



Spatial Autologistic Model with Generalized Dependent Parameter

Liang Fang¹ · Zaiying Zhou² · Yiping Hong³ 

Received: 28 September 2022 / Revised: 3 December 2022 / Accepted: 19 November 2023

© School of Mathematical Sciences, University of Science and Technology of China and Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract

In the spatial autologistic model, the dependence parameter is often assumed to be a single value. To construct a spatial autologistic model with spatial heterogeneity, we introduce additional covariance in the dependence parameter, and the proposed model is suitable for the data with binary responses where the spatial dependency pattern varies with space. Both the maximum pseudo-likelihood (MPL) method for parameter estimation and the Bayesian information criterion (BIC) for model selection are provided. The exponential consistency between the maximum likelihood estimator and the maximum block independent likelihood estimator (MBILE) is proved for a particular case. Simulation results show that the MPL algorithm achieves satisfactory performance in most cases, and the BIC algorithm is more suitable for model selection. We illustrate the application of our proposed model by fitting the Bur Oak presence data within the driftless area in the midwestern USA.

Keywords Spatial autologistic model · Maximum pseudolikelihood · Dependence parameter · Spatial heterogeneity

✉ Yiping Hong
yiping.hong@bit.edu.cn

Liang Fang
fl2021@bjfu.edu.cn

Zaiying Zhou
zyzhou@mail.tsinghua.edu.cn

¹ Department of Statistics, School of Economics and Management, Beijing Forestry University, No. 35 Qinghua East Road, Haidian District, Beijing 100083, People's Republic of China

² Center for Statistical Science, Tsinghua University, No. 1, Qinghua Garden, Haidian District, Beijing 100084, People's Republic of China

³ School of Mathematics and Statistics, Beijing Institute of Technology, No. 5 Zhongguancun South Street, Haidian District, Beijing 100081, People's Republic of China

1 Introduction

Spatial data with discrete responses appear very frequently in research on the distribution of species [14, 16, 23], the transmission of diseases [5, 10, 21], and land-use change [22]. For instance, binary response data can be used to describe the occurrence of certain wildlife species (e.g., one if present and zero if absent) [23]. For binary data, the spatial autologistic covariance model is one of the most well-known [4, 13]. Besag [4] proposed the spatial autologistic covariance model by modeling the conditional probability of the spatial observations with respect to their neighbors. In this model, the log-odds of the conditional probability have a linear form relating to the neighboring observations, where the dependency is described by spatial dependence parameters. Caragea and Kaiser [8] found that when the spatial dependence parameters in Besag's model are large, the observed values will be very close to 1, which makes it difficult to interpret the practical meaning of the explanatory variable term in applications. To resolve this problem, Caragea and Kaiser [8] proposed the centered autologistic model by transforming the neighboring observations.

In a spatial autologistic model, the spatial dependence parameter η_{ij} describes the influence of Y_j on the conditional probability of Y_i . This parameter is often assumed to be invariant with locations, that is, $\eta_{ij} = \eta \cdot I\{i \sim j\}$, where $I\{\cdot\}$ is the indicator function, and $i \sim j$ means that the observation location s_i corresponding to Y_i is the neighborhood of the location s_j corresponding to Y_j . For extensions on the dependence parameter, Augustin et al. [1] assumed that η_{ij} is inversely proportional to the distance between these two locations. They fitted the data on the spatial distribution of wildlife by introducing an extra covariate. Bo et al. [5] used the same method to fit the record data on hand-foot-and-mouth disease in Mainland China. To fit a product recommendation model using the past purchase behavior data of customers, Moon and Russell [17] assumed that η_{ij} has an exponential distribution structure, e.g., $\eta_{ij} = \sigma^2 \exp(-\lambda \|s_i - s_j\|)$, where σ^2 and λ are unknown parameters. For anisotropy structure, Bardos [2] considered a spatial anisotropic neighboring structure, where η_{ij} can have two different values, relating to the horizontal and vertical autocorrelation, respectively, within the set of neighbors. Caragea and Berg [7] introduced the spatial anisotropy structure by allowing η_{ij} to be dependent on the direction of $s_i - s_j$. However, these extensions of η_{ij} are only dependent on $s_i - s_j$, not on separate s_i and s_j . In other words, the spatial autologistic models considered in these works are only those with spatial autocovariance structures invariant with spatial observation locations. Bo et al. [5] pointed out that this assumption of stationarity may not be suitable when the range of observations is extensive. For instance, let Y_i be a binary response to the occurrence of a specific disease. The covariate representing the speed of spread may be related to the climate and socio-economic factors. In a heavily polluted area, the disease may be more likely to spread due to residents' lack of immunity. A stationary spatial model may not fit the data when these factors vary in different locations.

When the observed values come from a continuous random variable, the data with spatial heterogeneity can be fitted by various non-stationary models. In the literature,

the kernel convolution model is intuitive, which can be considered a Matérn covariance model with spatially variant parameters [19, 20]. Parker et al. [20] constructed the non-stationary model by dividing the observation region into small subregions and assigning parameters for each subregion with various values. The spatial autoregressive model is also a popular non-stationary model, including simultaneous and conditional autoregressive models. These two models are suitable for the cases where the data are observed from certain predetermined regions, e.g., the administrative divisions [12]. The construction of the conditional autoregressive model is similar to that of the spatial autologistic model. Let Z_1, \dots, Z_n be continuous variables. Disregarding the covariates and the observation errors, the conditional distribution of Z_i satisfies

$$Z_i \mid \{Z_j, j \neq i\} \sim \mathcal{N} \left(\sum_{j \sim i} c_{ij} Z_j, m_{ii} \right), \quad (1.1)$$

where the notation $j \sim i$ is defined before, c_{ij} is the spatial autocovariance part, and m_{ii} is the variance part. Hoef et al. [12] pointed out that one can include a covariate term in c_{ij} , such as the ecological covariate and connectivity structure for each observation location, which helps make the model suitable for more delicate autocovariance structures. However, this inclusion is not popular in the literature. Hanks and Hooten [11] introduced the landscape characteristic covariate \mathbf{x}_i and assumed that when $i \sim j$,

$$c_{ij} = \exp \left\{ \frac{1}{\|\mathbf{s}_i - \mathbf{s}_j\|} \left(\frac{\mathbf{x}_i^\top + \mathbf{x}_j^\top}{2} \right) \boldsymbol{\beta} \right\}, \quad (1.2)$$

where $\boldsymbol{\beta}$ is the vector of the unknown parameter. For the simultaneous autoregressive model, Bera et al. [3] assumed that the element of the weight matrix \mathbf{W} comes from extra spatial data satisfying a linear model. However, to our best knowledge, no previous work provides a spatial model with binary observations capable of fitting data with spatial heterogeneity.

Inspired by the work stated above, such as Hanks and Hooten's work [11] for the continuous case, we propose in this article a non-stationary model for binary responses, the spatial autologistic covariance model with spatial heterogeneity by generalizing the autocorrelation parameter η_{ij} . The Gibbs sampling method for data generation, the maximum likelihood estimation (MLE), two kinds of maximum pseudolikelihood estimation (MPL) for parameter estimation, and the Akaike information criterion (AIC) and Bayesian information criterion (BIC) for parameter selection are all included. For the maximum block independent likelihood estimation (MBILE), an estimation similar to the MPL, we prove the exponential consistency for a particular case where the explanatory variable term is an unknown constant. It is shown from numerical simulations that our proposed model can characterize the binary response data with spatial heterogeneity well. The performance of the MPL estimator and the criteria of variable selection for the simulation are also investigated. To illustrate the application of our proposed model, we also fit the Bur Oak presence dataset in the driftless area of the midwestern USA. Results show that, compared to Caragea and Kaiser's [8] model,

our proposed model is more capable of describing the spatial varying autocorrelation relationships and can provide better prediction results for this dataset.

The rest of the article is organized as follows. In Sect. 2, we define the spatial autologistic model with spatial heterogeneity. In Sect. 3, we give the methods of data generation, estimation, and variable selection for our proposed model. In Sect. 4, we prove the exponential consistency of the MBILE estimation. In Sect. 5, we provide the simulation results. In Sect. 6, we provide the applications for fitting the Bur Oak presence dataset. Conclusions and discussions are given in Sect. 7.

2 Definition of Our Proposed Model

In this section, for the spatial dataset with binary responses, we propose our model by generating the spatial autologistic model of Besag [4] and Caragea and Kaiser [8]. In real applications, the spatial dataset may have spatially varying autocorrelation structures. For instance, consider a disease with infectivity in the epidemic area stronger than in the non-epidemic area. Then assume that the autocovariance within the epidemic area is stronger. However, the autocorrelation parameter in the spatial autologistic covariance model is usually irrelevant to the spatial locations [24]. We propose the spatial autologistic model with spatial heterogeneity by introducing the correlation covariate. We extend Caragea and Kaiser [8]'s model by allowing the autocorrelation parameter to be dependent on the spatial locations. The methods of data generation, estimation, and selection are also provided for our proposed model.

Consider the binary-valued response $Y_i = Y(s_i)$, $i \in \{1, \dots, n\}$ observed on $s_1, \dots, s_n \in \mathbb{R}^2$, where the value of Y_i can be zero or one. Besag [4] first proposed the spatial autologistic model, but the covariate may not reflect the mean value of the response. To tackle this problem, Caragea and Kaiser [8] proposed the following centered spatial autologistic model:

$$\text{logit}\{\Pr(Y_i = 1 \mid Y_j, j \neq i)\} = \mathbf{x}_i^\top \boldsymbol{\beta} + \eta \sum_{i \sim j} \left\{ Y_j - \frac{\exp(\mathbf{x}_j^\top \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_j^\top \boldsymbol{\beta})} \right\}, \quad (2.1)$$

where $\mathbf{x}_i^\top \boldsymbol{\beta}$ is the covariate term or the explanatory variable term, $\boldsymbol{\beta}$ is an unknown p -dimensional vector of covariate parameters, η is the spatial autologistic parameter, $i \sim j$ means that $i \in A_j$ and $j \in A_i$, where $A_i \subset \{1, \dots, n\}$ is the set of neighbors for the location s_i . This model assumes that the conditional distribution of Y_i with respect to the other observed values is only related to the responses observed in the locations within A_i ; in other words, $\Pr(Y_i = 1 \mid Y_j, j \neq i) = \Pr(Y_i = 1 \mid Y_j, j \in A_i)$. Thus A_i is called the set of neighbors. One can set that $A_i = \{s_j : 0 < \|s_i - s_j\| \leq h\}$, where h is the radius of the neighbor. For instance, when the data are observed on a regular grid, Besag [4] would set h as the distance between two locations, so there can be at most four locations for A_i . Such a setting is called the nearest neighborhood setting. Note that the set of neighbor A_i for location s_i does not include s_i itself.

As we have mentioned in Introduction, although this condition allows the spatial autologistic parameter to change with space, this parameter is usually assumed to be

a constant or has a simple form with respect to $s_i - s_j$ in the literature. To introduce the heterogeneity to this model, we allow that η is dependent on i and j and propose the following spatial autologistic model with spatial heterogeneity:

$$\text{logit}\{\Pr(Y_i = 1 \mid Y_j, j \neq i)\} = \mathbf{x}_i^\top \boldsymbol{\beta} + \sum_{i \sim j} \eta_{ij} \left\{ Y_j - \frac{\exp(\mathbf{x}_j^\top \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_j^\top \boldsymbol{\beta})} \right\}, \quad (2.2)$$

$$\eta_{ij} = \frac{1}{2} \boldsymbol{\gamma}^\top (\mathbf{z}_i + \mathbf{z}_j) I\{i \sim j\}, \quad (2.3)$$

where $\mathbf{x}_i^\top \boldsymbol{\beta}$ is the covariate term, $i \sim j$ means that $i \in A_j$ and $j \in A_i$, A_i is the set of neighbors for location s_i , \mathbf{z}_i is a predetermined q -dimensional vector, called the correlation covariate, $\boldsymbol{\gamma}$ is an unknown q -dimensional vector of parameters, called the intrinsic correlation parameter. Here \mathbf{z}_i describes how the autocovariance term η_{ij} relates to space. Besag [4] proved that the joint distribution of (Y_1, \dots, Y_n) exists if and only if:

- $i \in A_j$ if and only if $j \in A_i$;
- $\eta_{ij} = \eta_{ji}$, and $\eta_{ij} = 0$ when $i \notin A_j$.

For the simulation, we consider the case of $q = 2$, where $\mathbf{z}_i = (1, z_i)^\top$. Now every location corresponds to a one-dimensional correlation covariate $z_i = z(s_i) \in \mathbb{R}$. For instance, if the response means the occurrence of a certain disease, then z_i can represent whether the location lies within the epidemic area; if the response means the existence of a certain wild animal, then z_i can represent the land type, such as whether the observation located in woodland. In this case,

$$\eta_{ij} = \{\gamma_0 + \gamma_1(z_i + z_j)/2\} \cdot I\{i \sim j\}, \quad (2.4)$$

where $\boldsymbol{\gamma} = (\gamma_0, \gamma_1)^\top$. In this setting, γ_1 represents the difference in autocorrelation in different spatial locations.

In applications, one can choose the correlation covariate \mathbf{z}_i by application backgrounds and the information of the dataset, such as the observation locations and the covariate \mathbf{x}_i . For instance, in Sect. 6, we will fit the Bur Oak presence data using our proposed model, where \mathbf{z}_i includes the spatial coordinates and the White Oak presence information. More precisely, we will choose $\mathbf{z}_i = (1, s_{i,1}, s_{i,2}, W_i)$, where $(s_{i,1}, s_{i,2})$ is the coordinate of observation locations, and W_i is the White Oak presence data. To remove possibly redundant correlation covariate terms in \mathbf{z}_i , one can perform the model selection by computing the BIC criterion for models with different \mathbf{z}_i , which will be introduced in Sect. 3. The model with the smallest BIC value is the model selection result.

3 Data Generation, Estimation, and Model Selection Methods

Next, we will provide the methods of data generation, estimation, and model selection for our proposed spatial autologistic model.

Since model (2.2) is defined by conditional probability, we can use Gibbs sampling to get a sample of (Y_1, \dots, Y_n) . The algorithm can be stated as follows:

Input: Covariates $\mathbf{x}_i, i \in \{1, \dots, n\}$, correlation covariates $\mathbf{z}_i, i \in \{1, \dots, n\}$, parameters $\boldsymbol{\beta}, \boldsymbol{\gamma}$.

1. First generate a suitable initial sample $\mathbf{y}^{(0)} = (y_1^{(0)}, \dots, y_n^{(0)})$, where $y_i^{(0)} \in \{0, 1\}$, $i \in \{1, \dots, n\}$.
2. When $\mathbf{y}^{(i)} = (y_1^{(i)}, \dots, y_n^{(i)})$, $i \geq 0$ is generated, generate $\mathbf{y}^{(i+1)} = (y_1^{(i+1)}, \dots, y_n^{(i+1)})$ by the following steps:
 - Generate $y_1^{(i+1)} \sim \text{Binom}(1, p_1^{(i+1)})$, where $p_1^{(i+1)} = \Pr(Y_1 = 1 \mid Y_2 = y_2^{(i)}, \dots, Y_n = y_n^{(i)})$ is computed by (2.2);
 - Generate $y_2^{(i+1)} \sim \text{Binom}(1, p_2^{(i+1)})$, where $p_2^{(i+1)} = \Pr(Y_2 = 1 \mid Y_1 = y_1^{(i+1)}, Y_3 = y_3^{(i)}, \dots, Y_n = y_n^{(i)})$ is generated by (2.2);
 - ...
 - Generate $y_n^{(i+1)} \sim \text{Binom}(1, p_n^{(i+1)})$, where $p_n^{(i+1)} = \Pr(Y_n = 1 \mid Y_1 = y_1^{(i+1)}, \dots, Y_{n-1} = y_{n-1}^{(i+1)})$ is generated by (2.2).
3. Repeat step 2 until the number of generations i reaches a sufficiently large number M .

Output: A generated sample $\mathbf{y}^{(M)}$.

In the simulation, the initial sample $y_1^{(0)}, \dots, y_n^{(0)}$ is generated from a random sample of size n from Bernoulli distribution $\text{Binom}(1, 0.5)$, and we set $M = 1000$.

The unknown parameters of our proposed model can be estimated by the maximum pseudolikelihood (MPL) proposed by Hughes et al. [13]. This method approximates the logarithm of the probability mass function $\log\{\Pr(Y_1, \dots, Y_n)\}$ by the sum of the logarithm of the conditional probability mass function on each location, e.g.,

$$\log\{\Pr(Y_1, \dots, Y_n)\} \approx \sum_{i=1}^n \log\{\Pr(Y_i \mid Y_j, j \neq i)\}. \quad (3.1)$$

For spatial autologistic model (2.2), the log-pseudolikelihood function is

$$\begin{aligned} \ell(\boldsymbol{\theta} \mid Y_1, \dots, Y_n) &= \sum_{i=1}^n \left[Y_i \logit\{\Pr(Y_i \mid Y_j, j \neq i)\} \right. \\ &\quad \left. - \log\left[1 + \exp\left[\logit\{\Pr(Y_i \mid Y_j, j \neq i)\}\right]\right] \right] \\ &= \sum_{i=1}^n \left[Y_i \left[\mathbf{x}_i^\top \boldsymbol{\beta} + \sum_{i \sim j} \eta_{ij} \left\{ Y_j - \frac{\exp(\mathbf{x}_j^\top \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_j^\top \boldsymbol{\beta})} \right\} \right] \right. \\ &\quad \left. - \log \left[1 + \exp \left[\mathbf{x}_i^\top \boldsymbol{\beta} + \sum_{i \sim j} \eta_{ij} \left\{ Y_j - \frac{\exp(\mathbf{x}_j^\top \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_j^\top \boldsymbol{\beta})} \right\} \right] \right] \right]. \end{aligned} \quad (3.2)$$

The maximum point of this function is defined as the MPL estimate of the parameter. The MPL can be computed by iterative optimization algorithms, such as the Nelder-Mead algorithm [18]. In Sects. 5 and 6, we will adopt the R function `optim()` (version 4.1.1) for implementing the Nelder-Mead algorithm. The iteration stops when the improvement of one iteration is smaller than $\epsilon_r(|\ell| + \epsilon_r)$, where ℓ is the log-pseudolikelihood function value, and ϵ_r is the relative convergence tolerance, for which the default value is $\epsilon_r = 1.49 \times 10^{-8}$, or the square root of the machine epsilon in double precision.

We can use the following Akaike information criterion (AIC) or Bayesian information criterion (BIC) to deal with the variable selection problem for our proposed model:

$$\text{AIC} = -2\ell(\hat{\theta} \mid Y_1, \dots, Y_n) + 2p; \quad \text{BIC} = -2\ell(\hat{\theta} \mid Y_1, \dots, Y_n) + \log(n)p, \quad (3.3)$$

where ℓ is the log-pseudolikelihood function, $\hat{\theta}$ is the MPL estimate, p is the number of unknown parameters, and n is the number of observations. The selection procedure prefers the model with the smallest AIC or BIC value. In Sect. 5, we will show by numerical simulations that the BIC criterion is more suitable for variable selection.

For the variable selection procedure, we adopt the forward selection algorithm. Denote the unknown parameters of our proposed model by $\theta = (\beta, \gamma) = (\theta_1, \dots, \theta_{p+q})$ and the feature set $F \subset \{1, \dots, p + q\}$, which represents the indices of selected parameters. We define the model with feature F as our proposed model which θ_j is unknown if $j \in F$ and θ_j is fixed at zero if $j \notin F$. The algorithm is as follows:

1. Let $F = \emptyset$ and $\text{BIC}_F = +\infty$.
2. For i in $\{1, \dots, p + q\} \setminus F$:
 - Let $F_{\text{new}} = F$.
 - Fit the proposed model with feature $F \cup \{i\}$ and compute the corresponding BIC value $\text{BIC}_{F \cup \{i\}}$;
 - If $\text{BIC}_{F \cup \{i\}} < \text{BIC}_F$, then let $F_{\text{new}} = F \cup \{i\}$ and $\text{BIC}_F = \text{BIC}_{F \cup \{i\}}$.
3. If $F_{\text{new}} = F$ still holds or $F_{\text{new}} = \{1, \dots, p + q\}$, then stop the algorithm. The model with feature F_{new} is the final selected model. If not, let $F = F_{\text{new}}$ and continue running Step 2 until the algorithm stops.

4 Exponential Consistency of the MPL Estimator

With the results of Comets [9], for the spatial autologistic model without centering, the exponential consistency of the maximum likelihood estimation (MLE) and the maximum block independent likelihood estimation (MBILE) can be proved under a specific setting. Here the MBILE is an estimator similar to the MPL, which splits the observation region into several squares and ignores the correlation between different squares. We assume that the covariate term is a single parameter β_0 and the covariate z_i in (2.4) has periodicity.

Assume that the data are observed on $\mathbf{i} \in \mathbb{Z}^d$ and the neighborhood structure is the nearest neighborhood. We further assume that the observations Y_i follow

$$\text{logit}(\Pr\{Y_i = 1 \mid Y_j : \mathbf{j} \neq \mathbf{i}\}) = \beta_0 + \sum_{\mathbf{i} \sim \mathbf{j}} \eta_{ij} Y_j; \tag{4.1}$$

$$\eta_{ij} = \{\gamma_0 + \gamma_1(z_i + z_j)/2\} \cdot I\{\mathbf{i} \sim \mathbf{j}\}, \tag{4.2}$$

where $\boldsymbol{\theta} = (\beta_0, \gamma_0, \gamma_1)$ is the unknown parameter, $\mathbf{i} \sim \mathbf{j}$ means that the observed locations \mathbf{i} and \mathbf{j} are adjacent to the nearest neighborhood definition.

We consider the consistency under increasing domain regime, and the observation region can be divided into several squares. Without loss of generality, we set $d = 2$. In this case, for any $\mathbf{i} = (i_1, i_2) \in \mathbb{Z}^2$, we denote the square by

$$R(\mathbf{i}) = \{(K_1 i_1 + j_1, K_2 i_2 + j_2) \mid j_1 = 0, \dots, K_1 - 1, j_2 = 0, \dots, K_2 - 1\}, \tag{4.3}$$

and denote $R(\mathbf{0}) = R$. Note that the square is a regular lattice with $K_1 \times K_2$ locations. The observed value in each square is denoted by

$$\mathbf{Y}_i^{[R]} := Y_{R(\mathbf{i})} = \{Y_{(K_1 i_1 + j_1, K_2 i_2 + j_2)} \mid j_1 = 0, \dots, K_1 - 1, j_2 = 0, \dots, K_2 - 1\}, \tag{4.4}$$

and we denote $[\mathbf{Y}_i^{[R]}]_{(j_1, j_2)} = Y_{(K_1 i_1 + j_1, K_2 i_2 + j_2)}$. Similarly, we define

$$\mathbf{z}_i^{[R]} := z_{R(\mathbf{i})} = \{z_{(K_1 i_1 + j_1, K_2 i_2 + j_2)} \mid j_1 = 0, \dots, K_1 - 1, j_2 = 0, \dots, K_2 - 1\}. \tag{4.5}$$

When $\mathbf{z}_i^{[R]}$ is not dependent on \mathbf{i} for all \mathbf{i} , the correlation covariate has periodicity with respect to the squares. In this case, $\mathbf{Y}_i^{[R]}$ can be regarded as a stationary process with the observed values relying on $\mathcal{X} = \{0, 1\}^R$. Thus, we can prove the consistency using the results in Comets [9]. Note that we denote $\{0, 1\}^A$ as the set of mappings from A to $\{0, 1\}$ for a countable set A .

We set the observed region as $\Lambda_n = \cup_{\mathbf{i}=(i_1, i_2): |i_1| \leq n, |i_2| \leq n} R(\mathbf{i}) = [-K_1 n, (K_1 + 1)n - 1] \times [-K_2 n, (K_2 + 1)n - 1]$, where $n \in \mathbb{N}$, and denote \mathbf{Y}_n as the response variable in this region. We set the observed values outside this region by $\mathbf{y}_b \in \{0, 1\}^{\Lambda_n^c}$. In this case, the conditional log-likelihood function for the observed values is

$$l_{\Lambda_n, \mathbf{y}_b}(\boldsymbol{\theta}, \mathbf{Y}_n) = f_{\Lambda_n}(\boldsymbol{\theta}, \mathbf{Y}_n; \mathbf{y}_b) - \log \left[\sum_{\tilde{\mathbf{Y}}_n \in \{0, 1\}^{\Lambda_n}} \exp\{f_{\Lambda_n}(\boldsymbol{\theta}, \tilde{\mathbf{Y}}_n; \mathbf{y}_b)\} \right], \tag{4.6}$$

where

$$f_{\Lambda_n}(\boldsymbol{\theta}, \mathbf{Y}_n; \mathbf{y}_b) = \beta_0 \sum_{\mathbf{i} \in \Lambda_n} \tilde{Y}_i + \sum_{\mathbf{i} \sim \mathbf{j}, \{\mathbf{i}, \mathbf{j}\} \cap \Lambda_n \neq \emptyset} \eta_{ij} \tilde{Y}_i \tilde{Y}_j,$$

$\tilde{Y} = \{\tilde{Y}_i : i \in \mathbb{Z}^2\}$ is the vector consisting of Y_n and y_b , where the elements are

$$\tilde{Y}_i = \begin{cases} Y_i & i \in \Lambda_n, \\ [y_b]_i & i \notin \Lambda_n. \end{cases}$$

Thus, the value of $l_{\Lambda_n, y_b}(\theta, Y_n)$ is only related to θ, Y_n , and the value of y_b on $\partial\Lambda_n := \{i \in \Lambda_n^C \mid \exists j \in \Lambda_n, i \sim j\}$. Here $\partial\Lambda_n$ is called the boundary of Λ_n .

The maximum block independent likelihood estimation (MBILE) is proposed by Lim et al. [15]. The main idea is to approximate the conditional log-likelihood function $l_{\Lambda_n, y_b}(\theta, Y_n)$ by the sum of conditional log-likelihood functions on different blocks, e.g.,

$$pl_{\Lambda_n, y_b}(\theta, Y_n) = \sum_{i=(i_1, i_2): |i_1| \leq n, |i_2| \leq n} l_{R(i), y_b^{[R]}(i)}(\theta, Y_i^{[R]}),$$

where $y_b^{[R]}(i)$ is the vector of Y_n and y_b , restricted in the region $R(i)^C$. For the MLE and MBILE, we have the following result of exponential consistency.

Theorem 4.1 *Consider the spatial autologistic model with spatially dependent autocorrelation parameters defined by (4.1) and (4.2), where the autocovariance covariate z_i is periodic with respect to the square R , e.g., $z_i^{[R]}$ defined by (4.5) is not related to i . Assume that the model is identifiable, denote*

$$\hat{\theta}_{n, y_b} = \underset{\theta}{\operatorname{argmax}} l_{\Lambda_n, y_b}(\theta, Y_n), \quad \tilde{\theta}_{n, y_b} = \underset{\theta}{\operatorname{argmax}} pl_{\Lambda_n, y_b}(\theta, Y_n),$$

are the MLE and MBILE of the parameter θ , respectively. Then $\forall \epsilon > 0$, there exists $c > 0, \delta > 0$, which are not related to y_b , such that

$$\begin{aligned} Pr_{n, \theta_0} \left(\|\hat{\theta}_{n, y_b} - \theta_0\| > \epsilon \right) &\leq c \exp(-|\Lambda_n| \delta); \\ Pr_{n, \theta_0} \left(\|\tilde{\theta}_{n, y_b} - \theta_0\| > \epsilon \right) &\leq c \exp(-|\Lambda_n| \delta), \end{aligned}$$

where θ_0 is the true value of the parameter, Pr_{n, θ_0} is the probability distribution of Y_n , $|\Lambda_n|$ is the number of observations when the observation region is Λ_n .

Proof Let $\mathcal{X} = \{0, 1\}^R$, where $R = R(\mathbf{0})$ is defined by (4.3). For all $i = (i_1, i_2) \in \mathbb{Z}^2$, the observations on the square $R(i)$ are denoted by $Y_i^{[R]} \in \mathcal{X}$, see (4.4). For all $\Lambda \subset \mathbb{Z}^2$, denote by $Y_\Lambda^{[R]} = \{Y_i^{[R]} : i \in \Lambda\} \in \mathcal{X}^\Lambda$ the observations on the region $\cup_{i \in \Lambda} R(i)$, $y_b^{[R]} = \{Y_i^{[R]} : i \notin \Lambda\} \in \mathcal{X}^{\Lambda^C}$ the response value outside the region, and $Y^{[R]} \in \mathcal{X}^{\mathbb{Z}^2}$ the combinations of the observations $Y_\Lambda^{[R]}$ and $y_b^{[R]}$. Due to the periodicity of z_i , let

$$\begin{aligned} I_{\{i\}}^{(1)}(Y^{[R]}) &= \sum_{j \in R} [Y_i^{[R]}]_j, \\ I_{\{i\}}^{(2)}(Y^{[R]}) &= \sum_{j, \ell \in R} [Y_i^{[R]}]_j [Y_i^{[R]}]_\ell, \end{aligned}$$

$$\begin{aligned}
I_{\{i, i+(1,0)\}}^{(2)}(\mathbf{Y}^{[R]}) &= \sum_{j=0}^{K_2-1} \left[\mathbf{Y}_i^{[R]} \right]_{(K_1-1, j)} \left[\mathbf{Y}_{i+(1,0)}^{[R]} \right]_{(0, j)}, \\
I_{\{i, i+(0,1)\}}^{(2)}(\mathbf{Y}^{[R]}) &= \sum_{j=0}^{K_1-1} \left[\mathbf{Y}_i^{[R]} \right]_{(j, K_2-1)} \left[\mathbf{Y}_{i+(0,1)}^{[R]} \right]_{(j, 0)}, \\
I_{\{i\}}^{(3)}(\mathbf{Y}^{[R]}) &= \frac{1}{2} \sum_{j, \ell \in R} (z_j + z_\ell) \left[\mathbf{Y}_i^{[R]} \right]_j \left[\mathbf{Y}_i^{[R]} \right]_\ell, \\
I_{\{i, i+(1,0)\}}^{(3)}(\mathbf{Y}^{[R]}) &= \frac{1}{2} \sum_{j=0}^{K_2-1} (z_{(K_1-1, j)} + z_{(0, j)}) \left[\mathbf{Y}_i^{[R]} \right]_{(K_1-1, j)} \left[\mathbf{Y}_{i+(1,0)}^{[R]} \right]_{(0, j)}, \\
I_{\{i, i+(0,1)\}}^{(3)}(\mathbf{Y}^{[R]}) &= \frac{1}{2} \sum_{j=0}^{K_1-1} (z_{(j, K_2-1)} + z_{(j, 0)}) \left[\mathbf{Y}_i^{[R]} \right]_{(j, K_2-1)} \left[\mathbf{Y}_{i+(0,1)}^{[R]} \right]_{(j, 0)},
\end{aligned}$$

then the log-likelihood function $l_{\Lambda_n, \mathbf{y}_b}(\boldsymbol{\theta}, \mathbf{Y}_n)$ of the model satisfies (4.6), where

$$\begin{aligned}
f_{\Lambda_n}(\boldsymbol{\theta}, \mathbf{Y}_n; \mathbf{y}_b) &= \beta_0 \sum_{i \in I} I_{\{i\}}^{(1)}(\mathbf{Y}^{[R]}) \\
&\quad + \gamma_0 \left[\sum_{i \in I} I_{\{i\}}^{(2)}(\mathbf{Y}^{[R]}) + \sum_{i \in I \text{ or } i+(1,0) \in I} I_{\{i, i+(1,0)\}}^{(2)}(\mathbf{Y}^{[R]}) \right. \\
&\quad \left. + \sum_{i \in I \text{ or } i+(0,1) \in I} I_{\{i, i+(0,1)\}}^{(2)}(\mathbf{Y}^{[R]}) \right] \\
&\quad + \gamma_1 \left[\sum_{i \in I} I_{\{i\}}^{(3)}(\mathbf{Y}^{[R]}) + \sum_{i \in I \text{ or } i+(1,0) \in I} I_{\{i, i+(1,0)\}}^{(3)}(\mathbf{Y}^{[R]}) \right. \\
&\quad \left. + \sum_{i \in I \text{ or } i+(0,1) \in I} I_{\{i, i+(0,1)\}}^{(3)}(\mathbf{Y}^{[R]}) \right], \\
I &= \{i = (i_1, i_2) \in \mathbb{Z}^2 : |i_1| \leq n, |i_2| \leq n\}.
\end{aligned}$$

Next, we verify that our model meets the assumptions of Theorem 2.1 of Comets [9]. Define the following distance on \mathcal{X} :

$$d(\mathbf{a}, \mathbf{b}) = \begin{cases} 0, & \mathbf{a} \text{ is the same as } \mathbf{b}; \\ 1, & \text{otherwise,} \end{cases}$$

then \mathcal{X} and this distance constitute a Polish space. One can verify that $I_{\{i\}}^{(1)}$, $I_{\{i\}}^{(2)}$, $I_{\{i, i+(1,0)\}}^{(2)}$, $I_{\{i, i+(0,1)\}}^{(2)}$, $I_{\{i\}}^{(3)}$, $I_{\{i, i+(1,0)\}}^{(3)}$, $I_{\{i, i+(0,1)\}}^{(3)}$ are all bounded continuous functions on $\Omega = \mathcal{X}^{\mathbb{Z}^2}$, and $I_{\{i\}}^{(1)}$, $I_{\{i\}}^{(2)}$, $I_{\{i\}}^{(3)}$ are only dependent on $\mathbf{Y}_i^{[R]}$; $I_{\{i, i+(0,1)\}}^{(2)}$, $I_{\{i, i+(0,1)\}}^{(3)}$ are only dependent on $\mathbf{Y}_i^{[R]}$ and $\mathbf{Y}_{i+(0,1)}^{[R]}$; $I_{\{i, i+(1,0)\}}^{(2)}$, $I_{\{i, i+(1,0)\}}^{(3)}$ are only

dependent on $Y_i^{[R]}$ and $Y_{i+(1,0)}^{[R]}$. Moreover, define the shift operator

$$\tau_i : \Omega \rightarrow \Omega, (\tau_i Y^{[R]})_j = Y_{i+j}^{[R]},$$

then for the $I_V^{(l)}, l = 1, 2, 3$ defined before, one can verify the following properties of shifting invariance and additivity:

$$I_V^{(l)} \circ \tau_i = I_{i+V}^{(l)} \text{ for all } i, V, l;$$

$$\left[\sum_{m=1}^3 \left\{ \sum_{V:(0,0) \in V} \|I_V^{(l)}\|_\infty \right\} \right] < \infty.$$

Finally, the model is identifiable. Then the theorem is proved, since the assumptions of Theorem 2.1 in Comets [9] hold. □

5 Numerical Simulations

In this section, numerical simulations will be performed to compare the cases where the autocorrelation parameter in the spatial logistic model is independent or dependent on spatial locations. We will consider two settings of the autocorrelation parameter for the dependent cases, linearly dependent on spatial locations or related to a circular area. The properties of the MPL estimation and model selection methods will also be investigated.

5.1 The Case Where the Autocorrelation Parameter is Linearly Dependent

Here we set the autocorrelation parameter as being linearly dependent, e.g., the z_i 's in (2.4) have a linear trend with spatial location $s_i = (s_{i,1}, s_{i,2})^\top$ with locations $s_{i,1}, s_{i,2} \in \{1, \dots, 30\}$, and the number of observations is $n = 30^2 = 900$. The observed values follow the autocovariance model

$$\text{logit}\{ \Pr(Y_i = 1 \mid Y_j, j \neq i) \} = \beta_0 + \beta_1 x_i + \sum_{i \sim j} \eta_{ij} \{Y_j - \mu_j\}, \tag{5.1}$$

where

$$\mu_j = \frac{\exp(\beta_0 + \beta_1 x_j)}{1 + \exp(\beta_0 + \beta_1 x_j)}, \tag{5.2}$$

and the neighborhood structure is the nearest-neighbor scheme. We generate each x_i from independent samples assuming the distribution $\mathcal{N}(s_{i,1} + s_{i,2} + 4, 1), i \in \{1, 2, \dots, n\}$ and choose $\beta_0 = -2, \beta_1 = 0.1$. The values of the covariate x_i 's and the approximated mean μ_i 's are shown in Fig. 1.

In this simulation, we have the following two choices of η_{ij} :

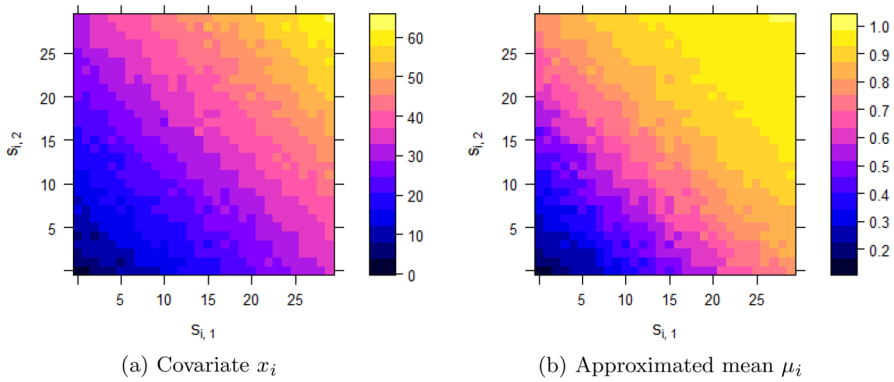
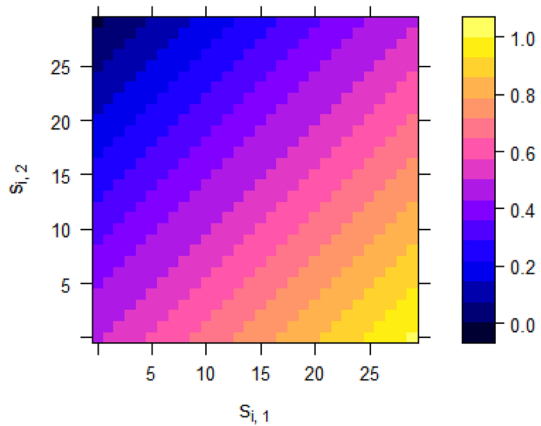


Fig. 1 Illustrations of the covariate x_i 's and the approximated mean μ_i 's of the observations in the case of linearly dependent autocorrelation parameter cases

Fig. 2 Illustration of the covariate z_i in (5.4) in the simulation of linearly dependent autocorrelation parameter cases



1. The autocorrelation parameter is independent of spatial locations:

$$\eta_{ij} = \gamma_0 \cdot I\{i \sim j\}, \tag{5.3}$$

where γ_0 can be 0.5, 1, 1.5.

2. The autocorrelation parameter is linearly dependent on spatial locations:

$$\eta_{ij} = \gamma_1(z_i + z_j)/2 \cdot I\{i \sim j\}, \tag{5.4}$$

where $z_i = (s_{i,1} - s_{i,2} + 29)/58$. Here, z_i has a linear relationship with spatial locations, ranging from 0 to 1. The values of z_i are shown in Fig. 2, indicating that this model has a smaller spatial autocovariance in the upper left corner and a larger autocovariance in the bottom right corner. We choose $\gamma_1 \in \{0.5, 1, 1.5\}$.

In summary, we consider six parameter combinations. For each combination, we use the MPL estimation to fit model (5.1), where the autocorrelation parameter can be (5.3), (5.4), or more generally, (2.4). The estimation properties are investigated in

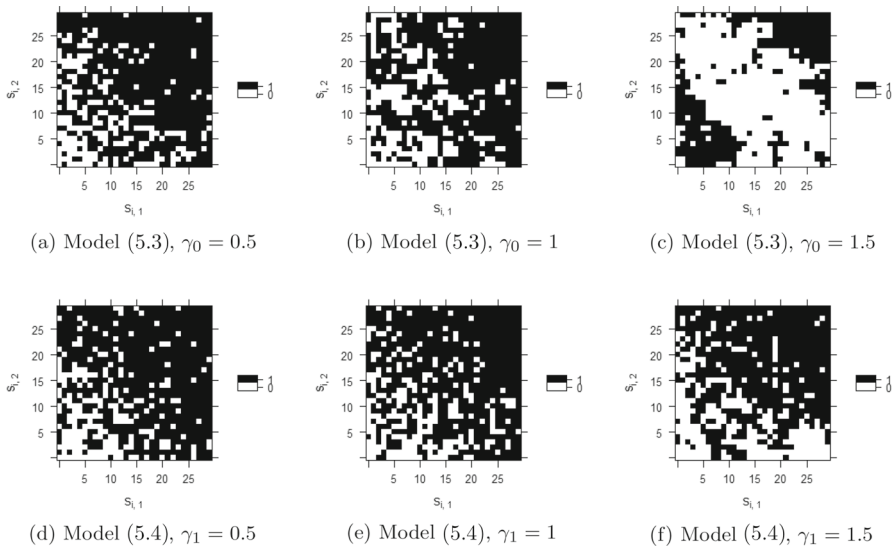


Fig. 3 Illustrations of a generated sample for each parameter combination in the simulation, where the autocorrelation parameter is linearly dependent with observation locations

terms of the sample mean and the standard deviation. For model selection, we compute the AIC and BIC for these models, select the model with the smallest criterion values, and explore their accuracy via the frequency of correct selections.

First, we show a randomly generated sample for each parameter combination, shown in Fig. 3. The first row is model (5.3), where the autocorrelation parameter does not change with space, with $\gamma_0 \in \{0.5, 1.0, 1.5\}$; the second row is model (5.4), where the autocorrelation parameter is linearly related to the locations, with $\gamma_1 \in \{0.5, 1.0, 1.5\}$. One can find that when the autocorrelation parameter does not change with space, the samples have different patterns for different γ_0 's. For instance, when $\gamma_0 = 0.5$, the white parts in the figure are small, but they become larger when $\gamma_0 = 1.0$ and $\gamma_0 = 1.5$. This indicates that the samples are smoother for larger γ_0 's. In these figures, for a sample with a larger autocorrelation parameter, the block of the same color is larger and has a smoother boundary. When the autocorrelation parameter is linearly related to the locations, one can inspect different smoothness only for $\gamma_1 = 1.5$. Figures of samples from different spatial logistic models with large autocorrelation parameters are distinct. However, if the autocorrelation parameter is small, it is difficult to tell whether the autocorrelation parameter is related to locations. In this case, one may need to adopt statistical methods such as model selection to inspect it.

Next, we compare the performance of the maximum pseudolikelihood (MPL) estimates, where the autocorrelation parameter can be: irrelevant to locations (5.3), linearly correlated with locations (5.4), or linearly correlated with locations with unknown intercept (2.4), respectively. Results are shown in Table 1. In Table 1, γ_0 and γ_1 in the true model are the values of γ_0 and γ_1 in models (5.3) and (5.4), respectively. The estimation results under the correct model are in bold. The number of replicates for the simulation is 500. Results show that, if the model is correctly specified, then the

Table 1 Sample mean and standard deviation (in brackets) of the MPL estimation in the simulation, where the autocorrelation parameters are linearly correlated with locations

True model	The autocorrelation is (5.3)			The autocorrelation is (5.4)		
	$\gamma_0 = 0.5$	$\gamma_0 = 1.0$	$\gamma_0 = 1.5$	$\gamma_1 = 0.5$	$\gamma_1 = 1.0$	$\gamma_1 = 1.5$
<i>Fit the model determined by (5.3)</i>						
β_0	-2.0156 (0.3302)	-1.9525 (0.7423)	-2.0825 (0.4722)	-2.0223 (0.2783)	-2.0488 (0.3625)	-2.2642 (0.5029)
β_1	0.1009 (0.0105)	0.0995 (0.0189)	0.1007 (0.0176)	0.1009 (0.0089)	0.1009 (0.0113)	0.1044 (0.0137)
γ_0	0.4961 (0.1345)	1.0092 (0.1074)	1.5100 (0.0620)	0.2280 (0.1387)	0.5081 (0.1374)	0.8734 (0.1099)
<i>Fit the model determined by (5.4)</i>						
β_0	-2.0130 (0.3282)	-2.2828 (0.4906)	-1.0928 (0.9090)	-2.0193 (0.2764)	-2.0349 (0.3489)	-2.0350 (0.4061)
β_1	0.1003 (0.0103)	0.1011 (0.0131)	0.0158 (0.0207)	0.1009 (0.0088)	0.1011 (0.0111)	0.1012 (0.0120)
γ_1	0.8476 (0.2519)	1.7357 (0.2199)	3.2147 (0.2938)	0.4704 (0.2522)	0.9899 (0.2353)	1.5005 (0.1560)
<i>Fit the more general model determined by (2.4)</i>						
β_0	-2.0187 (0.3319)	-2.0266 (0.7681)	-2.0964 (0.5668)	-2.0224 (0.2773)	-2.0345 (0.3545)	-2.0677 (0.5065)
β_1	0.1010 (0.0105)	0.1012 (0.0189)	0.1027 (0.0158)	0.1011 (0.0088)	0.1012 (0.0112)	0.1022 (0.0144)
γ_0	0.4812 (0.3523)	1.0180 (0.2501)	1.5028 (0.1237)	-0.0675 (0.3414)	-0.0421 (0.3588)	-0.0258 (0.3310)
γ_1	0.0278 (0.6463)	-0.0215 (0.4499)	0.0125 (0.2159)	0.5842 (0.6177)	1.0567 (0.6200)	1.5440 (0.4973)

Results when the model is correctly specified are in bold

estimates of β_0 , β_1 , and γ_0 have no apparent biases for the actual model determined by (5.3). When the autocorrelation increases, the standard deviations of the estimates for β_0 and β_1 first increase and then decrease, whereas the standard deviation of the estimates for γ_0 decreases. If the autocorrelation structure is misspecified as (5.4), then the estimates of β_0 and β_1 are biased when the autocorrelation is large. The estimates are unbiased if the autocorrelation structure is set as (2.4). However, the standard deviation of the estimates for γ_0 is significantly larger than that of the model determined by (5.3). For the true model determined by (5.4), the estimates of β_0 , β_1 , and γ_1 are also unbiased. When the autocorrelation increases, the estimates for β_0 and β_1 increase, whereas the standard deviation of the estimates for γ_1 decreases. If the autocorrelation structure is misspecified as (5.3), then the estimate of β_0 is biased when the autocorrelation is large. If the autocorrelation structure is set as (2.4), then the parameter estimates are also unbiased and have a larger variance than the correct

Table 2 Frequency of the model selection results by the AIC and BIC in the simulation, where the autocorrelation is linearly related to the locations

True model	Model selection results by the AIC			Model selection results by the BIC		
	Constant	Linear	General	Constant	Linear	General
$\gamma_0 = 0.5$	0.618	0.240	0.142	0.712	0.250	0.038
$\gamma_0 = 1.0$	0.798	0.008	0.194	0.960	0.024	0.016
$\gamma_0 = 1.5$	0.920	0.000	0.080	0.990	0.000	0.010
$\gamma_1 = 0.5$	0.284	0.524	0.192	0.304	0.668	0.028
$\gamma_1 = 1.0$	0.188	0.628	0.184	0.190	0.772	0.038
$\gamma_1 = 1.5$	0.010	0.702	0.288	0.052	0.908	0.040

The model selection frequencies corresponding to the correct results are shown in bold

model. In conclusion, the estimation behaves well when the model is correctly specified. If one adopts more general autocorrelation structure (2.4), the estimates may have a larger standard deviation. Therefore, removing redundant parameters in the model by a model selection procedure can be helpful.

The data generation and estimation are computationally efficient in our simulation settings. We adopt the 64-bit 20-core Intel Xeon Gold 6248 CPU running at 2.50 GHz for our computation. The time for generating one sample from the model considered in this section ranges from 12.62 to 19.84 s. The MPL estimation times for models (5.3) and (5.4) range from 1.21 to 5.09 s; and 1.21 to 5.75 s, respectively, whereas the MPL estimation time for model (2.4) ranges from 2.41 to 9.75 s. Therefore, removing redundant parameters can also cut the computational time.

Finally, we investigate the model selection performance of the AIC and BIC procedures defined by (3.3). For each parameter combination, we compute the AIC and BIC for the data, where the autocorrelation parameters in the model can be (5.3), (5.4), or (2.4). Results on the frequency of determined models are shown in Table 2. In this table, the "Constant," "Linear," and "General" mean that the autocorrelation parameter is irrelevant to locations (5.3), linearly correlated with locations (5.4), and linearly correlated with locations with unknown intercept (2.4). The frequency corresponding to the correct model is in bold. For the true model, γ_0 and γ_1 represent the true parameter values in the model determined by (5.3) and (5.4), respectively. Table 2 shows that the model selection works better when the autocorrelation parameter is larger. Moreover, the BIC procedure is more likely to choose the correct model than the AIC because the BIC procedure tends to select the model with fewer parameters. In conclusion, the AIC and BIC both perform well in model selection, where the BIC is recommended as the better choice.

5.2 The Case Where the Autocorrelation Parameter is Dependent with a Circle

Next, we consider the case where the autocorrelation parameter depends on a circular region. The covariance z_i in (2.4) is an indicator function of a circular region. In this case, if $\gamma_1 > 0$, then the autocovariance within the circle is stronger than that outside.

Table 3 List of models for the simulation where the autocorrelation parameter is dependent on a circular region

Number of models	$\text{logit}\{\Pr(Y_i = 1 \mid Y_j, j \neq i)\}$	η_{ij}
O1	$\beta_0 + \sum_{i \sim j} \eta_{ij} \{Y_j - \mu_j\}$	$\{\gamma_0 + \gamma_1(z_i + z_j)/2\} \cdot I\{i \sim j\}$
O2	$\beta_0 + \sum_{i \sim j} \eta_{ij} \{Y_j - \mu_j\}$	$\gamma_0 \cdot I\{i \sim j\}$
O3	$\beta_0 + \sum_{i \sim j} \eta_{ij} \{Y_j - \mu_j\}$	$\{\gamma_1(z_i + z_j)/2\} \cdot I\{i \sim j\}$
O4	$\beta_0 + \sum_{i \sim j} \eta_{ij} \{Y_j - \mu_j\}$	0
E1	$\beta_0 + \beta_1 z_i + \sum_{i \sim j} \eta_{ij} \{Y_j - \tilde{\mu}_j\}$	$\{\gamma_0 + \gamma_1(z_i + z_j)/2\} \cdot I\{i \sim j\}$
E2	$\beta_0 + \beta_1 z_i + \sum_{i \sim j} \eta_{ij} \{Y_j - \tilde{\mu}_j\}$	$\gamma_0 \cdot I\{i \sim j\}$
E3	$\beta_0 + \beta_1 z_i + \sum_{i \sim j} \eta_{ij} \{Y_j - \tilde{\mu}_j\}$	$\{\gamma_1(z_i + z_j)/2\} \cdot I\{i \sim j\}$
E4	$\beta_0 + \beta_1 z_i + \sum_{i \sim j} \eta_{ij} \{Y_j - \tilde{\mu}_j\}$	0

Assume that the observation region is $[0, 1]^2$, and the observation locations are $((k_1 - 0.5)/\sqrt{n}, (k_2 - 0.5)/\sqrt{n})^\top$, $k_1, k_2 \in \{1, \dots, \sqrt{n}\}$, which is denoted by $s_i = (s_{i,1}, s_{i,2})^\top$, $i \in \{1, \dots, n\}$. Choose the number of locations as $n \in \{30^2, 40^2, 50^2\}$, and the observations follow the following autologistic model:

$$\text{logit}\{\Pr(Y_i = 1 \mid Y_j, j \neq i)\} = \beta_0 + \sum_{i \sim j} \eta_{ij} \{Y_j - \mu_j\}, \quad (5.5)$$

where

$$\mu_j = \frac{\exp(\beta_0)}{1 + \exp(\beta_0)}, \quad (5.6)$$

with the neighborhood being the nearest neighborhood structure. The autocorrelation parameter in this model follows (2.3), where $z_i = I\{(s_{i,1} - 0.5)^2 + (s_{i,2} - 0.5)^2 \leq 0.3^2\}$. Thus, the autocorrelation parameters in the circle with center (0.5, 0.5) and radius 0.3 are larger than that outside the circle. Set $\beta_0 = \log(0.6/0.4) \approx 0.4055$, $\gamma_0 \in \{0.5, 1.0\}$, $\gamma_1 \in \{0, 0.5\}$, where $E\{Y_i\} \approx 0.6$. Note that when $\gamma_1 = 0$, the autocorrelation parameter is independent of space. We also take simulations for the case where $E\{Y_i\} \approx 0.4$ or $\beta_0 = \log(0.4/0.6) \approx -0.4055$. The results are omitted here due to the resemblance of $E\{Y_i\} \approx 0.6$.

Similar to the previous simulation, we first generate one sample and inspect the properties of the generated data. Then we perform maximum pseudolikelihood estimation to the models with different settings of autocorrelation parameters, using AIC and BIC (3.3) for model selection, and investigate the results. Consider eight models in Table 3, where the model for data generation can be model O1 (when $\gamma_1 = 0.5$) or model O2 ($\gamma_1 = 0$), $\mu_j = e^{\beta_0}/(1 + e^{\beta_0})$, $\tilde{\mu}_j = e^{\beta_0 + \beta_1 z_j}/(1 + e^{\beta_0 + \beta_1 z_j})$. Besides the correct settings, for the autocorrelation parameters in Table 3, we also consider the case where the parameter is zero if outside the circle and the case where the parameter is always zero. For the covariance term, we also consider the case where z_i is included. The number of replicates is 200 here in this simulation.

First, we inspect one generated data for each case where $n = 50^2$, $\beta_0 = \log(0.6/0.4)$, $\gamma_0 \in \{0.5, 1.0\}$, $\gamma_1 \in \{0, 0.5\}$. The results are shown in Fig. 4, where the

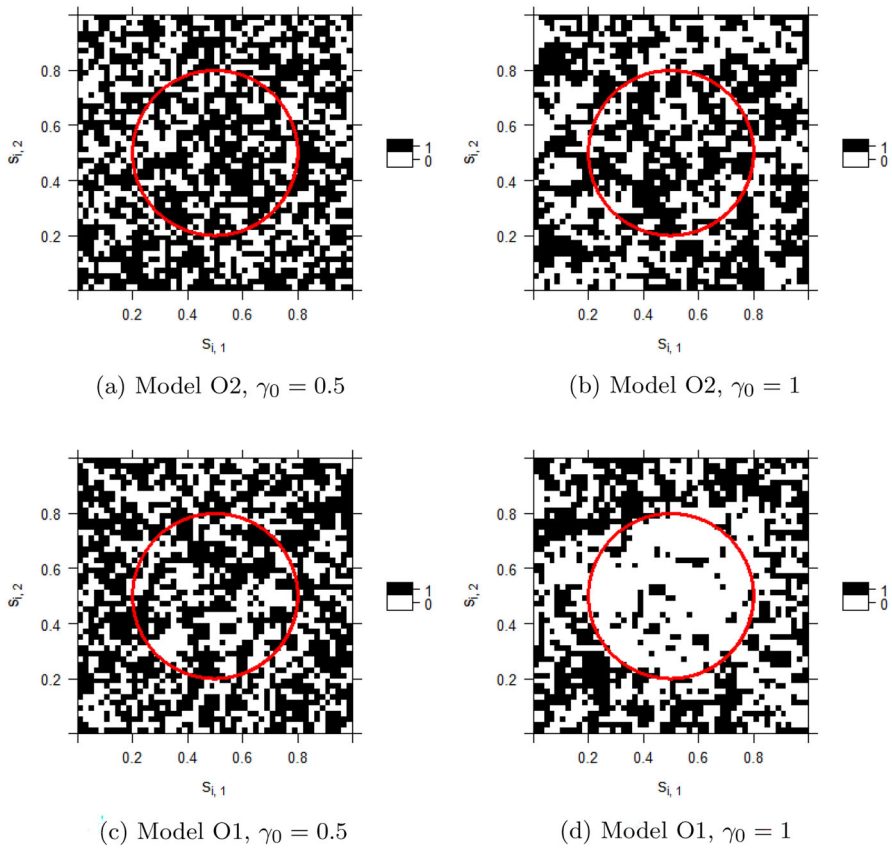


Fig. 4 Illustration of four generated data in simulations where the autocorrelation parameter is dependent on a circular region

first row represents the case with the autocorrelation parameter being independent of space (Model O2) and γ_0 being 0.5 or 1.0; the second row represents the case with the autocorrelation parameter being dependent on a circular region (Model O1), γ_0 being 0.5 or 1.0 and $\gamma_1 = 0.5$. Red curves show the circular region. Figure 4 shows that when the autocorrelation parameter is independent of the space, the white or black parts are more concentrated when the autocorrelation parameter is larger. When the autocorrelation parameter depends on a circular region, the patterns inside and outside the circle differ. The difference can be more drastic when γ_0 is large. Also, note that the data have more zero-valued observations inside the circle when $\gamma_0 = 1$. Thus we add an extra z_i , rendering Models E1–E4 in Table 3.

Next, we investigate the performance of the maximum pseudolikelihood estimation when $\beta_0 = \log(0.6/0.4)$. The results are shown in Tables 4 and 5, where the correct models are in bold. The bias of β_0 is calculated as the sample means of the estimates of β_0 minus its true value, and n is the number of observations. Table 4 shows that when the autocorrelation parameter is dependent on a circular region (the correct model

Table 4 Sample mean and standard deviation (in parentheses) of the MPL estimates for fitting Models O1 and O2 in the simulation when the autocorrelation parameter is dependent on a circular region

True model	n	Fitting model O1			Fitting model O2	
		Bias of β_0	γ_0	γ_1	Bias of β_0	γ_0
O1, $\gamma_0 = 0.5$	900	0.0088	0.5007	0.4727	-0.0268	0.6506
		(0.1255)	(0.1182)	(0.1901)	(0.1428)	(0.0967)
O1, $\gamma_0 = 0.5$	1600	-0.0037	0.5012	0.4938	-0.0556	0.6546
		(0.0794)	(0.0891)	(0.1444)	(0.0926)	(0.0680)
O1, $\gamma_0 = 0.5$	2500	-0.0029	0.5007	0.4852	-0.0536	0.6544
		(0.0747)	(0.0659)	(0.1047)	(0.0787)	(0.0558)
O1, $\gamma_0 = 1.0$	900	-0.0320	1.0235	0.5187	0.0636	1.1908
		(0.3680)	(0.1114)	(0.2175)	(0.6772)	(0.0867)
O1, $\gamma_0 = 1.0$	1600	0.0394	1.0079	0.5023	0.2139	1.1760
		(0.3268)	(0.0767)	(0.1782)	(0.4785)	(0.0543)
O1, $\gamma_0 = 1.0$	2500	-0.0113	0.9957	0.5265	0.1165	1.1721
		(0.2875)	(0.0648)	(0.1570)	(0.6095)	(0.0476)
O2, $\gamma_0 = 0.5$	900	0.0028	0.4976	0.0048	0.0045	0.4983
		(0.0974)	(0.1175)	(0.1984)	(0.0995)	(0.1036)
O2, $\gamma_0 = 0.5$	1600	0.0027	0.4991	0.0166	0.0039	0.5032
		(0.0786)	(0.0933)	(0.1599)	(0.0761)	(0.0818)
O2, $\gamma_0 = 0.5$	2500	-0.0089	0.4957	0.0098	-0.0074	0.4979
		(0.0614)	(0.0749)	(0.1323)	(0.0600)	(0.0616)
O2, $\gamma_0 = 1.0$	900	-0.1377	1.0063	0.0024	-0.1180	1.0068
		(0.4201)	(0.1099)	(0.1995)	(0.4563)	(0.0943)
O2, $\gamma_0 = 1.0$	1600	-0.1361	0.9990	0.0095	-0.0843	1.0016
		(0.3782)	(0.0789)	(0.1568)	(0.4152)	(0.0639)
O2, $\gamma_0 = 1.0$	2500	-0.1403	1.0052	-0.0085	-0.1054	1.0030
		(0.3532)	(0.0666)	(0.1251)	(0.3754)	(0.0569)

The results in bold mean that the fitted model is the correct one

Table 5 Sample mean and standard deviation (in parentheses) of the MPL estimates for fitting Models E1 and E2 in the simulation when the autocorrelation parameter is dependent on a circular region

True model	<i>n</i>	Fitting model E1			Fitting model E2		
		Bias of β_0	β_1	γ_0	Bias of β_0	β_1	γ_0
O1, $\gamma_0 = 0.5$	900	0.0085	-0.0876	0.5015	0.0232	-0.2266	0.6410
		(0.1255)	(0.5079)	(0.1181)	(0.1423)	(0.3208)	(0.0998)
O1, $\gamma_0 = 0.5$	1600	-0.0040	-0.0199	0.5018	0.0057	-0.2642	0.6457
		(0.0804)	(0.5112)	(0.0893)	(0.0941)	(0.2498)	(0.0698)
O1, $\gamma_0 = 0.5$	2500	-0.0031	-0.0674	0.5011	0.0177	-0.2894	0.6454
		(0.0745)	(0.3960)	(0.0659)	(0.0831)	(0.1916)	(0.0565)
O1, $\gamma_0 = 1.0$	900	-0.0202	0.1210	1.0262	-0.1191	-0.0831	1.1821
		(0.4463)	(0.6626)	(0.1122)	(0.7214)	(0.8911)	(0.1215)
O1, $\gamma_0 = 1.0$	1600	0.0560	0.0738	1.0098	-0.0084	-0.0813	1.1756
		(0.3671)	(0.5493)	(0.0768)	(0.7093)	(0.8337)	(0.0785)
O1, $\gamma_0 = 1.0$	2500	-0.0114	0.0761	0.9960	-0.0607	-0.0653	1.1719
		(0.3274)	(0.4308)	(0.0653)	(0.7324)	(0.8110)	(0.0776)

Table 5 continued

True model	n	Fitting model E1			Fitting model E2		
		Bias of β_0	γ_0	γ_1	Bias of β_0	β_1	γ_0
$O_2,$ $\gamma_0 = 0.5$	900	0.0049 (0.1160)	0.0064 (0.2268)	0.0040 (0.2043)	0.0034 (0.1167)	0.0020 (0.2008)	0.4968 (0.1039)
$O_2,$ $\gamma_0 = 0.5$	1600	0.0023 (0.0855)	0.0136 (0.1770)	0.0149 (0.1629)	0.0023 (0.0849)	0.0050 (0.1607)	0.5018 (0.0832)
$O_2,$ $\gamma_0 = 0.5$	2500	-0.0076 (0.0670)	0.0075 (0.1432)	0.0087 (0.1346)	-0.0082 (0.0662)	0.0020 (0.1295)	0.4970 (0.0619)
$O_2,$ $\gamma_0 = 1.0$	900	-0.1206 (0.4343)	-0.0616 (0.5610)	0.0081 (0.2019)	-0.1335 (0.4435)	-0.0590 (0.5574)	1.0114 (0.0935)
$O_2,$ $\gamma_0 = 1.0$	1600	-0.1209 (0.4110)	-0.0332 (0.5120)	0.0143 (0.1599)	-0.1071 (0.4016)	-0.0320 (0.4998)	1.0050 (0.0639)
$O_2,$ $\gamma_0 = 1.0$	2500	-0.1261 (0.3720)	-0.0435 (0.4773)	-0.0061 (0.1266)	-0.1150 (0.3655)	-0.0329 (0.4507)	1.0047 (0.0563)

is Model O1 in Table 3), the maximum pseudolikelihood estimation performs well, for it is unbiased with a small standard deviation for a large number of observations. When we fit the misspecified model (Model O2 in Table 3), β_0 has a larger bias, and γ_0 is overestimated. When the autocorrelation parameter is independent of the space (Model O2), the maximum pseudolikelihood estimation of γ_0 for the correct model works well. The estimation for β_0 in this case also performs well if $\gamma_0 = 0.5$ but can be underestimated if $\gamma_0 = 1$. When we fit Model O1, which has an excessive parameter, the estimation of β_0 and γ_0 is similar to that of Model O2, and the estimation of β_1 is close to zero with a small standard deviation for a large number of observations. Thus, if the model we fit is correct, the maximum pseudolikelihood estimation performs well unless for Model O2 and $\gamma_0 = 1.0$.

Then, we investigate the performance of the maximum pseudolikelihood estimation for the model with an excessive covariance term $\beta_1 z_i$. Table 5 shows that when the true model is Model O1. The fitted model is Model E1; the estimation results of β_0 , γ_0 and γ_1 are similar to the case of fitting the correct model, whereas the estimation results of β_1 are close to zero with a large variance. When the fitted model is Model E2, the estimation of β_0 is slightly larger than that of Model O2 when the autocorrelation parameter is small; and smaller when the autocorrelation parameter is large. The estimation of β_1 is likely to be smaller than zero, with a large standard deviation. The estimation of γ_0 behaves similarly to the case of fitting Model O2, with a large standard deviation when the autocorrelation parameter is large. When the true model is Model O2, and the fitted model is Model E2, the estimation results of β_0 and γ_0 are similar to the case of fitting the correct model. The mean value of the estimates of β_1 is close to zero, but the standard deviation is larger when $\gamma_0 = 1.0$. When fitting Model E1, the estimation results of β_0 , γ_0 , and γ_1 are similar to the case of fitting Model O1; the mean value of the estimates for β_1 is slightly smaller than zero, with the standard deviation larger when $\gamma_0 = 1.0$. Thus, if the covariance term has an excessive parameter, such as β_1 in Models E1 and E2, then the estimates of this excessive parameter may have a large variance, making the model selection necessary for fitting.

For the case when $\beta_0 = \log(0.4/0.6) \approx -0.4055$, the estimation results are similar to the case stated above, except for the true model being Model O2 and $\gamma_0 = 1.0$. In this latter case, β_0 is overestimated.

Note that the data generation and MPL estimation are also computationally efficient in the simulation in this section. The maximum generation time for the 50×50 non-stationary dataset is 52.82 s, and the maximum MPL estimation time for all cases is 31.85 s. The model with lesser parameters also has a shorter computational time.

Finally, we investigate how AIC and BIC (3.3) perform the model selection. The results are shown in Tables 6 and 7. Here n is the number of observations, and the correct selection results are in bold. Tables 6 and 7 show that if the correct model is Model O1, then the frequencies of correct selections are large for a large n . If the correct model is Model O2, then such frequencies are nearly invariant to n . Since the frequency of choosing the correct model via the BIC is larger than that of the AIC, we recommend the BIC for model selection. When $\beta_0 = \log(0.4/0.6) \approx -0.4055$, the AIC and BIC simulation results are similar to the case of $\beta_0 \approx 0.4055$.

In conclusion, simulation results show that the distribution of the generated data from our proposed model significantly differs from the standard autologistic model,

Table 6 The frequencies of model selection results obtained via the AIC in the simulation where the autocorrelation parameter is dependent on a circular region

True model	n	Frequency of correct results obtained via AIC							
		O1	E1	O2	E2	O3	E3	O4	E4
O1, $\gamma_0 = 0.5$	900	0.760	0.140	0.080	0.020	0.000	0.000	0.000	0.000
	1600	0.785	0.200	0.015	0.000	0.000	0.000	0.000	0.000
	2500	0.855	0.145	0.000	0.000	0.000	0.000	0.000	0.000
O1, $\gamma_0 = 1.0$	900	0.760	0.155	0.030	0.055	0.000	0.000	0.000	0.000
	1600	0.785	0.200	0.005	0.010	0.000	0.000	0.000	0.000
	2500	0.820	0.175	0.000	0.005	0.000	0.000	0.000	0.000
O2, $\gamma_0 = 0.5$	900	0.260	0.020	0.670	0.050	0.000	0.000	0.000	0.000
	1600	0.245	0.020	0.685	0.050	0.000	0.000	0.000	0.000
	2500	0.290	0.040	0.640	0.030	0.000	0.000	0.000	0.000
O2, $\gamma_0 = 1.0$	900	0.220	0.030	0.650	0.100	0.000	0.000	0.000	0.000
	1600	0.165	0.085	0.630	0.120	0.000	0.000	0.000	0.000
	2500	0.205	0.050	0.645	0.100	0.000	0.000	0.000	0.000

The results corresponding to the correct selections are in bold

Table 7 The frequencies of model selection results obtained via the BIC in the simulation where the autocorrelation parameter is dependent on a circular region

True model	n	Frequency of correct results obtained via BIC							
		O1	E1	O2	E2	O3	E3	O4	E4
O1, $\gamma_0 = 0.5$	900	0.550	0.005	0.425	0.010	0.010	0.000	0.000	0.000
	1600	0.860	0.015	0.125	0.000	0.000	0.000	0.000	0.000
	2500	0.990	0.000	0.010	0.000	0.000	0.000	0.000	0.000
O1, $\gamma_0 = 1.0$	900	0.670	0.015	0.275	0.040	0.000	0.000	0.000	0.000
	1600	0.905	0.000	0.085	0.010	0.000	0.000	0.000	0.000
	2500	0.990	0.000	0.005	0.005	0.000	0.000	0.000	0.000
O2, $\gamma_0 = 0.5$	900	0.030	0.000	0.970	0.000	0.000	0.000	0.000	0.000
	1600	0.045	0.000	0.950	0.005	0.000	0.000	0.000	0.000
	2500	0.045	0.000	0.955	0.000	0.000	0.000	0.000	0.000
O2, $\gamma_0 = 1.0$	900	0.015	0.000	0.975	0.010	0.000	0.000	0.000	0.000
	1600	0.030	0.000	0.945	0.025	0.000	0.000	0.000	0.000
	2500	0.015	0.005	0.980	0.000	0.000	0.000	0.000	0.000

The results corresponding to the correct selections are in bold

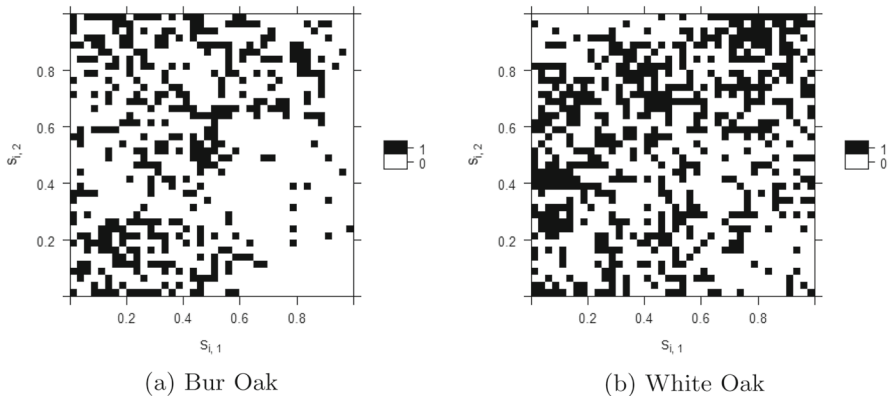


Fig. 5 Illustrations of the Bur Oak and White Oak species' existence data

where the autocorrelation parameter is a constant. Thus, for the data with the spatial autocorrelation being spatially dependent, we recommend fitting our proposed spatial autologistic model with generalized dependence parameters (2.2) and (2.3). The maximum pseudolikelihood estimation performs well for most cases in our simulation. For model selection, we recommend the BIC in (3.3).

6 Spatial Distribution of the Bur Oak Presence Within the Driftless Area

In this section, we fit the Bur Oak (*Quercus macrocarpa*) presence data within the Driftless Area of the Mississippi River, Midwestern USA, introduced by Caragea and Berg [7]. The data are collected from the Public Land Survey System (PLSS) dataset [6], which includes the forest competition information in an area of about 2500 square mile. In this section, we investigate the data within a 40×40 regular grid, where the area of one grid is one square mile. Figure 5 illustrates the observed data of two oak species, where black indicates presence and white indicates absence. We aim to fit the presence of the Bur Oak using the information of the White Oak (*Quercus alba*) and the locations.

6.1 Fitting the Model with Spatial Heterogeneity

Let Y_i and W_i be the presence data of the Bur Oak and the White Oak, respectively, observed on a regular grid within $[0, 1]^2$. More precisely, the observation locations are $((i_1 - 0.5)/40, (i_2 - 0.5)/40)$, $i_1, i_2 \in \{1, \dots, 40\}$, which are ordered lexicographically. Denote $s_i = (s_{i,1}, s_{i,2})$ as the observation location of Y_i . We will fit the following spatial autologistic model with spatial heterogeneity:

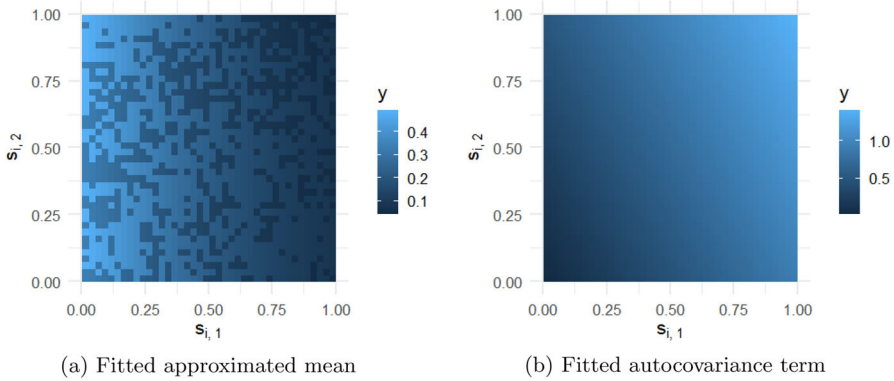


Fig. 6 Illustrations of the fitted approximated mean and the autocovariance term of the Bur Oak presence data

$$\text{logit}\{\Pr(Y_i = 1 \mid Y_j, j \neq i)\} = \mathbf{x}_j^\top \boldsymbol{\beta} + \sum_{i \sim j} \eta_{ij} \left\{ Y_j - \frac{\exp(\mathbf{x}_j^\top \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_j^\top \boldsymbol{\beta})} \right\}, \quad (6.1)$$

$$\eta_{ij} = \frac{1}{2} \boldsymbol{\gamma}^\top (\mathbf{z}_i + \mathbf{z}_j) I\{i \sim j\}, \quad (6.2)$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)$, $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \gamma_2, \gamma_3)$,

$$\mathbf{x}_i^\top \boldsymbol{\beta} = \beta_0 + \beta_1 s_{i,1} + \beta_2 s_{i,2} + \beta_3 W_i, \quad \boldsymbol{\gamma}^\top \mathbf{z}_i = \gamma_0 + \gamma_1 s_{i,1} + \gamma_2 s_{i,2} + \gamma_3 W_i,$$

W_i is the White Oak presence data, and $i \sim j$ means that $i \in A_j$ and $j \in A_i$, where A_i is the set of neighbors defined in Sect. 2, following the nearest neighborhood setting. The MPL estimation result of this model is $\hat{\boldsymbol{\beta}} = (-0.2092, -2.4866, 0.5768, -0.6805)$, $\hat{\boldsymbol{\gamma}} = (0.2119, 0.6474, 0.2294, 0.1838)$, and the computational time of this fitting is 57.45 s.

We also adopt the BIC criterion and the forward selection to reduce redundant parameters in our proposed autologistic model. The fitted model is also defined by (6.1) and (6.2), where

$$\mathbf{x}_i^\top \boldsymbol{\beta} = \beta_1 s_{i,1} + \beta_3 W_i, \quad \boldsymbol{\gamma}^\top \mathbf{z}_i = \gamma_1 s_{i,1} + \gamma_2 s_{i,2}.$$

The estimation values are $(\hat{\beta}_1, \hat{\beta}_3) = (-2.4448, -0.6518)$, $(\hat{\gamma}_1, \hat{\gamma}_2) = (0.9076, 0.5244)$. The computational time of the forward selection procedure is 109.34 seconds.

Figure 6 illustrates the fitted approximated mean $\hat{\beta}_1 s_{i,1} + \hat{\beta}_3 W_i$ and the fitted autocovariance $\hat{\gamma}_1 s_{i,1} + \hat{\gamma}_2 s_{i,2}$. Results show that the approximated mean of the Bur Oak presence is larger when $s_{i,1}$ is small, and the presence of the White Oak has a negative impact on the mean. The Bur Oak presence is more closely related to its neighborhood when $s_{i,1}$ is large.

Table 8 BIC and Brier skill scores of the fitted models for the Bur Oak presence data

Model type	BIC	Brier skill score
Proposed model (full)	1629.816	0.1314
Proposed model (BIC selected)	1603.503	0.1302
CK model (full)	1615.707	0.1252
CK model (BIC selected)	1609.795	0.1206

6.2 Comparison of Caragea and Kaiser’s Model

Next, we fit the data to Caragea and Kaiser’s [8] centered model (CK model) and compare the corresponding BIC value and the prediction performance, assessed by the Brier skill score. Let \hat{p}_i be the estimated conditional probability of $Y_i = 1$ with respect to its neighbors. The Brier score is defined by

$$BS = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{p}_i)^2,$$

whereas the Brier skill score is defined by comparing BS to the Brier score for a constant mean model, or more precisely,

$$BSS := 1 - BS / \{\bar{Y}(1 - \bar{Y})\},$$

where \bar{Y} is the mean value of Y_1, \dots, Y_n . A larger BSS indicates a better model prediction performance.

In this section, the fitted CK model is

$$\text{logit}\{\Pr(Y_i = 1 \mid Y_j, j \neq i)\} = \mathbf{x}_i^\top \boldsymbol{\beta} + \eta \sum_{i \sim j} \left\{ Y_j - \frac{\exp(\mathbf{x}_j^\top \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_j^\top \boldsymbol{\beta})} \right\},$$

where

$$\mathbf{x}_i^\top \boldsymbol{\beta} = \beta_0 + \beta_1 s_{i,1} + \beta_2 s_{i,2} + \beta_3 W_i,$$

and $i \sim j$ is defined similarly to model (6.1). The MPL estimate of this model is $\hat{\boldsymbol{\beta}} = (-0.3898, -2.5096, 1.1114, -0.6196)$ and $\hat{\eta} = 0.6778$, and the computational time is 21.04 s. We also use the BIC criterion and the forward selection procedure to reduce the number of parameters of this model. The reduced $\mathbf{x}_i^\top \boldsymbol{\beta}$ is

$$\mathbf{x}_i^\top \boldsymbol{\beta} = \beta_1 s_{i,1} + \beta_3 W_i,$$

where the parameter estimates are $(\hat{\beta}_1, \hat{\beta}_3) = (-2.1983, -0.5974)$, $\hat{\eta} = 0.6847$. The computational time of this selection procedure is 43.67 s.

Table 8 shows the BIC and the Brier skill score of different fitted models for the Bur Oak presence data, where “BIC selected” and “full” mean the model fitting results with and without the forward selection, respectively. Results show that our proposed model outperforms the CK model counterparts in the BIC and the Brier skill score,

indicating that our model has a better prediction performance compared to the CK model in this dataset.

In conclusion, our proposed spatial autologistic model is suitable for fitting the data in which the spatial dependency varies with space. Our model provides a better prediction performance in the sense of the Brier skill score compared to Caragea and Kaiser's [8] model for the Bur Oak dataset. The final fitting model of this dataset is model (6.1), where

$$\mathbf{x}_i^\top \boldsymbol{\beta} = -2.4448s_{i,1} - 0.6518W_i, \quad \boldsymbol{\gamma}^\top \mathbf{z}_i = 0.9076s_{i,1} + 0.5244s_{i,2}.$$

7 Conclusion and Discussion

In this article, we have generalized the centered spatial autologistic model by allowing the autocorrelation parameter to vary with space. This model can be used to fit the spatial binary dataset where the dependency varies with space. Simulation shows that whether the autocorrelation parameter varies with space can significantly affect the pattern of the generated spatial dataset. The maximum pseudolikelihood estimation performs well in estimating unknown parameters in most cases. Our suggested BIC can select the correct spatial autologistic models, verifying whether the autocorrelation parameter is dependent on space.

The results of stochastic simulations reveal the excellent properties of maximum pseudo-likelihood estimation. In the future, we could derive some asymptotic properties of the estimation, such as asymptotic normality, to provide further theoretical support for estimating spatial logistic models with spatial heterogeneity. Moreover, the maximum pseudolikelihood estimation may be biased when the autocorrelation parameter is large and independent of space, so an estimation method with a smaller bias may be called for. Future studies can also extend the model to fit observations with respect to time and space and spatial data with an unknown correlation structure of observed locations.

Acknowledgements This work is supported by the National Social Science Fund of China (No. 21BGL164) and partially by the National Natural Science Foundation of China (Nos. 12771286 and 11931001) and the Academic Initiation Program for Young Scholars at Beijing Institute of Technology (Grant No. XSQD-6120220282). Our authors greatly appreciate the sincere help of the journal's chief editors and editors, as well as the valuable comments and suggestions of the journal reviewers on many occasions, which have greatly improved our work.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Augustin, N.H., Muggleston, M.A., Buckland, S.T.: An autologistic model for the spatial distribution of wildlife. *J. Appl. Ecol.* **33**(2), 339–347 (1996)
2. Bardos, D.C., Guillera-Arroita, G., Wintle, B.A.: Valid auto-models for spatially autocorrelated occupancy and abundance data. *Methods Ecol. Evol.* **6**(10), 1137–1149 (2015)

3. Bera, A.K., Dogan, O., Taspinar, S.: Testing spatial dependence in spatial models with endogenous weights matrices. *SSRN Electron. J.* (2018)
4. Besag, J.E.: Spatial interaction and the statistical analysis of lattice systems. *J. R. Stat. Soc. Ser. B (Methodol.)* **36**(2), 192–236 (1974)
5. Bo, Y.-C., Song, C., Wang, J.-F., Li, X.-W.: Using an autologistic regression model to identify spatial risk factors and spatial risk patterns of hand, foot and mouth disease (HFMD) in Mainland China. *BMC Public Health* **14**, 358 (2014)
6. Bourdo, E.A.: A review of the general land office survey and of its use in quantitative studies of former forests. *Ecology* **37**(4), 754–768 (1956)
7. Caragea, P.C., Berg, E.: A centered bivariate spatial regression model for binary data with an application to presettlement vegetation data in the midwestern United States. *J. Agric. Biol. Environ. Stat.* **19**(4), 453–471 (2014)
8. Caragea, P.C., Kaiser, M.S.: Autologistic models with interpretable parameters. *J. Agric. Biol. Environ. Stat.* **14**(3), 281–300 (2009)
9. Comets, F.: On consistency of a class of estimators for exponential families of Markov random fields on the lattice. *Ann. Stat.* **20**(1), 455–468 (1992). <https://doi.org/10.1214/aos/1176348532>
10. Gumpertz, M.L., Graham, J.M., Ristaino, J.B.: Autologistic model of spatial pattern of phytophthora epidemic in bell pepper: Effects of soil variables on disease presence. *J. Agric. Biol. Environ. Stat.* **2**(2), 131–156 (1997). <https://doi.org/10.4148/2475-7772.1350>
11. Hanks, E.M., Hooten, M.B.: Circuit theory and model-based inference for landscape connectivity. *J. Am. Stat. Assoc.* **108**(501), 22–33 (2013)
12. Hoef, J.M.V., Peterson, E.E., Hooten, M.B., Hanks, E.M., Fortin, M.-J.: Spatial autoregressive models for statistical inference from ecological data. *Ecol. Monogr.* **88**(1), 36–59 (2018)
13. Hughes, J., Haran, M., Caragea, P.C.: Autologistic models for binary data on a lattice. *Environmetrics* **22**(7), 857–871 (2011)
14. Komori, O., Eguchi, S., Ikeda, S., Okamura, H., Ichinokawa, M., Nakayama, S.: An asymmetric logistic regression model for ecological data. *Methods Ecol. Evol.* **7**(2), 249–260 (2016)
15. Lim, J., Lee, K., Yu, D., Liu, H., Sherman, M.: Parameter estimation in the spatial auto-logistic model with working independent subblocks. *Comput. Stat. Data Anal.* **56**(12), 4421–4432 (2012)
16. Ma, Z., Zuckerberg, B., Porter, W.F., Zhang, L.: Spatial Poisson models for examining the influence of climate and land cover pattern on bird species richness. *For. Sci.* **58**(1), 61–74 (2012)
17. Moon, S., Russell, G.: Predicting product purchase from inferred customer similarity: an autologistic model approach. *Manag. Sci.* **54**(1), 71–82 (2008). <https://doi.org/10.1287/mnsc.1070.0760>
18. Nelder, J.A., Mead, R.: A simplex method for function minimization. *Comput. J.* **7**(4), 308–313 (1965)
19. Paciorek, C.J., Schervish, M.J.: Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics* **17**(5), 483–506 (2006)
20. Parker, R.J., Reich, B.J., Eidsvik, J.: A fused Lasso approach to nonstationary spatial covariance estimation. *J. Agric. Biol. Environ. Stat.* **21**(3), 569–587 (2016)
21. Shin, Y.E., Sang, H., Liu, D., Ferguson, T.A., Song, P.X.K.: Autologistic network model on binary data for disease progression study. *Biometrics* **75**(4), 1310–1320 (2019). <https://doi.org/10.1111/biom.13111>
22. Tepe, E., Guldmann, J.-M.: Spatio-temporal multinomial autologistic modeling of land-use change: a parcel-level approach. *Environ. Plan. B Urban Anal. City Sci.* **47**(3), 473–488 (2020). <https://doi.org/10.1177/2399808318786511>
23. Wang, Z., Zheng, Y.: Analysis of binary data via a centered spatial-temporal autologistic regression model. *Environ. Ecol. Stat.* **20**(1), 37–57 (2013)
24. Wolters, M.A.: Better autologistic regression. *Front. Appl. Math. Stat.* **3**, 24 (2017). <https://doi.org/10.3389/fams.2017.00024>

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.