Autologistic Regression Models for Spatio-Temporal Binary Data

Jun Zhu, Yanbing Zheng
Published online on: 21 Dec 2015
17

Autologistic Regression Models for Spatio-Temporal Binary Data

Jun Zhu and Yanbing Zheng

CONTENTS
17.1 Introduction ...................................................................................367
  17.1.1 Markov Random Field and Autologistic Model .........................368
  17.1.2 Spatio-Temporal Autologistic Model .........................................370
17.2 Spatio-Temporal Autologistic Regression Model ................................371
  17.2.1 Model ...................................................................................371
  17.2.2 Statistical Inference ................................................................372
    17.2.2.1 Maximum Pseudo-Likelihood Estimation .........................373
    17.2.2.2 Monte Carlo Maximum Likelihood Estimation .................373
  17.2.3 Bayesian Inference ................................................................374
  17.2.4 Prediction ..............................................................................375
17.3 Centered Autologistic Regression Model .................................................376
  17.3.1 Model with Centered Parameterization ....................................377
  17.3.2 Statistical Inference ................................................................378
    17.3.2.1 Expectation–Maximization Pseudo-Likelihood Estimator ...378
    17.3.2.2 Monte Carlo Expectation–Maximization Likelihood Estimator ...378
  17.3.3 Bayesian Inference .................................................................379
17.4 Data Example ..................................................................................379
17.5 Discussion ......................................................................................383
References ............................................................................................384

17.1 Introduction
Binary data on a spatial lattice are often encountered in environmental and ecological studies. Spatial statistical methods have been developed for modeling spatial binary responses and their relations to covariates while properly accounting for spatial correlation. In this chapter, we review autologistic models in the class of Markov random fields that model spatial dependence via autoregression and consider extensions to autologistic regression models for spatio-temporal binary data. In particular, the introduction of autoregression in space and time results in an unknown normalizing constant in the likelihood function, which makes estimation and statistical inference challenging. We describe...
several approaches to the inference for spatio-temporal autologistic regression models and illustrate them by an ecological data example.

17.1.1 Markov Random Field and Autologistic Model

For site \(i = 1, \ldots, n\), let \(Y_i\) denote the response variable at site \(i\). Besag (1974) developed a Markov random field model that conditionally specifies the distribution of \(Y_i\). Let \(Y = (Y_1, \ldots, Y_n)'\) and \(Y_{-i} = (Y_j : j \neq i)'\) denote the response variables with \(y = (y_1, \ldots, y_n)\) and \(y_{-i} = (y_j : j \neq i)'\) denoting a corresponding realization at all \(n\) sites and all except site \(i\), respectively. In a Markov random field model for \(Y\), the full conditional probability density of \(Y_i\) (conditional on all other sites) is assumed to depend only on the responses at neighboring sites; that is, \(p(y_i|y_{-i}) = p(y_i|y_j : j \in N_i)\), where \(N_i\) denotes a prespecified neighborhood of site \(i\). The conditional probability density is generally specified in an exponential form

\[
P(y_i|y_{-i}) = p(y_i|y_j : j \in N_i) = \exp\left\{A_i(y_{-i})y_i - B_i(y_{-i}) + C_i(y_i)\right\} \tag{17.1}
\]

where \(A_i\) is a natural parameter function, \(B_i\) is a function of the model parameters and \(y_{-i}\) but free of \(y_i\), and \(C_i\) is a function of \(y_i\) but free of the model parameters.

To ensure that the resulting joint distribution of \(Y\) is valid, Besag (1974) defined a negpotential function

\[
Q(y) = \ln \left\{ \frac{p(y)}{p(0)} \right\}, \tag{17.2}
\]

which is essentially the logarithm of the joint probability density function \(p(y)\) up to a normalizing constant since

\[
p(y) = \frac{\exp\{Q(y)\}}{\sum_{z \in \Omega} \exp\{Q(z)\}}, \tag{17.3}
\]

where \(\Omega\) denotes a suitable space of responses. It has been shown that \(Q(y)\) in (17.2) can be uniquely expanded on \(\Omega\) and the expansion is made up of the conditional probabilities \(p(y_i|y_{-i})\) in (17.1) under a positivity condition (Besag, 1974; Cressie, 1993). The Hammersley–Clifford Theorem and its corollary establish the sparsity of the expansion and most importantly, the validity of the joint probability \(p(y)\) through the negpotential function \(Q(y)\).

For binary data on a spatial lattice, Besag (1972) developed an autologistic model in the framework of Markov random fields. In particular, the binary response variable \(Y_i \in \{0, 1\}\) has a conditional Bernoulli distribution. The pairwise-only dependence is among neighboring sites according to the neighborhood \(N_i\). Thus, the natural parameter function is of the form

\[
A_i(y_{-i}) = \alpha_i + \sum_{j \in N_i} \theta_{ij}y_j, \tag{17.4}
\]
Autologistic Regression Models for Spatio-Temporal Binary Data

\[ B_i(y_{-i}) = \ln[1 + \exp(A_i(y_{-i}))], \] and \( C_i(y_i) = 0, \) where \( \alpha_i \) is a constant, \( \theta_{ij} \)s are spatial dependence parameters such that \( \theta_{ij} = \theta_{ji} \) for \( j \neq i \) and \( \theta_{ii} = 0 \) for site \( i = 1, \ldots, n. \) The corresponding negpotential function is

\[
Q(y) = \sum_{i=1}^{n} y_i \left( \alpha_i + (1/2) \sum_{j \in N_i} \theta_{ij} y_j \right). \tag{17.5}
\]

By the Hammersley–Clifford Theorem, the joint probability density is

\[
p(y) = \frac{\exp \left[ \sum_{i=1}^{n} y_i \left( \alpha_i + (1/2) \sum_{j \in N_i} \theta_{ij} y_j \right) \right]}{\sum_{z \in \Omega} \exp \left[ \sum_{i=1}^{n} z_i \left( \alpha_i + (1/2) \sum_{j \in N_i} \theta_{ij} z_j \right) \right]}.	ag{17.6}
\]

In (17.6), the normalizing constant in the denominator involves the model parameters and generally does not have an analytical form, which makes it a challenge to directly maximize the likelihood function.

The traditional parameterization of autologistic models may not be intuitive when incorporating regression. Similar to the parametrization used for auto-Gaussian models, a centered parameterization of autologistic models was proposed recently (Caragea and Kaiser, 2009; Kaiser et al., 2012), which is perhaps more suitable for regression purposes. In the centered parameterization,

\[
A_i(y_{-i}) = \ln\left( \kappa_i/(1 - \kappa_i) \right) + \sum_{j \in N_i} \theta_{ij}(y_j - \kappa_j),
\]

where \( \kappa_i \in (0,1), i = 1, \ldots, n. \) A detailed description of the centered parameterization of autologistic regression models is given in Section 17.3.

A special case of the autologistic model is the Ising model. The Ising model was first developed by Ernst Ising in his doctoral thesis as an attempt to describe phase transitions in ferromagnets (Ising, 1924, 1925). The basic idea is that microscopic magnets are arranged on a square lattice such that there is one magnet at each lattice site. Each magnet is assumed to have two possible spin directions, generally labeled as up \((y_i = +1)\) or down \((y_i = -1),\) and is assumed to only interact with its four nearest neighbors. In the Ising model, the total energy, also known as the Hamiltonian, of the configuration is given by

\[
H(y) = -\sum_{i=1}^{n} \sum_{j \in N_i, j < i} \theta_{ij} y_i y_j - \sum_{i=1}^{n} \alpha y_i, \tag{17.7}
\]

where \( N_i \) denotes the neighborhood of site \( i \) comprising the four nearest neighbors, the coefficient \( \theta \) represents the strength of interactions among the nearest neighbors, and the coefficient \( \alpha \) represents an external magnetic field. The cases \( \theta > 0 \) and \( \theta < 0 \) correspond to ferromagnetism and antiferromagnetism, respectively. The joint probability density of a configuration is given by the so-called Boltzmann factor

\[
Z^{-1}_\beta \exp\{-\beta H(y)\}, \tag{17.8}
\]

© 2016 by Taylor & Francis Group, LLC
where \( \beta = (k_B T)^{-1} \geq 0 \) with \( T \) denoting the Kelvin temperature and \( k_B \) denoting the Boltzmann constant, while \( Z_\beta = \sum_y \exp(-\beta H(y)) \) is a partition function (or, normalizing constant). When the parameter \( \beta \) in (17.8) surpasses a threshold value, a phase transition from short-range to long-range interactions would occur, resulting in an ordered phase with nonzero limiting correlation (see, e.g., Pickard, 1976, 1977).

17.1.2 Spatio-Temporal Autologistic Model

For spatio-temporal binary data, let \( Y_{i,t} \in \{ \pm 1 \} \) denote the binary response variable with \( y_{i,t} \) denoting a realization at site \( i = 1, \ldots, n \) and time \( t \). Let \( Y_t = (Y_{1,t}, \ldots, Y_{n,t})' \) denote the vector of binary responses with realizations \( y_t = (y_{1,t}, \ldots, y_{n,t})' \) at all sites and a given time point \( t \). Bartlett (1971, 1972) developed a Markov process with

\[
P(Y_{i,t+\Delta t} = y_{i,t}|y_t) = 1 - \lambda(\Delta t)\{1 - F_i(y_t)\},
\]

(17.9)

where \( \lambda \geq 0, \Delta t \geq 0, \) and \( F_i(y_t) \) is a function of \( y_t \). The joint probability density of \( Y_t \), when the Markov process is at equilibrium, is

\[
p(y_t) = c(\alpha, \theta) \exp \left( -\alpha \sum_{i=1}^n y_{i,t} - \theta \sum_{i=1}^n y_{i,t} z_{i,t}^* \right),
\]

(17.10)

where \( \alpha \) and \( \theta \) are two coefficients, under the condition that

\[
\sum_{i=1}^n \{1 - F_i(y_t)\} = \sum_{i=1}^n \{1 - F_i(\tilde{y}_{i,t})\} \exp(2\alpha y_{i,t} + 4\theta y_{i,t} z_{i,t}^*),
\]

(17.11)

where \( z_{i,t} \) denotes a linear combination of \( y_{j,t} \) for \( j \neq i \), \( z_{i,t}^* \) is a symmetrized form of \( z_{i,t} \) (e.g., in the one-dimensional space, if \( z_{i,t} = y_{i-1,t} \), then \( 2z_{i,t}^* = y_{i-1,t} + y_{i+1,t} \), and \( \tilde{y}_{i,t} = (y_{1,t}, \ldots, y_{i-1,t}, -y_{i,t}, y_{i+1,t}, \ldots, y_{n,t})' \). A direct and symmetric solution to Equation (17.11) is

\[
1 - F_i(y_t) = \exp(\alpha y_{i,t} + 2\theta y_{i,t} z_{i,t}^*) f(\alpha y_{i,t}, \theta y_{i,t} z_{i,t}^*),
\]

(17.12)

where \( f \) is a suitable, positive function, and even in both \( y_{i,t} \) and \( y_{i,t} z_{i,t}^* \).

Now, on a square lattice, let \( Y_{i',t} \in \{ 0, 1 \} \) denote the binary response with a realization \( y_{i',t} \) at row \( i \), column \( i' \), and time \( t \). Besag (1972) proposed a Markov process of binary responses developing through time on the square lattice, which can be viewed as a special case of Bartlett (1971, 1972). In particular, for fixed \( \alpha_y, \theta_{y,1}, \) and \( \theta_{y,2} \),

\[
P(Y_{i',t+\Delta t} = y|Y_{i',t} = y, Y_{i',t'}:t' \leq t+\Delta t, \text{ excluding } y_{i',t+\Delta t})
\]

\[
= -\Delta t \exp \left\{ \alpha_y + \theta_{y,1}(y_{i-1',t} + y_{i+1',t}) + \theta_{y,2}(y_{i-1',t} + y_{i+1',t}) \right\}^{-1} + o(\Delta t)
\]

\[
= 1 - \Delta t \exp \left\{ \alpha_y + \theta_{y,1}(y_{i-1',t} + y_{i+1',t}) + \theta_{y,2}(y_{i-1',t} + y_{i+1',t}) \right\} + o(\Delta t),
\]

(17.13)
gives the probability that \( Y_{i,t'} \), remains unchanged in the time interval \((t, t + \Delta t)\), given all other values at or before time \( t + \Delta t \). Further, it can be shown that its stationary distribution is an autologistic model with the full conditional distribution

\[
p(Y_{i,t'}|Y_{i-1,t'}, Y_{i+1,t'}, Y_{i,t-1,t}, Y_{i,t+1,t}) = \frac{\exp\left[\alpha + \theta_1(Y_{i-1,t'} + Y_{i+1,t'}) + \theta_2(Y_{i,t-1,t} + Y_{i,t+1,t})\right]}{1 + \exp\left[\alpha + \theta_1(Y_{i-1,t'} + Y_{i+1,t'}) + \theta_2(Y_{i,t-1,t} + Y_{i,t+1,t})\right]},
\]

(17.14)

where \( \alpha = \alpha_0 - \alpha_1, \theta_1 = \theta_{0,1} - \theta_{1,1}, \) and \( \theta_2 = \theta_{0,2} - \theta_{1,2} \).

### 17.2 Spatio-Temporal Autologistic Regression Model

#### 17.2.1 Model

For the analysis of spatio-temporal binary data in practice, it is often of interest to account for possible effects of covariates. For example, Gumpertz et al. (1997) and Huffer and Wu (1998) incorporated covariates in an autologistic model by replacing the constant \( \alpha_i \) in (17.6) with a linear regression term and the spatial lattice can be either regular or irregular. The resulting model is referred to as an autologistic regression model. Zhu et al. (2005) and Zheng and Zhu (2008) extended the autologistic regression model to a spatio-temporal autologistic regression model that accounts for covariates and spatio-temporal dependence simultaneously for binary responses measured repeatedly over discrete time points on a spatial lattice.

As earlier, let \( i = 1, \ldots, n \) denote sites on a spatial lattice. Further, let \( t \in \mathbb{Z} \) index discrete time points and \( Y_{i,t} \in \{0, 1\} \) denote the binary response variable at site \( i \) and time \( t \). Let \( x_{0,i,t} = 1 \) and let \( x_{k,i,t} \) denote the \( k \)th covariate at site \( i \) and time \( t \), for \( k = 1, \ldots, p \) and a total of \( p \) covariates. Zhu et al. (2005) developed a spatio-temporal autologistic regression model via the full conditional distributions:

\[
p\left(Y_{i,t}|y_{i,t'}: (i', t') \neq (i, t)\right) = p\left(Y_{i,t}|y_{i,t'}: (i', t') \in N_{i,t}\right) = \frac{\exp\left\{\sum_{k=0}^{p} \theta_k x_{k,i,t} y_{i,t} + \sum_{j 
\in N_i} \theta_{p+1} y_{j,i,t} y_{j,t} + \theta_{p+2} y_{j,i,t-1} y_{j,t+1}\right\}}{1 + \exp\left\{\sum_{k=0}^{p} \theta_k x_{k,i,t} + \sum_{j \in N_i} \theta_{p+1} y_{j,i,t} + \theta_{p+2} y_{j,i,t-1} y_{j,t+1}\right\}},
\]

(17.15)

where \( N_{i,t} = \{(j, t) : j \in N_i\} \cup \{(i, t-1), (i, t+1)\} \) denotes a spatio-temporal neighborhood for site \( i \) and time \( t \) and recall that \( N_i = \{j : \text{site } j \text{ is a neighbor of site } i\} \). The model parameters are the intercept \( \theta_0 \), slope \( \theta_k \) for covariate \( x_k \) with \( k = 1, \ldots, p \), a spatial autoregressive coefficient \( \theta_{p+1} \), and a temporal autoregressive coefficient \( \theta_{p+2} \). Let \( \theta = (\theta_0, \ldots, \theta_{p+2})' \) denote the vector of parameters in the model (17.15).
Let $Y_t = (Y_{1,t}, \ldots, Y_{n,t})'$ denote the binary responses at all sites and a given time point $t$ for $t = 1, \ldots, T$ and a total of $T$ sampling time points. Then, the joint distribution of $Y_2, \ldots, Y_{T-1}$ conditional on $Y_1$ and $Y_T$ is

$$p(y_2, \ldots, y_{T-1} | y_1, y_T; \theta)$$

$$= c(\theta)^{-1} \exp \left\{ \sum_{t=2}^{T-1} \left( \sum_{i=1}^{n} \sum_{k=0}^{p} \theta_{k,x_{i,t}} y_{i,t} + \frac{1}{2} \sum_{i=1}^{n} \sum_{j \in N_i} \theta_{p+1} y_{i,t} y_{j,t} + \sum_{l=1}^{S} \theta_{p+1+s} y_{i,t-l} \right) \right\},$$

(17.16)

where $c(\theta)$ is a normalizing constant and generally is intractable as it does not have an analytical form.

The full conditional distribution (17.15) is symmetric in time and thus depends on both past and future time points. For prediction at future time points, however, it would be more sensible to have the conditional distributions depend only on the past. For example, Zhu et al. (2008) proposed the following conditional distributions:

$$p(y_{i,t} | y_{j,t} : j \neq i, y_{t'} : t' = t-1, t-2, \ldots)$$

$$= p(y_{i,t} | y_{j,t} : j \in N_i, y_{t'} : t' = t-1, t-2, \ldots, t-S)$$

$$= \frac{\exp \left( \sum_{l=0}^{p} \theta_{k,x_{i,t}} y_{i,t} + \sum_{j \in N_i} \theta_{p+1} y_{i,t} y_{j,t} + \sum_{s=1}^{S} \theta_{p+1+s} y_{i,t-s} \right)}{1 + \exp \left( \sum_{l=0}^{p} \theta_{k,x_{i,t}} + \sum_{j \in N_i} \theta_{p+1} y_{i,t} + \sum_{s=1}^{S} \theta_{p+1+s} y_{i,t-s} \right)},$$

(17.17)

where $i = 1, \ldots, n$, $t = S+1, \ldots, T$, and $S$ is the maximum temporal lag. The term in (17.17) is a full conditional distribution for a given time point $t$, even though it is not a full conditional distribution for all $i$ and $t$. The spatial neighborhood $N_i$ may be further partitioned into different orders of neighborhood. In particular, let $N_i = \sum_{l=1}^{L} \sum_{k=1}^{S} N_i^{(l)}$, where $N_i^{(l)}$ denotes the $l$th-order neighborhood that comprises the $l$th nearest neighbors for $l = 1, \ldots, L$. Similar to the model specified via (17.16), the transition probability $p(y_{i,t} | y_{t'} : t' = t-1, \ldots, t-S)$ and the subsequent joint distribution function can be obtained. For ease of presentation, we focus on (17.16).

### 17.2.2 Statistical Inference

The intractable normalizing constant in the joint distribution function poses challenges in the statistical inference for the autologistic model with or without regression, an area of active research in the last couple of decades. While Besag (1975) originally proposed maximum pseudo-likelihood estimates (MPLEs), Huffer and Wu (1998) used Markov chain Monte Carlo (MCMC) methods to approximate the unknown normalizing constant and developed Monte Carlo maximum likelihood estimates (MCMLE) for spatial autologistic models. Further, Huang and Ogata (2002) generalized the pseudo-likelihood function and showed better performance of the resulting estimates than MPLE in terms of standard errors and efficiency relative to maximum likelihood estimates (MLEs). Berthelsen and Møller (2003) developed path sampling to approximate the ratio of unknown normalizing...
constants in spatial point processes, which Zheng and Zhu (2008) used for computing the MCMLE. Friel et al. (2009) proposed a fast computation method for the estimation of the normalizing constant based on a reduced dependence approximation of the likelihood function. Later, we describe statistical inference based on MPLE, MCMLE, and Bayesian hierarchical modeling.

17.2.2.1 Maximum Pseudo-Likelihood Estimation

Maximum pseudo-likelihood, first introduced by Besag (1975) for autologistic models, is a popular approach to the statistical inference for autologistic regression models. The MPLE is the value of $\theta$ that maximizes the product of the full conditional distributions,

$$\hat{\theta} = \arg \max_{\theta} L_{PL}(Y; \theta),$$

where the pseudo-likelihood function for a spatio-temporal autologistic model is

$$L_{PL}(Y; \theta) = \prod_{i,t} p(y_{i,t} | y_{i',t'} : (i', t') \neq (i, t)) = \prod_{i,t} \left[ \frac{\exp\left(\sum_{k=0}^{p} \theta_k x_{k,i,t} y_{i,t} + \sum_{j \in N_i} \theta_{p+1} y_{i,t} y_{j,t} + \theta_{p+2} y_{i,t}(y_{i,t-1} + y_{i,t+1})\right)}{1 + \exp\left(\sum_{k=0}^{p} \theta_k x_{k,i,t} + \sum_{j \in N_i} \theta_{p+1} y_{j,t} + \theta_{p+2}(y_{i,t-1} + y_{i,t+1})\right)} \right].$$

Although the pseudo-likelihood function (17.18) is not the true likelihood except in the trivial case of spatio-temporal independence, it can be shown that MPLEs are consistent and asymptotically normal under suitable regularity conditions (Guyon, 1995).

To maximize the pseudo-likelihood function and obtain the MPLE of $\theta$, it is straightforward to apply the standard logistic regression that assumes independence, which can be implemented by, for example, proc logistic in SAS or the function glm in R. The corresponding standard errors and approximate confidence intervals can be obtained by a parametric bootstrap. Specifically, in the parametric bootstrap, $M$ resamples of spatio-temporal binary responses are drawn according to the spatio-temporal autologistic regression model using Gibbs sampling or perfect sampling. For each resample, an MPLE is computed and the $M$ resampled MPLEs are used to obtain an estimate of the variance of the MPLE based on the original data. In particular, perfect sampling uses coupling and upon coalescence of the coupled Markov chains, the resulting Monte Carlo samples are guaranteed to be from the target distribution (e.g., Propp and Wilson, 1996; Møller, 1999).

17.2.2.2 Monte Carlo Maximum Likelihood Estimation

The maximum pseudo-likelihood approach is computationally efficient, but is statistically less efficient than maximum likelihood (Gumpertz et al., 1997; Wu and Huffer, 1997; Zheng and Zhu, 2008). An alternative approach is Monte Carlo maximum likelihood (MCML), where the normalizing constant is approximated using MCMC and thus direct maximization of likelihood function can be obtained.
The likelihood function can be rewritten as

\[ L(Y; \theta) = p(y_2, \ldots, y_{T-1} | y_1, y_T; \theta) = c(\theta)^{-1} \exp(\theta' z), \]

where

\[ z = \left( \sum_{i,t} y_{i,t}, \sum_{i,t} x_{1,i,t} y_{i,t}, \ldots, \sum_{i,t} x_{p,i,t} y_{i,t}, \frac{1}{2} \sum_{i,t} \sum_{i' \in N_i} y_{i,t} y_{i',t}, \sum_{i,t} \theta_{p+1} y_{i,t} y_{i,t-1} \right)'. \]

Based on a preselected parameter vector \( \psi = (\psi_0, \ldots, \psi_{p+2})' \), approximate the ratio of two normalizing constants via importance sampling by

\[ \frac{c(\theta)}{c(\psi)} = E_\psi \left\{ \frac{\exp(\theta' z)}{\exp(\psi' z)} \right\} \approx M^{-1} \sum_{m=1}^{M} \frac{\exp(\theta' z^m)}{\exp(\psi' z^m)} = M^{-1} \sum_{m=1}^{M} \exp((\theta - \psi)' z^m), \]

where \( z^m \) is \( z \) evaluated at the \( m \)th Monte Carlo sample of \( Y \) for \( m = 1, \ldots, M \). Monte Carlo samples of \( Y \) are generated from the joint distribution evaluated at \( \psi \). Then the MLE can be approximated by maximizing a rescaled version of the likelihood function

\[ c(\psi)L(Y; \theta) = \frac{c(\psi)}{c(\theta)} \exp(\theta' z) = \left[ M^{-1} \sum_{m=1}^{M} \exp((\theta - \psi)' z^m) \right]^{-1} \exp(\theta' z). \]

The variances of the estimates can be estimated by using the diagonal elements of the inverse of the observed Fisher information matrix (Huffer and Wu, 1998; Geyer, 1994).

The MCMLE provides a good approximation of the MLE of the model parameters when the reference parameter \( \psi \) is close to the truth (Geyer and Thompson, 1992). The MPLE is a natural choice for the reference parameter. However, when the spatial or temporal dependence is strong, MPLE can be far away from the MLE, whereas MCMLE with MPLE as the reference parameter may not exist and the iteration may lead to a sequence of estimates that drift off to infinity. In this case, we select \( \psi \) to be an approximation obtained by a stochastic approximation algorithm. This is a two-stage MCMC stochastic approximation algorithm proposed by Gu and Zhu (2001) for computing the MLEs of model parameters for a class of spatial models. In the first stage, the estimates are moved into a feasible region quickly by using large gain constants in the stochastic approximation and in the second stage, an optimal procedure is implemented with a stopping criterion chosen so that a desired precision can be obtained. By the first stage of the algorithm, \( \psi \) can be obtained.

### 17.2.3 Bayesian Inference

Bayesian hierarchical modeling can be applied for the inference about spatio-temporal autologistic regression models. Möller et al. (2006) presented an auxiliary variable MCMC algorithm that allows the construction of a proposal distribution so that the normalizing constants cancel out in the Metropolis–Hastings (MH) ratio. Zheng and Zhu (2008) proposed a Bayesian approach for both model parameter inference and prediction at future
time points using MCMC. They proposed an MH algorithm to generate Monte Carlo samples from the posterior distribution of the parameter $\theta$, where the likelihood ratio in the acceptance probability is approximated by

$$\frac{p(y_{2}, \ldots, y_{T-1}|y_{1}, y_{T}; \theta^{'})}{p(y_{2}, \ldots, y_{T-1}|y_{1}, y_{T}; \theta)} = \frac{\exp(\theta' z)}{\exp(\theta z)} \times \frac{c(\theta)}{c(\theta')} \approx \exp((\theta^{'} - \theta)' z) \times \frac{\sum_{m=1}^{M} \exp((\theta - \psi)z^m)}{\sum_{m=1}^{M} \exp((\theta^{'} - \psi)z^m)}. $$

Here, $M$ Monte Carlo samples of $Y$ need to be generated from the joint distribution $p(y_{2}, \ldots, y_{T-1}|y_{1}, y_{T}, \psi)$ evaluated at $\psi$, but only once at the beginning of the MH algorithm, which makes the algorithm efficient. For the MH algorithm, a good choice of the parameter vector $\psi$ helps to speed up the convergence process. The closer $\psi$ is to the posterior mode of $\theta$, the better the results are. Further, the variance of the proposal distribution needs to be adjusted to ensure a reasonable acceptance probability in the MH algorithm (Gelman et al., 2003).

Path sampling is an alternative way to calculate the ratio of two normalizing constants and is based on the following identity:

$$\ln \left\{ \frac{c(\theta)}{c(\theta')} \right\} = \int_{0}^{1} E_{\theta(s)} \left\{ \frac{d}{ds} \theta(s)' z \right\} ds$$

where the expectation is with respect to the joint distribution evaluated at the parameter $\theta(s)$ along a path of $\theta(s) = s\theta$ for $s \in [0, 1]$ from $\emptyset$ to $\theta$. However, the computation can be costly because multiple Monte Carlo samples of $Y$ are required for computing the expectation.

For the spatio-temporal autologistic regression model, Zheng and Zhu (2008) compared the performance of MPL, MCML, and Bayesian inference. They demonstrated that parameter inference via MPL can be statistically inefficient when spatial and/or temporal dependence is strong, whereas the statistical properties of the MCML are comparable to the Bayesian approach and the computation of MCML estimates is faster. Further, using Bayesian inference, the posterior distribution of the model parameters can be obtained and it becomes straightforward to construct credible bands at desired levels.

17.2.4 Prediction

Let $\tilde{Y} = (Y_{T+1}, \ldots, Y_{T+T^{'}})'$ denote the responses at future time points $T + 1, \ldots, T + T'$ with $T' \geq 1$. For prediction of $\tilde{Y}$ based on model parameter estimates from MPL and MCML, a Gibbs sampler can be used to obtain the Monte Carlo samples of $\tilde{Y}$ from

$$p(\tilde{y}|y_{T}, y_{T+T^{'}-1}; \theta) \propto \exp \left\{ \sum_{t=T+1}^{T+T'} \left( \sum_{i=1}^{n} \sum_{k=0}^{p} \theta_{k} x_{k,i,t} y_{i,t} + \frac{1}{2} \sum_{i=1}^{n} \sum_{j \in N_{i}} \theta_{p+1} y_{i,t} y_{j,t} \right) + \sum_{t=T+1}^{T+T^{'}+1} \sum_{i=1}^{n} \theta_{p+2} y_{i,t} y_{i,t-1} \right\}. $$
For prediction of $\tilde{Y}$ in the Bayesian framework, the posterior predictive distribution of $\tilde{Y}$ is

$$p(\tilde{y}|y, y_{T+T^*+1}) = \int p(\tilde{y}|y, y_{T+T^*+1}; \theta)p(\theta|y)d\theta.$$ 

To draw Monte Carlo samples of $\tilde{Y}$ from $p(\tilde{y}|y, y_{T+T^*+1})$, first draw $\theta$ from its posterior distribution $p(\theta|y)$ and then for each given $\theta$, draw $\tilde{Y}$ from $p(\tilde{y}|y, y_{T+T^*+1}; \theta)$ using a Gibbs sampler (Zheng and Zhu, 2008).

### 17.3 Centered Autologistic Regression Model

In the aforementioned autologistic regression models, the interpretation of model parameters is not straightforward (Caragea and Kaiser, 2009; Kaiser and Caregea, 2009). In the presence of positive spatial and temporal dependence, under the uncentered parameterization, the conditional expectation of $Y_{i,t}$ given its neighbors is

$$E(Y_{i,t}|Y_{i',t'}: (i', t') \in \mathcal{N}_{i,t}) = \exp \left\{ \sum_{k=0}^{p} \theta_k x_{k,i,t} + \sum_{j \in \mathcal{N}_i} \theta_{p+1} y_{j,t} + \theta_{p+2} (y_{j,t-1} + y_{j,t+1}) \right\}.$$ 

(17.19)

The expectation (17.20) is larger than the expectation of $Y_{i,t}$ under independence,

$$\frac{\exp \left\{ \sum_{k=0}^{p} \theta_k x_{k,i,t} \right\}}{1 + \exp \left\{ \sum_{k=0}^{p} \theta_k x_{k,i,t} \right\}}$$

as long as $Y_{i,t}$ has nonzero spatial and/or temporal neighbors, but is never smaller. This may not be reasonable when most neighbors are zeros and thus can bias the realizations toward 1. Hence, the interpretation of dependence parameters is difficult. Further, the marginal expectation of $Y_{i,t}$ (i.e., $E(Y_{i,t}|x_{k,i,t}, k = 1, \ldots, p)$) is greater than the expectation of $Y_{i,t}$ under independence. A simulation study in Wang (2013) showed that $E(Y_{i,t}|x_{k,i,t}, k = 1, \ldots, p)$ varies across different levels of spatial and temporal dependence for fixed regression coefficients. These make the interpretation of regression coefficients unclear since these coefficients are to reflect the effects of covariates and should have a consistent interpretation across varying dependence levels.

For non-Gaussian Markov random field models of spatial lattice data, the idea of centered parameterization was first proposed by Kaiser and Cressie (1997) for a Winsorized Poisson conditional model. More recently, Kaiser and Caregea (2009) explored the centered parameterization for a general exponential family of Markov random field models. In particular, Caragea and Kaiser (2009) studied the centered parameterization for spatial autologistic regression models and showed that the centered parameterization overcomes the interpretation difficulties. Wang and Zheng (2013) extended this work to the case of spatio-temporal autologistic regression models.
17.3.1 Model with Centered Parameterization

For site \( i = 1, \ldots, n \) and time \( t \), let \( \pi_{i,t} \) denote the probability of \( Y_{i,t} = 1 \) under spatio-temporal independence. That is,

\[
\pi_{i,t} = \frac{\exp \left( \sum_{k=0}^{p} \theta_k x_{k,i,t} \right)}{1 + \exp \left( \sum_{k=0}^{p} \theta_k x_{k,i,t} \right)}.
\]

Let \( y_{i,t}^* = y_{i,t} - \pi_{i,t} \) denote a centered response at site \( i \) and time \( t \), centering around \( \pi_{i,t} \). For pairwise-only dependence, Wang and Zheng (2013) defined a centered spatio-temporal autologistic regression model via the following full conditional distributions:

\[
p(y_{i,t} | y_{i',t'} : (i', t') \neq (i, t)) = p(y_{i,t} | y_{i',t'} : (i', t') \in \mathcal{N}_{i,t}) = \frac{\exp \left( \sum_{k=0}^{p} \theta_k x_{k,i,t} y_{i,t} + \sum_{j \in \mathcal{N}_i} \theta_{p+1} y_{j,t} y_{i,t}^* + \theta_{p+2} y_{i,t-1} + y_{i,t+1} \right)}{1 + \exp \left( \sum_{k=0}^{p} \theta_k x_{k,i,t} + \sum_{j \in \mathcal{N}_i} \theta_{p+1} y_{j,t} + \theta_{p+2} (y_{i,t-1} + y_{i,t+1}) \right)}.
\]

(17.20)

By the Hammersley–Clifford theorem and its corollary, the joint likelihood function of \( Y_2, \ldots, Y_{T-1} \) conditioned on \( Y_1 \) and \( Y_T \) is

\[
L(Y; \theta) = p(y_2, \ldots, y_{T-1} | y_1, y_T; \theta)
= c^*(\theta)^{-1} \exp \left( \sum_{t=2}^{T-1} \left( \sum_{i=1}^{n} \sum_{k=0}^{p} \theta_k x_{k,i,t} y_{i,t}^* + \frac{1}{2} \sum_{i=1}^{n} \sum_{j \in \mathcal{N}_i} \theta_{p+1} y_{j,t} y_{i,t}^* + \sum_{t=2}^{T} \sum_{i=1}^{n} \theta_{p+2} y_{i,t-1} y_{i,t+1} \right) \right),
\]

(17.21)

where \( c^*(\theta) \) is the normalizing constant. When the temporal autocorrelation coefficient is zero (i.e., \( \theta_{p+2} = 0 \)), the model reduces to a spatio-only autologistic regression model (Caragea and Kaiser, 2009; Hughes et al., 2011).

Thus, the conditional expectation of \( Y_{i,t} \) given its neighbors is

\[
E(Y_{i,t} | Y_{i',t'} : (i', t') \in \mathcal{N}_{i,t}) = \frac{\exp \left( \sum_{k=0}^{p} \theta_k x_{k,i,t} + \sum_{j \in \mathcal{N}_i} \theta_{p+1} y_{j,t} + \theta_{p+2} (y_{i,t-1} + y_{i,t+1}) \right)}{1 + \exp \left( \sum_{k=0}^{p} \theta_k x_{k,i,t} + \sum_{j \in \mathcal{N}_i} \theta_{p+1} y_{j,t} + \theta_{p+2} (y_{i,t-1} + y_{i,t+1}) \right)},
\]

which we denote as \( \pi_{i,t}^* \). Suppose that the spatial autoregressive coefficient \( \theta_{p+1} \) and the temporal autoregressive coefficient \( \theta_{p+2} \) are positive. Then, \( \pi_{i,t}^* > \pi_{i,t} \) when

\[
\theta_{p+1} \sum_{j \in \mathcal{N}_i} y_{j,t} + \theta_{p+2} (y_{i,t-1} + y_{i,t+1}) > \theta_{p+1} \sum_{j \in \mathcal{N}_i} \pi_{j,t} + \theta_{p+2} (\pi_{i,t-1} + \pi_{i,t+1}),
\]

where \( \sum_{j \in \mathcal{N}_i} \pi_{j,t} \) and \( \pi_{i,t-1} + \pi_{i,t+1} \) are the expected numbers of nonzero spatial and temporal neighbors under the independence model, respectively. Specifically, if \( \theta_{p+2} = 0 \), then
\( \pi_{i,t}^* > \pi_{i,t} \) only when the observed number of nonzero spatial neighbors is greater than the expected number of nonzero spatial neighbors under independence. That is, \( \sum_{j \in N_i} y_{j,t} > \sum_{j \in N_i} \pi_{j,t} \). If \( \theta_{p+1} = 0 \), then \( \pi_{i,t}^* > \pi_{i,t} \) only when the observed number of nonzero temporal neighbors is greater than the expected number of nonzero temporal neighbors under independence. That is, \( y_{i,t-1} + y_{i,t+1} > \pi_{i,t-1} + \pi_{i,t+1} \). Thus, the interpretation of \( \theta_{p+1} \) and \( \theta_{p+2} \) as local dependence parameters is more sensible. Further, the simulation study in Wang (2013) showed that the marginal expectation of \( Y_{i,t} \) under the centered parameterization remains constant over moderate levels of spatial and temporal dependence (i.e., \( E(Y_{i,t}|x_{k,i,t}, k=1, \ldots, p) \approx \pi_{i,t} \)). The interpretation of regression coefficients as effects of covariates is more sensible as well.

### 17.3.2 Statistical Inference

For the model with centered parameterization, its statistical inference has been developed based on expectation–maximization pseudo-likelihood, Monte Carlo expectation–maximization likelihood, and Bayesian inference (Wang and Zheng, 2013).

#### 17.3.2.1 Expectation–Maximization Pseudo-Likelihood Estimator

To obtain the maximum pseudo-likelihood estimates of the model parameters, the combination of an expectation–maximization (EM) algorithm and a Newton–Raphson algorithm, called the expectation–maximization pseudo-likelihood estimator (EMPLE), is considered. Specifically, update \( \pi_{i,t} \), the expectation of \( Y_{i,t} \) under the independent model, at the E step and then at the M step, update \( \hat{\theta}^* \) by maximizing

\[
\prod_{i,t} \frac{\exp \left\{ \sum_{k=0}^p \theta_k x_{k,i,t} y_{i,t} + \sum_{j \in N_i} \theta_{p+1} y_{j,t} y_{j,t}^*(l-1) + \theta_{p+2} y_{i,t} y_{i,t}^*(l-1) + y_{i,t}^*(l-1) \right\}}{1 + \exp \left\{ \sum_{k=0}^p \theta_k x_{k,i,t} y_{i,t} + \sum_{j \in N_i} \theta_{p+1} y_{j,t} y_{j,t}^*(l-1) + \theta_{p+2} y_{i,t} y_{i,t}^*(l-1) + y_{i,t}^*(l-1) \right\}},
\]

where \( y_{j,t}^*(l) \) is the centered response at the \( l \)th iteration. The M step can be carried out by a Newton–Raphson algorithm using the standard logistic regression and the E and M steps are repeated until convergence. A parametric bootstrap can be used to compute the standard error of the EMPLE. For the starting value \( \hat{\theta} \) at the start of the algorithm, different starting points can impact how long it takes to convergence. The maximum MPLE from the uncentered autologistic regression model is a natural choice.

#### 17.3.2.2 Monte Carlo Expectation–Maximization Likelihood Estimator

Let \( z_{\hat{\theta}}^* = (\sum_{i,t} x_{0,i,t} y_{i,t}^*, \ldots, \sum_{i,t} x_{p,i,t} y_{i,t}^*, \frac{1}{2} \sum_{i,t} \sum_{j \in N_i} y_{j,t}^* y_{j,t}^* y_{i,t}^*, \sum_{i,t} y_{i,t}^* y_{i,t}^* y_{i,t}^*)' \). We consider a rescaled version of the likelihood function

\[
c^*(\psi)L(Y; \Theta) = \frac{c^*(\psi)}{c^*(\hat{\theta})} \exp(\theta^* z_{\hat{\theta}}^*) = \left[ E_{\psi} \left\{ \frac{\exp(\theta^* z_{\hat{\theta}}^*)}{\exp(\psi^* z_{\hat{\theta}}^*)} \right\} \right]^{-1} \exp(\theta^* z_{\hat{\theta}}^*),
\]
where $\psi$ is a reference parameter and $z^*_{\psi}$ is $z^*$ with centers evaluated at $\psi$. Monte Carlo expectation–maximization likelihood (MCEML) estimator can be used by combining an EM algorithm and a Newton–Raphson algorithm. Specifically, first choose a reference parameter vector $\psi$ and generate $M$ Monte Carlo samples of $Y$ from the likelihood function evaluated at $\psi$. Then for the $l$th iteration, at the E step, we update $\pi^{(l-1)}_{i,t}$ and set $y_{i,t}^{(l-1)} = y_{i,t} - \pi^{(l-1)}_{i,t}$. At the M step, we maximize the rescaled version of the likelihood function

$$
\exp(\theta'z^*_{\theta^{(l-1)}}) \left[ M^{-1} \sum_{m=1}^{M} \exp \left( \theta'z^*(m)_{\theta^{(l-1)}} - \psi'z^*(m)_{\psi} \right) \right]^{-1},
$$

where $z^*_{\theta^{(l-1)}}$ is $z^*$ with centered responses $y_{i,t}^{(l-1)}$ and $z^*(m)_{\theta^{(l-1)}}$ and $z^*(m)_{\psi}$ are $z^*$ evaluated at the $m$th Monte Carlo sample of $Y$ generated at the beginning of the algorithm with centers computed at $\hat{\theta}^{(l-1)}$ and $\psi$, respectively. The M step can be carried out using a Newton–Raphson algorithm. We compute the observed Fisher information matrix and obtain the standard errors of the MCEMLE as a by-product of the MCEML estimation.

### 17.3.2.3 Bayesian Inference

We consider an MH algorithm to generate Monte Carlo samples of $\theta$ from the posterior distribution $p(\theta|y)$ (Zheng and Zhu, 2008), where the likelihood ratio in $\alpha(\theta^*|\theta)$ in the acceptance probability is approximated as

$$
p(y_2, \ldots, y_{T-1}|y_1, y_T, \theta^*) \frac{\exp(\theta'z^*_{\theta^*})}{\exp(\theta'z^*_{\theta})} \approx \frac{\exp(\theta'z^*_{\theta^*})}{\exp(\theta'z^*_{\theta})} \times \frac{c^*(\theta)}{c^*(\theta^*)} \sum_{m=1}^{M} \frac{\exp(\theta'z^*(m)_{\theta^*} - \psi'z^*(m)_{\psi})}{\exp(\theta'z^*(m)_{\theta})},
$$

where $z^*(m)_{\theta}$, $z^*(m)_{\theta^*}$, and $z^*(m)_{\psi}$ are $z^*$ evaluated at the $m$th Monte Carlo sample of $Y$ with centers computed based on $\theta$, $\theta^*$, and $\psi$, respectively.

### 17.4 Data Example

For illustration, we consider the outbreak of southern pine beetle (SPB) in North Carolina. The data consist of indicators of outbreak or not ($0 =$ no outbreak; $1 =$ outbreak) in the 100 counties of North Carolina from 1960 to 1996. Figure 17.1 gives a time series of the county-level outbreak maps. The average precipitation in the fall (in cm) will be the covariate and is mapped in Figure 17.2. We use the data from 1960 to 1991 for model fitting and set aside the data from 1992 to 1996 for model validation. Two counties are considered to be neighbors if the corresponding county seats are within 30 miles of each other.
Maps of southern pine beetle outbreak from 1960 to 1996 in the counties of North Carolina. A county is filled black if there was an outbreak and is unfilled otherwise.
FIGURE 17.2
Map of mean fall precipitation in the counties of North Carolina.
TABLE 17.1
Comparison of Model Parameter Estimation for the Spatio-Temporal Autologistic Model with Uncentered Parameterization Using Maximum Pseudo-likelihood (MPL), Monte Carlo Maximum Likelihood (MCML), and Bayesian Inference

| Parameters | MPL | | MCML | | Bayesian | |
|------------|-----|-------------------------------|-----|-------------------------------|-----|
|            | Estimate | SE | Estimate | SE | Mean | SD |
| Intercept $\theta_0$ | -5.16 | 0.66 | -2.71 | 0.20 | -2.71 | 0.20 |
| Slope $\theta_1$ | 0.25 | 0.18 | -0.14 | 0.05 | -0.14 | 0.05 |
| Spatial $\theta_2$ | 1.45 | 0.14 | 0.91 | 0.05 | 0.91 | 0.06 |
| Temporal $\theta_3$ | 1.71 | 0.24 | 1.02 | 0.11 | 1.03 | 0.13 |

TABLE 17.2
Comparison of Model Parameter Estimation for the Spatio-Temporal Autologistic Model with Centered Parameterization Using Expectation–Maximization Pseudo-likelihood (EMPL), Monte Carlo Expectation–Maximization Likelihood (MCEML), and Bayesian Inference

| Parameters | EMPL | | MCEML | | Bayesian | |
|------------|-----|-----------------------------|-----|-----------------------------|-----|
|            | Estimate | SE | Estimate | SE | Mean | SD |
| Intercept $\theta_0$ | -4.96 | 0.32 | -2.40 | 0.16 | -2.86 | 0.29 |
| Slope $\theta_1$ | 0.21 | 0.08 | -0.13 | 0.05 | -0.13 | 0.08 |
| Spatial $\theta_2$ | 1.47 | 0.13 | 0.95 | 0.05 | 0.95 | 0.06 |
| Temporal $\theta_2$ | 1.75 | 0.18 | 0.89 | 0.07 | 0.89 | 0.10 |

Tables 17.1 and 17.2 give the model parameter estimates and standard errors from fitting the spatio-temporal autologistic regression models with uncentered and centered parameterization, respectively (see Wang and Zheng, 2013; Zheng and Zhu, 2008). The parameter estimates and the corresponding standard errors for both models using all of the three inference approaches, MPLE, MCMLE, and Bayesian inference, are quite close. One possible reason for this is that for this data set, the influence of the center is small relative to the strength of spatio-temporal dependence. The average of the centers $\pi_i$ evaluated at the MCEMLE is only 0.05, and thus, the spatio-temporal autoregressive terms dominate the outbreak probabilities.

For comparison among various statistical inference approaches, the results suggest that the inference for the model parameters using the posterior distribution matches well with MCML, but the inference from pseudo-likelihood is different from both Bayesian inference and MCML for both uncentered and centered parameterization models. In addition, estimation based on pseudo-likelihood results in higher variance than the other approaches. In terms of computing time, Bayesian inference is more time consuming compared with the other two approaches. Further, we predict the SPB outbreak from 1992 to 2001 in North Carolina (Table 17.3). The responses at the end time point (here, $y_{i,2002}$) are generated from independent Bernoulli trials with probability of outbreak $\sum_{t=1991}^{1991} y_{i,t}/31$ for $i = 1, \ldots, 100$. The prediction performances based on models with uncentered and centered parameterization are comparable. Overall, our recommendation is to use a model with uncentered parameterization if prediction is of primary interest, since the two
TABLE 17.3
Comparison of the Prediction Performance between Models with Centered and Uncentered Parameterization

<table>
<thead>
<tr>
<th>Year</th>
<th>Centered</th>
<th>Uncentered</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EMPL</td>
<td>MCEML</td>
</tr>
<tr>
<td>1992</td>
<td>0.65</td>
<td>0.18</td>
</tr>
<tr>
<td>1993</td>
<td>0.72</td>
<td>0.19</td>
</tr>
<tr>
<td>1994</td>
<td>0.70</td>
<td>0.20</td>
</tr>
<tr>
<td>1995</td>
<td>0.63</td>
<td>0.23</td>
</tr>
<tr>
<td>1996</td>
<td>0.62</td>
<td>0.24</td>
</tr>
</tbody>
</table>

**Note:** For centered parameterization, the prediction is based on statistical inference obtained using expectation-maximization pseudo-likelihood (EMPL), Monte Carlo expectation-maximization likelihood (MCEML), and Bayesian inference. For uncentered parameterization, the prediction is based on statistical inference obtained using maximum pseudo-likelihood (MPL), Monte Carlo maximum likelihood (MCML), and Bayesian inference. Reported are the prediction error rates for each year in 1992–1996.

Parameterizations provide comparable performance in prediction but the centered parameterization is computationally more intensive. If the focus is on the interpretation of the regression coefficients, however, the centered parameterization is recommended.

17.5 Discussion

In this chapter, we have reviewed spatio-temporal autologistic regression models for spatio-temporal binary data. Alternatively, a generalized linear mixed model (GLMM) framework can be adopted for modeling such spatial data (Diggle and Ribeiro, 2007; Holan and Wikle [2015; Chapter 15 in this volume]). The response variable is modeled by a distribution in the exponential family and is related to covariates and spatial random effects in a link function. Thus, GLMM is flexible, as it is suitable for both Gaussian responses and non-Gaussian responses such as binomial and Poisson random variables. Statistical inference can be carried out using Bayesian hierarchical modeling, which is flexible as more complex structures can be readily placed on the model parameters. With suitable reduction of dimensionality for the spatio-temporal random effects, computation is generally feasible. In particular, faster computational algorithms are emerging such as integrated nested Laplace approximations (INLA) (Rue et al., 2009). Although likelihood-based approaches are suitable, it is sometimes a challenge to attain a full specification of the likelihood function, due to a lack of sufficient information and complex interactions among responses. In this case, an estimating equation approach may be attractive. For spatial binary data, Lin et al. (2008) developed a central limit theorem for a random field under various $L_p$ metrics and derived the consistency and asymptotic normality of quasi-likelihood estimators. Lin (2010) further developed a generalized estimating equation (GEE) method for spatio-temporal binary data, but only a single binary covariate was considered and the spatio-temporal dependence is limited to be separable. Moreover, variable selection methods for identifying the suitable set of covariates are of interest. For example, Fu et al. (2013) developed adaptive Lasso for the selection of covariates in an autologistic regression model and extension to
spatio-temporal autologistic regression model is discussed. Further research on this and other related topics will be worthwhile.

References


