inference functions. *Biometrika* **87**, 823–836.

Sener, I., Pendyala, R., and Bhat, C. (2011). Accommodating spatial correlation across choice alternatives in discrete choice models: An application to modeling residential location

choice behavior. *Journal of Transport Geography* **19**, 294–303.

Sherman, M. (1996). Variance estimation for statistics computed from spatial lattice data. *Journal of the Royal Statistical Society, Series B* **58**, 509–523.

Sherman, M. and Carlestein, E. (1994). Nonparametric estimation of the moments of a general statistics computed from spatial data. *Journal of the American Statistical Association* **89**, 496–500.

Song, P. X.-K. (2000). Multivariate dispersion models generated from Gaussian copula. *Scandinavian Journal of Statistics* **27**, 305–320.

Song, P. X.-K. (2007). *Correlated Data Analysis*. New York: Springer.

Thomson, M., Connor, S., D'Alessandro, U., Rowlingson, B., Diggle, P., Creswell, M., and Greenwood, B. (1999). Predict-ing malaria infection in Gambian children from satellite data and bed net use surveys: The importance of spatial correla-tion in the interpretation of results. *The American Journal of Tropical Medicine and Hygiene* **61**, 2–8.

Varin, C. (2008). On composite marginal likelihoods. *Advances in Statistical Analysis* **92**, 1–28.

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

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