Statistica Sinica Preprint No: SS-2022-0401	
Title	A Mixture Generalized Estimating Equations Approach for Complex Spatially-Dependent Data
Manuscript ID	SS-2022-0401
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202022.0401
Complete List of Authors	Huichen Zhu, Fangzheng Lin, Huixia Judy Wang and Zhongyi Zhu
Corresponding Authors	Huichen Zhu
E-mails	huichenzhu@cuhk.edu.hk
Notice: Accepted version subject to English editing.	

A Mixture Generalized Estimating Equations Approach for Complex Spatially-Dependent Data

Huichen Zhu[1]*,Fangzheng Lin[2], Huixia Judy Wang[3], Zhongyi Zhu[4]

[1] *Department of Statistics, The Chinese University of Hong Kong, China*

* *Corresponding author*

[2] *Department of Statistics, Fudan University, China*

[3] *Department of Statistics, The George Washington University*

[4] *Department of Statistics, Fudan University, China*

Abstract: The generalized estimating equations (GEE) method is a popular approach for analyzing dependent data of various types. While GEE estimators are robust against the misspecification of the correlation matrix, their estimation efficiency can be seriously affected by the choice of the working correlation matrix. For spatially correlated data, it is difficult to specify the true spatial correlation structure due to the complexity of dependence and the high dimension of spatial correlation matrices. To achieve estimation efficiency while allowing flexibility to capture complex spatial dependence, we propose a new GEE-type approach based on a mixture of spatial working correlation matrices, referred to as mix-GEE. We show that the mix-GEE estimator is asymptotically efficient without any parametric assumption on the distribution as long as one of the candidate

correlation structures or some linear combination is correctly specified. Moreover, to overcome challenges in obtaining the inverse of the high-dimensional spatial correlation matrix for the large data set, we develop a tapered mix-GEE algorithm to reduce the computational cost and guarantee the estimation efficiency when the spatial correlation can be captured by the mixture correlation structure. The advantages of the proposed methods are demonstrated through simulations and the analysis of soil chemistry data.

Key words and phrases: Estimation efficiency; generalized estimating equations; misspecification; spatial data.

1. Introduction

The generalized estimating equation (GEE) method is a widely used approach for analyzing data with various types of dependence due to several appealing features (Liang and Zeger, 1986). Firstly, for valid inference, the method requires no assumption on the distribution but only the correct specification of the mean function, allowing the misspecification of the covariance function. Secondly, it can deal with various data types, e.g., continuous, binary, or count data. Lastly, the estimating equations often have simple forms and thus are easy to solve. In the last several decades, there are a large number of works on GEE for analyzing longitudinal data; see Hardin and Hilbe (2012) for an overview. The GEE has also been used for

analyzing spatially dependent data; see Oman et al. (2007), Lin (2008), Lin (2010), Thurman et al. (2015), Adegboye et al. (2018) and Cattelan and Varin (2018). We aim to develop a flexible and computationally feasible GEE method to deal with spatial data with complex spatial dependence, which is commonly seen in practice (Guan et al., 2004).

It is known that while GEE estimators are robust against the misspecification of the correlation matrix, their estimation efficiency can be seriously affected by the choice of the working correlation matrix (Wang and Carey, 2003). Recently, there have been some studies to address this issue with longitudinal data; see Leung et al. (2009), Xu et al. (2012), Fang et al. (2019) and Tang and Wang (2019). For spatial data, it is more challenging to specify a correlation structure with good approximations since the spatial correlation matrix is high-dimensional with the dimension equalling the sample size, and the correlation structure is often complex. One common practice in spatial analysis is selecting a spatial correlation function among a class of known parametric functions, e.g., Matérn correlation function, based on the shape of the empirical semivariogram (Lin et al., 2005; Oman et al., 2007; Lin, 2008). However, this approach relies on the assumption that a single correlation structure applies across space. The assumption may not be flexible to accommodate complex spatial dependence, such as het-

erogeneous spatial correlation or spatial correlation from multiple sources.

We propose a new GEE method that is flexible and computationally feasible to accommodate complex and possibly heterogeneous spatial dependence and also results in theoretical estimation optimality. The proposed method is based on a single GEE with the spatial working correlation matrix represented by a linear combination of candidate spatial correlation matrices. The unknown parameters in the combination are estimated by maximizing the Gaussian pseudo-likelihood with the mixed correlation matrix. We refer to the proposed method as "mix-GEE" since it is based on a mixture of spatial correlation matrices.

In the literature, there exist two approaches that also use the idea of the mixture to accommodate correlation from multiple sources: Xu et al. (2012), and Adegboye et al. (2018). Our proposed work has clear distinctions from these two works. Xu et al. (2012) focuses on longitudinal data assuming that the data is from a mixture distribution. The method requires repeated measurements and a finite dimension of the correlation matrix, which is not feasible in analyzing spatial data. In addition, the efficiency of the estimator relies on the Gaussian mixture distribution assumption. Adegboye et al. (2018) proposed a hybrid GEE approach for analyzing count data. They combined a series of GEEs corresponding to

different spatial correlation structures using the generalized method of moments (Hansen, 1982). The sample analog of the optimal weight function of each GEE is often singular, which is caused by the high dimension of the stacked estimating equations and the lack of repeated measurements at each spatial location. The singularity leads to computational problems, so an identity matrix is added in Adegboye et al. (2018) to make the weight matrix positive definite. However, the resulting weight matrix is not optimal and may have adverse effects on the estimation efficiency (Wang and Carey, 2003; Lin, 2008). Recently, Lu et al. (2023) proposed a functional-coefficient autoregressive spatio-temporal model, employing a weight matrix fusion technique to integrate spatial weight matrices. Although our approach shares certain similarities, it is fundamentally distinct in both concept and methodology. We characterize spatial correlation by using a mixture of spatial working correlations for approximation, and our estimation technique leverages the robustness of the GEE estimator to misspecification of the correlation structure.

Unlike previous works, our proposed mix-GEE method can handle high-dimensional and complex spatial correlation, and it can achieve estimation efficiency without requiring strong assumptions. In particular, for cases where the spatial dependence comes from a single source, the mix-GEE

estimator is efficient as long as the dependence is correctly captured by one candidate correlation structure. For spatial dependence that has either a complex form or is caused by multiple sources, the mix-GEE estimator can achieve theoretical efficiency if the dependence can be described as some linear combination of correlation structures. It also shows competitive performance even when the structures are misspecified.

Despite its advantages, the proposed method also has some challenges, which call for the development of new algorithms and theories. First, to ensure the positive definiteness of the estimated correlation matrix, we pose some constraints on the weights and correlation parameters in the optimization problem; see (2.6). The constraints are also important for achieving the efficiency of GEE estimators (Sutradhar and Das, 1999). To solve the constrained optimization problem, the popular Expectation-Maximization (EM) algorithm used in Xu et al. (2012) is not applicable in our setup since the pseudo-likelihood function used in our method can not be decomposed into a sum of density functions. To overcome this computational challenge, we develop an adaptive barrier algorithm by utilizing the idea in Lange (2010). We show that the algorithm can strictly drive the negative pseudo-log-likelihood function downhill at each iteration step and guarantee the positive definiteness for the estimated correlation matrix. Secondly, we

consider the asymptotic properties of the estimator with different situations of the true correlation, i.e., the true correlation is a linear combination of all the candidate working correlations, a linear combination of a subset of the candidate working correlations, or a single candidate working correlation. Lastly, the analysis of large spatial data could be time-consuming due to the inversion of a large correlation matrix. To address this computational issue, we propose a tapered mix-GEE algorithm and show that the method can still guarantee estimation efficiency under proper assumptions.

The rest of the paper is organized as follows. We present the proposed mix-GEE method and the estimation algorithm in detail in Section 2. The asymptotic properties of the estimator are introduced in Section 3. The tapered mix-GEE method for large spatial data is introduced in Section 4. We present the simulation study and analysis of soil chemistry data in Section 5 and 6. All the technical proofs are provided in a separate supplementary material.

2. Proposed Method

2.1 Notation and model setup

Let $Y(\boldsymbol{s}_i)$ be the response measured at location $\boldsymbol{s}_i \in \mathbb{R}^2$, and $\boldsymbol{X}(\boldsymbol{s}_i)$ be the corresponding p-dimensional covariate vector, $i = 1, \cdots, n$. If the ef-

fect of covariates $\boldsymbol{X}(\boldsymbol{s}_i)$ on $Y(\boldsymbol{s}_i)$ is of interest, classical spatial regression methods are often based on maximum likelihood estimation by assuming that the data follow some parametric distributions, for instance, Gaussian (Cressie, 2015), transformed Gaussian (Palacios and Steel, 2006), skew-Gaussian (Zhang and El-Shaarawi, 2010), and skew-t (Bevilacqua et al., 2021). However, these assumptions of parametric distributions may be restrictive in practice. Besides, due to the complexity of spatial dependence, it is often challenging to maximize the likelihood which takes a complicated form (McCullagh, 1983; Cressie, 2015; Sun and Stein, 2016).

One alternative approach is the GEE estimator (Liang and Zeger, 1986). For the dependent measures of spatial data, the joint distribution and the resulting likelihood are not easy to fully specify. We, rather, rely on a marginal mean model with a quasi-likelihood (Wedderburn, 1974), where only the first two moments are required to be specified. Compared to the likelihood-based methods, the GEE-type approach is simple to construct, has no requirement on distributions, and the validity only requires the correct specification of the mean function. We make the following model assumption. Without notation ambiguity, hereafter we write $Y(\boldsymbol{s}_i)$ and $\boldsymbol{X}(\boldsymbol{s}_i)$ as Y_i and $\boldsymbol{X}_i$, respectively. We assume that

$$\mathrm{E}(Y_i|\boldsymbol{X}_i) = \mu(\boldsymbol{X}_i^T\boldsymbol{\beta}), \quad \mathrm{Var}(Y_i|\boldsymbol{X}_i) = V\{\mu(\boldsymbol{X}_i^T\boldsymbol{\beta})\}, \tag{2.1}$$

where $\boldsymbol{\beta}$ are the unknown parameter of interest, and $\mu(\cdot)$ and $V(\cdot)$ are the link and variance functions. Denote $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$, $\boldsymbol{X} = (\boldsymbol{X}_i^T, \ldots, \boldsymbol{X}_n^T)^T$, $\boldsymbol{\mu}(\boldsymbol{\beta}) = (\mu(\boldsymbol{X}_1^T\boldsymbol{\beta}), \ldots, \mu(\boldsymbol{X}_n^T\boldsymbol{\beta}))^T$, $\boldsymbol{D}(\boldsymbol{\beta}) = \partial\boldsymbol{\mu}(\boldsymbol{\beta})/\partial\boldsymbol{\beta}^T$, and $\boldsymbol{A}(\boldsymbol{\beta}) = \text{diag}(V[\mu\{\boldsymbol{X}_1^T\boldsymbol{\beta}\}], \ldots, V[\mu\{\boldsymbol{X}_n^T\boldsymbol{\beta}\}])$.

2.2 GEE estimator based on a mixed correlation structures

We define the variance-covariance matrix of $\boldsymbol{Y}$ in the form $A(\boldsymbol{\beta})^{1/2}R(\boldsymbol{\alpha})A(\boldsymbol{\beta})^{1/2}$, where $\boldsymbol{R}(\boldsymbol{\alpha})$ is a $n \times n$ spatial working correlation matrix of the response $\boldsymbol{Y}$, characterized by the nuisance parameter vector $\boldsymbol{\alpha}$. The conventional spatial GEE estimator of $\boldsymbol{\beta}$ can be obtained by solving the following estimating equation,

$$\boldsymbol{D}(\boldsymbol{\beta})^T\boldsymbol{A}(\boldsymbol{\beta})^{-1/2}\boldsymbol{R}(\boldsymbol{\alpha})^{-1}\boldsymbol{A}(\boldsymbol{\beta})^{-1/2}\{\boldsymbol{Y} - \boldsymbol{\mu}(\boldsymbol{\beta})\} = 0.$$

The GEE-type estimator is consistent even when the working correlation structure is misspecified. However, many studies show that its estimation efficiency can be adversely affected under the misspecification of the correlation matrix (Wang and Carey, 2003; Xu et al., 2012). In the current literature of spatial regression, $\boldsymbol{R}(\boldsymbol{\alpha})$ is often constructed by choosing from a class of parametric families such as the Matérn correlation structure (Lin et al., 2005; Oman et al., 2007; Lin, 2008). In many applications, it is challenging to approximate the complex and high-dimensional spatial

correlation matrix by using a single spatial correlation structure.

To accommodate spatial data with complex dependence, we present a new GEE-based approach based on a mixture of spatial working correlation matrices. The new mixture correlation can capture complex dependence coming from multiple sources. We define the following estimating equations,

$$\boldsymbol{D}(\boldsymbol{\beta})^T\boldsymbol{A}(\boldsymbol{\beta})^{-1/2}\left\{\sum_{k=1}^{K}\pi_k\boldsymbol{R}^{(k)}(\boldsymbol{\alpha}_k)\right\}^{-1}\boldsymbol{A}(\boldsymbol{\beta})^{-1/2}\{\boldsymbol{Y}-\boldsymbol{\mu}(\boldsymbol{\beta})\}=0, \quad (2.2)$$

where K is the number of the candidate correlation structures, the weights $\pi_k \geq 0$, $\sum_{k=1}^{K}\pi_k = 1$, $\boldsymbol{R}^{(k)}(\boldsymbol{\alpha}_k)$ is kth spatial working correlation matrix with unknown nuisance parameters $\boldsymbol{\alpha}_k = (\alpha_{k,1}, \cdots, \alpha_{k,p_k})^T$, where p_k is the dimension of $\boldsymbol{\alpha}_k$. For the commonly used spatial correlation structures, we have $p_k = 1$ for exponential, Gaussian, and spherical correlation structures and $p_k = 2$ for Matérn.

The issue of potential anisotropy always happens in spatial data. For example, the spatial correlation between observations can vary in different directions. Thus, we construct the mixture working correlation matrix based on geometric anisotropy, which is one of the most common forms of anisotropy (Budrikaite and Ducinskas, 2005). It is defined by a linear transformation of the coordinate system. Specifically, we define the (i, j)-th element of $\boldsymbol{R}^{(k)}(\boldsymbol{\alpha}_k)$ as

$$\boldsymbol{R}_{ij}^{(k)}(\boldsymbol{\alpha}_k) = \rho_k(\|\boldsymbol{B}_k(\lambda_k, \psi_k)(\boldsymbol{s}_i - \boldsymbol{s}_j)\|, \boldsymbol{\alpha}_k), \quad (2.3)$$

where $\|\cdot\|$ represents the L_2 norm,

$$\boldsymbol{B}_k(\lambda_k,\psi_k)=\begin{pmatrix}1 & 0\\ 0 & \lambda_k\end{pmatrix}\begin{pmatrix}\cos(\psi_k) & -\sin(\psi_k)\\ \sin(\psi_k) & \cos(\psi_k)\end{pmatrix},\quad 0<\lambda_k\le 1,\ \psi_k\in[0,\pi),$$

λ_k and ψ_k represent stretching and rotation parameters, and $\rho_k(\cdot)$ is a spatial correlation function. The isotropic spatial correlation structure corresponds to $\lambda_k=1$, for which we have $\|\boldsymbol{B}_k(\lambda_k,\psi_k)(\boldsymbol{s}_i-\boldsymbol{s}_j)\|=\|\boldsymbol{s}_i-\boldsymbol{s}_j\|$.

Remark 1. Throughout the numerical studies of this paper, we consider the working mixture correlation structure with $K=3$, where each component takes the form (2.3) with an exponential function,

$$\rho_k(||\boldsymbol{B}_k(\lambda_k,\psi_k)(\boldsymbol{s}_i-\boldsymbol{s}_j)||,\alpha_k)=e^{-\alpha_k||\boldsymbol{B}_k(\lambda_k,\psi_k)(\boldsymbol{s}_i-\boldsymbol{s}_j)||},\quad \alpha_k>0, k=1,2,3, \tag{2.4}$$

where $(\lambda_1,\psi_1)=(1,0)$, $(\lambda_2,\psi_2)=(1/6,0)$, and $(\lambda_3,\psi_3)=(1/6,\pi/2)$. The exponential correlation function is popular in spatial analysis. The combination $(\lambda_1,\psi_1)=(1,0)$ represents the isotropic exponential correlation structure, and $(\lambda_2,\psi_2)=(1/6,0)$ and $(\lambda_3,\psi_3)=(1/6,\pi/2)$ represent two anisotropic correlation structures, reflecting stronger spatial correlations in the directions of south-north and east-west, respectively. While our proposed method works and the theoretical results apply for any fixed K, a large K may cause slow convergence due to the large unknown parameter space. For practical applications, we found that $K=3$ is often sufficient

to provide flexible approximations.

2.3 Estimation procedure

Our primary interest is to estimate the regression coefficients $\boldsymbol{\beta}$, but the nuisance parameters $\boldsymbol{\psi} = (\boldsymbol{\alpha}_k^T, \pi_k)_{k=1}^K$ are still unknown. In this section, we construct an alternating iterative procedure to simultaneously estimate $\boldsymbol{\beta}$ and $\boldsymbol{\psi}$. We also show that the proposed procedure can automatically select the most suitable subset of the candidate correlation structures to describe the true spatial correlation. The procedure is as follows.

Given $\boldsymbol{\psi} = (\boldsymbol{\alpha}_k^T, \pi_k)_{k=1}^K$, we estimate $\boldsymbol{\beta}$ by solving (2.2). Then given $\boldsymbol{\beta}$, we update the nuisance parameter by the pseudo-likelihood estimator defined as follows. Denote $\boldsymbol{\epsilon}(\boldsymbol{\beta}) = \boldsymbol{A}(\boldsymbol{\beta})^{-1/2}\{\boldsymbol{Y} - \boldsymbol{\mu}(\boldsymbol{\beta})\}$. We adopt the Gaussian pseudo-likelihood with the mixed working correlation structure, for which the corresponding twice negative pseudo-log-likelihood is

$$\ell_n(\boldsymbol{\psi}, \boldsymbol{\beta}) = \frac{1}{n}\log\left\{\left|\sum_{k=1}^K \pi_k \boldsymbol{R}^{(k)}(\boldsymbol{\alpha}_k)\right|\right\} + \frac{1}{n}\boldsymbol{\epsilon}(\boldsymbol{\beta})^T\left\{\sum_{k=1}^K \pi_k \boldsymbol{R}^{(k)}(\boldsymbol{\alpha}_k)\right\}^{-1}\boldsymbol{\epsilon}(\boldsymbol{\beta}) \quad (2.5)$$

The maximum pseudo-likelihood estimator of $\boldsymbol{\psi}$ can be obtained by minimizing (2.5) under the constraints,

$$\sum_{k=1}^K \pi_k = 1,\ \pi_k \geq 0, \alpha_{k,l} > 0,\ k = 1, \cdots, K, l = 1, \cdots, p_k. \quad (2.6)$$

The constraints $\alpha_{k,l} > 0$ guarantee the positive definiteness of commonly-used working spatial correlation matrices, such as the exponential corre-

lation matrices in (2.4) (Furrer et al., 2016). The positive definiteness of the mixed correlation matrix is guaranteed when the constraints in (2.6) are satisfied. When the working correlation structure is correctly specified, maximizing the Gaussian pseudo-likelihood over the nuisance parameters is equivalent to solving an unbiased estimating equation (Wu et al., 2001; Sun et al., 2009). Thus, the maximum pseudo-likelihood estimator of $\boldsymbol{\psi}$ still converges to the true value even if the Gaussian assumption is violated with a correctly specified correlation structure.

Suppose that the objective function (2.5) with linear constraints (2.6) is twice continuously differentiable, which commonly-used spatial working correlation matrices satisfy, such as exponential and Gaussian correlation matrices. Lange (2010) proposed a general effective algorithm to minimize a convex function subject to linear constraints, called the adaptive barrier algorithm. We adapt the algorithm to minimize the objective function in (2.5) with linear constraints (2.6), and refer to it as the pseudo-likelihood adaptive barrier (PLAB) algorithm. The adaptive barrier algorithm has the general property that the objective function is ensured to not increase after each iteration. In contrast, it can be proved that our proposed algorithm will drive the pseudo-likelihood function in (2.5) strictly downhill, which will be discussed in detail later.

Below we present the steps of the PLAB algorithm. First, we construct the adaptive barrier function, which involves $\ln\pi_k$ and $\ln\alpha_{k,l}$, $k = 1, \cdots, K$, $l = 1, \cdots, p_k$, and $\pi_K = 1 - \sum_{k=1}^{K-1} \pi_k$. The adaptive barrier function allows the convergence to a boundary point of the feasible region defined by the constrained (2.6). The proposed objective function (2.5) also ensures that the updated parameters are feasible if the initial values are in the feasible region (Lange, 2010). Let $\boldsymbol{\psi}^{(t)}$ represent the value of $\boldsymbol{\psi}$ at the t-th step, and δ be a small positive constant. Then the adaptive barrier function at the t-th step is defined as

$$\mathcal{F}(\boldsymbol{\psi}|\boldsymbol{\psi}^{(t)}) = \delta \left\{ \sum_{k=1}^{K} \left(\pi_k^{(t)} \ln\pi_k \right) + \sum_{k=1}^{K} \sum_{l=1}^{p_k} \left(\alpha_{k,l}^{(t)} \ln\alpha_{k,l} - \alpha_{k,l} \right) \right\}. \tag{2.7}$$

Following the general theory of the adaptive barrier algorithm in Section 16.3 of Lange (2010), $\mathcal{F}(\boldsymbol{\psi}|\boldsymbol{\psi}^{(t)})$ is concave and attains its maximum at $\boldsymbol{\psi} = \boldsymbol{\psi}^{(t)}$. Therefore, the minimization of $\ell_n(\boldsymbol{\psi}, \boldsymbol{\beta})$ with constraints can be transformed to the problem of minimizing the surrogate function without constraints, that is,

$$\boldsymbol{\psi}^{(t+1)} = \arg\min_{\boldsymbol{\psi}} \mathcal{S}(\boldsymbol{\psi}|\boldsymbol{\psi}^{(t)}) \quad \text{and} \quad \mathcal{S}(\boldsymbol{\psi}|\boldsymbol{\psi}^{(t)}) = \ell_n(\boldsymbol{\psi}, \boldsymbol{\beta}) - \mathcal{F}(\boldsymbol{\psi}|\boldsymbol{\psi}^{(t)}), \tag{2.8}$$

because $\ell_n(\boldsymbol{\psi}^{(t+1)}, \boldsymbol{\beta}) = \mathcal{S}(\boldsymbol{\psi}^{(t+1)}|\boldsymbol{\psi}^{(t)}) + \mathcal{F}(\boldsymbol{\psi}^{(t+1)}|\boldsymbol{\psi}^{(t)}) \leq \mathcal{S}(\boldsymbol{\psi}^{(t)}|\boldsymbol{\psi}^{(t)}) + \mathcal{F}(\boldsymbol{\psi}^{(t)}|\boldsymbol{\psi}^{(t)}) = \ell_n(\boldsymbol{\psi}^{(t)}, \boldsymbol{\beta})$. The updated value $\boldsymbol{\psi}^{(t+1)}$ can be obtained by using existing optimization methods such as Newton, quasi-Newton, or

Nelder-Mead methods. The iteration is repeated until the difference between $\boldsymbol{\psi}^{(t+1)}$ and $\boldsymbol{\psi}^{(t)}$ is sufficiently small, and the estimate of $\boldsymbol{\psi}$ is taken to be the value at the last iteration step. For the Gaussian pseudo-likelihood, the following theorem illustrates that the PLAB algorithm drives $\ell_n(\boldsymbol{\psi}^{(t+1)}, \boldsymbol{\beta})$ strictly downhill because we can show that $\mathcal{F}(\boldsymbol{\psi}|\boldsymbol{\psi}^{(t)})$ in (2.7) is strictly concave.

Theorem 1. *Suppose that the initial value $\boldsymbol{\psi}^{(0)}$ is within the feasible region and $\boldsymbol{\psi}^{(t+1)} \neq \boldsymbol{\psi}^{(t)}$, then $\ell_n(\boldsymbol{\psi}^{(t+1)}, \boldsymbol{\beta}) < \ell_n(\boldsymbol{\psi}^{(t)}, \boldsymbol{\beta})$ for every t, and the value of $\boldsymbol{\psi}$ at every iteration step satisfies the constraints.*

Theorem 1 indicates that if $\ell_n(\boldsymbol{\psi}, \boldsymbol{\beta})$ has a unique minimum under the constraints in (2.6), then the PLAB algorithm may converge to the global minimum. Our numerical experience suggests that the PLAB algorithm is not sensitive to the initial value $\boldsymbol{\psi}^{(0)}$. In practice, we can take a random initial value within the region of constraints or use the initial value that minimizes $\ell_n(\boldsymbol{\psi}, \boldsymbol{\beta})$ over a small number of grid points. The choice of δ in (2.7) is not essential, and we simply take $\delta = 10^{-4}$. We summarize the proposed alternating iterative algorithm to estimate $\boldsymbol{\beta}$ and $\boldsymbol{\psi}$ in Algorithm 1. We denote the final iterative values of $(\boldsymbol{\beta}, \boldsymbol{\psi})$ as $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\psi}})$, and refer them to as the mix-GEE estimations.

In Algorithm 1, as the sample size n goes to infinity, the difference

between the $(m+s)$th and mth updates of $\boldsymbol{\beta}$ in Algorithm 1 vanishes in probability for any m and s. The initial values $\boldsymbol{\beta}^{(0)}$ based on the working independence correlation structure is $\sqrt{n}$-consistent (Lin, 2008). The vanishing difference can be verified by the consistency of $\boldsymbol{\beta}^{(0)}$ and Lemma 1 and 2 in the Supplementary material. It indicates that a good choice of the initial values leads to fast convergence.

Algorithm 1 Mix-GEE algorithm

Set the initial values of parameters of interest as $\boldsymbol{\beta}^{(0)}$.
while The convergence criterion does not satisfy **do**
In the mth iteration: fix $\boldsymbol{\beta}^{(m-1)}$
Compute $\boldsymbol{\epsilon}(\boldsymbol{\beta}^{(m-1)}) = \boldsymbol{A}(\boldsymbol{\beta}^{(m-1)})^{-1/2}\{\boldsymbol{Y} - \boldsymbol{\mu}(\boldsymbol{\beta}^{(m-1)})\}$.
Set the initial values of nuisance parameters as $\boldsymbol{\psi}^{(0)}$
while The convergence criterion does not satisfy **do**
In the tth iteration: update the nuisance parameters $\boldsymbol{\psi}^{(t)}$ by (2.8) with $\boldsymbol{\epsilon}(\boldsymbol{\beta}^{(m-1)})$ plugged in.
end while
The final iterative values of $\boldsymbol{\psi}$ are denoted as $\boldsymbol{\psi}^{(m)}$.
Update the parameter of interest $\boldsymbol{\beta}^{(m)}$ by solving the equation (2.5) with $\boldsymbol{\psi}^{(m)}$ plugged in.
end while

The initial values $\boldsymbol{\beta}^{(0)}$ are the GEE estimator based on the working independence correlation. The initial values $\boldsymbol{\psi}^{(0)}$ must be in the feasible region.

Remark 2. The PLAB algorithm has three main features that make it appealing to compute the mix-GEE estimator. Firstly, the monotonicity property, $\ell_n(\boldsymbol{\psi}^{(t+1)}, \boldsymbol{\beta}) \leq \ell_n(\boldsymbol{\psi}^{(t)}, \boldsymbol{\beta})$, promotes numerical stability. Secondly, this algorithm allows π_k, $k \in \{1, \cdots, K\}$, to approach the boundary of the feasible region constructed by the constraints (2.6), i.e., $\pi_k = 0$, be-

cause the barrier function is adaptive in the sense that the coefficient $\pi_k^{(t)}$ of the barrier term $ln\pi_k$ changes from one iteration to the next. Therefore, when $\pi_k^{(t)}$ tends to zero, the term $\pi_k^{(t)} ln\pi_k$ can be very small allowing π_k to approach the boundary $\pi_k = 0$. This property can be rigorously proven for convex programming problems (Lange, 1994). Thirdly, if we let the initial values of $\alpha_{k,l}$ strictly larger than 0 for $k = 1, \ldots, K$ and $l = 1, \ldots, p_k$, the updated value of $\alpha_{k,l}$ will also be strictly larger than 0 (Lange, 1994).

3. Asymptotic Properties

We establish the asymptotic properties of the mix-GEE estimator $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\psi}})$ under some conditions, including the following important identifiability conditions 1-2, and other regularity conditions (C1)-(C6) given in the supplementary material. All the properties are obtained under the increasing domain, i.e., the minimum distance between any two locations is bounded away from zero (Furrer et al., 2016).

Let $\boldsymbol{\Psi}$ be the nuisance parameter space, defined by $\boldsymbol{\Psi} = \big\{(\boldsymbol{\alpha}_k^T, \pi_k)_{k=1}^K : \boldsymbol{\alpha}_k \in \boldsymbol{\Psi}_{\boldsymbol{\alpha}_k},\ \pi_k \in [0,1],\ \sum_{k=1}^K \pi_k = 1\big\}$, where $\boldsymbol{\Psi}_{\boldsymbol{\alpha}_k}$ is a compact subset of $(0,\infty) \times \cdots \times (0,\infty)$. Suppose that $\mathrm{E}\{\ell_n(\boldsymbol{\psi}, \boldsymbol{\beta}_0)\}$ converges uniformly to $\bar{\ell}(\boldsymbol{\psi}, \boldsymbol{\beta}_0)$ over $\boldsymbol{\Psi}$, where $\boldsymbol{\beta}_0$ is the true value of the regression parameter.

Condition 1. *Let* $\boldsymbol{\Psi}_0 = \big\{\boldsymbol{\psi}_0 \in \boldsymbol{\Psi} : \bar{\ell}(\boldsymbol{\psi}_0, \boldsymbol{\beta}_0) = \inf_{\boldsymbol{\psi} \in \boldsymbol{\Psi}} \bar{\ell}(\boldsymbol{\psi}, \boldsymbol{\beta}_0)\big\}$. *For*

every $\xi > 0$ *and* $\boldsymbol{\psi}_0 \in \boldsymbol{\Psi}_0$, $\inf_{\boldsymbol{\psi}\in\boldsymbol{\Psi}, d(\boldsymbol{\psi},\boldsymbol{\Psi}_0)\geq\xi} \bar{\ell}(\boldsymbol{\psi},\boldsymbol{\beta}_0) > \bar{\ell}(\boldsymbol{\psi}_0,\boldsymbol{\beta}_0)$, *where* $d(\boldsymbol{\psi},\boldsymbol{\Psi}_0) = \inf_{\boldsymbol{\psi}_0\in\boldsymbol{\Psi}_0}\{||\boldsymbol{\psi}-\boldsymbol{\psi}_0||\}$ *represents the distance from point to set.*

Condition 2. $\boldsymbol{\Psi}_0$ *may either contain a unique point or more than one point. If it contains more than one point and for any two different points* $\boldsymbol{\psi}_1 = (\boldsymbol{\alpha}_{k,1}^T, \pi_{k,1})_{k=1}^K$, $\boldsymbol{\psi}_2 = (\boldsymbol{\alpha}_{k,2}^T, \pi_{k,2})_{k=1}^K \in \boldsymbol{\Psi}_0$, *it satisfies that* $\pi_{k,1} = \pi_{k,2}$ *for all* $k \in \{1,\cdots,K\}$, *and* $\pi_{k^*,1} = \pi_{k^*,2} = 0$ *with* k^* *corresponding to* $\boldsymbol{\alpha}_{k^*,1} \neq \boldsymbol{\alpha}_{k^*,2}$.

According to the definition in Condition 1, it is obvious that $\boldsymbol{\Psi}_0$ may not be a single point. When $\bar{\ell}(\boldsymbol{\psi},\boldsymbol{\beta}_0)$ obtains the minimum on the boundary of the feasible region of $\boldsymbol{\Psi}$, for example, when $\pi_k = 0$ for some $k \in \{1,\cdots,K\}$, the values of $\ell_n(\boldsymbol{\psi},\boldsymbol{\beta}_0)$ are the same for $\boldsymbol{\alpha}_k \in \boldsymbol{\Psi}_{\boldsymbol{\alpha}_k}$ with the corresponding $\pi_k = 0$. Under this situation, $\boldsymbol{\Psi}_0$ contains multiple points by its definition. The boundary situation when $\boldsymbol{\Psi}$ contains more than one point may happen when the true correlation structure is a linear combination of a proper subset of the candidate correlation structures. Condition 2 specifically states the different situations of $\boldsymbol{\Psi}_0$. When the two conditions and Condition (C6) in the supplementary material hold, it is obvious that if $\pi_k^0 = 0$ for some $k \in \{1,\cdots,K\}$, we have $\boldsymbol{\Psi}_0 = \{(\boldsymbol{\alpha}_k{}^T, \pi_k^0)_{k=1}^K : \boldsymbol{\alpha}_k = \boldsymbol{\alpha}_k^0$ *with* $\pi_k^0 > 0$, *and* $\boldsymbol{\alpha}_k \in \boldsymbol{\Psi}_{\boldsymbol{\alpha}_k}$ *with* $\pi_k^0 = 0\}$. Otherwise, we have $\boldsymbol{\Psi}_0 = \{(\boldsymbol{\alpha}_k^{0T}, \pi_k^0)_{k=1}^K\}$.

The regularity conditions (C1)-(C6) are detailed in the supplementary material. Conditions (C1)-(C3) are regular requirements to derive asymptotic properties in spatial analysis; similar conditions can be found in Shaby and Ruppert (2012), Furrer et al. (2016), and Bachoc et al. (2018). Condition (C1) asserts that the asymptotic properties are deduced under the increasing domain framework. Two primary asymptotic frameworks in spatial literature are the increasing domain asymptotics and the infill asymptotics, for which more details can be found in Jenish and Prucha (2009). Condition (C2) ensures certain eigenvalue properties of high-dimensional matrices and is met by many standard spatial correlation functions, such as exponential and Gaussian. Condition (C3) is a regular assumption on the true covariance matrix. Further, Conditions (C4)-(C5) are discussed in Lin (2008). Condition (C4) pertains to the GEE method, while Condition (C5) controls the rate of correlation decay for products of random variables. Lastly, Condition (C6) can be easily verified, as by Kaufman et al. (2008), we have $\partial \mathrm{E}\{\ell_n(\boldsymbol{\psi}, \boldsymbol{\beta}_0; \gamma_2)\}/\partial \boldsymbol{\psi} = 0$ at $\boldsymbol{\psi}_0 = (\boldsymbol{\alpha}_k^{0T}, \pi_k^0)_{k=1}^K$.

Theorem 2. *Under the model assumption* (2.1)*, and if Conditions 1 and (C1)-(C5) in the supplementary material hold, we have that for every* $\zeta > 0$*, as* $n \to \infty$*,* $P\left\{d(\widehat{\boldsymbol{\psi}}, \boldsymbol{\Psi}_0) \geq \zeta\right\} \to 0$.

Theorem 2 shows that the distance between $\widehat{\boldsymbol{\psi}}$ and $\boldsymbol{\Psi}_0$ will converge in

probability to zero. Since $\boldsymbol{\Psi}_0$ are allowed to contain multiple points, the asymptotic variance of $\widehat{\boldsymbol{\beta}}$ may not be unique. To guarantee the uniqueness of the asymptotic variance over $\boldsymbol{\Psi}_0$ and establish the asymptotic distribution of $\widehat{\boldsymbol{\beta}}$, the identifiable condition, i.e., Condition 2 is necessary.

Let $\boldsymbol{\Sigma}_0 = \mathrm{Cov}(\boldsymbol{Y})$ be the true variance-covariance matrix of $\boldsymbol{Y}$, and $\boldsymbol{R}(\boldsymbol{\psi}) = \sum_{k=1}^{K} \pi_k \boldsymbol{R}^{(k)}(\boldsymbol{\alpha}_k)$ be the working correlation matrix. If Condition 2 holds, then $\boldsymbol{R}(\boldsymbol{\psi}_0)$ for all $\boldsymbol{\psi}_0 \in \boldsymbol{\Psi}_0$ is a unique value. Furthermore, we denote

$$\boldsymbol{\Pi}(\boldsymbol{\beta}, \boldsymbol{\psi}) = \sqrt{n} \left\{ \boldsymbol{D}(\boldsymbol{\beta})^T \boldsymbol{A}(\boldsymbol{\beta})^{-1/2} \boldsymbol{R}(\boldsymbol{\psi})^{-1} \boldsymbol{A}(\boldsymbol{\beta})^{-1/2} \boldsymbol{D}(\boldsymbol{\beta}) \right\}^{-1} \boldsymbol{D}(\boldsymbol{\beta})^T \boldsymbol{A}(\boldsymbol{\beta})^{-1/2} \boldsymbol{R}(\boldsymbol{\psi})^{-1} \boldsymbol{A}(\boldsymbol{\beta})^{-1/2}.$$

The asymptotic distribution of $\widehat{\boldsymbol{\beta}}$ is as follows,

Theorem 3. *Under the model assumption* (2.1)*, and if Conditions 1-2 and (C1)-(C5) in the supplementary material hold, suppose that for* $\boldsymbol{\psi}_0 \in \boldsymbol{\Psi}_0$, $\boldsymbol{\Pi}(\boldsymbol{\beta_0}, \boldsymbol{\psi}_0) \boldsymbol{\Sigma}_0 \boldsymbol{\Pi}(\boldsymbol{\beta_0}, \boldsymbol{\psi}_0)^T$ *converges to a positive definite matrix* $\Xi(\boldsymbol{\beta}_0, \boldsymbol{\psi}_0)$*, then we have* $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{D} N\left\{0, \Xi(\boldsymbol{\beta}_0, \boldsymbol{\psi}_0)\right\}$.

The convergence condition of $\boldsymbol{\Pi}(\boldsymbol{\beta_0}, \boldsymbol{\psi}_0) \boldsymbol{\Sigma}_0 \boldsymbol{\Pi}(\boldsymbol{\beta_0}, \boldsymbol{\psi}_0)^T$ in Theorem 3 is mild. Condition (C5) controls the rate of correlation decay for products of random variables. Lin (2008) demonstrated that if Condition (C5) holds, it is easy to show that $\boldsymbol{\Pi}(\boldsymbol{\beta_0}, \boldsymbol{\psi}_0) \boldsymbol{\Sigma}_0 \boldsymbol{\Pi}(\boldsymbol{\beta_0}, \boldsymbol{\psi}_0)^T$ converges to a positive definite matrix $\Xi(\boldsymbol{\beta}_0, \boldsymbol{\psi}_0)$, as $n \to \infty$. There are many processes for which

Condition (C5) holds.

When establishing the asymptotic property in Theorem 3, it is not necessary that the true correlation of $\boldsymbol{Y}$ is of the form $\{\sum_{k=1}^{K}\pi_k\boldsymbol{R}^{(k)}(\boldsymbol{\alpha}_k) : (\pi_k,\boldsymbol{\alpha}_k)\in\boldsymbol{\Psi}\}$. When Condition (C6) holds, which means that the true correlation matrix of $\boldsymbol{Y}$ is $\sum_{k=1}^{K}\pi_k^0\boldsymbol{R}^{(k)}(\boldsymbol{\alpha}_k^0)$, if $\pi_k^0>0$ for all $k\in\{1,\cdots,K\}$, we have $\boldsymbol{\Psi}_0=\{({\boldsymbol{\alpha}_k^0}^T,\pi_k^0)_{k=1}^K\}$, otherwise $\boldsymbol{\Psi}_0=\{({\boldsymbol{\alpha}_k}^T,\pi_k^0)_{k=1}^K : \boldsymbol{\alpha}_k=\boldsymbol{\alpha}_k^0 \textit{ for } \pi_k^0>0, \textit{ and } \boldsymbol{\alpha}_k\in\boldsymbol{\Psi}_{\boldsymbol{\alpha}_k} \textit{ for } \pi_k^0=0\}$. Furthermore, when Condition (C6) holds, the working correlation matrix $\boldsymbol{R}(\boldsymbol{\psi}_0)$, $\boldsymbol{\psi}_0\in\boldsymbol{\Psi}_0$ is equal to the true correlation matrix. Consequently, the mix-GEE estimator $\widehat{\boldsymbol{\beta}}$ is asymptotically efficient, in the sense that the asymptotic variance of the mix-GEE estimator $\widehat{\boldsymbol{\beta}}$ is the smallest among the GEE estimators with all the possible working correlation matrices.

Remark 3. Our proposed approach assumes that the working correlations, represented by parameters $\boldsymbol{\alpha}$, are distinct from those in the marginal mean function, $\boldsymbol{\beta}$. However, in certain scenarios, the correlation and marginal mean functions might share commonalities, such as shared parameters, or be subject to specific constraints. For instance, (Huang and Pan, 2021) developed a GEE method for jointly modeling the mean and within-subject correlation in longitudinal binary data, ensuring that the correlation ad-

heres to upper bound constraints. It is possible to extend the joint modeling approach to spatially dependent data to incorporate commonality or prior information. This could improve estimation efficiency and warrants further investigation.

4. Tapered mix-GEE method for analyzing large spatial data

For large spatial data set, it is computationally challenging to solve the equation (2.2) and minimize (2.5) due to the difficulty of inverting the large $n \times n$ spatial working correlation matrix. To reduce the computational cost, we propose a tapered estimation procedure. Kaufman et al. (2008) proposed the covariance tapering for minimizing the Gaussian likelihood with large spatial data. The main idea is to make the original correlation matrix sparse by multiplying it with a sparse correlation matrix elementwise so that efficient sparse matrix algorithms can be used to reduce the computational cost.

There are two commonly used tapering methods: the one-taper approximation and the two-taper approximation. Kaufman et al. (2008) pointed out that the two-taper approximation is computationally slower than the one-taper approximation. However, the two-taper approximation is unbi-

ased for the estimation of covariance parameters when the correlation structure is correctly specified, and the bias for the one-taper approximation is usually larger. Furrer et al. (2016) refined the one-taper approximation by letting the tapering range increase to infinity, which can guarantee consistency for the estimation of covariance parameters.

Let $\boldsymbol{T}(\gamma)$ be a $n \times n$ tapering matrix, a sparse correlation matrix with elements equal zero if the distance between two locations exceeds a prespecified tapering range γ. For any two $n \times n$ matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, let $\boldsymbol{A} \circ \boldsymbol{B}$ define the Schur product, that is, element-wise matrix product. We propose the tapering version of (2.5) as

$$\boldsymbol{D}(\boldsymbol{\beta})^T \boldsymbol{A}(\boldsymbol{\beta})^{-1/2} \{\boldsymbol{R}(\boldsymbol{\psi}) \circ \boldsymbol{T}(\gamma_{1n})\}^{-1} \boldsymbol{A}(\boldsymbol{\beta})^{-1/2} \{\boldsymbol{Y} - \boldsymbol{\mu}(\boldsymbol{\beta})\} = 0, \quad (4.9)$$

where $\boldsymbol{R}(\boldsymbol{\psi}) = \sum_{k=1}^{K} \pi_k \boldsymbol{R}^{(k)}(\boldsymbol{\alpha}_k)$. In (4.9), the mixed working correlation matrix is tapered by $\boldsymbol{T}(\gamma_{1n})$ with the range parameter γ_{1n}. This tapering version is inspired by the refined one-taper approximation in Furrer et al. (2016), and through letting $\gamma_{1n} \to \infty$, we can obtain an efficient regression estimator under certain conditions (see Theorem 4). To estimate the nuisance parameter $\boldsymbol{\psi}$, we propose the tapering version of (2.5) as

$$\ell_n(\boldsymbol{\psi}, \boldsymbol{\beta}; \gamma_2) = \frac{1}{n} \log \left\{ \left| \boldsymbol{R}(\boldsymbol{\psi}) \circ \boldsymbol{T}(\gamma_2) \right| \right\} + \frac{1}{n} \boldsymbol{\epsilon}(\boldsymbol{\beta})^T \left[\{\boldsymbol{R}(\boldsymbol{\psi}) \circ \boldsymbol{T}(\gamma_2)\}^{-1} \circ \boldsymbol{T}(\gamma_2) \right] \boldsymbol{\epsilon}(\boldsymbol{\beta}), \quad (4.10)$$

where γ_2 is a range parameter of the tapering matrix $\boldsymbol{T}(\cdot)$. The tapering

version (4.10) is called two-taper approximation in Kaufman et al. (2008). When the working correlation structure $\boldsymbol{R}(\boldsymbol{\psi})$ is correctly specified, minimizing the two-taper approximation (4.10), w.r.t., $\boldsymbol{\psi}$, is equivalent to solve an unbiased estimating equation (Kaufman et al., 2008), so that the estimator of $\boldsymbol{\psi}$ from the two-taper approximation is still expected to be consistent.

There exist several commonly used tapering functions, such as Wendland$_1$ and Bohman; see Stein (2013). In this paper, we adopt the Wendland$_1$ tapering function $\boldsymbol{T}(\gamma)$ whose ij-th element is defined as

$$\boldsymbol{T}_{ij}(\gamma) = \left(1 - \frac{||\boldsymbol{s}_i - \boldsymbol{s}_j||}{\gamma}\right)^4 \left(1 + \frac{4||\boldsymbol{s}_i - \boldsymbol{s}_j||}{\gamma}\right) I(||\boldsymbol{s}_i - \boldsymbol{s}_j|| \leq \gamma).$$

The parameter γ controls the degree of approximation, with smaller values indicating more severe tapering. In the tapered mix-GEE method, through letting $\gamma_{1n} \to \infty$ in (4.9), we can pursue the estimation efficiency. Theoretically, there is no restriction on γ_{1n}, so we can take a small rate, i.e., severe tapering, to reduce the computational cost. Moreover, for the estimator of the nuisance parameter $\boldsymbol{\psi}$, we only need its consistency instead of efficiency. Thus we can fix γ_2 to largely reduce the computational burden. Under some regularity conditions, we can prove that the estimator of $\boldsymbol{\beta}$ under the tapering version is still efficient when the working correlation structure is correctly specified; see details in Theorem 4.

Denote $\widehat{\boldsymbol{\beta}}^{\mathcal{T}}$ and $\widehat{\boldsymbol{\psi}}^{\mathcal{T}}$ as the tapered mix-GEE estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\psi}$, obtained by using the similar iterative algorithm as in Section 2, with (2.5) and (2.6) replaced by (4.9) and (4.10), respectively. To establish the asymptotic properties of the tapered estimators, we assume the following additional conditions.

Condition 3. Tapering range condition: *Assume that* $\gamma_2 > 0$ *is a fixed constant, and* $\gamma_{1n} \to \infty$ *as* $n \to \infty$.

Condition 4. Identifiability condition under tapering*: Suppose that* $\mathrm{E}\{\ell_n(\boldsymbol{\psi}, \boldsymbol{\beta}_0; \gamma_2)\}$ *converges uniformly to* $\bar{\ell}(\boldsymbol{\psi}, \boldsymbol{\beta}_0; \gamma_2)$ *over* $\boldsymbol{\Psi}$. *Let* $\widetilde{\boldsymbol{\Psi}}_0 = \left\{\boldsymbol{\psi}_0 \in \boldsymbol{\Psi} : \bar{\ell}(\boldsymbol{\psi}_0, \boldsymbol{\beta}_0; \gamma_2) = \inf_{\boldsymbol{\psi} \in \boldsymbol{\Psi}} \bar{\ell}(\boldsymbol{\psi}, \boldsymbol{\beta}_0; \gamma_2)\right\}$. *For every* $\xi > 0$ *and* $\boldsymbol{\psi}_0 \in \widetilde{\boldsymbol{\Psi}}_0$, *assume that* $\inf_{\boldsymbol{\psi} \in \boldsymbol{\Psi}, d(\boldsymbol{\psi}, \widetilde{\boldsymbol{\Psi}}_0) \geq \xi} \bar{\ell}(\boldsymbol{\psi}, \boldsymbol{\beta}_0; \gamma_2) > \bar{\ell}(\boldsymbol{\psi}_0, \boldsymbol{\beta}_0; \gamma_2)$.

Condition 3 shows that there is no restriction on the rate of γ_{1n}, as long as it tends to infinity, which is consistent with the requirement in Furrer et al. (2016) for the refined one-taper approximation approach. In this paper, we simply take $\gamma_{1n} = \lfloor n^{2/5} \rfloor$, where $\lfloor \cdot \rfloor$ is the floor function. For the two-taper approximation, Kaufman et al. (2008) pointed out that even quite sparse tapered correlation matrix $\boldsymbol{R}(\boldsymbol{\psi}) \circ \boldsymbol{T}(\gamma_2)$ can produce decent results. Thus, we take the value of γ_2 corresponding to 4% sparse rate, defined as the proportion of nonzero elements in $\boldsymbol{R}(\boldsymbol{\psi}) \circ \boldsymbol{T}(\gamma_2)$.

Theorem 4. *Under the model assumption (2.1), Conditions 1-3 and (C1)-(C6), for every* $\zeta > 0$, *we have,* $P\left\{d(\widehat{\boldsymbol{\psi}}^{\mathcal{T}}, \widetilde{\boldsymbol{\Psi}}_0) \geq \zeta\right\} \to 0$, *as* $n \to \infty$. *Furthermore, suppose that the identifiability condition 2 holds for* $\widetilde{\boldsymbol{\Psi}}_0$, *and for* $\boldsymbol{\psi}_0 \in \widetilde{\boldsymbol{\Psi}}_0$, $\boldsymbol{\Pi}(\boldsymbol{\beta_0}, \boldsymbol{\psi}_0)\boldsymbol{\Sigma}_0\boldsymbol{\Pi}(\boldsymbol{\beta_0}, \boldsymbol{\psi}_0)^T$ *converges to a positive definite matrix* $\Xi(\boldsymbol{\beta}_0, \boldsymbol{\psi}_0)$, *then we have* $\sqrt{n}(\widehat{\boldsymbol{\beta}^{\mathcal{T}}} - \boldsymbol{\beta}) \xrightarrow{D} N\left\{0, \Xi(\boldsymbol{\beta}_0, \boldsymbol{\psi}_0)\right\}$, *as* $n \to \infty$.

Theorem 4 suggests that when Condition (C6) holds, the tapered mix-GEE estimator $\widehat{\boldsymbol{\beta}}^{\mathcal{T}}$ achieves the same asymptotic efficiency as the untapered mix-GEE estimator $\widehat{\boldsymbol{\beta}}$. Condition (C6) entails that the true correlation structure is a linear combination of the candidate correlation structures. Moreover, if Condition (C6) is satisfied with $\pi_k^0 \geq 0$ for $k = 1, \cdots, K$ and the sum $\sum_{k=1}^{K} \pi_k^0 = 1$, we have $\widetilde{\boldsymbol{\Psi}}_0$ equals $\boldsymbol{\Psi}_0$.

5. Simulation Study

We conduct a simulation study of both continuous and binary spatial data to assess the finite sample performance of the proposed methods. We consider $n = 225$ for studying the performance of the mix-GEE method and $n = 900$ for the tapered mix-GEE method. For each case, the simulation is repeated 200 times.

5.1 Assessment of the mix-GEE method

We compare the performance of the proposed mix-GEE method with four methods: (1) the GEE estimator assuming working independence (IND); (2) the GEE estimator assuming the anisotropic Matérn correlation structure (Matern); (3) the hybrid GEE method (HGEE Adegboye et al., 2018); and (4) the omniscient method (OMNI), which is the GEE estimator based on the true covariance matrix and serves as the benchmark. The Matern method is based on the anisotropic Matérn correlation structure, defined as

$$\rho_{\mathrm{MAT}}(||\boldsymbol{B}(\lambda,\psi)(\boldsymbol{s}_i-\boldsymbol{s}_j)||,\theta,\upsilon)=\frac{1}{\Gamma(\upsilon)}\left(\frac{\theta||\boldsymbol{B}(\lambda,\psi)(\boldsymbol{s}_i-\boldsymbol{s}_j)||}{2}\right)^{\upsilon}2K_{\upsilon}(\theta||\boldsymbol{B}(\lambda,\psi)(\boldsymbol{s}_i-\boldsymbol{s}_j)||),$$

where $\upsilon > 0$, $\theta > 0$ and $K_{\upsilon}(\cdot)$ is the modified Bessel function of the second kind of order υ, $\Gamma(\cdot)$ is the gamma function, and θ controls the strength of spatial dependence. IND and OMNI are implemented using their closed expression and the function `nleqslv` in the R package `nleqslv`. The Matern method is implemented using the function `likfit` in the R package `geoR`. The HGEE method is implemented using the program provided by the authors.

The locations are generated as $\boldsymbol{s} = \{r+U(-0.2,0.2), c+U(-0.2,0.2)\}$, where $r = 1,\cdots,15$, $c = 1,\cdots,15$,

and $U(-0.2, 0.2)$ is the uniform distribution in $[-0.2, 0.2]$. Thus, we have 225 irregularly spaced locations, denoted as $\boldsymbol{s}_i$, $i = 1, \cdots, 225$. For the continuous spatial data, we generate responses from $Y(\boldsymbol{s}_i) = \boldsymbol{X}(\boldsymbol{s}_i)^T\boldsymbol{\beta}_0 + \eta(\boldsymbol{s}_i)$, where $\boldsymbol{X}(\boldsymbol{s}_i) = (X_1(\boldsymbol{s}_i), X_2(\boldsymbol{s}_i))^T$, $X_1(\boldsymbol{s}_i), X_2(\boldsymbol{s}_i) \overset{i.i.d.}{\sim} N(0, 1)$, $\boldsymbol{\beta}_0 = (1, -1)$. We generate the binary spatial responses from the Probit regression model $Y(\boldsymbol{s}_i) = I\{\eta(\boldsymbol{s}_i) \leq \boldsymbol{X}(\boldsymbol{s}_i)^T\boldsymbol{\beta}_0\}$, where $X_1(\boldsymbol{s}_i), X_2(\boldsymbol{s}_i) \overset{i.i.d.}{\sim}$ Bernoulli(0.5) and $\boldsymbol{\beta}_0 = (0.2, -0.2)$. For both continuous and binary data, we consider four cases to generate $\eta(\boldsymbol{s_i})$ from the standard Gaussian random field with the following spatial correlation functions:

Case 1 the isotropic exponential correlation function $\rho_{\text{EX}}(||\boldsymbol{s}_i - \boldsymbol{s}_j||, \alpha)$ with $\alpha = -\ln(0.7)$, where $\rho_{\text{EX}}(||\boldsymbol{s}_i - \boldsymbol{s}_j||, \alpha) = e^{-\alpha||\boldsymbol{s}_i - \boldsymbol{s}_j||}$, $\alpha > 0$.

Case 2 the anisotropic exponential correlation function $\rho_{\text{EX}}(||\boldsymbol{B}(\lambda, \psi)(\boldsymbol{s}_i - \boldsymbol{s}_j)||, \alpha)$, with $\alpha = -\ln(0.7)$, stretching parameter $\lambda = 1/4$, and rotation parameter $\psi = \pi/2$.

Case 3 the anisotropic mixed correlation function formed by two anisotropic exponential correlation functions, i.e., $\pi_1\rho_{\text{EX}}(||\boldsymbol{B}_1(\lambda_1, \psi_1)(\boldsymbol{s}_i - \boldsymbol{s}_j)||, \alpha_1) + (1 - \pi_1)\rho_{\text{EX}}(||\boldsymbol{B}_2(\lambda_2, \psi_2)(\boldsymbol{s}_i - \boldsymbol{s}_j)||, \alpha_2)$, where $\pi_1 = 0.5$, $(\alpha_1, \lambda_1, \psi_1) = (-\ln(0.7), 1/5, 0)$ and $(\alpha_2, \lambda_2, \psi_2) = (-\ln(0.7), 1/5, \pi/2)$.

Case 4 the anisotropic mixed correlation function formed by an anisotropic exponential correlation function and an anisotropic spherical correlation function, i.e., $\pi_1\rho_{\rm EX}(||\boldsymbol{B}_1(\lambda_1,\psi_1)(\boldsymbol{s}_i-\boldsymbol{s}_j)||,\alpha_1)+(1-\pi_1)\rho_{\rm SP}(||\boldsymbol{B}_2(\lambda_2,\psi_2)(\boldsymbol{s}_i-\boldsymbol{s}_j)||,\alpha_2)$, where $\pi_1=0.5$, $(\alpha_1,\lambda_1,\psi_1)=(-\ln(0.7),1/5,\pi/2)$, $(\alpha_2,\lambda_2,\psi_2)=(5,1/5,0)$, and $\rho_{\rm SP}(||\boldsymbol{s}_i-\boldsymbol{s}_j||,\alpha)=(1-1.5||\boldsymbol{s}_i-\boldsymbol{s}_j||/\alpha+0.5||\boldsymbol{s}_i-\boldsymbol{s}_j||^3/\alpha^3)\,I(||\boldsymbol{s}_i-\boldsymbol{s}_j||<\alpha)$, $\alpha>0$.

For the mix-GEE method, we use the mixture of three candidate correlation structures given in Remark 1. Note that the correlation structures used in the mix-GEE are misspecified under Cases 2-4.

Tables 1 and 2 summarize the mean squared errors (MSE) of different methods for continuous and binary responses, respectively. The results show that the IND estimator has the lowest efficiency under all scenarios considered, suggesting that ignoring the spatial dependence could have a severe adverse effect on the estimation efficiency. The mix-GEE method performs similarly to the benchmark OMNI and generally has a smaller MSE than the other methods under all scenarios considered. Although both mix-GEE and HGEE can handle complex spatial dependence, simulation results show that the HGEE is more efficient than IND but less efficient than the mix-GEE method. There are two possible reasons. One is that the

estimated weight matrix in HGEE is not optimal; see discussions in Section 1. The other is that the HGEE method uses an empirical semivariogram to estimate the nuisance parameters, which could be unstable in finite samples.

Table 1: MSE$\times 10^3$ of different methods for continuous responses. The reported MSE is the average of all regression coefficients. Values in the parentheses are the standard errors.

Method	Case 1	Case 2	Case 3	Case 4
IND	6.50	5.63	5.98	6.47
	(0.48)	(0.39)	(0.42)	(0.47)
Matern	1.95	1.19	1.08	1.01
	(0.14)	(0.08)	(0.08)	(0.07)
HGEE	4.28	1.95	2.97	3.51
	(0.34)	(0.15)	(0.25)	(0.32)
mix-GEE	2.02	0.76	0.81	0.75
	(0.15)	(0.06)	(0.06)	(0.06)
OMNI	1.94	0.75	0.81	0.75
	(0.14)	(0.06)	(0.06)	(0.06)

mix-GEE: the proposed method; IND: the GEE estimator assuming working independence; Matern: the GEE estimator assuming the anisotropic Matérn correlation; HGEE: the hybrid GEE method from Adegboye et al. (2018); OMNI: the GEE estimator based on the true covariance matrix.

5.2 Assessment of the tapered mix-GEE method

In this subsection, we assess the finite sample performance of the proposed tapered mix-GEE (Tmix-GEE) method for analyzing large spatial data. We consider spatial data from 900 irregularly spaced locations $\boldsymbol{s} = \{r + U(-0.2, 0.2), c + U(-0.2, 0.2)\}$, where $r = 1, \cdots, 30$, and

Table 2: MSE×10 of different methods for binary responses. The reported MSE is the average of all regression coefficients. Values in the parentheses are the standard errors.

Method	Case 1	Case 2	Case 3	Case 4
IND	1.78	3.52	3.44	2.88
	(0.20)	(0.36)	(0.33)	(0.28)
Matern	0.37	0.35	0.38	0.34
	(0.04)	(0.04)	(0.03)	(0.03)
HGEE	0.97	2.82	2.61	2.56
	(0.29)	(0.74)	(0.65)	(0.67)
mix-GEE	0.31	0.25	0.29	0.29
	(0.02)	(0.02)	(0.02)	(0.02)
OMNI	0.31	0.24	0.28	0.29
	(0.02)	(0.02)	(0.02)	(0.02)

mix-GEE: the proposed method; IND: the GEE estimator assuming working independence; Matern: the GEE estimator assuming the anisotropic Matérn correlation; HGEE: the hybrid GEE method from Adegboye et al. (2018); OMNI: the GEE estimator based on the true covariance matrix.

$c = 1, \cdots, 30$. The responses are generated in the same way as that in Section 5.1. The mix-GEE and Matern methods are about five times slower than the Tmix-GEE method for analyzing the large spatial data generated in this subsection. Therefore, we only compare the results of Tmix-GEE, HGEE, and OMNI based on 200 simulation replicates.

Table 3 summarizes the estimation performance of the Tmix-GEE, HGEE, and OMNI methods and the average computing time of Tmix-GEE and HGEE. The computing time is averaged over 200 replicates and four different cases. From the computational aspect, HGEE is faster than

the Tmix-GEE, which is because HGEE estimates the nuisance parameters through the empirical semivariogram, which does not involve solving the inverse of a high dimensional matrix. However, the estimation efficiency of the HGEE estimator is lower than the Tmix-GEE estimator under all scenarios considered. The MSE of the Tmix-GEE method is closer to the benchmark method OMNI, indicating that the Tmix-GEE is an effective method to analyze large spatial data.

Table 3: The estimation performance of different methods for analyzing continuous and binary responses. The reported MSE is the average of all regression coefficients. The reported results are MSE$\times 10^4$ for continuous responses and MSE$\times 10^2$ for binary responses. Values in the parentheses are the standard errors.

Response	Method	Case 1	Case 2	Case 3	Case 4	Time/s
Continuous	Tmix-GEE	6.10	2.50	2.45	2.20	29.25
		(0.45)	(0.20)	(0.18)	(0.16)	
	HGEE	13.28	7.85	9.17	9.86	17.70
		(0.90)	(0.63)	(0.66)	(0.72)	
	OMNI	5.90	2.30	2.44	2.20	
		(0.43)	(0.18)	(0.17)	(0.16)	
Binary	Tmix-GEE	0.82	0.70	0.82	0.71	48.00
		(0.05)	(0.05)	(0.06)	(0.05)	
	HGEE	1.60	3.46	5.85	2.85	16.20
		(0.22)	(0.90)	(2.38)	(0.57)	
	OMNI	0.76	0.64	0.73	0.64	
		(0.05)	(0.04)	(0.05)	(0.05)	

Tmix-GEE: the tapering version of the mix-GEE; HGEE: the hybrid GEE method from Adegboye et al. (2018); OMNI: the GEE estimator based on the true covariance matrix.

6. Application to The Soil Chemistry Data

In this section, we apply the proposed mix-GEE method to analyze the soil chemistry data set, soild250, in the R package `geoR`. The data set was collected on a 25×10 regular grid, resulting in 250 observations on 22 variables measuring various soil chemistry properties. The locations are recorded in Linha (x-coordinate) and Coluna (y-coordinate). As in Lu et al. (2014), we consider eight variables, including Ca (calcium content), Mg (magnesium content), K (potassium content), Al (aluminum content), C (carbon content), N (nitrogen content), pHKCl (soil PH by KCl) and CEC (Cation exchange capacity). Zheng et al. (2010) analyzed the spatial spectral density for the CEC. We focus on studying the impacts of the soil contents of Ca, Mg, K, Al, C, N, and PHKCl on the CEC, where the CEC is an important soil characteristic that measures the soil's ability to hold nutrients, water, herbicides, and other soil amendments.

We assume the linear model:

$$\mathrm{E}\{Y(\boldsymbol{s}_i)|\boldsymbol{X}_i\} = \boldsymbol{X}(\boldsymbol{s}_i)^T\boldsymbol{\beta}, \quad i = 1, \ldots, 250,$$

where $Y(\boldsymbol{s}_i)$ denotes the CEC measurement at location $\boldsymbol{s}_i$, and $\boldsymbol{\beta}$ is an 8-dimensional vector representing the intercept and the coefficients of PHKCl, Ca, Mg, K, Al, C, and N, respectively.

Figure 1 plots the empirical semivariograms of the data against the distance calculated at four different directions, that is, north (0°), north-east (45°), east (90°), south-east (135°). Figure 1 shows that the empirical semivariograms at 45° and 135° are similar, but quite different from those at 0° and 90°, indicating the possibility of anisotropy. The anisotropy is further confirmed by the result of a nonparametric test for isotropy (Guan et al., 2004) with a p-value less than 0.0025. The potential anisotropy and the spatial dependence make it challenging to specify a proper correlation structure for the GEE-based analysis. The proposed mix-GEE approach would appeal to analyzing such data with complex spatial dependence.

Table 4 shows the results of coefficient estimation. Since the true spatial correlation structure is unknown, we calculate the standard errors of the estimates by IND, Matern, and mix-GEE from the sandwich formula with the potential correlation structure. The standard errors of the HGEE estimates are obtained by the resampling method suggested by Adegboye et al. (2018). The results from the mix-GEE method suggest that PHKCl has a significant negative effect. In contrast, the soil contents of Ca, Mg, and K have significant positive effects on CEC, and the effects of Al, C, and N are not significant. These findings are also supported by the results

in Lu et al. (2014). The CEC measures the volume of all the exchangeable cations (Rhoades, 1983; Ross and Ketterings, 1995), so the cations Ca, Mg, K, and Al are expected to have positive effects. Among these cations, Ca, Mg, and K are more likely to have significant impacts on CEC than Al since the former three cations widely exist in the soil while the content of Al is usually smaller; see Bear (1964). Among these four methods, the HGEE estimates have the largest standard errors, which makes the method miss the significance of Mg. Moreover, the estimates by IND and Matern show a significant negative effect of C, which is surprising since C only exists in the anions CO_3^{2-} and HCO_3^-, which are known not directly related to CEC (Rhoades, 1983).

Table 4: The estimated coefficients of different methods for analyzing the soil data.

Coefficient	IND	Matern	HGEE	mix-GEE
Intercept	15.77(1.51)	11.25(1.49)	15.79(2.23)	8.82(1.16)
PHKCl	-2.97(0.30)	-1.83(0.31)	-2.98(0.50)	-1.22(0.25)
Ca	1.61(0.12)	1.41(0.14)	1.61(0.15)	1.21(0.11)
Mg	1.27(0.47)	1.04(0.50)	1.27(0.80)	1.00(0.38)
K	1.16(0.38)	0.90(0.36)	1.16(0.46)	0.82(0.26)
Al	0.28(1.06)	1.28(1.04)	0.22(1.50)	1.18(0.79)
C	-0.96(0.33)	-0.62(0.29)	-0.96(0.48)	-0.13(0.22)
N	4.94(3.43)	1.10(3.07)	4.96(2.82)	-2.10(2.51)

Values in the parentheses are the standard errors.

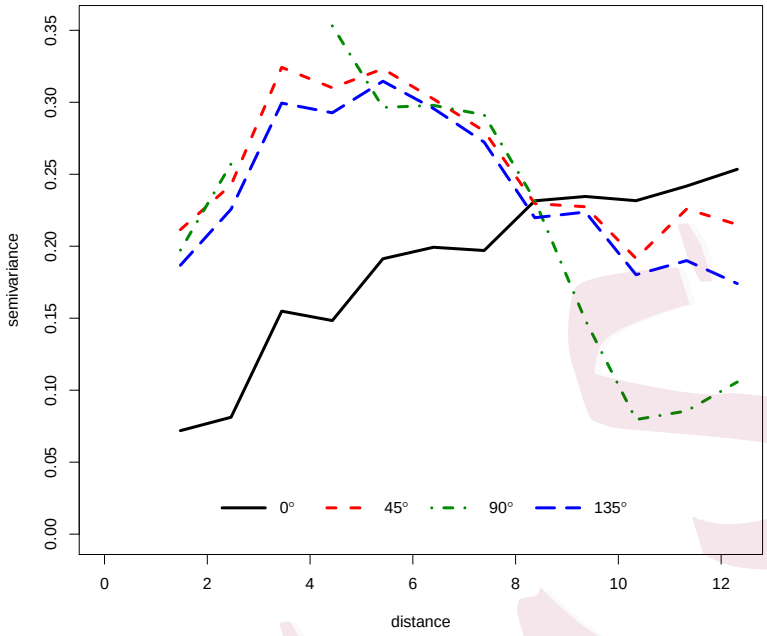

Figure 1: Empirical semivariograms against the distance calculated at four different directions of $0°, 45°, 90°, 135°$ for the soil data.

Supplementary Materials

The supplementary material contains the technical proofs of Theorems 1-4 in the main paper.

Acknowledgements

The research is partially supported by National Natural Science Foundation of China grants 12071087, 12226308, 12331009, General Research Fund

14308823, and Direct Grants for the Chinese University of Hong Kong 4053590.

References

Adegboye, O. A., D. H. Leung, and Y.-G. Wang (2018). Analysis of spatial data with a nested correlation structure. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 67*(2), 329–354.

Bachoc, F. et al. (2018). Asymptotic analysis of covariance parameter estimation for gaussian processes in the misspecified case. *Bernoulli 24*(2), 1531–1575.

Bear, F. E. (1964). Chemistry of the soil. *Soil Science 98*(1), 70.

Bevilacqua, M., C. Caamaño-Carrillo, R. B. Arellano-Valle, and V. Morales-Oñate (2021). Non-gaussian geostatistical modeling using (skew) t processes. *Scandinavian Journal of Statistics 48*(1), 212–245.

Budrikaite, L. and K. Ducinskas (2005). Modelling of geometric anisotropic spatial variation. *Mathematical Modelling and Analysis*, 361–366.

Cattelan, M. and C. Varin (2018). Marginal logistic regression for spatially clustered binary data. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 67*(4), 939–959.

Cressie, N. (2015). *Statistics for spatial data*. John Wiley & Sons.

Fang, F., J. Li, and J. Wang (2019). Optimal model averaging estimation for correlation structure in generalized estimating equations. *Communications in Statistics-Simulation and Computation 48*(5), 1574–1593.

Furrer, R., F. Bachoc, and J. Du (2016). Asymptotic properties of multivariate tapering for estimation and prediction. *Journal of Multivariate Analysis 149*, 177–191.

Guan, Y., M. Sherman, and J. A. Calvin (2004). A nonparametric test for spatial isotropy using subsampling. *Journal of the American Statistical Association 99*(467), 810–821.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the econometric society*, 1029–1054.

Hardin, J. W. and J. M. Hilbe (2012). *Generalized estimating equations*. chapman and hall/CRC.

Huang, Y. and J. Pan (2021). Joint generalized estimating equations for longitudinal binary data. *Computational Statistics & Data Analysis 155*, 107110.

Jenish, N. and I. R. Prucha (2009). Central limit theorems and uniform laws of large numbers for arrays of random fields. *Journal of econometrics 150*(1), 86–98.

Kaufman, C. G., M. J. Schervish, and D. W. Nychka (2008). Covariance tapering for likelihood-based estimation in large spatial data sets. *Journal of the American Statistical Association 103*(484), 1545–1555.

Lange, K. (1994). An adaptive barrier method for convex programming. *Methods and Applications of Analysis 1*(4), 392–402.

Lange, K. (2010). *Numerical analysis for statisticians*. Springer Science & Business Media.

Leung, D. H., Y.-G. Wang, and M. Zhu (2009). Efficient parameter estimation in longitudinal data analysis using a hybrid gee method. *Biostatistics 10*(3), 436–445.

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

Lin, P.-S. (2008). Estimating equations for spatially correlated data in multi-dimensional space. *Biometrika 95*(4), 847–858.

Lin, P.-S. (2010). A working estimating equation for spatial count data. *Journal of statistical planning and inference 140*(9), 2470–2477.

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

Lu, Z., X. Ren, and R. Zhang (2023). On semiparametrically dynamic functional-coefficient autoregressive spatio-temporal models with irregular location wide nonstationarity. *Journal*

of the American Statistical Association, 1–12.

Lu, Z., Q. Tang, L. Cheng, et al. (2014). Estimating spatial quantile regression with functional coefficients: A robust semiparametric framework. *Bernoulli 20*(1), 164–189.

McCullagh, P. (1983). Quasi-likelihood functions. *The Annals of Statistics 11*(1), 59–67.

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.

Palacios, M. B. and M. F. J. Steel (2006). Non-gaussian bayesian geostatistical modeling. *Journal of the American Statistical Association 101*(474), 604–618.

Rhoades, J. (1983). Cation exchange capacity. *Methods of Soil Analysis: Part 2 Chemical and Microbiological Properties 9*, 149–157.

Ross, D. S. and Q. Ketterings (1995). Recommended methods for determining soil cation exchange capacity. *Recommended soil testing procedures for the northeastern United States 493*, 62–69.

Shaby, B. and D. Ruppert (2012). Tapered covariance: Bayesian estimation and asymptotics. *Journal of Computational and Graphical Statistics 21*(2), 433–452.

Stein, M. L. (2013). Statistical properties of covariance tapers. *Journal of Computational and Graphical Statistics 22*(4), 866–885.

Sun, W., J. Shults, and M. Leonard (2009). A note on the use of unbiased estimating equations to estimate correlation in analysis of longitudinal trials. *Biometrical Journal: Journal of Mathematical Methods in Biosciences 51*(1), 5–18.

Sun, Y. and M. L. Stein (2016). Statistically and computationally efficient estimating equations for large spatial datasets. *Journal of Computational and Graphical Statistics 25*(1), 187–208.

Sutradhar, B. C. and K. Das (1999). Miscellanea. on the efficiency of regression estimators in generalised linear models for longitudinal data. *Biometrika 86*(2), 459–465.

Tang, N. and W. Wang (2019). Robust estimation of generalized estimating equations with finite mixture correlation matrices and missing covariates at random for longitudinal data. *Journal of Multivariate Analysis 173*, 640–655.

Thurman, A. L., R. Fu, Y. Guan, and J. Zhu (2015). Regularized estimating equations for model selection of clustered spatial point processes. *Statistica Sinica*, 173–188.

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.

Wedderburn, R. W. (1974). Quasi-likelihood functions, generalized linear models, and the gauss—newton method. *Biometrika 61*(3), 439–447.

Wu, C.-t., M. L. Gumpertz, and D. D. Boos (2001). Comparison of gee, minque, ml, and reml estimating equations for normally distributed data. *The American Statistician 55*(2), 125–130.

Xu, L., N. Lin, B. Zhang, and N.-Z. Shi (2012). A finite mixture model for working correlation matrices in generalized estimating equations. *Statistica Sinica*, 755–776.

Zhang, H. and A. El-Shaarawi (2010). On spatial skew-gaussian processes and applications. *Environmetrics: The official journal of the International Environmetrics Society 21*(1), 33–47.

Zheng, Y., J. Zhu, and A. Roy (2010). Nonparametric bayesian inference for the spectral density function of a random field. *Biometrika 97*(1), 238–245.

[1] Department of Statistics, The Chinese University of Hong Kong, China
E-mail: zhuhuichenecho@gmail.com

[2] Department of Statistics, Fudan University, China

[3] Department of Statistics, The George Washington University

[3] Department of Statistics, Fudan University, China