2012

A Theoretical Development and Simulation-Based Comparison of Four Parameter Estimation Methods for the Spatio-Temporal Autologistic Model with Emphasis on Maximum Generalized and Block Generalized Pseudolikelihood

Jordan Earl Purdy
The University of Montana

A Theoretical Development and Simulation-Based Comparison of Four Parameter Estimation Methods for the Spatio-Temporal Autologistic Model with Emphasis on Maximum Generalized and Block Generalized Pseudolikelihood

By

Jordan Earl Purdy

B.S. George Fox University, Newberg, OR, 2006
M.A. The University of Montana, Missoula, MT, 2008

Dissertation

presented in partial fulfillment of the requirements
for the degree of

Doctorate of Philosophy
in Mathematics

The University of Montana
Missoula, MT

May 2012

Approved by:

Dr. Sandy Ross, Associate Dean
Graduate School

Dr. Jon Graham, Chair
Mathematical Sciences

Dr. John Bardsley
Mathematical Sciences

Dr. Solomon Harrar
Mathematical Sciences

Dr. Dave Patterson
Mathematical Sciences

Dr. Jesse Johnson
Computer Science

Purdy, Jordan E. Doctorate of Philosophy, May 2012 Mathematics

A Theoretical Development and Simulation-Based Comparison of Four Parameter Estimation Methods for the Spatio-Temporal Autologistic Model with Emphasis on Maximum Generalized and Block Generalized Pseudolikelihood

Committee Chair: Jon Graham, Ph.D.

A regular lattice of spatially dependent binary observations is often modeled using the autologistic model. It is well known that likelihood-based inference methods cannot be employed in the usual way to estimate the parameters of the autologistic model due to the intractability of the normalizing constant for the corresponding joint likelihood. Two popular and vastly contrasting approaches to parameter estimation for the autologistic model are maximum pseudolikelihood (PL) and Markov Chain Monte Carlo Maximum Likelihood (MCMCML). Two newer and less understood approaches are maximum generalized pseudolikelihood (GPL) and maximum block generalized pseudolikelihood (BGPL). Both of these newer methods represent varying degrees of compromise between maximum pseudolikelihood and MCMCML. The research performed in this dissertation focuses on these four estimation methods, with particular emphasis given to GPL and BGPL, and incorporates theoretical, simulation-based, and application-based components.

The theoretical components of this dissertation are three-fold. First, when employing GPL or BGPL, the need to distinguish between types of neighbors within the neighborhood set of a lattice site is formally developed. Such a distinction ultimately affects the functional forms of the generalized pseudolikelihood and block generalized pseudolikelihood functions. Second, extensions of generalized and block generalized pseudolikelihood for use in the space-time domain are proposed. As GPL and BGPL were initially only developed for use in the spatial domain, these extensions are the first of their kind. Third, and finally, the basic asymptotic property of strong consistency is established for the estimates obtained via maximum generalized and maximum block generalized pseudolikelihood.

In addition to the aforementioned theoretical components, this dissertation also includes two simulation studies. In particular, a large scale purely spatial simulation study using the autologistic model was conducted comparing the performances of PL, MCMCML, GPL, and BGPL. This was the first such study to simultaneously compare GPL and BGPL, and it was also the first such study to simultaneously incorporate a covariate, spatial anisotropy, and higher-order neighborhood systems. The results of this study indicate that GPL tends to outperform BGPL, and that both of these newer methods tend to achieve their intended performance-based compromise between PL and MCMCML, particularly in situations where estimation is notoriously difficult. Additionally, a small-scale space-time simulation study using the spatio-temporal autologistic model was conducted comparing the performances of PL, MCMCML, and the proposed space-time extensions of GPL and BGPL. Such a space-time simulation study has never before been conducted. The results of this study suggest that the proposed extensions of GPL and BGPL are indeed appropriate, and that the relative performances of the four estimation methods in the space-time domain are largely analogous

to those from the purely spatial domain.

The final component of this dissertation is application-based. More specifically, fire occurrence data from Oregon and Washington state are modeled, while accounting for the Departure from Average fire potential metric as a covariate, using the spatio-temporal autologistic model. All four estimation methods (PL, MCMCML, GPL, and BGPL) are employed to estimate the parameters of several proposed spatio-temporal autologistic models, and a Monte Carlo-based sum of absolute error (SAE) statistic is used as a model selection criterion.

Acknowledgments

I would like to thank the Department of Mathematical Sciences at The University of Montana for giving me this opportunity, and the members of my Dissertation Committee for their involvement and patience. I would also like to thank my advisor, Dr. Jon Graham, for all of the time and effort he has invested in me, and for his continued guidance throughout my time as a graduate student. I am extremely fortunate to have had such a mentor. Finally, I would like to thank my wife, Amber, for her unwavering support, patience, and love. With her by my side, I am, and always will be, able to put the proper perspective on whatever difficulties arise in life.

Contents

Abstract **ii**

Acknowledgments **iv**

List of Tables **x**

List of Figures **xiii**

1 Introduction to the Autologistic Model **1**

1.1 Introduction 1

1.2 Summary of Parameter Estimation Methods 5

1.2.1 The Coding Method 6

1.2.2 Pseudolikelihood 8

1.2.3 Markov Chain Monte Carlo Maximum Likelihood 9

1.2.4 Generalized Pseudolikelihood 12

1.2.5 Block Generalized Pseudolikelihood 16

1.2.6 Recursion Method 20

1.2.7 Additional Estimation Methods 24

1.3 Applications of Estimation Methods 25

1.4 Simulations of Estimation Methods 29

1.5 Spatio-Temporal Autologistic Model (STAM) 36

1.6 Research Objectives 42

2 Origins of the Autologistic Model **44**

2.1 Introduction 44

2.2 Development of the Autologistic Model 44

2.2.1 Difficulties with a Conditional Specification 45

2.2.2 Obtaining a Consistent Conditional Specification: Neighbors, Cliques, Negpotential Function, and the Hammersley-Clifford Theorem 47

2.2.3 Simplifying Assumptions: Pairwise-Only Dependence and Conditional Exponential Distributions 58

2.2.4 Subtle Complications: Uniqueness and Edge Sites 60

2.2.5 Binary Data and the Autologistic Model 63

2.3 The Autologistic Model and Covariates 68

3 Estimation Methods for the Autologistic Model **71**

3.1 Introduction 71

3.2 Maximum Pseudolikelihood Estimation 74

3.2.1 PL: Background 75

3.2.2 PL: Application to the ALM 76

3.3 Markov Chain Monte Carlo Maximum Likelihood Estimation 77

3.3.1 MCMCML: Background 78

3.3.2 MCMCML: Gibbs Sampler 81

3.3.3 MCMCML: Application to the ALM 83

3.4 Maximum Generalized Pseudolikelihood Estimation 86

3.4.1 GPL: Background 87

3.4.2 GPL: Application to ALM 91

3.5 Maximum Block Generalized Pseudolikelihood Estimation 94

3.5.1 BGPL: Background 94

3.5.2 BGPL: Application to ALM 99

4 Strong Consistency of the MGPLEs and MBGPLEs **102**

4.1 Introduction 102

4.2 Notation 103

4.3 Lemmas 107

4.3.1 Lemma 1 108

4.3.2 Lemma 2 111

4.3.3 Lemma 3 115

4.3.4 Lemma 4 120

4.3.5 Lemma 5 123

4.4 Proof of MGPLE's Strong Consistency 124

4.5 Strong Consistency of MBGPLEs 125

5 Autologistic Model Simulation Study 130

5.1 Introduction – Three Questions of Interest 130

5.2 Cases of Simulation Study 133

5.2.1 First-Order Isotropy without a Covariate – FI2p 133

5.2.2 First-Order Two-Way Anisotropy without a Covariate – FAhv3p 137

5.2.3 Second-Order Isotropy without a Covariate – SI2p 138

5.2.4 First-Order Isotropy with a Covariate – FI3p 139

5.2.5 First-Order Two-Way Anisotropy with a Covariate – FAhv4p 140

5.2.6 Second-Order Isotropy with a Covariate – SI3p 141

5.2.7 The Ising Model and Phase Transition 142

5.3 GPL vs. BGPL – Question 1 142

5.4 Reference Points for MCMCML – Question 2 152

5.5 PL vs. MCMCML vs. GPL vs. BGPL – Question 3 161

5.6 Concluding Remarks 172

6 Spatio-Temporal Autologistic Model 176

6.1 Introduction 176

6.2 The Spatio-Temporal Autologistic Model 178

6.3 Spatio-Temporal PL 182
6.4 Spatio-Temporal MCMCML 184
6.5 Spatio-Temporal GPL 186
6.6 Spatio-Temporal BGPL 189
6.7 Spatio-Temporal Autologistic Model Simulation Study 193
6.7.1 Scope 193
6.7.2 Procedure 195
6.7.3 Results 199

7 Spatio-Temporal Autologistic Model Application 205

7.1 Introduction 205
7.2 Description of Data Set 206
7.3 Approach to Model Fitting and Model Selection 209
7.4 Results 213

8 Concluding Remarks 221

Bibliography 225

A Supplementary Tables and Figures for Chapter 5 228

A.1 Supplemental Items for Section 5.3 228
A.2 Supplemental Items for Section 5.4 228
A.3 Supplemental Items for Section 5.5 228

List of Tables

5.1 Sample Means and Standard Errors of α and θ when the Model is FI2p and the Method of Estimation is GPL for Group Sizes of 5 (gs5) and 9 (gs9) 148

5.2 Sample Means and Standard Errors of α and θ when the Model is FI2p and the Method of Estimation is BGPL for Block Sizes of 2x2 (bs2), 3x3 (bs3), and 4x4 (bs4) 149

5.3 Relative MAEs of α, θ_h, θ_v, and β (Sine Wave Covariate) when the Model is FAhv4p and the Methods of Estimation are GPL for a Group Size of 9 (gs9) and BGPL for Block Sizes of 2x2 (bs2), 3x3 (bs3), and 4x4 (bs4) 152

5.4 Sample Means and Standard Errors of α and θ when the Model is an Ising Model Near Phase Transition and the Method of Estimation is MCMCML Under the Following Reference Points: MPLE (PL), MGPLE for a Group Size of 9 (gs9), and MBGPLE for a Block Size of 4x4 (bs4) 157

5.5 Proportion of Trials Converging/Non-Outlying when the Model is an Ising Model Near Phase Transition with $(\alpha, \theta) = (-3.2, 1.6)$ and the Method of Estimation is MCMCML Under the Following Reference Points: MPLE (MCMCML_{PL}), MGPLE for a Group Size of 9 (MCMCML_{gs9}), and MBGPLE for a Block Size of 4x4 (MCMCML_{bs4}) 158

5.6 Proportion of MCMCML Trials Converging as Non-Outliers for at Least One of the Three Reference Point Sets (MPLE, MGPLE for a Group Size of 9, or MBGPLE for a Block Size of 4x4) when the Model is an Ising Model Near Phase Transition 158

5.7 Sample Means and Standard Errors of α, θ_h, and θ_v when the Model is FAhv3p and the Methods of Estimation are PL, GPL for a Group Size of 9 (gs9), BGPL for a Block Size of 4x4 (bs4), and MCMCML (26×26, 50×50, and 74×74 Lattices) 167

5.8 Relative MAEs of α, θ, and β (Fire Covariate) when the Model is FI3p and the Methods of Estimation are PL, GPL for a Group Size of 9 (gs9), BGPL for a Block Size of 4x4 (bs4), and MCMCML 169

5.9 Proportion of Trials (out of 500) Used in Comparisons of MPLEs, MGPLEs, MBGPLEs, and MCMCMLEs when the Model is an Ising Model Near Phase Transition 169

6.1 Sample Means and Standard Errors of α, θ, and τ when the Model is FIAR1 (No Covariate) and the Methods of Estimation are PL, GPL, BGPL, and MCMCML 200

6.2 Relative MAEs of α, θ, and τ when the Model is FIAR1 (No Covariate) and the Methods of Estimation are PL, GPL, BGPL, and MCMCML 201

6.3 Sample Means and Standard Errors of α, θ, τ, and β (Unif(0,1) or Fire (DA) Covariate) when the Model is FIAR1 (With a Covariate) and the Methods of Estimation are PL, GPL, BGPL, and MCMCML 203

6.4 Relative MAEs of α, θ, τ, and β (Unif(0,1) or Fire (DA) Covariate) when the Model is FIAR1 (With a Covariate) and the Methods of Estimation are PL, GPL, BGPL, and MCMCML 204

7.1 PL-AIC Values for the Five Fire Application Models 215

7.2 Parameter Estimates and Standard Errors for Fire Application Models M3 and M5 where the Methods of Estimation are PL, GPL for a three-site L-shaped Group Size, BGPL for a 2x2x1 Block Size, and MCMCML 218

7.3 SAEs for Fire Application Models M3 and M5 where the Methods of Estimation are PL, GPL for a three-site L-shaped Group Size, BGPL for a 2x2x1 Block Size, and MCMCML 219

A.1 MAEs of α and θ when the Model is FI2p and the Method of Estimation is GPL for Group Sizes of 5 (gs5) and 9 (gs9) 229

A.2 MAEs of α and θ when the Model is FI2p and the Method of Estimation is BGPL for Block Sizes of 2x2 (bs2), 3x3 (bs3), and 4x4 (bs4) 230

A.3 Sample Means and Standard Errors of α, θ_h, and θ_v when the Model is FAhv3p and the Methods of Estimation are GPL for a Group Size of 9 (gs9) and BGPL for Block Sizes of 2x2 (bs2), 3x3 (bs3), and 4x4 (bs4) (26×26, 50×50, and 74×74 Lattices) 231

A.4 Proportion of Trials Converging when the Model is an Ising Model with $(\alpha, \theta) = (-3.2, 1.6)$ and the Method of Estimation is MCMCML Under the Following Reference Points: MPLE (MCMCML_{PL}), MGPLE for a Group Size of 9 (MCMCML_{gs9}), and MBGPLE for a Block Size of 4x4 (MCMCML_{bs4}) 235

A.5 Proportion of MCMCML Trials Converging for At Least One of the Three Reference Point Sets (MPLE, MGPLE for a Group Size of 9, or MBGPLE for a Block Size of 4x4) when the Model is an Ising Model Near Phase Transition . 235

A.6 Sample Means and Standard Errors of α, θ, and β (Fire Covariate) when the Model is FI3p and the Methods of Parameter Estimation are PL, GPL for a Group Size of 9 (gs9), BGPL for a Block Size of 4x4 (bs4), and MCMCML . . 237

A.7 Sample Means and Standard Errors of α and θ when the Model is an Ising Model Near Phase Transition and the Methods of Estimation are PL, GPL for Group Sizes of 5 (gs5) and 9 (gs9), BGPL for Block Sizes of 2x2 (bs2), 3x3 (bs3), and 4x4 (bs4), and MCMCML 238

List of Figures

1.1 First-Order Coding Pattern 7

1.2 GPL Structure 13

1.3 BGPL Structure 17

1.4 Recursion Method Visual Aid 22

2.1 1^{st}- and 2^{nd}-Order Neighborhood Structures 49

2.2 5x5 Lattice Example 49

2.3 4x4 Local vs. Global Markovian Example 57

3.1 "L-Shape" Group Structure 87

3.2 2×2 Block Structure 95

4.1 GPL Border Sites 109

4.2 Coloring Schematic Example for a Group and its Boundary Neighbors 112

5.1 Covariate Types 140

5.2 Standard Errors and Estimated Biases of α and θ when the Model is FI2p and the Method of Estimation is GPL for a Group Size of 9 (gs9) 150

5.3 Relative MAEs of α and θ when the Model is FI2p and the Methods of Estimation are GPL for Group Sizes of 5 (gs5) and 9 (gs9) and BGPL for Block Sizes of 2x2 (bs2), 3x3 (bs3), and 4x4 (bs4) 151

5.4 Relative MAEs of α and θ when the Model is an Ising Model Near Phase Transition and the Method of Estimation is MCMCML Under the Following Reference Points: MPLE (MCMCML(PL)), MGPLE for a Group Size of 9 (MCMCML(gs9)), and MBGPLE for a Block Size of 4x4 (MCMCML(bs4)) . . 159

5.5 Proportion of Convergent Trials ((a), (c), (e)) and Proportion of Convergent and Non-Outlying Trials ((b), (d), (f)) when the Model is an Ising Model Near Phase Transition and the Method of Estimation is MCMCML Under the Following Reference Points: MPLE (MCMCML(PL)), MGPLE for a Group Size of 9 (MCMCML(gs9)), and MBGPLE for a Block Size of 4x4 (MCMCML(bs4)) 160

5.6 Relative MAEs of α, θ_h, and θ_v when the Model is FAhv3p and the Methods of Estimation are PL, GPL for a Group Size of 9 (gs9), BGPL for a Block Size of 4x4 (bs4), and MCMCML 170

5.7 Relative MAEs of α and θ when the Model is FI2p and the Methods of Estimation are PL, GPL for a Group Size of 9 (gs9), BGPL for a Block Size of 4x4 (bs4), and MCMCML 172

5.8 Relative MAEs of α and θ when the Model is an Ising Model Near Phase Transition and the Methods of Estimation are PL, GPL for a Group Size of 9 (gs9), BGPL for a Block Size of 4x4 (bs4), and MCMCML 173

6.1 Space-Time Domain Group Structure Example 187

6.2 Space-Time Domain Block Structure Example 190

6.3 Cross-Sections 1 and 11 for Both Covariate Types 196

7.1 Four Terrestrial Ecosystems of Oregon and Washington 207

7.2 Fire Occurrence and Departure from Average (DA) Readings for the Temperate Grasslands Ecosystem – Cross-Sections 1 and 11 of 2002 210

A.1 Relative MAEs of α, θ_h, and θ_v when the Model is FAhv3p and the Methods of Estimation are GPL for a Group Size of 9 (gs9) and BGPL for Block Sizes of 2x2 (bs2), 3x3 (bs3), and 4x4 (bs4) 233

A.2 Relative MAEs of α and θ when the Model is an Ising Model Near Phase Transition and the Method of Estimation is MCMCML Under the Following Reference Points: MPLE (MCMCML(PL)), MGPLE for a Group Size of 9 (MCMCML(gs9)), and MBGPLE for a Block Size of 4x4 (MCMCML(bs4)) . . 236

Chapter 1

Introduction to the Autologistic Model

1.1 Introduction

Consider a forested region that is annually afflicted by wild fires. Envision partitioning a large subsection of that forested region into equally sized pixels. Now associate with each pixel of that subsection a binary random variable which is equal to one if any portion of the region within that pixel experiences fire ignition in some time period, and zero otherwise. Generally, these random variables should not be viewed as independent since pixels within the subsection that are closer together are more apt to share a similar fire status than pixels that are farther apart. Under such a scenario, the subsection can be viewed as a regular lattice of spatially dependent binary random variables. Now suppose, for a cross section in time, observations over the lattice were collected along with pertinent covariate information corresponding to each pixel within the region (i.e. direction of the wind, average wind speed over the region, a measure of relative greenness, etc.). The intuitive statistical question then

becomes: How is the probability structure under such a paradigm modeled? The canonical work by Besag ([2]) not only provided an answer to this question, but also elaborated on how the probability structure of spatially dependent Gaussian, binomial, Poisson, and exponential (i.e. exponential family of distributions) random variables on a regular lattice can intuitively be defined. For the situation encompassing spatially dependent binary random variables on a regular lattice, the model resulting from Besag's work is aptly named the autologistic model. A thorough theoretical treatment of the origins of the autologistic model is provided in Chapter 2, but the simplest type of autologistic model, known as the Ising model, is introduced here, without formal development, to motivate and illustrate the content of this first chapter.

Let $S = \{(i,j) : i = 1,2,\ldots,m_r; j = 1,2,\ldots,m_c\}$ be a finite two-dimensional subset of an infinite-dimensional regular lattice, where m_r and m_c denote the number of rows and columns, respectively, of S. Thus, (i,j) is the site located on the i^{th} row and j^{th} column of S. For notational simplicity, let $n \equiv m_r \times m_c$ be the number of sites and $S = \{i : i = 1,\ldots,n\}$, where the sites of S are labeled top to bottom within columns and left to right across columns, although such an ordering is arbitrary. Let Z_i be a binary random variable located on S, where $\Omega_i = \{0,1\}$ is the corresponding support set. Furthermore, let $\boldsymbol{Z} = (Z_1,\ldots,Z_n)'$ denote the random vector of binary variables over the entire lattice (S), where the joint support is $\boldsymbol{\Omega}^n = \Omega_1 \times \ldots \times \Omega_n$. Define $\mathrm{P}(z_i) \equiv \mathrm{P}(Z_i = z_i)$ to be the probability of the random variable Z_i taking on the value $z_i \in \Omega_i$, and similarly, define $\mathrm{P}(\boldsymbol{z}) \equiv \mathrm{P}(\boldsymbol{Z} = \boldsymbol{z})$ to be the probability of the random vector $\boldsymbol{Z}$ taking on the value $\boldsymbol{z} \in \boldsymbol{\Omega}^n$. The conditional probability form of the Ising model is then given by the following equation:

$$\mathrm{P}(z_i|\{z_j : j \in N_i\}) \;=\; \frac{\exp\{\alpha z_i + \theta \sum_{j\in N_i} z_i z_j\}}{1+\exp\{\alpha + \theta \sum_{j\in N_i} z_j\}}, \qquad i = 1,\ldots,n, \tag{1.1}$$

where N_i is the set of sites neighboring site i (see Definition 2.2), α is the spatial trend parameter, and θ is the spatial dependence parameter. The corresponding joint Ising model

can be shown to have form:

$$\mathrm{P}(\boldsymbol{z}) = \frac{\exp\{\alpha \sum_{i=1}^{n} z_i + \frac{1}{2}\theta \sum_{i=1}^{n} \sum_{j \in N_i} z_i z_j\}}{\sum_{\boldsymbol{y} \in \Omega^n} \exp\{\alpha \sum_{i=1}^{n} y_i + \frac{1}{2}\theta \sum_{i=1}^{n} \sum_{j \in N_i} y_i y_j\}}, \quad \boldsymbol{z} \in \boldsymbol{\Omega}^n. \tag{1.2}$$

Note that the functional form of the Ising model, as given in (1.1), bears a strong resemblance to that of the standard logistic model that is commonly employed with independent binary random variables. This is not a coincidence. If the lattice sites were in fact spatially independent, i.e. $\theta = 0$, then (1.1) reduces to a simple intercept-only logistic model. Thus, the functional form of the autologistic model is identical to that of the logistic model, except that it also conditions on neighboring values of the binary response variable (i.e. the random variable at a given site conditions on *itself* through it neighbors' values). This explains prefacing the term "logistic" with "auto" (meaning "self") in naming the model.

To help solidify an understanding of this relationship between the logistic and autologistic models and to help nurture intuition for the forthcoming exploration of the autologistic model throughout this thesis, consider the fire example discussed at the beginning of this chapter. If the spatial dependence is momentarily ignored, then the conditional probability that a specified region experiences a fire ignition, given pertinent covariate values such as wind direction, average wind speed, and relative greenness, can be modeled using the usual logistic regression model. If, however, the pixel regions of the forest are instead viewed as a Markov random field (i.e. they are assumed to have a spatial dependence structure: see Definition 2.3), then the conditional probability that a specified region experiences a fire ignition, given the site's covariate values and the fire ignition status of all other pixels in the region, depends spatially only on the fire ignition status of the pixels within some neighborhood of the specified site. The effect of these neighboring sites' fire ignition status on the conditional probability is then quantified through one or more spatial dependence parameters corresponding to the number of neighboring sites with fire ignition. Hence, the usual logistic model is extended to

an autologistic model.

The fact that the autologistic model is a seemingly intuitive extension of the logistic model might seem to suggest that its development is void of difficulty and that its implementation is straightforward. In reality, however, its development requires a powerful theorem to make such a formulation even possible, and its implementation often requires the use of computationally expensive estimation methods. More specifically, with the probability formulation coming from a conditional point of view, the fact that the sites of a lattice are arbitrarily labeled means that severe restrictions on the conditional distribution's functional form exist. Such restrictions could have prevented the development of the autologistic model if not for the arrival of the Hammersley-Clifford theorem (see Theorem 2.3), which specifies the requisite restrictions for obtaining a well-defined joint probability structure equivalent to the conditional probability formulation. But even with this powerful result that enables a viable conditional approach to modeling the probability structure, a closed form for the resulting joint density can only be obtained up to a constant of proportionality. To see this, note that the denominator of (1.2) consists of 2^n summands. Even for small lattice samples of size 20×20, this corresponds to $2^{400} \approx 10^{120}$ summands! Consequently, with a generally intractable normalizing constant for the likelihood function, methods other than exact maximum likelihood must be developed and/or implemented to estimate the model parameters, most of which are computationally expensive. In fact, since the autologistic model's first appearance in the literature ([2]), the nearly forty years of subsequent research on the autologistic model has primarily been focused on methods of parameter estimation and their corresponding properties. The central interest of this thesis continues this trend by exploring theoretical properties and comparing practical results of several such estimation methods, including two recently developed methods that attempt to offer a compromise between historical methods. The strong consistency of the estimators for both of these newer methods is shown in Chapter 4 and comprises the major theoretical component of this dissertation.

The remainder of this chapter summarizes relevant, with respect to this dissertation, re-

search on the autologistic model. Section 1.2 will briefly introduce numerous methods for estimating the parameters of the autologistic model; a thorough treatment of four such estimation methods will be covered in Chapter 3. Section 1.3 provides a brief account of numerous data sets the autologistic model has been applied to in the literature, while section 1.4 summarizes the results of the various simulation studies in the literature concerned with parameter estimation for the autologistic model. Section 1.5 briefly summarizes two approaches considered for extending the autologistic model into the space-time domain, as well as the approach employed for this dissertation, and discusses the few applications and simulation studies revolving around such spatio-temporal autologistic models. A formal presentation of spatio-temporal autologistic models is provided in Chapter 6, along with associated methods of parameter estimation. Finally, section 1.6 outlines the objectives of this dissertation for both the autologistic model and the spatio-temporal autologistic model.

1.2 Summary of Parameter Estimation Methods

The autologistic model is an intuitive extension of the logistic model, but estimating its corresponding parameters is nontrivial since the inherent spatial dependence among the sites of the lattice renders maximum likelihood estimation computationally intractable. Over the past four decades numerous methods of parameter estimation have been developed to handle, to varying extents, the intractable normalizing constant of the likelihood function. This section summarizes, in some mathematical detail, many of these estimation methods as well as pertinent, with regard to this dissertation, established properties of the resulting estimators. Four of these methods are presented in full mathematical detail in Chapter 3.

1.2.1 The Coding Method

The coding method was developed by Besag ([1],[2]) in 1974 and was the first estimation method created for autologistic model parameters. The essence of this method is to partition the sites of the lattice into conditionally independent sets, construct and maximize conditional likelihoods for each such set of sites, and then average across these estimates to obtain the corresponding coding method estimates for the parameters of the autologistic model. A more precise description of this method is given below.

The coding method hinges on observing that the lattice sites can be minimally partitioned, via a spatial Markov assumption (see Definition 2.3 and (2.3)), into conditionally independent sets of sites, where the partitioning is dictated by the established neighborhood structure. For instance, Figure 1.1 depicts a 5×5 lattice under a first-order neighborhood system where all of the "X" sites are conditionally independent, given the values at the "O" sites. Similarly, the "O" sites are conditionally independent, given the values at the "X" sites. Under a second-order neighborhood system, there would be four such sets of sites, where every site of a set depends functionally on at least two variable values from each of the three remaining sets. For each set of conditionally independent sites, a conditional likelihood function is then obtained by forming the product of the autologistic conditional probabilities across all sites within the conditionally independent set. For instance, assuming the Ising model (equation (1.1)), and letting S_k denote the k^{th} set of conditionally independent sites (so $S = \bigcup_{k=1}^{K} S_k$), the conditional likelihood for S_k, denoted $\mathrm{CL}_k(\cdot|\cdot)$, is as follows:

$$\mathrm{CL}_k(\alpha, \theta|\boldsymbol{z}) \quad = \quad \prod_{i \in S_k} \frac{\exp\{\alpha z_i + \theta \sum_{j \in N_i} z_i z_j\}}{1 + \exp\{\alpha + \theta \sum_{j \in N_i} z_j\}}, \qquad k = 1, 2, \tag{1.3}$$

where $K = 2$ since the Ising model adheres to a first-order neighborhood system (see Figure 1.1). For each of the K conditional likelihoods, standard likelihood methods are then used to obtain parameter estimates as with ordinary logistic regression. Finally, the K estimates are

Figure 1.1: First-Order Coding Pattern

X	O	X	O	X
O	X	O	X	O
X	O	X	O	X
O	X	O	X	O
X	O	X	O	X

Two sets of conditionally independent sites under a first-order neighborhood system, i.e. all of the "X" sites are conditionally independent and all of the "O" sites are conditionally independent.

then combined (usually averaged) to obtain the coding method estimates for the parameters of the autologistic model.

The coding method really does nothing more than use conditional maximum likelihood several times and combine the resulting set of estimates. The primary appeal of this method is its straightforward implementation and the fact that its estimates are consistent ([3]). Unfortunately, each coding likelihood used to obtain the parameter estimates necessarily models no more than 50% of the data at once. For instance, the conditional likelihood formed from the "X" sites in Figure 1.1 necessarily does not include the conditional distributions of the "O" sites, which comprise half of the lattice sites. To emphasize this limitation, consider a 20 by 20 lattice. Then each coding likelihood is necessarily only modeling at most 200 of the 400 random variables in producing an estimate of the parameters. This results in more variable estimators than could be achieved if all data were simultaneously modeled in the likelihood. Hence, despite the fact that the coding method is easy to implement, its use of a likelihood based on just a subset of the data is most unsatisfactory.

1.2.2 Pseudolikelihood

In 1975, only roughly three years after the coding method was first introduced, the method of pseudolikelihood was developed by Besag ([3]). The driving idea behind pseudolikelihood is to ignore the spatial correlation that exists between the random variables on the lattice, and thus artificially treat the observations as independent and identically distributed realizations. Effectively, the separate conditional likelihoods in the coding method are combined to form a single "pseudo"likelihood. For the Ising model, the pseudolikelihood function, denoted $\mathrm{PL}(\cdot|\cdot)$, is given as:

$$\begin{aligned} \mathrm{PL}(\alpha, \theta|\boldsymbol{z}) &= \prod_{k=1}^{2} \mathrm{CL}_k(\alpha, \theta|\boldsymbol{z}) \\ &= \prod_{i=1}^{n} \frac{\exp\{\alpha z_i + \theta \sum_{j \in N_i} z_i z_j\}}{1 + \exp\{\alpha + \theta \sum_{j \in N_i} z_j\}}. \end{aligned} \tag{1.4}$$

Straightforward maximum likelihood techniques can then be employed on the pseudolikelihood function to determine the parameter estimates. However, because the likelihood function that is being optimized is not the true likelihood function, unless the lattice sites are in fact mutually independent, the method is called maximum pseudolikelihood (MPL) estimation. This renders the asymptotic standard errors generally reported as part of standard logistic regression output invalid. The standard errors for the maximum pseudolikelihood estimates (MPLEs) may be obtained using a resampling technique such as parametric bootstrapping since the observations are not actually independent ([19]).

The obvious advantages of maximum pseudolikelihood estimation include its use of all of the data and the ease with which maximum pseudolikelihood estimates are obtained using any basic statistical software program that has a generalized linear models package, which includes logistic regression, such as R. The primary disadvantage of maximum pseudolikelihood estimation is the fact that it blatantly ignores the spatial dependence structure of the lattice. It should be intuitive, though, that the performance of pseudolikelihood improves

with decreased spatial dependency, since the "closer" the data are to being independent, the "closer" the pseudolikelihood function is to the true likelihood function. In other words, the pseudolikelihood function is the true likelihood function only when all n sites of the lattice are mutually independent. Finally, while Besag non rigorously argued the almost sure convergence of the MPLEs to the true parameter values, many authors have rigorously proved the strong consistency and/or the asymptotic normality of the maximum pseudolikelihood estimators under suitable conditions, including Geman and Graffigne ([13]), Jensen and Møller ([24]), Comets ([7]), and Guyon and Kunsch ([20]).

1.2.3 Markov Chain Monte Carlo Maximum Likelihood

While the coding method and pseudolikelihood established approaches to estimating spatially dependent binary data, these methods did so by somewhat unsatisfactorily circumventing the problem of the intractable normalizing constant. It was not until Geyer and Thompson ([15]) developed the Markov Chain Monte Carlo Maximum Likelihood (MCMCML) method in 1992 that the pesky normalizing constant issue was approached head-on and resolved, at least to some extent. A rigorous presentation of MCMCML is given in section 3.3, but a brief overview of the method, without any theoretical justification, is given here. In particular, MCMCML starts by selecting an arbitrary reference point in the parameter space. The Gibbs sampler, or Metropolis-Hastings algorithm, is then employed to generate a Markov chain of lattice realizations from the joint density of interest (e.g. (1.2)), but with the parameters set equal to the aforementioned reference point. Ideally, this reference point is "close" to the true parameter vector. These realizations are then used to create a Monte Carlo approximate ratio of normalizing constants, where the numerator is evaluated at the true, but unknown, parameter values and the denominator is evaluated at the reference point values. Finally, this Monte Carlo approximate ratio is then used to construct a Monte Carlo approximate likelihood, which is then numerically maximized to obtain the Markov chain Monte Carlo maximum likelihood estimates (MCMCMLEs). The following outline of MCMCML applied

to the Ising model, which comes largely from [18], should help solidify the basic understanding of the method.

Let $\boldsymbol{\psi} = (\alpha, \theta)'$ denote the true, but unknown, parameter values and let $c(\boldsymbol{\psi})$ represent the normalizing constant of the joint Ising model, i.e.

$$c(\boldsymbol{\psi}) \equiv \sum_{\boldsymbol{z} \in \Omega^n} \exp\{\alpha \sum_{i=1}^{n} z_i + \frac{1}{2}\theta \sum_{i=1}^{n} \sum_{j \in N_i} z_i z_j\}, \quad \boldsymbol{\psi} \in \boldsymbol{\Psi}, \tag{1.5}$$

where $\boldsymbol{\Psi} = \mathbb{R} \times \mathbb{R}$ is the two dimensional parameter space. Additionally, let $\boldsymbol{\phi} = (\hat{\alpha}, \hat{\theta})'$ be an initial estimate of $\boldsymbol{\psi}$. The MPLEs are typically a convenient choice for $\boldsymbol{\phi}$, i.e. $\boldsymbol{\phi} \equiv \hat{\boldsymbol{\psi}}_{PL} = (\hat{\alpha}_{PL}, \hat{\theta}_{PL})'$ ([18]). Either the Gibbs sampler or the Metropolis-Hastings algorithm, via the Ising model (see (1.1)), is then used to generate an ergodic Markov chain of sample lattice realizations, $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m$, for m large, from the joint probability mass function given by (1.2), but with $\boldsymbol{\phi}$ playing the role of $\boldsymbol{\psi}$. These m Monte Carlo samples are then used, after sufficient burn-in, to obtain a Monte Carlo approximation for the ratio of normalizing constants, $\frac{c(\boldsymbol{\psi})}{c(\boldsymbol{\phi})}$, which is given by the following equation:

$$\frac{c(\boldsymbol{\psi})}{c(\boldsymbol{\phi})} = \frac{1}{m} \sum_{k=1}^{m} \exp\{S_{\boldsymbol{y_k}}(\alpha - \hat{\alpha}) + N_{\boldsymbol{y_k}}(\theta - \hat{\theta})\}, \tag{1.6}$$

where $\boldsymbol{y_k} = (y_{k1}, \ldots, y_{kn})'$ is the vector of responses on the k^{th} generated lattice, $k = 1, \ldots, m$, and $S_{\boldsymbol{y_k}} = \sum_{i=1}^{n} y_{ki}$ and $N_{\boldsymbol{y_k}} = \frac{1}{2} \sum_{i=1}^{n} \sum_{j \in N_i} y_{ki} y_{kj}$ are the jointly sufficient statistics for the k^{th} realization from the Markov chain. The Monte Carlo approximate negative log-likelihood function, denoted $l_{m,\boldsymbol{z}}(\boldsymbol{\psi})$, is then given by the following equation:

$$\begin{aligned} l_{m,\boldsymbol{z}}(\boldsymbol{\psi}) &= -\log\{\frac{c(\boldsymbol{\psi})}{c(\boldsymbol{\phi})}\} + \alpha S_{\boldsymbol{z}} + \theta N_{\boldsymbol{z}} \\ &= -\log\left\{\frac{1}{m} \sum_{k=1}^{m} \exp\{S_{\boldsymbol{y_k}}(\alpha - \hat{\alpha}) + N_{\boldsymbol{y_k}}(\theta - \hat{\theta})\}\right\} + \alpha S_{\boldsymbol{z}} + \theta N_{\boldsymbol{z}}, \end{aligned} \tag{1.7}$$

where $S_{\boldsymbol{z}}$ and $N_{\boldsymbol{z}}$ are defined analogously to $S_{\boldsymbol{y}_k}$ and $N_{\boldsymbol{y}_k}$, respectively, and represent the

sufficient statistics for the true lattice data. Finally, the Markov chain Monte Carlo maximum likelihood estimates (MCMCMLEs) are obtained by numerically minimizing $l_{m,\boldsymbol{z}}(\boldsymbol{\psi})$ with respect to $\boldsymbol{\psi} = (\alpha, \beta)'$. Asymptotic standard errors of the MCMCMLEs can be obtained by inverting the approximate Markov Chain Monte Carlo observed information matrix ([18]), which is just the Hessian matrix of $l_{m,\boldsymbol{z}}(\boldsymbol{\psi})$ evaluated at the MCMCMLEs.

The impossible-to-overstate advantage of the previously described Markov Chain Monte Carlo maximum likelihood method is the fact that estimates are obtained by approximating the true log-likelihood function, not by ignoring the spatial dependence or by employing conditional likelihoods that model a subset of the data. The resulting MCMCML estimates converge almost surely to the true, but unknown, parameter values as a result of the Mean Ergodic Theorem ([15]), and have been shown in simulation studies to be asymptotically normally distributed ([23]). Furthermore, Graham ([18]) verified via a simulation study that "both the Monte Carlo likelihood ratio test statistic and the Monte Carlo Wald test statistic have their usual asymptotic chi-squared distributions" ([23]). While the method of MCMCML appears to be optimal, it is, however, not without difficulties. Specifically, the primary disadvantages of the MCMCML method are its computational expense, which can be quite substantial, and the fact that while the reference point is in theory arbitrary, it needs to be relatively close, in practice, to the true parameter values or the Markov chain may not converge ([36]). If the spatial dependence is strong, the maximum pseudolikelihood estimates may be poor, and the MCMCML method may not converge with the MPLEs as reference points.

Spurred by the intuitive appeal of MCMCML and the computational efficiency of MPL, researchers began seeking a compromise between the two methods which addressed the spatial dependence, yet did so in a computationally efficient manner. It can be inferred from some of the subsequent research ([11], [22], [32]) that two objectives were to find a method that would either estimate the model parameters almost as effectively as MCMCML, but with reduced computational expense, and/or to improve the reference values for the MCMCML algorithm in situations were pseudolikelihood was performing poorly.

1.2.4 Generalized Pseudolikelihood

The method of generalized pseudolikelihood (GPL) was developed in 2002 by Huang and Ogata ([22]) as a compromise between pseudolikelihood and MCMCML. Ultimately this method is a straightforward extension of pseudolikelihood. A thorough presentation of GPL is given in section 3.4, but a brief overview of the method, as described by Sherman et al. ([32]), is given here; the Ising model will serve as an example throughout the forthcoming discussion.

For each site of the lattice, let $g(i)$ denote the group of sites "associated" with site i, $i = 1, \ldots, n$. Typically, each site's group consists of that site and its nearest neighbors, but in theory the modeler is free to implement any sort of group configuration imaginable and different sites can even have different group configurations. For example, under the Ising model, a common group configuration is $g(i) = \{i \cup N_i\}\ \forall i = 1, \ldots, n$, where N_i are the four first-order neighbors of site i. This creates a 5-site group in the shape of a cross. The set of variables corresponding to the sites of group i are denoted by $\boldsymbol{Z}_{g(i)} = \{Z_k : k \in g(i)\}$. Similarly, $\boldsymbol{Z}^{g(i)} = \{Z_k : k \notin g(i)\}$ denotes the set of variables corresponding to the sites not in the i^{th} group. Additionally, let $N_k^{g(i)}$ denote the neighbors of site k that are members of $g(i)$ and $N_k^{\partial g(i)}$ denote the neighbors of site k that are not members of $g(i)$, i.e. the sites of $N_k^{\partial g(i)}$ are boundary neighbors of group i. Furthermore, $N_k = N_k^{g(i)} \cup N_k^{\partial g(i)}$, where $N_k^{g(i)} \cap N_k^{\partial g(i)} = \emptyset$. For example, in Figure 1.2, under a first-order neighborhood system, $N_{X_0}^{g(X_0)} = \{X_1, X_2, X_3, X_4\}$ and $N_{X_0}^{\partial g(X_0)} = \emptyset$, while $N_{X_1}^{g(X_0)} = \{X_0\}$ and $N_{X_1}^{\partial g(X_0)} = \{B_1, B_2, B_3\}$. The need for such a distinction among the neighbors of a site is not expressed in the literature, but was discovered over the course of this thesis work; more will be said about this crucial distinction in Chapter 3. For each group, $g(i)$, the corresponding joint likelihood function conditioned on the group boundary sites is then constructed. For example, under the Ising model, the joint likelihood for group i, where $\boldsymbol{z}_{g(i)} \in \boldsymbol{\Omega}^{|g(i)|}$, is the following:

Figure 1.2: GPL Structure

e	e	e	e	e	e	e
e	e	e	B_4	e	e	e
e	e	B_2	X_2	B_6	e	e
e	B_1	X_1	X_0	X_4	B_8	e
e	e	B_3	X_3	B_7	e	e
e	e	e	B_5	e	e	e
e	e	e	e	e	e	e

A group, its boundary neighbors, and the corresponding external sites under the Ising model on a $7x7$ lattice subset of S. Sites X_1, X_2, X_3, and X_4 are the neighbors and additional group members corresponding to site X_0, while the 8 B sites represent the fixed boundary neighbors of the group. The 36 e sites represent the external sites of the group, i.e., the sites that are conditionally independent of the group sites under the spatial Markov assumption.

$$\mathrm{P}(\boldsymbol{z}_{g(i)}|\boldsymbol{z}^{g(i)}) = \frac{\exp\left\{\alpha\sum_{k\in g(i)} z_k + \theta\left\{\frac{1}{2}\sum_{k\in g(i)}\sum_{j\in N_k^{g(i)}} z_k z_j + \sum_{k\in g(i)}\sum_{j\in N_k^{\partial g(i)}} z_k z_j\right\}\right\}}{\sum_{\boldsymbol{y}\in\Omega^{|g(i)|}} \exp\left\{\alpha\sum_{k\in g(i)} y_k + \theta\left\{\frac{1}{2}\sum_{k\in g(i)}\sum_{j\in N_k^{g(i)}} y_k y_j + \sum_{k\in g(i)}\sum_{j\in N_k^{\partial g(i)}} y_k y_j\right\}\right\}}, \tag{1.8}$$

where $|g(i)|$ is the number of sites in group i. For the Ising model, $|g(i)| = 5\ \forall i = 1, \ldots, n$, when $g(i)$ is as defined above. By the spatial Markov assumption, the conditioned sites of the group joint distributions are reduced from all other sites of the lattice (i.e. $\boldsymbol{z}^{g(i)}$) to just the neighboring boundary sites of the group, i.e., the sites that are neighbors of at least one group site but are not themselves a member of the group (see Figure 1.2). The sites that are neither a member of group i nor a boundary neighbor of group i are referred to as the external sites of group i. Since the number of sites within a group is relatively small and the boundary neighbor sites are viewed as fixed, the normalizing constant of a group's joint

likelihood can be obtained fairly easily via brute force. For example, with the Ising model, there are five sites in a group (and eight fixed boundary neighbor sites), which means that the corresponding normalizing constant (denominator of right hand side of (1.8)) contains 2^5 summands, which can be computed relatively quickly. The generalized pseudolikelihood function is then constructed by taking the product over all n group joint likelihood functions. For example, the generalized pseudolikelihood function, denoted $\mathrm{GPL}(\cdot|\cdot)$, for the Ising model is the following:

$$\begin{aligned}\mathrm{GPL}(\alpha,\theta|\boldsymbol{z}) &= \prod_{i=1}^{n}\mathrm{P}(\boldsymbol{z}_{g(i)}|\boldsymbol{z}^{g(i)}) \\ &= \prod_{i=1}^{n}\frac{\exp\left\{\alpha\sum\limits_{k\in g(i)}z_k+\theta\left\{\frac{1}{2}\sum\limits_{k\in g(i)}\sum\limits_{j\in N_k^{g(i)}}z_kz_j+\sum\limits_{k\in g(i)}\sum\limits_{j\in N_k^{\partial g(i)}}z_kz_j\right\}\right\}}{\sum\limits_{\boldsymbol{y}\in\Omega^{|g(i)|}}\exp\left\{\alpha\sum\limits_{k\in g(i)}y_k+\theta\left\{\frac{1}{2}\sum\limits_{k\in g(i)}\sum\limits_{j\in N_k^{g(i)}}y_ky_j+\sum\limits_{k\in g(i)}\sum\limits_{j\in N_k^{\partial g(i)}}y_ky_j\right\}\right\}}. \quad (1.9)\end{aligned}$$

Note that for a 20×20 lattice (with a neighboring buffer), there are a total of 400 groups each with 32 normalizing constant summands, for a total of $12,800$ pieces to compute. This offers a significant reduction to the 10^{120} summands required by the true likelihood. The corresponding maximum generalized pseudolikelihood estimates (MGPLEs) are obtained by numerically maximizing the generalized pseudolikelihood function with respect to the model parameters. As was the case with the MPLEs, standard errors for the MGPLEs can be obtained using a resampling technique, such as parametric bootstrapping, since the generalized pseudolikelihood function, under any reasonable group size, still does not fully account for the spatial dependency.

Generalized pseudolikelihood is a compromise between PL and MCMCML for the following reasons. If each site's group consists only of that site, i.e. $g(i)=\{i\}\ \forall i=1,\ldots,n$, then GPL reduces to PL since the generalized pseudolikelihood function becomes the pseudolikelihood function. For example, under the Ising model, if $g(i)=\{i\}$, then (1.4) and (1.9) are equivalent.

On the other hand, if each site's group consists of the entire lattice, i.e. $g(i) = \{1, \ldots, n\}$ $\forall i = 1, \ldots, n$, then each group's joint likelihood is actually the true likelihood. For example, under the Ising model, if $g(i) = \{1, \ldots, n\}$, then (1.2) and (1.8) are equivalent. Thus, while PL makes no adjustments to the normalizing constant and MCMCML cleverly approximates the normalizing constant to fully accommodate the spatial dependency, GPL partially adjusts (the extent depends on group size) the normalizing constant by accommodating the spatial dependency within groups of the lattice, but without the computational expense.

The primary advantage of the method of generalized pseudolikelihood is the fact that the user can dictate the amount of implemented compromise between PL and MCMCML by altering the size of the groups. Hence, in situations of relatively strong spatial dependency, where PL is relatively inefficient and MCMCML is perhaps too computationally expensive, GPL should in theory produce estimates that are more efficient than the corresponding MPLEs, but at a cheaper computational expense than the MCMCMLEs. The noteworthy disadvantages of GPL include the following. First, if the group size is too large, regardless of the strength of the spatial dependency, the computational expense of GPL will swiftly increase as a function of group size. Second, since a resampling method is generally implemented to obtain valid standard errors for the MGPLEs, the computational expense in obtaining standard errors could become substantial for even moderately sized groups. Therefore, MCMCML will generally be preferable to GPL when employed using large group sizes. Despite these limitations of GPL, Huang and Ogata ([22]) suggested that in the presence of very strong spatial dependency, the MGPLEs that are obtained using a reasonable group size can still possibly be used as the reference point for the MCMCML algorithm when the MPLEs may be too poor for the Markov chain to converge. This claim will be analyzed as part of a simulation study in Chapter 5.

It should also be mentioned that Sherman et al. ([32]) observed an interesting phenomenon with regard to the MGPLEs that was attributed to edge effects (see section 2.2.4 for a discussion of edge effects). Specifically, while one might intuitively speculate that the improvement

in efficiency the MGPLEs yield over the MPLEs is a monotone increasing function of group size, simulations revealed that if the group size becomes too large, the increase in efficiency is outweighed by the increasing influence of edge effects. Finally, while GPL is a relatively simple extension of PL, the strong consistency and asymptotic normality of the MGPLEs, under suitable conditions, have hitherto remained unproven in the literature. A proof of the strong consistency of the MGPLEs is provided in Chapter 4.

1.2.5 Block Generalized Pseudolikelihood

Another method that serves as a compromise between PL and MCMCML, that was presented in an unpublished paper by Lim et al. ([27]) and in a Bayesian setting by Friel et al. ([11]), is the method of block generalized pseudolikelihood (BGPL). As with GPL, this method is a relatively straightforward extension of PL. A thorough presentation of BGPL is given in section 3.5, but a brief overview of the method, as described by [11], but in a classical (not Bayesian) setting, is given here. Once again the Ising model will serve as an example throughout the forthcoming discussion.

The lattice of interest is first partitioned into $L \leq n$ disjoint blocks, where $b(l)$ denotes the sites "associated" with block l, $l = 1, \ldots, L$ (so $S = \bigcup_{l=1}^{L} b(l)$). Typically, the lattice is partitioned so that each block has an equal number of sites (if possible), but in theory the modeler is free to partition in such a way that blocks have different numbers of sites. More practical blocking mechanisms include forming $b \times b$ blocks across the lattice for $b = 2, 3,$ or 4, or letting multiple adjacent columns (or rows) represent a block. For illustrative purposes, an Ising model in which a 3×3 block structure has been established will be assumed. The set of variables corresponding to the sites of block l are denoted by $\boldsymbol{Z}_{b(l)} = \{Z_k : k \in b(l)\}$, $l = 1, \ldots, L$. Similarly, $\boldsymbol{Z}^{b(l)} = \{Z_k : k \notin b(l)\}$ denotes the set of variables corresponding to the sites not in the l^{th} block. Additionally, let $N_k^{b(l)}$ denote the neighbors of site k that are members of block l and $N_k^{\partial b(l)}$ denote the neighbors of site k that are on the boundary of block

Figure 1.3: BGPL Structure

e	e	e	e	e	e	e
e	e	B_4	B_6	B_8	e	e
e	B_1	X_1	X_4	X_7	B_{10}	e
e	B_2	X_2	X_5	X_8	B_{11}	e
e	B_3	X_3	X_6	X_9	B_{12}	e
e	e	B_5	B_7	B_9	e	e
e	e	e	e	e	e	e

A 3×3 block, its boundary neighbors, and the corresponding external sites under the Ising model on a $7x7$ lattice subset of S. The 9 X sites are the members of the block, while the 12 B sites represent the fixed boundary neighbors of the block. The 28 e sites represent the external sites of the block, i.e., the sites that are conditionally independent of the block sites according to the spatial Markov assumption.

l, i.e. the sites of $N_k^{\partial g(i)}$ are boundary neighbors of block l. Furthermore, $N_k = N_k^{b(l)} \cup N_k^{\partial b(l)}$, where $N_k^{b(l)} \cap N_k^{\partial b(l)} = \emptyset$. For example, suppose the block given in Figure 1.3 is the l^{th} block in the partitioning of S, then under a first-order neighborhood system, $N_{x_5}^{b(l)} = \{X_2, X_4, X_6, X_8\}$ and $N_{x_5}^{\partial b(l)} = \emptyset$, while $N_{x_1}^{b(l)} = \{X_2, X_4\}$ and $N_{x_1}^{\partial b(l)} = \{B_1, B_4\}$, and $N_{x_4}^{b(l)} = \{X_1, X_5, X_7\}$ and $N_{x_4}^{\partial b(l)} = \{B_6\}$. As with GPL, the need for such a distinction among the neighbors of a site is not expressed in the literature, but its importance was realized over the course of this thesis work; more will be said about this crucial distinction in Chapter 3. For each of these blocks, the corresponding joint likelihood function is then constructed. For example, under the Ising model, the joint likelihood for block l, where $\boldsymbol{z}_{b(l)} \in \boldsymbol{\Omega}^{|b(l)|}$, is the following:

$$\mathrm{P}(\boldsymbol{z}_{b(l)}|\boldsymbol{z}^{b(l)}) = \frac{\exp\left\{\alpha \sum_{k\in b(l)} z_k + \theta\left\{\frac{1}{2}\sum_{k\in b(l)}\sum_{j\in N_k^{b(l)}} z_k z_j + \sum_{k\in b(l)}\sum_{j\in N_k^{\partial b(l)}} z_k z_j\right\}\right\}}{\sum_{\boldsymbol{y}\in\Omega^{|b(l)|}} \exp\left\{\alpha \sum_{k\in b(l)} y_k + \theta\left\{\frac{1}{2}\sum_{k\in b(l)}\sum_{j\in N_k^{b(l)}} y_k y_j + \sum_{k\in b(l)}\sum_{j\in N_k^{\partial b(l)}} y_k y_j\right\}\right\}}, \quad (1.10)$$

where $|b(l)|$ is the number of sites in the l^{th} block. For the Ising model, $|b(l)| = 9\ \forall l = 1, \ldots, L$, when $b(l)$ is as defined above. By a spatial Markov assumption, the conditioned sites of the block joint distributions are reduced from all other sites of the lattice (i.e. $\boldsymbol{z}^{b(l)}$) to just the neighboring boundary sites of the block, i.e., the sites that are neighbors of at least one block site but are not themselves a member of the block (see Figure 1.3). The sites that are neither a member of block l nor a boundary neighbor of block l are referred to as the external sites of block l. Since the number of sites within a block is relatively small and the boundary neighbor sites are viewed as fixed, the normalizing constant of a block's joint likelihood can be obtained fairly easily via brute force or via a recursion method (see section 1.2.6). For example, with the Ising model, there are nine sites in a block (and twelve fixed boundary neighbor sites), which means that the corresponding normalizing constant (denominator of right hand side of (1.10)) contains 2^9 summands, which can be computed relatively quickly. The block generalized pseudolikelihood function is then constructed by taking the product over all L block joint conditional likelihood functions. For example, the block generalized pseudolikelihood function, denoted BGPL$(\cdot|\cdot)$, for the Ising model is the following:

$$\begin{aligned}
\mathrm{BGPL}(\alpha, \theta|\boldsymbol{z}) &= \prod_{l=1}^{L} \mathrm{P}(\boldsymbol{z}_{b(l)}|\boldsymbol{z}^{b(l)}) \\
&= \prod_{l=1}^{L} \frac{\exp\left\{\alpha \sum_{k\in b(l)} z_k + \theta\left\{\frac{1}{2}\sum_{k\in b(l)}\sum_{j\in N_k^{b(l)}} z_k z_j + \sum_{k\in b(l)}\sum_{j\in N_k^{\partial b(l)}} z_k z_j\right\}\right\}}{\sum_{\boldsymbol{y}\in\Omega^{|b(l)|}} \exp\left\{\alpha \sum_{k\in b(l)} y_k + \theta\left\{\frac{1}{2}\sum_{k\in b(l)}\sum_{j\in N_k^{b(l)}} y_k y_j + \sum_{k\in b(l)}\sum_{j\in N_k^{\partial b(l)}} y_k y_j\right\}\right\}}. \quad (1.11)
\end{aligned}$$

The corresponding maximum block generalized pseudolikelihood estimates (MBGPLEs) are obtained by numerically maximizing the block generalized pseudolikelihood function with respect to the model parameters. As was the case with the MPLEs and the MGPLEs, standard errors for the MBGPLEs can be obtained using a resampling technique, such as parametric bootstrapping, since the block generalized pseudolikelihood function, under any reasonable block size, still does not fully account for the spatial dependency.

Block generalized pseudolikelihood is a compromise between PL and MCMCML for essentially the same reasons as GPL. Specifically, if each block consists only of a single site, i.e. $b(l) = \{l\}\ \forall l = 1, \ldots, n$ (so $L = n$), then BGPL reduces to PL since the block generalized pseudolikelihood function becomes the pseudolikelihood function. For example, under the Ising model, if $b(l) = \{l\}$, then (1.4) and (1.11) are equivalent. Furthermore, if there is only one block that consists of the entire lattice, i.e. $b(1) = \{1, \ldots, n\}$ (so $L = 1$), then that block's joint likelihood is actually the true likelihood. For example, under the Ising model, if $b(1) = \{1, \ldots, n\}$, then (1.2) and (1.10) are equivalent. Thus, just as with GPL, while PL makes no adjustments to the normalizing constant and MCMCML cleverly approximates the normalizing constant to fully accommodate the spatial dependency, BGPL partially adjusts (the extent depends on block size) the normalizing constant by accommodating the spatial dependency within blocks of the lattice.

Before discussing properties of the MBGPLEs, as well as pros and cons of BGPL, the difference between GPL and BGPL should be clarified. Under GPL each site has a group associated with it, but each site can belong to multiple groups; for instance, under the Ising model with $g(i) = \{i \cup N_i\}$, any given internal site of the lattice is a neighbor to four other sites on the lattice and, thus, that internal site belongs to five distinct groups. On the other hand, under BGPL, each site belongs to only one block. Hence, under BGPL each site only appears in one of the block joint likelihoods that are used to form the BGPL function. Under GPL, each site can appear in several of the group joint likelihoods that are used to form the GPL function. In other words, groups overlap, but blocks do not. This subtlety is the only difference between the two methods and if both the block and the group size are one, then GPL and BGPL both reduce to PL. A thorough performance-based comparison of these two methods does not currently exist in the literature, but is a topic of Chapter 5.

Returning to points of interest with BGPL, one substantial advantage of BGPL, just as with GPL, is the fact that it allows the user to dictate the amount of compromise between PL and MCMCML. Specifically, the extent of the compromise that is expressed through the BGPL

function is determined by the block size. Hence, both BGPL and GPL can be used in place of PL or MCMCML in situations of moderate to strong spatial dependency, where the unadjusted normalizing constant of the PL function causes efficiency issues in the method and the approximate normalizing constant obtained through MCMCML can make the method computationally expensive. The disadvantages of BGPL parallel those of GPL. First and foremost, when and if large block sizes are used to model lattices exhibiting moderate to strong spatial dependency, the computational expense of BGPL will markedly increase. Do note, however, that if block size is equal to group size, everything else being held equal, BGPL should be less computationally expensive than GPL due to the redundancy or overlap present in the GPL function. The second limitation of BGPL, as with GPL, centers on obtaining standard errors through a resampling method. Since such a method of obtaining standard errors can itself be computationally expensive, the standard errors of the MBGPLEs can be costly to ascertain under a relatively large block size. Thus, in situations where large block sizes are needed, MCMCML will presumably be preferable to BGPL. Furthermore, although it is not suggested in the literature, in the presence of strong spatial dependence, the MBGPLEs obtained using a reasonable block size can still be used, just as the MGPLEs can, as the reference point for the MCMCML algorithm when the MPLEs are ineffective at producing a convergent Markov chain. Finally, as with the MGPLEs, the strong consistency and asymptotic normality of the MBGPLEs remain unproven. A proof of the strong consistency of the MBGPLEs is provided in Chapter 4.

1.2.6 Recursion Method

To efficiently calculate the group/block normalizing constants using GPL/BGPL, a method known as the Recursion method can be employed. This method was originally presented in Reeves et al. ([30]) as a means for computing the exact maximum likelihood estimates for small lattices, but was adapted for supplementary use with BGPL by Pettitt et al. ([11]). The essence of the original Recursion method, as described in Pettitt et al. ([11]), is to use a

recursive formula to calculate the normalizing constant so that maximum likelihood methods can be directly implemented. A fairly detailed overview of this method for calculating an intractable normalizing constant, as presented in [30], is given here; the Ising model will again serve as an example throughout the forthcoming development.

Let $q(\boldsymbol{z})$ denote the unnormalized joint likelihood function for the discrete random vector $\boldsymbol{z}$. For the Ising model, $q(\boldsymbol{z})$ would be the numerator of (1.2):

$$q(\boldsymbol{z}) = \exp\{\alpha \sum_{i=1}^{n} z_i + \frac{1}{2}\theta \sum_{i=1}^{n} \sum_{j \in N_i} z_i z_j\}, \quad \boldsymbol{z} \in \boldsymbol{\Omega}^n. \tag{1.12}$$

Let $r < n$ denote the *lag* for the algorithm and let $k = n - r$. For the Ising model, $r = \min\{m_r, m_c\}$, where m_r is the number of lattice rows and m_c is the number of lattice columns. The following factorization of $q(\cdot)$, referred to as the *lag-r model*, can then be obtained:

$$q(\boldsymbol{z}) = q_1(z_1, z_2, \ldots, z_{r+1}) q_2(z_2, z_3, \ldots, z_{r+2}) \ldots q_k(z_k, z_{k+1}, \ldots, z_n). \tag{1.13}$$

In the context of the Ising model, where for illustrative purposes it is assumed that the lattice of interest is $3x4$ (so $n = 12$, $r = 3$, and $k = 9$) as in figure 1.4, there are nine such functions, $q_1(\cdot), \ldots, q_9(\cdot)$, which are defined as follows:

$$q_1(z_1, \ldots, z_4) = \exp\{\alpha \sum_{i=1}^{4} z_i + \beta(z_1 z_2 + z_2 z_3 + z_1 z_4)\}, \tag{1.14}$$

$$q_u(z_u, \ldots, z_{u+3}) = \exp\{\alpha z_{u+3} + \beta z_{u+3}(z_u + z_{u+3-1})\}, \quad u = 2, 3, 5, 6, 8, 9, \tag{1.15}$$

$$q_w(z_w, \ldots, z_{w+3}) = \exp\{\alpha z_{w+3} + \beta z_{w+3} z_w\}, \quad w = 4, 7. \tag{1.16}$$

The visual intuition behind (1.14), (1.15), and (1.16) is aided by the following explanation, when coupled with Figure 1.4. Specifically, $q_1(z_1, \ldots, z_4)$ accounts for the $\sum_{i=1}^{4} z_i$ portion of the jointly sufficient statistic associated with the spatial trend parameter ($\sum_{i=1}^{12} z_i$), while the $(z_1 z_2 + z_2 z_3 + z_1 z_4)$ terms account for the vertical neighbor relations between sites 1 and 2 and sites 2 and 3, and the one horizontal neighbor relation between sites 1 and 4; all

Figure 1.4: Recursion Method Visual Aid

Z_1	Z_4	Z_7	Z_{10}
Z_2	Z_5	Z_8	Z_{11}
Z_3	Z_6	Z_9	Z_{12}

Visual aid for understanding how the $q(\cdot)$ functions of the Recursion method are defined when working with the Ising model.

three of which are part of the jointly sufficient statistic for the spatial dependence parameter ($\frac{1}{2}\sum_{i=1}^{12}\sum_{j\in N_i} z_i z_j$). Next, $q_2(z_2,\ldots,z_5)$ accounts for the z_5 component of the jointly sufficient statistic associated with the spatial trend parameter, while the $z_5(z_2+z_4)$ term accounts for the one vertical neighbor relation between sites 4 and 5, and the one horizontal neighbor relation between sites 2 and 5; both of which are part of the jointly sufficient statistic for the spatial dependence parameter. Similar statements apply to $q_u(z_u,\ldots,z_{u+3})$, $u = 3, 5, 6, 8,$ and 9, and analogous statements can be made for $q_4(z_4,\ldots,z_7)$ and $q_7(z_7,\ldots,z_{10})$. Thus, the $q(\cdot)$ functions are defined in such a way that when they are all multiplied together, each neighbor relation will have been accounted for and the jointly sufficient statistics will have been correctly manufactured.

Now let $(z_i, z_{i+1},\ldots,z_j)$ be concisely denoted by $\boldsymbol{z}_i^j$. The normalizing constant for $q(\boldsymbol{z})$, $c(\boldsymbol{\psi})$, where $\boldsymbol{\psi}$ is the parameter vector, is then given by the following equation:

$$\begin{aligned} c(\boldsymbol{\psi}) &= \sum_{\boldsymbol{z}} q(\boldsymbol{z}) \\ &= \sum_{\boldsymbol{z}_{k+1}^n}\sum_{z_k} q_k(\boldsymbol{z}_k^n)\sum_{z_{k-1}} q_{k-1}(\boldsymbol{z}_{k-1}^{n-1})\ldots\sum_{z_1} q_1(\boldsymbol{z}_1^{r+1}). \end{aligned} \tag{1.17}$$

Returning to the $3x4$ Ising model example, (1.17) becomes the following:

$$c(\alpha,\beta) = \sum_{\boldsymbol{z}\in\Omega^{12}} q(\boldsymbol{z})$$

$$= \sum_{\boldsymbol{z}_{10}^{12} \in \Omega^3} \sum_{z_9 \in \{0,1\}} q_9(\boldsymbol{z}_9^{12}) \sum_{z_8 \in \{0,1\}} q_8(\boldsymbol{z}_8^{11}) \ldots \sum_{z_1 \in \{0,1\}} q_1(\boldsymbol{z}_1^4). \tag{1.18}$$

Now define the following two functions:

$$Q_1(\boldsymbol{z}_2^{r+1}) = \sum_{z_1} q_1(\boldsymbol{z}_1^{r+1}), \tag{1.19}$$

$$Q_t(\boldsymbol{z}_{t+1}^{r+t}) = \sum_{z_t} q_t(\boldsymbol{z}_t^{r+t}) Q_{t-1}(\boldsymbol{z}_t^{r+t-1}), \quad t = 2, \ldots, k. \tag{1.20}$$

In the context of the 3×4 Ising model example, (1.19) and (1.20) become the following:

$$Q_1(\boldsymbol{z}_2^4) = \sum_{z_1 \in \{0,1\}} q_1(\boldsymbol{z}_1^4), \tag{1.21}$$

$$Q_t(\boldsymbol{z}_{t+1}^{3+t}) = \sum_{z_t \in \{0,1\}} q_t(\boldsymbol{z}_t^{3+t}) Q_{t-1}(\boldsymbol{z}_t^{3+t-1}), \quad t = 2, \ldots, k. \tag{1.22}$$

Finally, the intractable normalizing constant can then be recursively evaluated as follows:

$$c(\boldsymbol{\psi}) = \sum_{\boldsymbol{z}_{k+1}^n} Q_k(\boldsymbol{z}_{k+1}^n). \tag{1.23}$$

In light of (1.23), the normalizing constant corresponding to the $3x4$ Ising model example is then recursively evaluated as follows:

$$c(\alpha, \beta) = \sum_{\boldsymbol{z}_{10}^{12} \in \Omega^3} Q_9(\boldsymbol{z}_{10}^{12}). \tag{1.24}$$

Hence, the Recursion method provides a recursive approach to calculating the intractable normalizing constant of the likelihood function associated with a lattice of spatially dependent random variables, which then makes exact maximum likelihood methods possible.

In addition to providing a means for computing the normalizing constant of an intractable likelihood function, the recursion method also provides this approach at a computational expense that is less than that of brute force [30]. Unfortunately, the method is really only

viable for lattices where the smaller of the number of rows and the number of columns is at most 20, and the other dimension is not unreasonably large ([11]). Hence, the method was borrowed for supplementary use with BGPL where block sizes were sufficiently small ([11]). Specifically, the recursive formula for calculating the normalizing constant is used to more efficiently (relative to direct computations) calculate the normalizing constant of each block joint likelihood function, thus alleviating, in theory, some of the computational expense of BGPL. While it is not discussed in the literature, there is no reason, in theory, why the Recursion method couldn't also be used in a supplementary fashion with GPL to more efficiently calculate the normalizing constant of each group joint likelihood function.

Since the Recursion method, on its own, is only viable with relatively small lattices and the simulation studies carried out for this thesis (see Chapter 5) focused on lattices larger than 20×20, the method was only considered for use as a supplement to both BGPL and GPL. Once the Recursion method was implemented with GPL and BGPL, however, it became apparent that the method did not offer a substantial improvement in computational expense when computing the normalizing constants associated with the group and block joint likelihoods. Consequently, the Recursion method will not be discussed further in this thesis.

1.2.7 Additional Estimation Methods

While the aforementioned estimation methods comprise the focus of this dissertation, minus the coding method (for reasons discussed in sections 1.3 and 1.4 below) and the Recursion method, other estimation methods do exist, two of which are briefly discussed here. The first is Markov Chain Monte Carlo Stochastic Approximation (MCMC-SA) which, as described by He et al. ([21]), is a "two-stage stochastic approximation" algorithm that is ultimately an iterative extension of MCMCML. In particular, once the MCMCML estimate of $\boldsymbol{\psi}$ is found, the Gibbs sampler is then run again, but with the MCMCMLEs as the reference point ($\boldsymbol{\phi}$), and the resulting chain of realizations is then used to obtain new MCMCMLEs. Note

that an important distinction between this method and MCMCML is that with MCMC-SA the numerical optimization is carried out using a stochastic approximation of the Newton-Raphson algorithm. This process continues in an iterative fashion until some pre-specified tolerance is reached; the resulting estimates signify the completion of the first stage of MCMC-SA. The second and final stage of MCMC-SA iteratively repeats the first stage, where the initial value for $\boldsymbol{\phi}$ in the Gibbs sampler is set equal to the final estimate of $\boldsymbol{\psi}$ obtained from stage I, but now the numerical optimization is performed by coupling the stochastic approximation of the Newton-Raphson algorithm with an averaging procedure. Once a pre-specified tolerance is achieved within stage II, the MCMC-SA estimate of $\boldsymbol{\psi}$ is then just the MCMCMLEs obtained from the final iteration of stage II. The method of MCMC-SA is not considered in this dissertation as it is substantially more expensive computationally and, more importantly, simulation results revealed little difference between this method and standard MCMCML ([21]). Finally, the reduced dependency approximation method, or RDA, developed by Pettitt et al. ([11]), aims to better approximate the normalizing constant for large lattices using the recursion method. RDA is not considered in this thesis since it is essentially just an extension of the Recursion method.

1.3 Applications of Estimation Methods

The estimation methods for the autologistic model discussed in section 1.2 have been applied to a wide range of real data sets in the literature. Arguably the most commonly implemented method in the literature, despite its documented limitations (see section 1.4), is PL, which is routinely either the only estimation method used or at least one of several methods used. The MPLEs are routinely calculated for, if no other reason, use as the reference values for MCMCML and/or the starting values in the numerical optimization of the GPL and BGPL functions. Pseudolikelihood's popularity is almost certainly due to its straightforward implementation, but it is also popular because the MPLEs are "adequate for most purposes"

when the spatial dependency is small ([36]). Markov chain Monte Carlo maximum likelihood is also an extremely popular estimation method in the literature and, from a purely empirical perspective, it seems to have become the "gold standard" since the late 1990s. Use of the coding method is seen predominately either right after the autologistic model was first introduced, or in papers conducting a simulation study comparing several estimation methods, which are also then typically applied to a real data set. The coding method's limited number of appearances in the literature is most likely attributable to both the inefficiency of the estimates obtained from the coding method ([3]) and the fact that PL, a more efficient method (see section 1.4), was presented in [3] shortly after the autologistic model was first introduced in 1974. As several simulation studies (see section 1.4) have demonstrated that the coding method is inferior to both PL and MCMCML ([35], [36], [18]), any use of the coding method in the literature is hard to find past the late 1990s. Finally, as GPL and BGPL are both relatively new estimation methods, their use in the literature is, to date, sparse. The remainder of this section recounts, in limited detail, a representative sampling of examples in the literature where the estimation methods for the autologistic model have been applied to real data sets.

The coding method was employed to model the spatial distribution (presence/absence) of the weed *Plantago lanceolata* in defunct mine workings in Flintshire, England ([2]). Pseudolikelihood, along with parametric bootstrapping, was implemented to model the spatial distribution (presence/absence) of a phytophthora epidemic in bell pepper plants, while incorporating soil water content and leaf disk assays as covariates ([19]). The spatial distribution (presence/absence) of mountain pine beetle in a stand of lodgepole pine trees in Oregon was modeled, while incorporating the covariates age, vigor, and log of diameter breast height, using pseudolikelihood to estimate the model parameters and a parametric bootstrap procedure to determine the corresponding standard errors ([29]). Pseudolikelihood was also used, under an anisotropic first-order neighborhood system, to model the spatial distribution (presence/absence) of the citrus sudden death (CSD) disease in a grove of citrus trees in Minas

Gerais State, Brazil ([25]). The spatial distribution (presence/absence) of the plant species *Rumex acetosella* (red sorrel) in abandoned agricultural fields in the Piedmont region of New Jersey, U.S., was modeled separately over 11 time points (annually from 1968 to 1978) using pseudolikelihood, while incorporating the presence/absence of the plant species *Lonicera japonica* (Japanese honeysuckle) as a cofactor; a parametric bootstrap was employed to obtain the corresponding standard errors ([5]).

Huffer and Wu applied the autologistic model, on several different occasions in the literature, to describe the spatial distribution of two Florida plant species ([35], [36], [23]). Initially, the spatial distribution (presence/absence) of *Castanea pumila* was modeled, while incorporating the climate covariate median freeze-free period (in days), or FZF, using the coding method, PL, and MCMCML to estimate the corresponding parameters ([35]). The coding method, PL, and MCMCML were again used to model the spatial distributions (presence/absence) of *Castanea pumila* and *Zanthoxylum Clava-herculis* while incorporating climate variables ([36]). Specifically, the presence/absence of *Castanea pumila* was again modeled, but this time incorporating three climate covariates: mean annual temperature (^{o}C), mean total annual precipitation (mm), and a moisture index. Furthermore, the presence/absence of *Zanthoxylum Clava-herculis* was also modeled while incorporating the climate covariates mean temperature (^{o}C) of coldest month, mean minimum temperature (^{o}C) of coldest month, and elevation (ft). Finally, the spatial distribution of *Castanea pumila* was additionally modeled in [23], while again incorporating the climate covariate FZF, but only using MCMCML to estimate the model parameters.

Pseudolikelihood and MCMC were employed, under a second-order neighborhood system, to model, separately, both the spatial distributions (presence/absence) of the subarctic evergreen woodland and the boreal evergreen forest, while incorporating climate covariates, in British Columbia, Canada. Generalized pseudolikelihood, under five possible group sizes, was used to model the spatial distribution of Wiebe's wheat yield data ([22]) once the original observations had been transformed into binary data according to whether a site's yield was at least as

large as the median site yield ([22]). Pseudolikelihood, GPL, and MCMC were implemented to model the spatial distribution (high or low) of liver cancer mortality for 2,003 counties of the eastern United States during the 1950's ([32]). Finally, PL and BGPL were used, in a Bayesian setting, to model the spatio-temporal distribution that corresponds to whether the 72 genes that make up the mitochondrial chromosome of the *Plasmodium falciparum* genome are down- or up-regulated over 46 1-hr. intervals ([11]). Note that while this data set is a 46x72 spatio-temporal lattice in one-dimensional space, it can still be treated as a two-dimensional lattice of spatial binary random variables, with time treated as a second spatial dimension.

For this thesis work, PL, MCMCML, GPL, and BGPL are used to model a cross-section in time of the spatial distribution (presence/absence) of fire ignition in Oregon and Washington state while incorporating a MODIS-derived (Moderate Resolution Imaging Spectroradiometer) fire potential metric, Departure from Average (DA), as a covariate. The data for this analysis, which cover the months between May and October for each of the four years from 2002 to 2005, were graciously shared by Cindy Leary, a researcher at The University of Montana, during the course of her work with the Fire Sciences Laboratory in Missoula, Montana. The following brief description of the data set is a summary of the more exhaustive description given in [26]. In particular, the spatial domain encompassed by Oregon and Washington that is modeled comprises an 870 km by 740 km regular lattice that consists of four distinct terrestrial ecosystems including Temperate Broadleaf and Mixed Forests, Temperate Coniferous Forests, Temperate Grasslands, Savannas, and Shrublands, and finally, Deserts and Xeric Shrublands. The Moderate Resolution Imaging Spectroradiometer (MODIS) satellite platform data, obtained from "NASA's Land Processes Distributed Active Archive Center (LP DAAC)", were used to derive the Enhanced Vegetation Index (EVI) that is ultimately used to produce the DA values for each 1 km^2 pixel. Furthermore, the binary fire data for the spatial domain were constructed by first "merging state and federal fire occurrence records for Oregon and Washington into a single historical database" and then converting these data into a lattice of binary values where a pixel (or site) receives a value of 1 if any location within the

1 km^2 pixel experiences fire ignition, and a value of 0 otherwise. A more detailed description of the data set, as well as the results of the analysis, is given in chapter 7.

1.4 Simulations of Estimation Methods

While it is clear that the autologistic model, under all of the aforementioned estimation methods, can be and has been readily applied to countless real-life data sets, perhaps the most pertinent questions that should then be addressed are: Which, if any, of the methods is best? and/or Under what circumstances is one method preferable to all other methods? In an attempt to address such questions, a handful of simulation studies have been conducted comparing some of the estimation methods for the autologistic model described in section 1.2. Some pertinent details of the seven such simulation studies in the literature are discussed below.

Graham ([17],[18]) performed a simulation study on the Ising model that compared the coding method, PL, and MCMCML. Throughout the simulation study, the intercept parameter (i.e. spatial trend parameter) was held fixed. Multiple spatial dependence parameter values were considered; these values correspond to spatial dependence strengths that sequentially increase from weak to strong. Three different lattice sizes were used for the study, including 20×20, 40×40, and 60×60. For each of the possible parameter vectors, under all three lattice sizes, 100 replicates were generated. The mean estimates between the three methods tended to "differ very little" for the two larger lattice sizes, but for the smallest lattice size, the MCMCMLEs were the most biased, particularly for "larger model values of (the spatial dependence parameter)." The coding method estimates were the least biased of the three methods, especially for the smallest lattice size. The estimated standard errors of the MCMCMLEs were "significantly smaller" than those of the coding method estimates and "slightly smaller" than those of the MPLEs; this relationship was most pronounced at the smallest

lattice size (20×20) and relatively negligible at the larger lattice sizes. Hence, MCMCML seemed to perform the best of the three estimation methods except in situations of strong spatial correlation, where "a significant percentage ($\approx 80\%$) of the data are ones or zeros (and) the bias (is) serious."

Using a first-order isotropic autologistic model that included a covariate term, Wu and Huffer ([35]) conducted a simulation study that compared the coding method, PL, and MCMCML. This was the first detailed study of an autologistic model which also included a covariate ([35]). Two possibilities for the covariate term were separately considered; the first was a "smooth" covariate, generated from a sine function, while the second covariate was randomly generated from a uniform distribution. Throughout the simulation study, both the intercept and covariate parameters were held fixed; the value of the covariate parameter that was used corresponds to a "strong covariate effect" ([23]). As in Graham's simulation study ([17],[18]), multiple spatial dependence parameter values were considered; these values again correspond to spatial dependence strengths that sequentially increase from weak to strong. The coding method and PL were compared under both 40×40 and 80×80 lattices, but PL and MCMCML were only compared under 40×40 lattices, presumably because of the computational expense involved in carrying out MCMCML for 80×80 lattices. For each of the spatial dependence values, for both covariate options, and for each lattice size (when applicable), 30 replicates were generated. The parameter estimates obtained from the coding method and PL were, for all situations, "very close" and "strongly correlated." The mean estimates obtained from PL and MCMCML were also very close across all situations, but the variances and mean absolute errors (MAE) for the MCMCMLEs were smaller than those for the MPLEs. Consequently, even though it was the most computationally expensive, the authors concluded that of the three estimation methods, MCMCML was the best since it was always either just as good or better than the other two estimation methods. The simulation study also revealed a tendency, in all three methods, "for the error of the estimates to increase as the spatial correlation increased."

Wu and Huffer ([36]) conducted another simulation study that compared the coding method, PL, and MCMCML for a first-order isotropic autologistic model that included a covariate term corresponding to a diagonal sine wave across the lattice. Throughout the simulation study, the covariate parameter was again held fixed at a value corresponding to a relatively strong covariate effect. As in Wu and Huffer's first simulation study ([35]), multiple spatial dependence parameter values were considered; once again, these values correspond to spatial dependence strengths that sequentially increase from weak to strong. The values for the intercept were then selected "in such a way as to balance the spatial interaction and avoid situations where the value "1" or "0" dominates the entire lattice." The reason for their imposing such a balance can surely be traced back to Besag, who asserted that parameter estimation for the autologistic model is numerically more efficient if the lattice has a roughly 50/50 split of 0's and 1's ([2]). A lattice size of 40×40 was used throughout the simulation study and 500 replicates were generated for each parameter vector considered. For all scenarios, the mean (of the 500) estimates were "very close" between the three methods. However, the mean estimates obtained using the coding method tended to be the least biased among the three methods while the mean estimate obtained using MCMCML tended to be the most biased. Among all scenarios, the standard errors and mean squared errors of the MCMCMLEs were "consistently smaller" than those for either the coding method or PL. Hence, even though it is the most computationally expensive of the three methods, Wu and Huffer concluded that the MCMCMLEs are an improvement over both the MPLEs and the estimates obtained via the coding method. As in the simulation study reported in [35], this study also revealed "a tendency in the three methods for the error of the estimates to increase as the spatial interaction increases." The authors speculate that this phenomenon is a result of a decrease in the "effective sample size" as the spatial dependency increases. To elaborate, if the lattice has n sites, then as the spatial correlation between sites increases, the "effective sample size" decreases further from n.

A simulation study primarily comparing PL and GPL (using multiple group sizes) for the

Ising model was carried out by Huang and Ogata ([22]). The primary objective of this study was to evaluate the performances of PL and GPL as the parameter vector neared the critical values ($(-2\theta, \theta)'$ where $\theta \approx \pm 1.76$) associated with phase transition (see section 2.2.5). Hence, multiple parameter vectors that sequentially approached one of the critical values were considered. A lattice size of 64×64 was used throughout the simulation study and 500 replicates were generated for each of the considered parameter vectors. From the presented results, it can be inferred that the mean estimates of the two methods were similar when the parameter vector was relatively far from the critical value, but as the parameter vector approached the critical value the performance of the methods differed. Specifically, the mean estimates for the MPLEs were more biased than the corresponding mean estimates for the MGPLEs (across all group sizes), and the bias of the MGPLEs generally decreased with increased group size.

He et al. ([21]) conducted a simulation study of a first-order anisotropic autologistic model, which also included a covariate term, by comparing PL, MCMCML, and MCMC-SA. The covariate term was the same diagonal sine wave used in the simulation studies carried out in [35] and [36]. Throughout this simulation study, as in several of the aforementioned simulation studies ([17], [18], [35], [36]), both the intercept and covariate parameters were held fixed; the value of the covariate parameter that was used again corresponds to a "strong covariate effect" ([23]). For the two spatial dependence terms, multiple parameter values were considered; in particular, these values correspond to different combinations of anisotropy (i.e. isotropy or two-way anisotropy) and spatial dependence strength (ranging from weak to moderately strong). A lattice size of 40×40 was used throughout the simulation study and 500 replicates were generated for each parameter vector considered. All scenarios yielded similar results. Specifically, while the mean estimates were negligibly different between the three methods, the standard deviations and the estimated standard errors for the MCMCMLEs and the MCMC-SA estimates were superior to those of the MPLEs. Furthermore, as the MCMCMLEs and the MCMC-SA estimates were "often too close to be distinguishable," MCMCML was determined

to be the best method of the three studied since it is less computationally expensive than MCMC-SA.

Sherman et al. ([32]) conducted a simulation study of the Ising model by comparing PL, MCMCML, and GPL (using multiple group sizes). The primary objective of this simulation study, as with the one conducted by Huang and Ogata ([22]), was to evaluate the performances of the three methods for parameter vectors near the critical values associated with phase transition. To that end, two parameter vectors were considered, one vector just "below" one of the critical values and one vector just "above" that same critical value. Corresponding to the region of the liver mortality data set discussed in section 1.3, the lattice used for this simulation study was irregular in shape and consisted of 2003 sites (essentially a 45×45 lattice). For both parameter vectors considered, 100 replicates were generated. When implementing MCMCML, multiple variations of the "control parameters" for the Gibbs sampler were considered, such as the number of Gibbs steps, the initial burn-in, and the spacing between retained sweeps (see section 3.3.2). For the parameter vector just below the selected critical value, all three methods (and their respective variations) had similar mean estimates, but the standard deviations for the MCMCMLEs were the largest while the standard deviations for the MPLEs and the MGPLEs were negligibly different. For the parameter vector just above the selected critical value, the MCMCMLEs obtained using relatively few sweeps of the Gibbs sampler "behave(d) very badly" relative to the MPLEs and MGPLEs. The mean estimates for the MPLEs, MGPLEs, and MCMCMLEs (using relatively many sweeps of the Gibbs sampler) are very similar, but once again the standard deviations for the MCMCMLEs were the largest while the standard deviations for the MPLEs and the MGPLEs were negligibly different. The authors concluded that MCMCML works relatively well for parameter values below the critical value, but for parameter values above that same critical value, "careful monitoring of the Markov chain is necessary to realize (its) benefits."

Finally, under a Bayesian paradigm, a simulation study comparing PL, BGPL, and RDA was carried out by Friel et al. ([11]). The lattice size was fixed at 50×50 throughout the

simulation and 50 replicates were generated for each scenario. The simulation performances of the three methods were compared using the average absolute bias of the posterior mean. The results of the simulations indicated that while both BGPL and RDA out-performed PL, RDA was possibly a better method since it produced results comparable to BGPL, but at a cheaper computational cost. As a Bayesian approach to the autologistic model is not of interest for this dissertation, no additional details for this simulation will be discussed. It is worth mentioning, however, that under a Bayesian paradigm, the parameters of the autologistic model are viewed as random values, whereas under a frequentist paradigm the parameters are viewed as fixed, but unknown, values. Hence, a Bayesian approach to implementing the autologistic model first requires specifying a prior probability density for the model parameters and then, given the observed lattice of binary response values, seeks through Bayes' rule to determine the posterior probability density for the model parameters. Requisite point estimates and/or confidence intervals are then constructed via this posterior distribution.

When aggregated, the results of the above simulations reveal several trends. First, although for small to moderately sized lattices the estimates obtained using the coding method typically have a smaller bias than the estimates obtained using PL or MCMCML, they are consistently, regardless of the lattice size, the most variable and have the most error. Such drawbacks of the coding method certainly yield insight into why it is rarely ever implemented in practice, and justify why it is not considered further in this thesis. Second, in comparing PL and MCMCML, the MPLEs tend to have a smaller bias than the MCMCMLEs, especially as the spatial dependence increases, but as the lattice size increases the difference in the mean estimates between the two methods appears to become negligible, as long as the parameter value is not close to the critical values where phase transition occurs in which case both methods exhibit problems. Furthermore, the variability in the MCMCMLEs is consistently smaller than the variability in the MPLEs, except possibly when the lattice sites are spatially independent, or the parameter value is near the critical values associated with phase transition. Third, and finally, the mean estimates for PL and GPL are comparable in cases of strong to

moderate spatial dependence, but in cases where the parameter value is near the critical values associated with phase transition, the MGPLEs tend to have a smaller bias than the corresponding MPLEs.

Despite the varying amounts of insight the aforementioned simulation studies have provided with respect to how PL, GPL, BGPL, and MCMCML perform and, to some extent, compare, many questions remain either unconvincingly answered or altogether yet unaddressed. Many of these questions fall into two general categories that are of particular interest for this thesis. The first such category revolves around whether or not the two newest methods, GPL and BGPL, accomplish their intended purposes. In particular, do the MGPLEs and MBGPLEs strike the intended compromise between PL and MCMCML with respect to bias, variability, and other performance-based measures such as the mean absolute error (MAE)? Does increasing the block/group size dictate the amount of observed compromise? Is there even a practical difference between the MGPLEs and the MBGPLEs? And finally, do the MGPLEs and/or the MBGPLEs provide better reference points than the MPLEs for MCMCML in some situations? The second category of questions revolves around whether or not the performances of the four methods under the "standard" scenarios (e.g. the Ising model) can be extended to the "nonstandard" scenarios. In particular: how do GPL and BGPL perform when covariates are incorporated into the autologistic model? How does allowing for neighborhood structures other than a first-order system, and/or allowing for anisotropy, affect the performance of the estimation methods? And finally, to what extent do the answers to the previous questions depend on the lattice size? An extensive simulation study aimed at addressing these questions has been conducted as a component of this thesis work. The specifics of this simulation, as well as the results, are given in Chapter 5.

1.5 Spatio-Temporal Autologistic Model (STAM)

Recall the example given at the beginning of this chapter where we had observed a cross-section (in time) of data, along with pertinent covariate information, that indicated the presence or absence of wild fire over a regular lattice within a forested region. Now imagine the same scenario, but instead of one cross-sectional observation in time, we have regularly spaced observations, such as weekly composites, of the lattice over time. The natural extension of the question asked at the beginning of section 1.1, that was answered in sections 1.2, 1.3, and 1.4 above, is then the following: Can the purely spatial autologistic model be generalized naturally to a spatio-temporal autologistic model? Fortunately, the answer to this question is yes, but how this extension is accomplished depends upon the type of temporal dependency, if any, that the lattice exhibits with regard to the binary response variables measured on the observational units. For example, an absorbing state is a type of temporal dependency in which a site remains permanently in one state (say diseased), for the duration of the study, if or once it becomes diseased. To elaborate within the context of the wild fire data, once the trees of a region become burned, if they cannot return to an unburned status (i.e. cannot burn again) before the time of observation has elapsed, then the sites of the lattice belong to an absorbing state with regard to the fire presence variable. Such a temporal dependency produces a generalization of the autologistic model into the space-time domain that is much different than that of a non-absorbing state temporal dependency, as will be pointed out below. Thus, the type of temporal dependency ultimately affects how the autologistic model is generalized into the space-time domain.

Under non-absorbing states, Zhu et al. ([37]) proposed treating the sequence of observed lattices as a space-time Markov random field. Such a proposal allows for the temporal parameters to be added to the spatial autologistic model in a natural way that allows for both the spatial and temporal dependencies to be accounted for simultaneously. In particular, the resulting spatio-temporal autologistic model, for a given site at a given time, conditions spa-

tially on the nearest neighbor values, as under the purely spatial model, and temporally on the auto-lagged (first-order) values, both from the past and in the future if desired. However, as with the normalizing constant of the purely spatial autologistic model's joint likelihood function, the normalizing constant of the spatio-temporal autologistic model's joint likelihood function is intractable, rendering parameter estimation nontrivial. Zhu et al. ([37]) proposed using pseudolikelihood to estimate the parameters of this model and a parametric bootstrap to obtain the corresponding standard errors; they additionally proposed a means for using the Gibbs sampler to forecast future realizations of the lattice. Finally, note that this model assumes (1) that the parameters are space and time invariant, i.e. the effects of space and time on the binary response variable are the same across all spatial locations and all time steps, and (2) that space and time do not interact, i.e. the effects of space and time are separable.

Under absorbing states, Besag ([4]) first proposed a method that was later extended by Chadoeuf et al. ([6]) to allow for observing multiple binary random variables at each site and storing the resulting information as a count ([6]). In this approach, the sequence of observed lattices is once again viewed as a space-time Markov random field, but now the resulting spatio-temporal autologistic model conditions spatially on the nearest neighbor values and temporally on both the nearest neighbor values (at all possible lags) and the auto-lag-one value (past first-order). Furthermore, because an absorbing state dictates that the probability a site has a value of 1 at the current time step, given that it had a value of 1 at the preceding time step, is necessarily one, only the sites of the MRF having value 0 contribute directly (i.e. rather than only indirectly as fixed spatial/temporal neighbor values for sites with value 0) to the likelihood. As is the case with the normalizing constant of the joint likelihood function for the spatio-temporal autologistic model proposed by Zhu et al. ([37]), the joint likelihood function for this spatio-temporal autologistic model has an intractable normalizing constant, and parameter estimation is again nontrivial. Chadeouf et al. ([6]) proposed using an extension of the coding method to estimate the parameters of this spatio temporal autologistic model; they additionally proposed a means for using the Gibbs sampler to forecast future

realizations of the lattice. Finally, note that as with the aforementioned non-absorbing state model, this absorbing state model assumes that the parameters are space and time invariant and that space and time do not interact, i.e. space and time are separable.

Ultimately the context of the data will determine the type of temporal dependency (e.g. absorbing vs. non-absorbing state), while the objectives of the research, as well as the corresponding questions of interest, will determine the particular temporal neighborhood structure (e.g. past lags vs. past and future lags) employed. Hence, for the purposes of this dissertation, we too will view the sequence of lattices as a space-time MRF and our focus will be limited to a non-absorbing state temporal dependency since the data set of interest in Chapter 7 qualifies as such. Furthermore, it will also be assumed that the temporal neighborhood structure will be autoregressive (with past lags) in nature. In other words, only values of the binary response variable for a particular site that were observed in the past can be temporally conditioned on to help model the current value of the binary response variable at that site. The reasons for making such an assumption here can be attributed to how any model fit to the data set of interest in Chapter 7 will ultimately be utilized. Note that this temporal component of the space-time Markov assumption is akin to that proposed by Chadeouf et al. ([6]) in the aforementioned absorbing state model above. It is worth mentioning that one could condition on the values at sites in the past that are at nearby sites, but this higher order spatio-temporal conditioning was not explored here. Finally, we too will assume that the parameters are space and time invariant and that space and time do not interact, i.e. space and time are separable. The spatio-temporal model resulting from such stipulations, along with the necessary background required to arrive at such a model, is formally presented in Chapter 6. To help facilitate a better understanding later in this dissertation, however, such a spatio-temporal extension of the Ising model is presented here, without formal development.

Let $S = \{(i, j, t) : i = 1, 2, \ldots, m_r; j = 1, 2, \ldots, m_c; t = 1, \ldots, m_T\}$ now be a finite three-dimensional array subset of an infinite-dimensional regular lattice, where m_r, m_c, and m_T denote the number of rows, columns, and (two-dimensional) lattices (through time), respec-

tively, of S. Let $S_t = \{(i,j,t) : i = 1,\ldots,m_r; j = 1,\ldots,m_c\}$ denote the lattice of S at time t. Thus, (i,j,t) is the site located on the i^{th} row and j^{th} column of S_t. For notational simplicity, we will again let $n \equiv m_r \times m_c$, and we will numerically label sites of S_t from top to bottom within its columns, and from left to right across its columns, $\forall\ t = 1,\ldots,m_T$, although such an ordering is again arbitrary. Hence, $S_t = \{(i,t) : i = 1,\ldots,n\}$, where (i,t) denotes site i of the t^{th} lattice of the array. Let $Z_{i,t}$ now be a binary random variable located on S_t, with support set $\Omega_{i,t} = \{0,1\}$. Analogous to the previous sections of this chapter, let $\boldsymbol{Z} = (\boldsymbol{Z}_1', \boldsymbol{Z}_2', \ldots, \boldsymbol{Z}_{m_T}')'$ denote the random vector of binary variables over the entire array S, where $\boldsymbol{Z}_t' = (Z_{1,t}, Z_{2,t}, \ldots, Z_{n,t})\ \forall\ t = 1,\ldots,m_T$. The joint support set for $\boldsymbol{Z}$, however, is now $\boldsymbol{\Omega}^{n\cdot m_T} = \{0,1\} \times \ldots \times \{0,1\} = \{0,1\}^{n\cdot m_T}$. Furthermore, $\mathrm{P}(z_{i,t}) \equiv \mathrm{P}(Z_{i,t} = z_{i,t})$ now denotes the probability of the random variable $Z_{i,t}$ taking on the value $z_{i,t} \in \Omega_{i,t}$, and $\mathrm{P}(\boldsymbol{z}) \equiv \mathrm{P}(\boldsymbol{Z} = \boldsymbol{z})$ now denotes the probability of the random vector $\boldsymbol{Z}$ taking on the value $\boldsymbol{z} \in \Omega^{n\cdot m_T}$.

The conditional probability form of the Ising model (from (1.1)), when extended into the space-time domain dictated by the above assumptions, is then given by the following equation:

$$\mathrm{P}(z_{i,t}|\{z_{j,t'} : (j,t') \in N_{i,t} \cup \Gamma_{i,t}\}) \quad =$$

$$= \quad \frac{\exp\left\{\alpha z_{i,t} + \theta z_{i,t} \sum\limits_{(j,t')\in N_{i,t}} z_{j,t'} + \tau z_{i,t} \sum\limits_{(j,t')\in \Gamma_{i,t}} z_{j,t'}\right\}}{1 + \exp\left\{\alpha + \theta \sum\limits_{(j,t')\in N_{i,t}} z_{j,t'} + \tau \sum\limits_{(j,t')\in \Gamma_{i,t}} z_{j,t'}\right\}}, \quad \forall\ (i,t) \in S, \tag{1.25}$$

where, analogous to before, $N_{i,t}$ is the set of sites which are spatial neighbors of site (i,t) (i.e. $(j,t') \in N_{i,t} \Rightarrow t' = t$), α is the intercept parameter, and β is the spatial dependence parameter. Additionally, $\Gamma_{i,t}$ is the set of sites which are temporal neighbors of site (i,t) (i.e. $(j,t') \in \Gamma_{i,t} \Rightarrow t' \leq t-1$), and τ is the corresponding temporal dependence parameter. Technically, no formal notational distinction is necessary for spatial versus temporal neighbors (see Chapter

6), but such a distinction will be made in this dissertation for clarity of presentation. Note that the functional form of the spatio-temporal extension of the Ising model, as given in (1.25), still bears a strong resemblance to that of the standard logistic model that is commonly employed with independent binary random variables. In fact, if the sites of the array were spatially and temporally independent, i.e. $\theta = 0$ and $\tau = 0$, then (1.25) reduces to a simple intercept-only logistic model. Thus, the functional form of such a spatio-temporal autologistic model, as with the purely spatial autologistic model, is identical to that of the logistic model, except that it also conditions on neighboring values, both spatially and temporally, of the binary response variable.

The corresponding extension of the joint Ising model (from (1.2)) into the space-time domain can be shown to have form:

$$\mathrm{P}(\boldsymbol{Z} = \boldsymbol{z})$$

$$= \frac{\exp\left\{\alpha \sum_{t=1}^{m_T}\sum_{i=1}^{n} z_{i,t} + \frac{1}{2}\theta \sum_{t=1}^{m_T}\sum_{i=1}^{n} z_{i,t} \sum_{(j,t')\in N_{i,t}} z_{j,t'} + \tau \sum_{t=1}^{m_T}\sum_{i=1}^{n} z_{i,t} \sum_{(j,t')\in \Gamma_{i,t}} z_{j,t'}\right\}}{\sum_{\boldsymbol{y}\in\boldsymbol{\Omega}^{n\cdot m_T}} \exp\left\{\alpha \sum_{t=1}^{m_T}\sum_{i=1}^{n} y_{i,t} + \frac{1}{2}\theta \sum_{t=1}^{m_T}\sum_{i=1}^{n} y_{i,t} \sum_{(j,t')\in N_{i,t}} y_{j,t'} + \tau \sum_{t=1}^{m_T}\sum_{i=1}^{n} y_{i,t} \sum_{(j,t')\in \Gamma_{i,t}} y_{j,t'}\right\}}, \quad (1.26)$$

$\boldsymbol{z} \in \boldsymbol{\Omega}^{n\cdot m_T}$. Note that the spatio-temporal extension of the Ising model given by (1.25) and (1.26) is not substantially different, at least in functional form, than the extensions that would result from the models proposed by Zhu et al. ([37]) or Chadeouf et al. ([6]). In particular, if τ was pre-multiplied by $\frac{1}{2}$, like θ, in (1.26), then this modified joint likelihood, along with (1.25), would yield the joint and conditional functional forms of the model proposed by Zhu et al. ([37]). Such a modification allows for both past and future auto-lagged temporal neighbors. Furthermore, if (1.25) and (1.26) are modified so that only those sites, (i,t), of S which have variable value 0 (i.e. $z_{i,t} \neq 1$) contribute directly (i.e. rather than only indirectly as fixed spatial/temporal neighbor values for sites with value 0), then these functions yield the joint

and conditional functional forms of the model proposed by Chadeouf et al. ([6]).

Estimating the parameters of the spatio-temporal autologistic model given by (1.25) and (1.26) is, as the case with the models proposed by Zhu et al. ([37]) and Chadeouf et al. ([6]), non-trivial. More specifically, the normalizing constant, or rather the denominator of (1.26), has $2^{n \cdot m_T}$ summands and, therefore, cannot generally be computed in any reasonable amount of time. Hence, as with the purely spatial autologistic model, methods other than maximum likelihood are needed to estimate the parameters of the spatio-temporal autologistic model. Fortunately, most, if not all, of the methods introduced in section 1.2 can be straightforwardly extended into the space-time domain. Four such methods, including PL, MCMCML, GPL, and BGPL, are formally presented in Chapter 6. It should be noted that while (1) PL's use in the space-time domain is already established ([37]) and (2) MCMCML's initial theoretical development naturally permits its use in the space-time domain ([15]), the development of GPL and BGPL presented in Chapter 6 is novel.

Regardless of the type of temporal dependency or the type of temporal neighborhood structure, spatio-temporal autologistic models have been applied to relatively few real data sets. In fact, only two such applications were found in the literature. In particular, Chadoeuf et al. ([6]) used their spatio-temporal ALM under an absorbing state, along with the coding method, to model the simultaneous spread of *Phellinus noxius* and *Rigidoporus lignosus* (soil based mushrooms) in a field of Hevea brasiliensis trees over the course of 3.5 years. Additionally, they implemented a modification of the Gibbs sampler to simulate the spatio-temporal process into the future as a means of forecasting. Finally, Zhu et al. ([37]) used their spatio-temporal ALM under a non-absorbing state, along with maximum pseudolikelihood estimation, to model the presence and absence of southern pine beetle in the 100 counties of North Carolina from 1960 to 1996 while simultaneously accounting for 11 covariates that had previously been identified as important. They too employed the Gibbs sampler to simulate the process into the future as a means of forecasting.

While applications of a spatio-temporal autologistic model in the literature are indeed sparse, simulation studies comparing estimation methods for such space-time models are even rarer. In fact, the literature is void of even a single simulation study evaluating the performances of estimation methods for the aforementioned spatio-temporal autologistic models. This is one of the primary motivations behind the spatio-temporal autologistic model simulation study carried out for this dissertation, the results of which are presented in Chapter 6.

1.6 Research Objectives

Having presented roughly 40 years of research and development for the autologistic model, my research objectives for this dissertation can now be summarized. First, I have set out to prove the strong consistency of the MGPLEs and the MBGPLEs. The background for these proofs, as well as the proofs themselves, are presented in Chapter 4. Second, using my own R-code, I have sought to conduct a thorough simulation study comparing the performances of PL, MCMCML, GPL, and BGPL for the purely spatial autologistic model. This simulation study has aimed to accommodate varying lattice size, levels of anisotropy, group/block size, neighborhood size, spatial and covariate dependence parameter values, and alternative reference points for MCMCML (using MPLE, MGPLE, and MBGPLE). The specifics of the simulation study, as well as the results, are presented in Chapter 5. Third, I have endeavored to implement PL, MCMCML, GPL, and BGPL as estimation methods for a non-absorbing state spatio-temporal autologistic model. The required modifications to the purely spatial methods and the resultant spatio-temporal methods are detailed in Chapter 6. Fourth, I have sought to conduct a small scale simulation study comparing the performances of PL, MCMCML, GPL, and BGPL for the spatio-temporal autologistic model under a non-absorbing state. The specifics of this simulation study, as well as the results, are presented in Chapter 6. Fifth and finally, I have aimed to apply both the purely spatial ALM and the spatio-temporal ALM to a real data set using PL, MCMCML, GPL, and BGPL. The particular data set con-

sists of fire occurrence information for multiple regions across Oregon and Washington state. The objective is to model the presence/absence of fire ignitions over these regions while simultaneously accounting for spatial, temporal, and covariate effects. The covariate of interest in this application is the departure from average variable, a Normalized Difference Vegetation Index (NDVI)-based measure of relative changes in vegetation across the sites of the regions. A thorough description of this fire occurrence data set, as well as a discussion of the results obtained from the autologistic models that were fit, is provided in Chapter 7.

Chapter 2

Origins of the Autologistic Model

2.1 Introduction

While it is ultimately the case that the autologistic model is an intuitive extension of the standard logistic model ([19]), the theoretical development which justifies arriving at such a conclusion is both important and enlightening. In particular, trudging through such theoretical terrain will provide the justification needed to obtain the viable functional form of the autologistic model, which in turn will provide clarity on why the autologistic model is just the logistic model with the addition of spatial dependence terms in the deterministic portion of the model. Section 2.2 rigorously presents the development of the autologistic model while section 2.3 fleshes out the relationship between the logistic and autologistic models.

2.2 Development of the Autologistic Model

The initial development of automodels to characterize the probability structure of spatially dependent random variables on a regular lattice is due to Besag ([2]), but Cressie gives a

detailed treatment of [2] in sections 6.4 and 6.5 of [8] and provides some useful naming conventions for several items that Besag had left otherwise untitled. Consequently, the following presentation of the origins of the autologistic model is largely, but not exclusively, a detailed recounting of pertinent portions of Cressie's ([8]) presentation of [2]. When necessary, details as well as examples have been added for clarity. All direct quotes come from sections 6.4 or 6.5 of [8] unless otherwise indicated. Finally, note that all proofs of theorems given in this section are omitted as they can be found in [8].

2.2.1 Difficulties with a Conditional Specification

When initially considering how to model spatially dependent data observed on a regular lattice, an important question was whether a joint or conditional probability formulation should be used. A conditional probability model describes the probability distribution of the random variable at one site on the lattice *conditioned* on the values of the random variable at neighboring sites, as formally defined below. Due to its "intuitive appeal" a conditional probability formulation became the common modeling choice ([2]). However, in developing a conditional specification, which necessitates constructing the likelihood function (i.e. the joint density) from the resulting conditional distributions, two problems surfaced: one malignant if unresolved and the other benign. Before discussing these problems, some notation and terminology must be established. The presentation will be in the discrete setting to mimic that of the autologistic model, but all results analogously hold in the continuous setting for general automodels with appropriate summations being systematically replaced by integrals.

For this dissertation, a "regular" lattice means a regularly spaced grid of points, or sites; in other words, the distance between any two vertically/horizontally consecutive sites is constant. Suppose the regular lattice of interest has n sites. Let Z_i denote a discrete random variable corresponding to site i, with countable support set Ω_i, $i = 1, \ldots, n$. The random vector over the entire lattice, $\boldsymbol{Z} = (Z_1, \ldots, Z_n)'$, has support set $\boldsymbol{\Omega}^n = \Omega_1 \times \ldots \times \Omega_n$. For example,

$\Omega_i = \{0, 1\}$ for spatially dependent binary random variables and $\boldsymbol{\Omega}^n = \{0, 1\} \times \ldots \times \{0, 1\}$. Finally, define $\mathrm{P}(z_i) \equiv \mathrm{P}(Z_i = z_i)$ to be the probability of the random variable Z_i taking on the value $z_i \in \Omega$, and similarly, define $\mathrm{P}(\boldsymbol{z}) \equiv \mathrm{P}(\boldsymbol{Z} = \boldsymbol{z})$ to be the joint probability of the random vector $\boldsymbol{Z}$ taking on the value $\boldsymbol{z} \in \boldsymbol{\Omega}^n$.

Definition 2.1 (Positivity Condition)**.** *If $P(Z_i = z_i) > 0$ for each $i, i = 1, \ldots, n$, then $P(\boldsymbol{Z} = (z_1, \ldots, z_n)') > 0$.*

The positivity condition simply asserts that if $z_1, z_2, \ldots, z_n$ can individually occur at sites $1, 2, \ldots, n$, respectively, then they can occur together. This condition does not hold in general, but it is usually satisfied in practice. The interested reader can find an example where the positivity condition fails on pages 411-412 of [8]. Assuming this positivity condition holds, the following theorem asserts a critical relationship between any conditional probability specification and the corresponding joint probability specification for a random vector $\boldsymbol{Z}$.

Theorem 2.1 (Factorization Theorem)**.** *Suppose $\{Z_i\}_{i=1}^n$ have joint probability mass function $P(\cdot)$, whose support $\boldsymbol{\Omega}^n$ satisfies the positivity condition. Then,*

$$\frac{P(\boldsymbol{x})}{P(\boldsymbol{y})} = \prod_{i=1}^{n} \frac{P(x_i | x_1, \ldots, x_{i-1}, y_{i+1}, \ldots, y_n)}{P(y_i | x_1, \ldots, x_{i-1}, y_{i+1}, \ldots, y_n)}, \ \boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{\Omega}^n \tag{2.1}$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ are any two realizations of $\boldsymbol{Z}$, with $P(\boldsymbol{y}) > 0$.

The positivity condition is required to ensure a positive denominator in the right hand side of (2.1).

In light of the Factorization Theorem, the two aforementioned problems with pursuing a conditional specification are readily apparent. First, because sites on a lattice are arbitrarily labeled, and because all possible $n!$ factorizations of (2.1) must ultimately produce the same joint probability structure, there exist severe restrictions on the possible functional form for the conditional probability distribution. In fact, this consistency problem has the potential to

prevent the construction of a viable conditional probability distribution. Fortunately, with the advent of the Hammersley-Clifford Theorem (Theorem 2.3 below), this possibly terminal problem is overcome. After developing some additional concepts in the first half of section 2.2.2, this essential result will then be presented. The second problem with pursuing a conditional specification, which is really a problem with the joint specification, pertains to the normalizing constant of the (joint) likelihood function. While Theorem 2.1 allows for the conversion of the conditional specification to its joint format, the joint form can only be expressed up to a proportionality constant. More specifically, through the conditional distributions, equation (2.1) provides an expression for the ratio of joint probabilities, but it doesn't provide a closed-form expression for the absolute joint probability, $P(\boldsymbol{z})$. This is because a conditional specification cannot in general yield a closed form for the normalizing constant of the likelihood function; a statement that will be made clear in section 2.2.2. Hence, even though a consistent conditional formulation can be obtained through the Hammersley-Clifford Theorem, it will generally be the case that the resulting likelihood function will have an intractable normalizing constant. This means that exact maximum likelihood methods cannot be employed and alternative parameter estimation methods will be required, which is the focus of Chapter 3.

2.2.2 Obtaining a Consistent Conditional Specification: Neighbors, Cliques, Negpotential Function, and the Hammersley-Clifford Theorem

In pursuit of a conditional probability specification that successfully resolves the consistency problem, two important concepts and one important function are foundational. The first crucial, but intuitive, concept is the following notion of a neighbor.

Definition 2.2 (Neighbor)**.** *Site k is defined to be a neighbor of site i if the conditional distribution of Z_i, given all other values, depends functionally on z_k, for $k \neq i$. Also define*

$$N_i \equiv \{k : k \text{ is a neighbor of } i\} \tag{2.2}$$

to be the neighborhood set of site i.

Note that a site cannot be a neighbor of itself (i.e.: $i \notin N_i$). In general then, if the lattice of interest has m_r rows and m_c columns (so $m_r \times m_c = n$), where the sites of the lattice are arbitrarily numbered from top to bottom, left to right, then the set of *first-order* neighbors of site i are defined to be this:

$$N_i = \{i-1, i+1, i-m_r, i+m_r\}, \quad i = 1, \ldots, n,$$

the sites immediately above, below, to the left, and to the right of site i. The sites marked with an X_1 in Figure 2.1 provide a visual illustration of the generic "cross" shape corresponding to the set of first-order neighbors of an arbitrary site, labeled Z in the figure. Hence, for a 5x5 lattice (see Figure 2.2), the set of first-order neighbors of site 7 is $N_7 = \{2, 6, 8, 12\}$. Similarly, the set of second-order neighbors of site i are defined to be this:

$$N_i = \{i-1, i+1, i-m_r, i+m_r, i-(m_r+1), i-(m_r-1), i+(m_r+1), i+(m_r-1)\}, \quad i = 1, \ldots, n.$$

This set includes the 4 first-order neighbors as well as the 4 immediate diagonal sites. The sites marked with an X_1 as well as the sites marked with an X_2 in Figure 2.1 provide a visual illustration of the generic "block" shape corresponding to the set of second-order neighbors of an arbitrary site. Thus, for a 5x5 lattice (see Figure 2.2), the set of second-order neighbors of site 7 is $N_7 = \{1, 2, 3, 6, 8, 11, 12, 13\}$. Higher-order Neighborhood systems are systematic extensions of the first- and second-order systems discussed above, but are not discussed in this thesis. The astute reader may have accurately surmised that the neighborhood set of an edge site is different than that of an internal site. Several approaches to defining neighborhood sets for edge sites have been proposed, but choosing which edge adjustment to implement is ultimately a matter of selecting the least objectionable option. More will be said about edge adjustments in section 2.2.4.

Regular lattices of random variables in which a neighborhood structure is either assumed or imposed (when modeling) are extremely commonplace and have led to a particular class of

Figure 2.1: 1^{st}- and 2^{nd}-Order Neighborhood Structures

X_2	X_1	X_2
X_1	Z	X_1
X_2	X_1	X_2

The sites marked with an X_1 represent the first-order neighbors of site Z, while both the sites marked with an X_1 and the sites marked with an X_2 represent the second-order neighbors of site Z.

Figure 2.2: 5x5 Lattice Example

Z_1	Z_6	Z_{11}	Z_{16}	Z_{21}
Z_2	Z_7	Z_{12}	Z_{17}	Z_{22}
Z_3	Z_8	Z_{13}	Z_{18}	Z_{23}
Z_4	Z_9	Z_{14}	Z_{19}	Z_{24}
Z_5	Z_{10}	Z_{15}	Z_{20}	Z_{25}

A 5x5 lattice of random variables arbitrarily labeled top to bottom, left to right.

spatial random vectors known as Markov random fields, which are formalized in the following definition.

Definition 2.3 (Markov Random Field)**.** *Any probability measure whose conditional distributions define a neighborhood structure* $\{N_i : i = 1, \ldots, n\}$ *is defined to be a Markov random field.*

Hence, under a Markov random field assumption, the conditional distribution of z_i only depends on site i's neighboring sites' values:

$$\mathrm{P}(z_i|\{z_j : j \neq i\}) \quad = \quad \mathrm{P}(z_i|\{z_j : j \in N_i\}), \qquad i = 1, \ldots, n. \tag{2.3}$$

Note that the relationship expressed in equation 2.3 applies throughout the remainder of this chapter, but will generally not be explicitly written so as to follow the presentation of this material given in [8].

The second crucial concept in the pursuit of a consistent conditional probability specification, which is a mechanism for grouping sites according to an established neighborhood set, is known as a clique.

Definition 2.4 (Clique)**.** *A clique is defined to be a set of sites that consists either of a single site or of sites that are all neighbors of each other.*

In referring to Figure 2.2 under a first-order neighborhood system, any single site forms a one element clique while $\{7, 12\}$ and $\{7, 8\}$ are examples of two element cliques. Three or larger element cliques under a first-order neighborhood system do not exist. Once again referring to Figure 2.2, but now under a second-order neighborhood system, $\{7, 12, 13\}$ and $\{13, 14, 19\}$ are examples of three element cliques, while $\{7, 8, 12, 13\}$ and $\{13, 14, 18, 19\}$ are examples of four element cliques. The one and two element cliques under a second-order neighborhood system are the same as those under a first-order neighborhood system. Five or larger element cliques under a second-order neighborhood system do not exist.

The final component that is foundational to constructing a consistent conditional probability specification is the negpotential function, which utilizes the relationship given by equation (2.1), where $\boldsymbol{y} \equiv \mathbf{0}$. Note that the term "negpotential," while seemingly obscure, has its origins in statistical physics, and to make sense of it we must momentarily look ahead. In particular, (2.4) can ultimately be expressed as a summation of the family of terms in, what in statistical physics is referred to as, a potential (the family of $V_\kappa(\cdot)$ terms in (2.18)), and the joint probability mass function, $\mathrm{P}(\cdot)$, is sometimes, though not in this dissertation, expressed as a function of the negative of this summation ([12]). Hence, (2.4) is referred to as the "negpotential" (i.e. negative potential) function.

Definition 2.5 (Negpotential Function)**.** *Without loss of generality, assume that zero can occur at each site; that is* $\mathbf{0} \in \boldsymbol{\Omega}^n$*. The negpotential function is then defined as follows:*

$$Q(\boldsymbol{z}) = log\left(\frac{P(\boldsymbol{z})}{P(\mathbf{0})}\right),\ \boldsymbol{z} \in \boldsymbol{\Omega}^n. \tag{2.4}$$

It is important to recognize that knowledge of the negpotential function, $Q(\cdot)$, is equivalent to knowledge of the joint density, $\mathrm{P}(\cdot)$. To see this equivalency, let $\boldsymbol{z} \in \boldsymbol{\Omega}^n$ and note that $\sum_{\boldsymbol{y}\in\boldsymbol{\Omega}^n} \mathrm{P}(\boldsymbol{y}) = 1$. Observe the following:

$$\begin{aligned}
\mathrm{P}(\boldsymbol{z}) &= \frac{\mathrm{P}(\boldsymbol{z})}{\sum_{\boldsymbol{y}\in\boldsymbol{\Omega}^n} \mathrm{P}(\boldsymbol{y})} \quad \left(\text{dividing by } \sum_{\boldsymbol{y}\in\boldsymbol{\Omega}^n} \mathrm{P}(\boldsymbol{y}) = 1\right) \\
&= \frac{\frac{\mathrm{P}(\boldsymbol{z})}{\mathrm{P}(\mathbf{0})}}{\sum_{\boldsymbol{y}\in\boldsymbol{\Omega}^n} \frac{\mathrm{P}(\boldsymbol{y})}{\mathrm{P}(\mathbf{0})}} \quad \left(\text{dividing all terms by } \mathrm{P}(\mathbf{0})\right) \\
&= \frac{\exp\left(\log\left[\frac{\mathrm{P}(\boldsymbol{z})}{\mathrm{P}(\mathbf{0})}\right]\right)}{\sum_{\boldsymbol{y}\in\boldsymbol{\Omega}^n} \exp\left(\log\left[\frac{\mathrm{P}(\boldsymbol{y})}{\mathrm{P}(\mathbf{0})}\right]\right)} \\
&= \frac{\exp(Q(\boldsymbol{z}))}{\sum_{\boldsymbol{y}\in\boldsymbol{\Omega}^n} \exp(Q(\boldsymbol{y}))}.
\end{aligned} \tag{2.5}$$

This relationship between the negpotential function and the joint density is critical and will be essential to the Hammersley-Clifford Theorem. It implies that if we can identify a form for the negpotential function $Q(\cdot)$, this is equivalent to specifying the joint density. Two important properties of the negpotential function, which allow for a simpler statement of (as well as a simpler proof of) the Hammersley-Clifford Theorem, are given in the following theorem.

Theorem 2.2. *The negpotential function satisfies the following two properties:*

(i) Let $\boldsymbol{z}_i \equiv (z_1, \ldots, z_{i-1}, 0, z_{i+1}, \ldots, z_n)'$ *and let* 0_i *denote the event* $Z_i = 0$. *Then*

$$\frac{P(z_i|\{z_j : j \neq i\})}{P(0_i|\{z_j : j \neq i\})} = \frac{P(\boldsymbol{z})}{P(\boldsymbol{z}_i)} = exp\{Q(\boldsymbol{z}) - Q(\boldsymbol{z}_i)\}. \tag{2.6}$$

(ii) For $\boldsymbol{z} \in \boldsymbol{\Omega}^n$, Q *can be uniquely expanded on* $\boldsymbol{\Omega}^n$ *as*

$$Q(\boldsymbol{z}) = \sum_{1 \leq i \leq n} z_i G_i(z_i) \; + \sum_{1 \leq i < j \leq n} \sum z_i z_j G_{ij}(z_i, z_j) \; + \sum_{1 \leq i < j < k \leq n} \sum \sum \sum z_i z_j z_k G_{ijk}(z_i, z_j, z_k)$$
$$+ \; \ldots \; + \; z_1 z_2 \cdots z_n G_{12\ldots n}(z_1, z_2, \ldots, z_n). \tag{2.7}$$

While the proof of Theorem 2.2 can be found in [8] and will not be discussed, it is important to note that the proof hinges on the following definitions:

$$z_i G(z_i) \;=\; Q(0, \ldots, 0, z_i, 0, \ldots, 0), \tag{2.8}$$

$$\begin{aligned} z_i z_j G_{ij}(z_i, z_j) \;=\; & Q(0, \ldots, 0, z_i, 0, \ldots, 0, z_j, 0, \ldots, 0) \\ & -Q(0, \ldots, 0, z_i, 0, \ldots, 0, 0, 0, \ldots, 0) \\ & -Q(0, \ldots, 0, 0, 0, \ldots, 0, z_j, 0, \ldots, 0), \end{aligned} \tag{2.9}$$

"and similar higher-order difference formulas for the rest of the Gs." The expansion of the negpotential function given in equation (2.7) is unique, but the $\{G_{ij\ldots}\}$ functions can only be uniquely specified "by defining $G_{ij\ldots}(z_i, z_j, \ldots) \equiv 0$ whenever $z_i = 0$, or $z_j = 0$, or ...."

Theorem 2.2 ultimately demonstrates how the negpotential function can be expanded in terms of conditional probabilities. For instance, consider $z_i G(z_i)$:

$$\begin{aligned} z_i G_i(z_i) &= Q(0,\ldots,0,z_i,0,\ldots,0) && (2.10)\\ &= \log\left\{\frac{\mathrm{P}(0,\ldots,0,z_i,0,\ldots,0)}{\mathrm{P}(0,\ldots,0,0,0,\ldots,0)}\right\} && (2.11)\\ &= \log\left\{\frac{\frac{\mathrm{P}(0,\ldots,0,z_i,0,\ldots,0)}{\mathrm{P}(0_1,\ldots,0_{j-1},0_{j+1},\ldots,0_n)}}{\frac{\mathrm{P}(0,\ldots,0,0,0,\ldots,0)}{\mathrm{P}(0_1,\ldots,0_{j-1},0_{j+1},\ldots,0_n)}}\right\} \\ &= \log\left\{\frac{\mathrm{P}(z_i|\{0_j : j\neq i\})}{\mathrm{P}(0_i|\{0_j : j\neq i\})}\right\}, && (2.12)\end{aligned}$$

where (2.10) follows from (2.8), (2.11) follows from (2.4), and (2.12) follows by the definition of conditional probability. Similarly, by defining $\mathbf{0}_{-ij} \equiv (0,\ldots,0,z_i,0,\ldots,0,z_j,0,\ldots,0)$, $\mathbf{0}_{-i} \equiv (0,\ldots,0,z_i,0,\ldots,0,0,0,\ldots,0)$, and $\mathbf{0}_{-j} \equiv (0,\ldots,0,0,0,\ldots,0,z_j,0,\ldots,0)$, we obtain the following result for the pairwise interaction term $z_i z_j G_{ij}(z_i,z_j)$:

$$\begin{aligned} z_i z_j G_{ij}(z_i,z_j) &= Q(\mathbf{0}_{-ij}) - Q(\mathbf{0}_{-j}) - Q(\mathbf{0}_{-i}) && (2.13)\\ &= Q(\mathbf{0}_{-ij}) - Q(\mathbf{0}_{-j}) + Q(\mathbf{0}) - Q(\mathbf{0}_{-i}) && (2.14)\\ &= \log\left\{\exp\left\{Q(\mathbf{0}_{-ij}) - Q(\mathbf{0}_{-j})\right\}\cdot\exp\left\{Q(\mathbf{0}) - Q(\mathbf{0}_{-i})\right\}\right\} \\ &= \log\left\{\frac{\mathrm{P}(\mathbf{0}_{-ij})}{\mathrm{P}(\mathbf{0}_{-j})}\cdot\frac{\mathrm{P}(\mathbf{0})}{\mathrm{P}(\mathbf{0}_{-i})}\right\} && (2.15)\\ &= \log\left\{\frac{\frac{\mathrm{P}(z_i|z_j,\{0_k:k\neq i,j\})}{\mathrm{P}(0_i|z_j,\{0_k:k\neq i,j\})}}{\frac{\mathrm{P}(z_i|\{0_k:k\neq i\})}{\mathrm{P}(0_i|\{0_k:k\neq i\})}}\right\}. && (2.16)\end{aligned}$$

where (2.13) follows from (2.9), (2.14) follows because $Q(\mathbf{0}) = 0$ by (2.4), and (2.15) follows from (2.6). Three-way interaction terms, such as $z_i z_j z_k G_{ijk}(z_i,z_j,z_k)$, as well as all possible higher-way interaction terms, can similarly be expressed as functions of conditional probabilities.

Equations (2.12) and (2.16) demonstrate that the negpotential function, as expressed in equation (2.7), can be expressed in terms of conditional probabilities. Then, since knowledge

of the negpotential function is equivalent to knowledge of the joint probability mass function, a resolution to the consistency problem can be pursued through the G functions making up the negpotential function. This means that the consistency problem can be concisely thought of as "those conditions needed to yield well-defined G functions." This simple restatement of the consistency problem is a direct consequence of equation (2.7), which usefully defines the negpotential function in terms of the G functions. Note that by well-defined it is meant that the conditional probability expansion of every interaction term should be invariant to the site whose conditional probabilities are specified. For example, the pairwise interaction term, $z_i z_j G_{ij}(z_i, z_j)$, should be the same if site j's conditional probabilities are specified rather than site i's in the right hand side of equation (2.16), i.e.

$$z_i z_j G_{ij}(z_i, z_j) \quad = \quad \log \left\{ \frac{\frac{\mathrm{P}(z_j | z_i, \{0_k : k \neq i,j\})}{\mathrm{P}(0_j | z_i, \{0_k : k \neq i,j\})}}{\frac{\mathrm{P}(z_j | \{0_k : k \neq j\})}{\mathrm{P}(0_j | \{0_k : k \neq j\})}} \right\}.$$

The requisite conditions for producing well-defined G functions are exactly those expressed by the Hammersley-Clifford Theorem.

Theorem 2.3 (Hammersley-Clifford Theorem)**.** *Suppose that* $\boldsymbol{Z}$ *is distributed according to a Markov random field on* $\boldsymbol{\Omega}^n$ *that satisfies the positivity condition. Then, the negpotential function* $Q(\cdot)$ *given by* (2.4) *must satisfy the property:*

$$\textit{if sites } i, j, \ldots, s \textit{ do not form a clique, then } G_{ij\ldots s}(\cdot) \equiv 0,$$

where the cliques are formed by some neighborhood structure $\{N_i : i = 1, \ldots, n\}$.

In other words, the Hammersley-Clifford Theorem asserts that as long as all G functions corresponding to sets of sites that don't form a clique are defined to be 0, then the resulting negpotential function and equivalent joint probability model will be well-defined from the conditional probability specification.

While Theorem 2.3 is crucial in that it reveals how the consistency issue can be successfully addressed, two corollaries of the Hammersley-Clifford Theorem are also important. The first is the converse of Theorem 2.3, which additionally exposes the intractable normalizing constant issue, while the second establishes a useful equivalency between the local and global Markovian properties.

Corollary 2.1. *Suppose $\boldsymbol{\Omega}^n$ is countable in Euclidean n-space (or $\boldsymbol{\Omega}^n$ is a Lebesgue measurable subset of Euclidean n-space) and that a well-defined set of G functions can be obtained from specified conditional probabilities and neighborhoods $\{N_i : i = 1, \ldots, n\}$. Then the resulting negpotential function Q defined by* (2.7) *yields a unique well-defined joint probability function proportional to $exp\{Q(\cdot)\}$, provided the summability condition*

$$\sum_{\boldsymbol{z} \in \boldsymbol{\Omega}^n} exp\{Q(\boldsymbol{z})\} \quad < \quad \infty \tag{2.17}$$

holds.

The importance of this corollary is two-fold. First, the development up to and including the Hammersley-Clifford Theorem assumed the lattice of interest was a Markov random field (MRF) and focused on constructing a valid conditional probability specification for this MRF, but this corollary affirms the converse relationship. In other words, as long as "the specified conditional probabilities yield well-defined G functions (of which, only those defined on the cliques are nonzero) and a summability condition is satisfied, then there is a unique Markov random field over $\{Z_i : i = 1, \ldots, n\}$." The second and final important consequence of the corollary is that the likelihood function, up to a constant of proportionality, is an explicit function of Q, which when viewed according to (2.7), depends only on the nonnull G functions. Unfortunately, the corresponding proportionality constant for the joint probability, $\sum_{\boldsymbol{y} \in \boldsymbol{\Omega}^n} \exp\{Q(\boldsymbol{y})\}$, rarely has a closed form, which is the previously referenced malignant issue with a conditional formulation. Because this normalizing constant is usually intractable and cannot be computed directly in a computationally feasible amount of time (e.g. for the

binary case, n sites on the lattice implies 2^n summands in the normalizing constant), the joint probability structure, and therefore, the likelihood function, generally have no closed form. Since likelihood-based inference methods cannot then be employed in the usual way, one complication in implementing a conditional probability formulation is having to develop and/or employ parameter estimation methods other than maximum likelihood, and, in fact, much of the roughly forty years of subsequent research on the autologistic model has focused on methods of parameter estimation.

The second important corollary of the Hammersley-Clifford Theorem, which must surely have been foundational in the development of the alternative estimation method known as block generalized pseudolikelihood (see section 3.5), is given in Besag ([2]).

Corollary 2.2. *For any given Markov random field,*

$$P(Z_i = z_i, Z_j = z_j, \ldots, Z_s = z_s | \{z_k : k \neq i, j, \ldots, s\})$$

depends only upon $z_i, z_j, \ldots, z_s$ and the values at sites neighboring sites $i, j, \ldots, s$.

In other words, Corollary 2.2 conveys that the local Markovian properties are equivalent to the global Markovian properties. To help explain what this actually means, consider the 4×4 lattice given in Figure 2.3, where the values in the 2×2 block of "z" sites shall be thought of as random and those in the remaining "o" sites shall be thought of as fixed. Then the information in the joint conditional distribution of the random variables in the 2×2 block, given the neighbors of this block, $\mathrm{P}(z_1, z_2, z_3, z_4 | o_1, \ldots, o_{12})$, is equivalent to the information in $\prod_{i=1}^{4} \mathrm{P}(z_i | \{z_j : j \neq i\}, o_1, \ldots, o_{12})$. The utility of this equivalency will be demonstrated in section 3.5.

Before presenting the actual functional form of the conditional distribution at each site, as well as how that form was reached, some additional notation must be developed to arrive at a useful simplified expression for the generic conditional probability corresponding to site

Figure 2.3: 4x4 Local vs. Global Markovian Example

O_1	O_5	O_7	O_9
O_2	Z_1	Z_3	O_{10}
O_3	Z_2	Z_4	O_{11}
O_4	O_6	O_8	O_{12}

A 4x4 MRF where only the variables at the "z" sites are viewed as random.

i. Let $\aleph$ denote the set of all cliques and let $\kappa \in \aleph$. Now define $\boldsymbol{z}_\kappa \equiv (z_i : i \in \kappa)'$, the vector of variable values corresponding to the sites of clique κ. Additionally, define $V_\kappa(\boldsymbol{z}_\kappa) \equiv \{\prod_{i\in\kappa} z_i\} G_\kappa(\boldsymbol{z}_\kappa)$, the portion of the negpotential function under (2.7) that corresponds to clique κ. Consequently, the expansion of the negpotential function given by (2.7) can more concisely be expressed as follows:

$$Q(\boldsymbol{z}) \quad = \quad \sum_{\kappa\in\aleph} V_\kappa(\boldsymbol{z}_\kappa). \tag{2.18}$$

Now recall from Theorem 2.2 that $\boldsymbol{z}_i = (z_1, \ldots, z_{i-1}, 0, z_{i+1}, \ldots, z_n)'$ and that $\boldsymbol{z} = (z_1, \ldots, z_{i-1}, z_i, z_{i+1}, \ldots, z_n)'$. Since $z_i = 0$ when considering $\boldsymbol{z}_i$, then for every $\kappa \in \aleph$ such that site i is a member of clique κ, $V_\kappa(\boldsymbol{z}_\kappa) = \left\{\prod_{i\in\kappa} z_i\right\} G_\kappa(\boldsymbol{z}_\kappa) = 0$. Thus, in taking the difference between $Q(\boldsymbol{z})$ and $Q(\boldsymbol{z}_i)$, the only $V_\kappa(\boldsymbol{z}_\kappa)$ terms of $Q(\boldsymbol{z})$ that are not canceled out by their corresponding term in $Q(\boldsymbol{z}_i)$ are those $V_\kappa(\boldsymbol{z}_\kappa)$ terms where site i is a member of clique κ, i.e.:

$$Q(\boldsymbol{z}) - Q(\boldsymbol{z}_i) \quad = \quad \sum_{\kappa: i\in\kappa} V_\kappa(\boldsymbol{z}_\kappa). \tag{2.19}$$

Substituting the right-hand side of (2.19) into the far right-hand side of (2.6) yields the following relationship between the conditional probability corresponding to site i, given the

variable values at all other sites of the lattice, and the cliques of which site i is a member:

$$\mathrm{P}(z_i|\{z_j : j \neq i\}) \quad \propto \quad \exp\{\sum_{\kappa: i \in \kappa} V_\kappa(\boldsymbol{z}_\kappa)\}, \qquad i = 1, \ldots, n. \tag{2.20}$$

Hence, "a conditional specification typically involves just a few nonzero functions $\{G_{ij\ldots s}(\cdot)\}$," and only those defined on cliques for which site i is a member.

2.2.3 Simplifying Assumptions: Pairwise-Only Dependence and Conditional Exponential Distributions

Having resolved the consistency problem with a conditional specification and having introduced the intractable normalizing constant problem that is ultimately the motivation for this dissertation, the next step is to determine a "specific form for the conditional distribution at each site," in accordance with an established neighborhood structure. In particular, we assume the conditional distributions are of exponential family. This implies that the conditional distribution corresponding to site i, given the values of the random variables at every other site of the lattice, has the following form:

$$\mathrm{P}(z_i|\{z_j : j \neq i\}) \quad = \quad \exp\left[A_i(\{z_j : j \neq i\})B_i(z_i) + C_i(z_i) + D_i(\{z_j : j \neq i\})\right], \tag{2.21}$$

$i = 1, \ldots, n$, where $\{A_i(\cdot)\}$ and $\{D_i(\cdot)\}$ are functions of site i's observed neighboring values and the forms for $\{B_i(\cdot)\}$ and $\{C_i(\cdot)\}$ are specified by the particular exponential family distribution. Besag states that "a valid choice of A_i determines the type of dependence upon neighboring site values and D_i is then the appropriate normalizing function (for the conditional distribution)" ([2]). In using the exponential family of distributions as the form of the conditional distributions, an extremely useful theorem provides the functional form of A_i, but before stating this result the notion of pairwise-only dependence between sites must first be explained.

Definition 2.6 (Pairwise-only Dependence)**.** *If in* (2.7) *the subset of G functions corresponding to cliques containing three or more sites are all equivalent to 0, then the sites of the MRF are said to exhibit pairwise-only dependence.*

In other words, pairwise-only dependence between sites means that for all $\kappa \in \aleph$ such that $|\kappa| \geq 3$, $G_\kappa(\cdot) \equiv 0$. Note that pairwise-only dependence does not mean that cliques containing 3 or more sites do not exist, but rather that every such clique ends up having its corresponding G function being set to 0. The rationale for making such an assumption is that it is believed to be adequate enough in practice to capture any spatial relationship ([2]) and the complexity of the resulting conditional distributions is substantially reduced; i.e. such an assumption "narrows down the search for sensible and interpretable models," as can be seen in the following theorem.

Theorem 2.4. *Assume* (2.21) *and pairwise-only dependence between sites. Then*

$$A_i(\{z_j : j \neq i\}) \quad = \quad \alpha_i + \sum_{j=1}^{n} \theta_{ij} B_j(z_j), \qquad i = 1, \ldots, n, \tag{2.22}$$

where $\theta_{ji} = \theta_{ij}$, $\theta_{ii} = 0$, *and* $\theta_{ik} = 0$ *for* $k \notin N_i$.

Theorem 2.4 reveals two important results. First, A_i is a *linear* function of the parameters corresponding to site i and second, spatial dependence is inherently conveyed through the $\{\theta_{ij} : i, j = 1, \ldots, n\}$ parameters since $\theta_{ik} = 0$ iff $k \notin N_i$. Consequently, by substituting the right hand side of equation (2.22) into the right hand side of equation (2.21), and in turn substituting this into the right hand side of equation (2.20), the following relationship is obtained:

$$\mathrm{P}(z_i|\{z_j : j \neq i\}) \quad \propto \quad \exp\left\{ \left(\alpha_i + \sum_{j=1}^{n} \theta_{ij} B_j(z_j) \right) B_i(z_i) + C_i(z_i) \right\}, \; i = 1, \ldots, n, \tag{2.23}$$

where $\theta_{ji} = \theta_{ij}$, $\theta_{ii} = 0$, and $\theta_{ik} = 0$ for $k \notin N_i$. In other words,

$$\sum_{\kappa: i \in \kappa} V_\kappa(\boldsymbol{z}_\kappa) \quad = \quad \left(\alpha_i + \sum_{j=1}^{n} \theta_{ij} B_j(z_j) \right) B_i(z_i) + C_i(z_i), \qquad i = 1, \ldots, n. \tag{2.24}$$

This general form for an exponential family conditional distribution is then exploited to obtain the autologistic model for a MRF of binary random variables (see section 2.2.5).

2.2.4 Subtle Complications: Uniqueness and Edge Sites

Before applying the previous development to derive a conditional specification for a Markov random field of binary random variables, two subtle complications must be briefly addressed. The first deals with properties of the probability measure induced by the specified conditional distributions while the second deals with the treatment of edge sites in specifying neighborhood sets.

Given a countably infinite Markov random field with all corresponding neighborhood sets having finite cardinality, two important properties that come into question with respect to the probability measure ($\mathrm{P}_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \boldsymbol{\Theta}$) induced by the conditional probabilities $\mathrm{P}(z_i | \{z_j : j \neq i\}, \boldsymbol{\theta})$, $i = 1, \ldots, n$, are uniqueness and identifiability. Recall that for $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \boldsymbol{\Theta}$, "uniqueness is characterized by the property (that) if $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$, then $\mathrm{P}_{\boldsymbol{\theta}_1} = \mathrm{P}_{\boldsymbol{\theta}_2}$," while "identifiability is characterized by the property (that) if $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$, then $\mathrm{P}_{\boldsymbol{\theta}_1} \neq \mathrm{P}_{\boldsymbol{\theta}_2}$." Although in general there are situations in which these properties fail, for the purposes of this dissertation, both uniqueness and identifiability will always be assumed, except in cases where phase transition is considered (see further below). The interested reader can consult [9], [10], and [16] for necessary and sufficient conditions under which uniqueness and/or identifiability hold.

In addition to possible theoretical complications with a conditional specification of a MRF, there is also a practical complication. In particular, as any sample of lattice sites is necessarily

finite and the conditional distributions given by (2.21) are functions of the observed neighboring values (recall equation (2.3)), the conditional distributions corresponding to edge sites are incomplete. In an attempt to produce complete conditional distributions for edge sites, many edge adjustment procedures have been suggested. Three approaches that permeate the literature include (1) viewing the lattice as a torus, (2) implementing a "guard region," or buffer, and (3) using a weighted average of the available neighboring sites' values. While all three methods yield complete conditional distributions for edge sites, each method carries with its use certain unavoidable concessions. To elaborate on this, each of the three aforementioned methods is explained in some detail below along with their corresponding limitations.

Arguably the most commonly implemented edge adjustment procedure in the literature is a toroidal edge adjustment. This approach views the lattice as a torus, meaning that the lattice is wrapped onto a torus so that there are no edges and every site has a complete neighborhood set ([8]). For example, under a toroidal edge adjustment the first-order neighborhood set of site 1 from Figure 2.2 would be $N_1 = \{2, 5, 6, 21\}$, while the second-order neighborhood set of site 1 would be $N_1 = \{2, 5, 6, 7, 10, 21, 22, 25\}$. The potential problem with using a toroidal edge adjustment, however, is that it assumes that an edge of the lattice is dependent on the opposite edge, which in reality is rarely the case. In the fire ignition example in Chapter 1, there may happen to have been more fire ignitions on one side of the region than the other; in this case, the use of a toroidal edge correction does not make sense.

Besides the toroidal edge adjustment, the other most commonly employed method in the literature for making edge adjustments is the guard region. This method treats the edge sites as fixed. This means that the observed values at the edge sites are used in the conditional distributions of the "internal" (i.e. non edge sites) sites whose neighborhood sets contain at least one edge site, but that the conditional distributions corresponding to the edge sites do not contribute to the likelihood function. In other words, guard region sites "contribute to the likelihood only through their neighborhood relations with internal sites not in the guard region." The obvious problem with a guard region adjustment is the potentially tremendous loss of

information. For example, in one of the four regions of interest from the Oregon/Washington fire ignitions data set, the corresponding lattice dimensions are $38x46$. Hence, under either a first- or second-order neighborhood system, only the conditional distributions corresponding to $\frac{36\times44}{38\times46} \approx 0.9062$ of the sites will be used in constructing the likelihood function. Thus, using a guard region does not require any potentially absurd assumptions, unlike the torus method, but it can come with a substantial information loss.

The final edge adjustment method that will be considered for this dissertation is a weighted edge adjustment. This method weights the functions (A_i and D_i) of the observed values of the existing neighboring sites by the reciprocal of the proportion of available neighbors for the given edge site ([18]). For example, under a first-order neighborhood system, the conditional distribution for site 1 in figure 2.2 would take the functions of the observed values at sites 2 and 6 (its only two existing neighbors) and multiply them both by $\frac{4}{2} = 2$ since there are four possible neighbors and site 1 only has 2 of its possible neighbors available; under a second-order neighborhood system the functions of the observed values at sites 2, 6, and 7 would both be multiplied by $\frac{8}{3}$ since there are eight possible neighbors and site 1 only has 3 of its possible neighbors available. The problem with a weighted edge adjustment is that it always assumes each unobserved neighboring value of an edge site is equal to the average of the observed neighboring values of that edge site, even if the average is based only on a small fraction of available neighbors. Thus, this approach produces "complete" neighborhood sets for edge sites by using the average of available neighboring values to estimate information about unavailable neighboring values.

Having described three common approaches, as well as their limitations, to creating complete neighborhood sets for edge sites, an obvious question lingers: Which method should be employed? Simulation studies comparing the three approaches discussed above demonstrated that, in practice, all three methods are comparable ([18]). Hence, it seems only rational when modeling real (not simulated) data to let the context dictate the selection of an edge adjustment procedure. Along this vein, both a toroidal edge adjustment and a guard region

edge adjustment were implemented for the simulation studies of Chapters 5 and 6, while a weighted edge adjustment was be implemented for the fire occurrence application of Chapter 7. Justification for these decisions will be provided at the appropriate time.

2.2.5 Binary Data and the Autologistic Model

Given the preceding theoretical development for a conditional probability specification of Markov random fields, the next step is to apply these results to binary data under a $\{0, 1\}$ parameterization. Before delving into the specifics, however, an observation should be noted. Specifically, since pairwise-only dependence between sites will be assumed, the negpotential function, as expanded in (2.7), simplifies to the following:

$$Q(\boldsymbol{z}) \quad = \quad \sum_{i=1}^{n} z_i G_i(z_i) + \sum_{1 \le i < j \le n} \sum z_i z_j G_{ij}(z_i, z_j). \tag{2.25}$$

Additionally, since $G_i(0) = G_{ij}(0,1) = G_{ij}(1,0) = 0$ implies that $G_i(1) = \alpha_i$ and $G_{ij}(1,1) = \theta_{ij}$, then the negpotential function for binary data further reduces to the following:

$$Q(\boldsymbol{z}) \quad = \quad \sum_{i=1}^{n} \alpha_i z_i + \sum_{1 \le i < j \le n} \sum \theta_{ij} z_i z_j, \tag{2.26}$$

where $\theta_{ij} = 0$ unless there exists a two-site clique containing sites i and j. Using (2.26) produces the following for $Q(\boldsymbol{z}) - Q(\boldsymbol{z}_i)$, where we recall that $\boldsymbol{z}_i = (z_1, \ldots, z_{i-1}, 0, z_{i+1}, \ldots, z_n)'$:

$$Q(\boldsymbol{z}) - Q(\boldsymbol{z}_i) \quad = \alpha_i z_i + \textstyle\sum_{j=1}^{n} \theta_{ij} z_i z_j, \tag{2.27}$$

where $\theta_{ij} = \theta_{ji}$ and we define $\theta_{ii} \equiv 0$ "to maintain identifiability of the parameters." When combined with equations (2.19) and (2.24), equation (2.27) reveals the following:

$$\left(\alpha_i + \sum_{j=1}^{n} \theta_{ij} B_j(z_j) \right) B_i(z_i) + C_i(z_i) \quad = \quad \alpha_i z_i + \sum_{j=1}^{n} \theta_{ij} z_i z_j, \qquad i = 1, \ldots, n. \tag{2.28}$$

Thus, $B_k(z_k) = z_k \ \forall k = 1, \ldots, n$ and $C_i(z_i) = 0 \ \forall i = 1, \ldots, n$. In other words, the following holds for a MRF of binary random variables:

$$P(z_i|\{z_j : j \neq i\}) \quad \propto \quad \exp\left\{\alpha_i z_i + \sum_{j=1}^{n} \theta_{ij} z_i z_j\right\}, \qquad i = 1, \ldots, n. \tag{2.29}$$

As $P(0_i|\{z_j : j \neq i\})$ is the constant of proportionality, or normalizing constant, for equation (2.29) (see equation (2.30)), the next step is to ascertain its specific form.

If the right hand side of (2.27) is substituted into the far right hand side of (2.6), then the following is obtained for the far left hand side of (2.6):

$$\frac{P(z_i|\{z_j : j \neq i\})}{P(0_i|\{z_j : j \neq i\})} \quad = \quad \exp\left\{\alpha_i z_i + \sum_{j=1}^{n} \theta_{ij} z_i z_j\right\}. \tag{2.30}$$

Setting $z_i = 1$ in the numerator of the left hand side of (2.30) produces the following:

$$\frac{P(1_i|\{z_j : j \neq i\})}{P(0_i|\{z_j : j \neq i\})} \quad = \quad \exp\left\{\alpha_i + \sum_{j=1}^{n} \theta_{ij} z_j\right\}. \tag{2.31}$$

Since z_i can only take on the values 0 or 1,

$$\begin{aligned} \mathrm{P}(1_i|\{z_j : j \neq i\}) + \mathrm{P}(0_i|\{z_j : j \neq i\}) &= 1 \\ &\Leftrightarrow \\ \mathrm{P}(1_i|\{z_j : j \neq i\}) &= 1 - \mathrm{P}(0_i|\{z_j : j \neq i\}). \end{aligned} \tag{2.32}$$

Replacing the numerator of the left hand side of (2.31) with the right hand side of (2.32) and solving for $\mathrm{P}(0_i|\{z_j : j \neq i\})$ yields the following for the normalizing constant of the conditional distribution of site i:

$$\mathrm{P}(0_i|\{z_j : j \neq i\}) \quad = \quad \frac{1}{1 + \exp\left\{\alpha_i + \sum_{j=1}^{n} \theta_{ij} z_j\right\}} \qquad i = 1, \ldots, n. \tag{2.33}$$

Similarly, it can be shown that:

$$\mathrm{P}(1_i|\{z_j : j \neq i\}) = \frac{\exp\left\{\alpha_i + \sum_{j=1}^{n} \theta_{ij} z_j\right\}}{1 + \exp\left\{\alpha_i + \sum_{j=1}^{n} \theta_{ij} z_j\right\}} \quad i = 1, \ldots, n. \tag{2.34}$$

Finally, if the right hand side of equation (2.29) is multiplied by the right hand side of equation (2.33), then the conditional form of the *autologistic model* is derived:

$$\mathrm{P}(z_i|\{z_j : j \neq i\}) = \frac{\exp\left\{\alpha_i z_i + \sum_{j=1}^{n} \theta_{ij} z_i z_j\right\}}{1 + \exp\left\{\alpha_i + \sum_{j=1}^{n} \theta_{ij} z_j\right\}}, \; z_i \in \Omega = \{0, 1\}, \; i = 1, \ldots, n. \tag{2.35}$$

As $\{\theta_{ij} : i, j = 1, \ldots, n\}$ are the spatial-dependence parameters, the binary random variables $Z_1, \ldots, Z_n$ are said to be independent if $\theta_{ij} = 0 \; \forall i, j = 1, \ldots, n$. Recall also that $\theta_{ij} = 0$ iff sites i and j do not form a two-site clique, which is to say that $\theta_{ij} = 0$ in (2.35) iff $j \notin N_i$. Note that the $\{\alpha_i\}$ are referred to as the spatial-trend parameters, but it is often helpful to think of them more as the "proportion of ones" parameters.

While (2.35) gives the conditional specification for a MRF of binary random variables, the joint specification is obtained by substituting the right hand side of (2.26) into the right hand side of (2.5):

$$\mathrm{P}(\boldsymbol{z}) = \frac{\exp\left\{\sum_{i=1}^{n} \alpha_i z_i + \sum\sum_{1 \leq i < j \leq n} \theta_{ij} z_i z_j\right\}}{\sum_{\boldsymbol{y} \in \Omega^n} \exp\left\{\sum_{i=1}^{n} \alpha_i y_i + \sum\sum_{1 \leq i < j \leq n} \theta_{ij} y_i y_j\right\}}, \quad \boldsymbol{z} \in \Omega^n. \tag{2.36}$$

Note that $\sum_{i=1}^{n} z_i$ and $\sum\sum_{1 \leq i \leq j \leq n} z_i z_j$ are sufficient statistics for this model. If (2.36) is viewed as a function of the parameters rather than as a function of the random vector, $\boldsymbol{Z}$, then (2.36) is the likelihood function. As was discussed previously for general Markov random fields, the normalizing constant (denominator of the right hand side of (2.36)) of the likelihood function

for a MRF of binary random variables is intractable. Consequently, exact maximum likelihood methods for parameter estimation cannot be readily employed since n sites corresponds to 2^n summands in the normalizing constant. Four alternative estimation methods will be rigorously presented in Chapter 3.

Since the right hand side of (2.36) exhibits as many as n spatial-trend parameters and $\frac{n(n-1)}{2}$ spatial-dependence parameters, but the MRF consists of just n sites, the model is overparameterized. However, if spatial homogeneity is assumed, then only a single spatial-trend parameter is required and the potential number of spatial-dependence parameters is drastically reduced, with the number of such parameters depending on the type of anisotropy. Recall that spatial homogeneity means that both the spatial trend and the spatial dependence are independent of site location on the lattice ([2]). For example, if the MRF is homogeneous and isotropic, then the parameter set $\{\alpha_i, \theta_{ij} : i, j = 1, \ldots, n\}$ reduces to $\{\alpha, \theta\}$. This two parameter autologistic model is generally referred to in the literature as the *Ising model*, with equation (2.35) simplifying to the following:

$$\mathrm{P}(z_i|\{z_j : j \neq i\}) = \frac{\exp\left\{\alpha z_i + \theta \sum_{j \in N_i} z_i z_j\right\}}{1 + \exp\left\{\alpha + \theta \sum_{j \in N_i} z_j\right\}}, \quad z_i \in \Omega = \{0, 1\}, \tag{2.37}$$

where N_i is the neighborhood set for site i, $i = 1, \ldots, n$. It should be noted here that while the Ising model is nice in its simplicity, it is not without difficulties. In fact, the Ising model on a two-dimensional infinite square lattice suffers from a phenomenon known as *phase transition*, which can create identifiability issues in the model ([28]). The following brief description of phase transition is summarized from Pickard's ([28]) detailed treatment of the matter within the context of the Ising model. In particular, there exist two critical parameter values ($\alpha = 0$ and $\theta = \pm\frac{1}{2}\sinh^{-1}(1) \approx \pm 0.44$ under a $\{-1, 1\}$ parameterization, and $\boldsymbol{\psi} = (-2\theta, \theta)'$ where $\theta \approx \pm 1.76$ under a $\{0, 1\}$ parameterization) in the parameter space such that for parame-

ter values between these critical values, sites of the lattice exhibit long range independence ("asymptotic independence"), but for parameter values beyond these critical values, sites of the lattice exhibit "long-range correlation," even as the distance between sites approaches infinity. A phase transition is the name reserved for such situations where there exists an "abrupt change in the qualitative behavior at a critical parameter value" ([28]). It is important to understand, however, that even though the Ising model is described by local properties through neighborhood relationships, phase transition is a global property of the lattice, not a local property. Thus, in the presence of long-range correlation, statistical inference becomes difficult since the model resulting from equation (2.37) suffers from identifiability problems. The Ising model is not the only autologistic model with phase transitions, but due to the difficulty inherent in identifying the critical values associated with phase transitions, it appears, at least from our literature review, to be the only such model whose critical values have been determined. Geman and Geman indicate, however, that the singularities of the normalizing constant, when viewed as a function of the parameters, are where the phase transitions occur ([12]).

For the purpose of clarity through illustration, one more example of an autologistic model under the assumption of a homogeneous MRF will be given. Specifically, if the MRF is homogeneous with two-way anisotropy under a first-order neighborhood system (suppose vertical and horizontal gradients where the spatial dependence in these two directions is different), then the parameter set is $\{\alpha, \theta_v, \theta_h\}$, where θ_v corresponds to the vertical gradient and θ_h corresponds to the horizontal gradient. This three-parameter autologistic model has the following altered form of equation (2.35):

$$\mathrm{P}(z_i|\{z_j : j \neq i\}) = \frac{\exp\left\{\alpha z_i + \theta_v \sum_{j \in N_i^v} z_i z_j + \theta_h \sum_{j \in N_i^h} z_i z_j\right\}}{1 + \exp\left\{\alpha + \theta_v \sum_{j \in N_i^v} z_j + \theta_h \sum_{j \in N_i^h} z_j\right\}}, \quad z_i \in \Omega = \{0, 1\}, \tag{2.38}$$

where N_i^v and N_i^h are, respectively, the vertical and horizontal neighborhood sets for site i, $i = 1, \ldots, n$. If $\theta_v = \theta_h$, then equation (2.38) reduces to the Ising model (equation (2.37)).

Clearly spatial homogeneity is a useful assumption with respect to parameter reduction, but is it reasonable? As it turns out, spatial homogeneity is essentially a requirement when working with the autologistic model, as is indicated by [2]: "it is unreasonable to apply (the autologistic model) in situations where there is evidence of gross heterogeneity over the lattice." Thus, unless otherwise explicitly stated, spatial homogeneity will be assumed throughout the remainder of this thesis. In some instances, however, spatial homogeneity is only attained after one or more covariates have been included in the autologistic model. In particular, if covariates that adequately model the trend have also been observed across the lattice of interest, then including them in the autologistic model can correct any spatial heterogeneity and ultimately yield a homogeneous system for which the autologistic model spatial parameters are once again appropriate. Therefore, spatial modelers who are seeking to employ the autologistic model should be aware of the fact that even if initial EDA of the response variable reveals the presence of spatial heterogeneity across the lattice, abandoning the autologistic model is not automatically the proper course of action.

2.3 The Autologistic Model and Covariates

Equations (2.37) and (2.38) are two specific examples of a first-order autologistic model, the first under isotropy and the second under two-way anisotropy. In general, however, the autologistic model can be succinctly expressed as follows:

$$\mathrm{P}(z_i|\{z_j : j \neq i\}) = \frac{\exp\left\{\boldsymbol{\psi}'\boldsymbol{T}_{z_i}\right\}}{\sum_{y\in\{0,1\}} \exp\left\{\boldsymbol{\psi}'\boldsymbol{T}_{y}\right\}}, \quad z_i \in \Omega = \{0,1\}, \quad i = 1, \ldots, n, \quad (2.39)$$

where $\boldsymbol{\psi}$ is the vector of parameters, including the single spatial trend parameter and all spatial dependence parameters, and $\boldsymbol{T}_{z_i}$ is a vector of statistics resulting from the specified value for z_i. With the Ising model, for example, $\boldsymbol{\psi} = (\alpha, \theta)'$ and $\boldsymbol{T}_{z_i} = (z_i, \sum_{j \in N_i} z_i z_j)'$. The corresponding succinct representation of the joint autologistic model, or rather the likelihood function, is the following:

$$\mathrm{P}(\boldsymbol{z}|\boldsymbol{\psi}) = \frac{\exp\left\{\boldsymbol{\psi}' \sum_{i=1}^{n} \boldsymbol{T}_{z_i}\right\}}{\sum_{\boldsymbol{y} \in \Omega^n} \exp\left\{\boldsymbol{\psi}' \sum_{i=1}^{n} \boldsymbol{T}_{y_i}\right\}}, \quad \boldsymbol{z} \in \Omega^n. \tag{2.40}$$

where $\sum_{i=1}^{n} \boldsymbol{T}_{z_i}$ is the vector of jointly sufficient statistics for a specified value of $\boldsymbol{z} = (z_1, \ldots, z_n)'$. For the Ising model, $\sum_{i=1}^{n} \boldsymbol{T}_{z_i} = (\sum_{i=1}^{n} z_i, \frac{1}{2} \sum_{i=1}^{n} \sum_{j \in N_i} z_i z_j)'$, where the $\frac{1}{2}$ in the second element of $\sum_{i=1}^{n} \boldsymbol{T}_{z_i}$ accounts for the double counting of $z_i z_j$ pairs.

Now suppose that covariates (and/or cofactors) need to be incorporated into the autologistic model to help explain any observed trend in the process across the region. Recall the fire ignition example given at the beginning of Chapter 1 with possible covariates including wind direction, wind speed, and a measure of relative greenness. The resulting autologistic model will have exactly the same form as (2.39) with $\boldsymbol{\psi}$ and $\boldsymbol{T}_{z_i}$ incorporating all covariate parameters and corresponding sufficient statistics, respectively ([19]). For example, suppose there are p spatial dependence parameters and k covariate parameters (with no interaction terms), then $\boldsymbol{\psi} = (\alpha, \theta_1, \ldots, \theta_p, \beta_1, \ldots, \beta_k)'$ and $\boldsymbol{T}_{z_i}$ will be of length $1 + p + k$ and consist of a sufficient statistic for the single spatial trend parameter, p sufficient statistics for the p spatial dependence parameters, and k sufficient statistics for the k covariate parameters. For the sake of clarity, consider a first-order neighborhood system, under vertical and horizontal anisotropy, with a single covariate term, $\boldsymbol{X} = (X_1, \ldots, X_n)' \in \mathbb{R}_1 \times \ldots \times \mathbb{R}_n = \mathbb{R}^n$, observed on the MRF.

Then the corresponding autologistic model would take the form:

$$\mathrm{P}(z_i|\{z_j : j \neq i\}, \boldsymbol{x}) = \frac{\exp\left\{\alpha z_i + \theta_v \sum_{j\in N_i^v} z_i z_j + \theta_h \sum_{j\in N_i^h} z_i z_j + \beta x_i z_i\right\}}{1 + \exp\left\{\alpha + \theta_v \sum_{j\in N_i^v} z_j + \theta_h \sum_{j\in N_i^h} z_j + \beta x_i\right\}}, z_i \in \Omega = \{0, 1\}, \quad (2.41)$$

where $\boldsymbol{\psi} = (\alpha, \theta_v, \theta_h, \beta)'$ and $\boldsymbol{T}_{z_i} = (z_i, \sum_{j\in N_i^v} z_i z_j, \sum_{j\in N_i^h} z_i z_j, x_i z_i)'$.

The main observation to be made from the preceding discussion is that the "autologistic model is a simple generalization to spatial data of the standard logistic model for independent binary data" ([19]). More specifically, if the sites of the lattice are spatially independent, then equation (2.39), with the addition of k covariates, is nothing more than a logistic regression model with $\boldsymbol{\psi} = (\alpha, \beta_1, \ldots, \beta_k)'$ and $\boldsymbol{T}_{z_i} = (z_i, z_i x_{1i}, \ldots, z_i x_{ki})'$, where x_{ji} is the value of the j^{th} covariate at site i. Hence, the functional form of the autologistic model is identical to that of the logistic model, except that it also conditions on neighboring values of the binary random variable (i.e. the random variable at a given site conditions on *itself* through its neighbors' values), which explains prefacing the term "logistic" with "auto" (meaning "self") in naming the model.

Despite the fact that the autologistic model is functionally just a simple extension of the logistic model, parameter estimation for the model, as discussed in section 2.2.2, is not straightforward like that for the logistic model. This fact is a direct consequence of the lack of independence between the sites of the lattice, which ultimately result in the intractable normalizing constant for the joint autologistic model. In particular, the denominator of the right hand side of equation (2.40) is intractable, which means methods other than exact maximum likelihood have to be employed. Numerous such methods were briefly discussed in section 1.2 of Chapter 1. Four of these methods will be presented in detail in Chapter 3.

Chapter 3

Estimation Methods for the Autologistic Model

3.1 Introduction

As introduced in Chapter 1 and thoroughly developed in Chapter 2, the normalizing constant for the likelihood function of the autologistic model is intractable, as is the case with most likelihood functions corresponding to Markov random field models ([18]). Consequently, methods other than maximum likelihood must be employed to estimate the model parameters. Several such alternative methods were briefly introduced, as well as compared and contrasted, in Chapter 1, and a more rigorous presentation of four of those methods is the topic of this chapter. Specifically, the methods of maximum pseudolikelihood (PL), Markov chain Monte Carlo maximum likelihood (MCMCML), generalized pseudolikelihood (GPL), and block generalized pseudolikelihood (BGPL) are presented in sections 3.2, 3.3, 3.4, and 3.5, respectively. Note that the strong consistency of the MGPLEs, as well as the strong consistency of the MBGPLEs, is proven in Chapter 4. Before these estimation methods can

be presented, however, some notation must first be established.

Let $D = \{(i, j, k, \ldots) : i = 1, 2, \ldots; j = 1, 2, \ldots; k = 1, 2, \ldots; \ldots\}$ be an infinite dimensional square lattice, where $(i, j, k, \ldots)$ denotes site $(i, j, k, \ldots)$ on the lattice. For all applications of the (purely spatial) autologistic model considered in this thesis, attention will be restricted to a finite two-dimensional subset, S, of D; i.e. $S = \{(i, j) : i = 1, \ldots, m_r; j = 1, \ldots, m_c\} \subset D$, where m_r is the number of rows of S and m_c the number of columns. In Chapter 6, S will be expanded to a finite three-dimensional subset of D when the spatio-temporal autologistic model is considered. For notational simplicity, further define $n \equiv m_r \times m_c$ and $S = \{i : i = 1, \ldots, n\}$, where i denotes site i of the lattice and sites are arbitrarily numerically labeled top to bottom within columns and left to right across columns.

Let $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_n)'$ be a vector of discrete random variables on S, with joint support $\Omega^n \equiv \Omega_1 \times \ldots \times \Omega_n$, where $\boldsymbol{Z}$ is distributed according to a Markov random field (MRF). Recall that the autologistic model has joint support $\Omega^n = \{0, 1\} \times \ldots \times \{0, 1\}$. The joint density (i.e. likelihood), or MRF distribution, of $\boldsymbol{Z}$ has the following form:

$$\begin{aligned} \mathrm{L}(\boldsymbol{\psi}) &= \mathrm{P}(\boldsymbol{Z} = \boldsymbol{z}; \boldsymbol{\psi}) \\ &= \frac{\exp\{\boldsymbol{T}'(\boldsymbol{z}) \cdot \boldsymbol{\psi}\}}{\sum_{\boldsymbol{y} \in \Omega^n} \exp\{\boldsymbol{T}'(\boldsymbol{y}) \cdot \boldsymbol{\psi}\}} \\ &= \frac{\exp\{\boldsymbol{T}'(\boldsymbol{z}) \cdot \boldsymbol{\psi}\}}{c(\boldsymbol{\psi})}, \end{aligned} \tag{3.1}$$

where $\boldsymbol{\psi} = (\alpha, \theta_1 \ldots, \theta_{p-1})'$ is the vector of parameters, with parameter space $\boldsymbol{\Psi}$, $\boldsymbol{T}(\boldsymbol{z}) = (T_1(\boldsymbol{z}), \ldots, T_p(\boldsymbol{z}))'$ is the corresponding vector of jointly sufficient statistics (determined by the particular realization $\boldsymbol{z}$), and $c(\boldsymbol{\psi})$ is the intractable normalizing constant. Recall that α is the spatial trend parameter while $\theta_1, \ldots, \theta_{p-1}$ are the spatial dependence parameters, with the number of such parameters, p, depending on the type of anisotropy, if any, present over the MRF. If one or more covariates (and/or cofactors) are also observed over the MRF, (3.1) is easily expanded. In particular, suppose there are k covariates, $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$, where

$\boldsymbol{x}_q = (x_{q1}, \ldots, x_{qn})'$, $1 \leq q \leq k$, and x_{qi} is the value of the q^{th} covariate at the i^{th} site. Let $\boldsymbol{X}$ be the nxk matrix of covariate values, where the q^{th} column of $\boldsymbol{X}$ is $\boldsymbol{x}_q$. Then the MRF distribution for $\boldsymbol{Z}$ is denoted as follows:

$$\begin{aligned} \mathrm{L}(\boldsymbol{\psi}) &= \mathrm{P}(\boldsymbol{Z} = \boldsymbol{z} | \boldsymbol{X}; \boldsymbol{\psi}) \\ &= \frac{\exp\{\boldsymbol{T}'(\boldsymbol{z}; \boldsymbol{X}) \cdot \boldsymbol{\psi}\}}{\sum_{\boldsymbol{y} \in \Omega^n} \exp\{\boldsymbol{T}'(\boldsymbol{y}; \boldsymbol{X}) \cdot \boldsymbol{\psi}\}} \\ &= \frac{\exp\{\boldsymbol{T}'(\boldsymbol{z}; \boldsymbol{X}) \cdot \boldsymbol{\psi}\}}{c(\boldsymbol{\psi})}, \end{aligned} \tag{3.2}$$

where $\boldsymbol{\psi} = (\alpha, \theta_1, \ldots, \theta_{p-1}, \beta_1, \ldots, \beta_k)'$ and $\boldsymbol{T}(\boldsymbol{z}; \boldsymbol{X}) = (T_1(\boldsymbol{z}; \boldsymbol{X}), \ldots, T_{p+k}(\boldsymbol{z}; \boldsymbol{X}))'$ is the vector of jointly sufficient statistics.

In addition to the joint density, the conditional probability density for Z_i, given $\{Z_j : j \neq i\}$, is also required. By the spatial Markov property, the form of the conditional distribution is the following:

$$\begin{aligned} \mathrm{P}(Z_i = z_i | \{z_j : j \neq i\}; \boldsymbol{\psi}) &= \mathrm{P}(Z_i = z_i | \{z_j : j \in N_i\}; \boldsymbol{\psi}) \\ &= \frac{\exp\{\boldsymbol{t}'(z_i, \{z_j : j \in N_i\}) \cdot \boldsymbol{\psi}\}}{\sum_{y_i \in \Omega_i} \exp\{\boldsymbol{t}'(y_i, \{y_j : j \in N_i\}) \cdot \boldsymbol{\psi}\}}, \quad i = 1, \ldots, n, \end{aligned} \tag{3.3}$$

where $\boldsymbol{\psi}$ is the same parameter vector as in (3.1), $\boldsymbol{t}(z_i, \{z_j : j \in N_i\}) = (t_1(z_i, \{z_j : j \in N_i\}), \ldots, t_p(z_i, \{z_j : j \in N_i\})'$ is the vector of jointly sufficient statistics for $\boldsymbol{\psi}$ corresponding to the i^{th} site conditional likelihood, and N_i is the neighborhood set (see definition 2.2) for the i^{th} site. The normalizing constant for the conditional distribution is not an issue like $c(\boldsymbol{\psi})$ is for the joint density since the normalizing constant of the conditional distribution has a closed form, consisting of just a few summands. Finally, just as covariates were incorporated into the joint density, covariates can analogously be incorporated into the conditional model specification. In particular, given the vector of k covariate values for site i, $\boldsymbol{x}^i = (x_{1i}, \ldots, x_{ki})'$,

the conditional distribution becomes the following for $i = 1, \ldots, n$:

$$\mathrm{P}(Z_i = z_i | \{z_j : j \in N_i\}, \boldsymbol{x}^i; \boldsymbol{\psi}) \quad = \quad \frac{\exp\{\boldsymbol{t}'(z_i, \{z_j : j \in N_i\}, \boldsymbol{x}^i) \cdot \boldsymbol{\psi}\}}{\sum_{y_i \in \Omega_i} \exp\{\boldsymbol{t}'(y_i, \{y_j : j \in N_i\}, \boldsymbol{x}^i) \cdot \boldsymbol{\psi}\}}, \tag{3.4}$$

where $\boldsymbol{\psi}$ is the same parameter vector as in (3.2) and $\boldsymbol{t}(z_i, \{z_j : j \in N_i\}, \boldsymbol{x}^i) = (t_1(z_i, \{z_j : j \in N_i\}; \boldsymbol{x}^i), \ldots, t_{p+k}(z_i, \{z_j : j \in N_i\}; \boldsymbol{x}^i)'$. Note that it can be inferred from Wu and Huffer ([35]) that the neighboring covariate values of site i are not used in (3.4) because the spatial dependence in the covariates is indirectly accounted for when the z_j in N_i are included in the model.

3.2 Maximum Pseudolikelihood Estimation

The method of pseudolikelihood (PL), developed by Besag ([3]), is a simple approach to estimating the parameters of a MRF model. The essence of the method is straightforward: construct a "pseudo"-likelihood function for the MRF by forming the product of the n site conditional densities given by (3.3), or by (3.4) if covariates are included. This pseudolikelihood function is then treated like a true likelihood function to obtain the corresponding maximum pseudolikelihood estimates (MPLEs). The remainder of this section is broken into two subsections. In section 3.2.1 a general treatment of PL is provided that is largely a summary of Besag's ([3]) development of the method, while in section 3.2.2 the method is demonstrated within the context of the Ising model, when extended to incorporate a covariate.

3.2.1 PL: Background

Assuming covariates have been observed over the MRF of interest, the PL function, denoted $\text{PL}(\boldsymbol{\psi})$, is just the product of the n site conditional distributions:

$$\text{PL}(\boldsymbol{\psi}|\boldsymbol{z},\boldsymbol{X}) = \prod_{i=1}^{n} \frac{\exp\{\boldsymbol{t}'(z_i,\{z_j : j \in N_i\},\boldsymbol{x}^i)\boldsymbol{\psi}\}}{\sum_{y_i\in\Omega_i}\exp\{\boldsymbol{t}'(y_i,\{y_j : j \in N_i\},\boldsymbol{x}^i)\boldsymbol{\psi}\}}, \quad \boldsymbol{z}\in\Omega^n, \boldsymbol{\psi}\in\boldsymbol{\Psi}. \quad (3.5)$$

Thus, the pseudolikelihood function is formed in the same fashion that the true likelihood function is formed for independent random variables, which explains prefacing the name of the method with "pseudo" since the random variables are in fact (almost always) dependent. As is typically the case with the likelihood function, the logarithm of the pseudolikelihood function is worked with in practice. In particular, the log of the PL function is the following:

$$\begin{aligned}\log\{\text{PL}(\boldsymbol{\psi}|\boldsymbol{z},\boldsymbol{X})\} &= \sum_{i=1}^{n}\boldsymbol{t}'(z_i,\{z_j : j\in N_i\},\boldsymbol{x}^i)\boldsymbol{\psi} \\ &\quad - \sum_{i=1}^{n}\log\left\{\sum_{y_i\in\Omega_i}\exp\{\boldsymbol{t}'(y_i,\{y_j : j\in N_i\},\boldsymbol{x}^i)\boldsymbol{\psi}\}\right\}. \quad (3.6)\end{aligned}$$

The MPLEs, $\hat{\boldsymbol{\psi}}_{PL}$, are then obtained by numerically maximizing (3.6). Any statistical software package that has a generalized linear models (GLIM) package, such as R, can obtain the MPLEs in a matter of seconds; when the MRF consists of binary response variables (i.e. the ALM), the pseudolikelihood is just a logistic regression model likelihood. However, since the pseudolikelihood function is not a true likelihood function, the asymptotic standard errors generally reported as part of standard logistic regression output are invalid since the observations are not actually independent. As was mentioned in Chapter 1, the standard errors for the MPLEs may be obtained using a resampling technique such as parametric bootstrapping ([19]); this technique will be presented in Chapter 7 when PL, as well as the other three estimation methods discussed in this current chapter, are applied to the fire data from Oregon and Washington.

As for properties of the MPLEs, all is known with respect to asymptotic efficiency, strong consistency, and asymptotic normality. In particular, the MPLEs are not asymptotically efficient, "but the inefficiency will often be slight, and compensated by dramatic computational simplifications" ([33]). Strauss ([33]) attributes the inefficiency of the MPLEs to the fact that they are generally not a function of the minimal sufficient statistics. Finally, although the MPLEs are inefficient, they do possess the desirable properties of strong consistency and asymptotic normality, one or both of which have been proven, under suitable conditions, by several authors, including Geman and Graffigne ([13]), Jensen and Møller ([24]), Comets ([7]), and Guyon and Kunsch ([20]). The proof of the strong consistency of the MPLEs given by Geman and Graffigne ([13]) is instrumental in the proof of the strong consistency of the MGPLEs and the MBGPLEs given in Chapter 4.

3.2.2 PL: Application to the ALM

To demonstrate the use of PL with a concrete example, consider the Ising model with a single covariate, $\boldsymbol{x} = (x_1, \ldots, x_n)'$ measured on the n sites (so $\boldsymbol{\psi} = (\alpha, \theta, \beta)'$):

$$\mathrm{P}(Z_i = z_i | \{z_j : j \in N_i\}, \boldsymbol{x}; \boldsymbol{\psi}) = \frac{\exp\left\{\alpha z_i + \theta z_i \sum_{j \in N_i} z_j + \beta z_i x_i\right\}}{1 + \exp\left\{\alpha + \theta \sum_{j \in N_i} z_j + \beta x_i\right\}}, i = 1, \ldots, n. \quad (3.7)$$

Note that the joint probability mass function corresponding to the ALM given in (3.7) has the following vector of jointly sufficient statistics:

$$\boldsymbol{T}(\boldsymbol{z}; \boldsymbol{x}) = (\sum_{i=1}^{n} z_i, \frac{1}{2} \sum_{i=1}^{n} \sum_{j \in N_i} z_i z_j, \sum_{i=1}^{n} z_i x_i)'.$$

The corresponding pseudolikelihood function is then just the product of the n site conditional distributions given by (3.7):

$$\mathrm{PL}(\boldsymbol{\psi}|\boldsymbol{z},\boldsymbol{x}) = \prod_{i=1}^{n} \frac{\exp\left\{\alpha z_i + \theta z_i \sum_{j\in N_i} z_j + \beta z_i x_i\right\}}{1+\exp\left\{\alpha + \theta \sum_{j\in N_i} z_j + \beta x_i\right\}}. \tag{3.8}$$

The logarithm of (3.8) is then taken to obtain the log PL function:

$$\begin{aligned}\log\{\mathrm{PL}(\boldsymbol{\psi}|\boldsymbol{z},\boldsymbol{X})\} &= \alpha\sum_{i=1}^{n} z_i + \theta\sum_{i=1}^{n}\sum_{j\in N_i} z_i z_j + \sum_{i=1}^{n}\beta z_i x_i \\ &\quad - \sum_{i=1}^{n}\log\left[1+\exp\left\{\alpha+\theta\sum_{j\in N_i} z_j + \beta x_i\right\}\right]. \end{aligned}\tag{3.9}$$

The MPLEs, $\hat{\boldsymbol{\psi}}_{\mathrm{PL}} = (\hat{\alpha}_{PL}, \hat{\theta}_{PL}, \hat{\beta}_{PL})'$, are then obtained by numerically maximizing (3.9) with respect to $\boldsymbol{\psi}$. As was mentioned in section 3.2.1, the corresponding standard errors must be obtained using a resampling method such as parametric bootstrapping.

3.3 Markov Chain Monte Carlo Maximum Likelihood Estimation

The method of Markov chain Monte Carlo maximum likelihood (MCMCML), developed by Geyer and Thompson ([15]), is used to obtain an approximation of the maximum likelihood estimate (MLE) in situations where the MLE is unavailable. In particular, when the true likelihood is known only up to a constant of proportionality, i.e. its normalizing constant is intractable, MCMCML can be employed to determine an estimate of the true MLE ([15]). For the purposes of this thesis, the detailed development of the MCMCML method is three-fold. First, in section 3.3.1, an in-depth description of MCMCML is provided and largely

parallels the development of the method that is given in [23]. This background information first demonstrates how a ratio of normalizing constants at two distinct parameter values is sufficient to obtain the MCMC approximate log likelihood function, and then briefly discusses properties of the corresponding MCMCMLEs. Second, in section 3.3.2, the Gibbs sampler is reviewed since it provides the necessary machinery for generating a Markov chain of lattice realizations necessary to construct the MCMC approximate log likelihood. Finally, in section 3.3.3, the use of MCMCML methods within the context of the autologistic model is demonstrated, and in particular, its use with the Ising model, when generalized to incorporate a covariate term, is illustrated. Note that the theory behind Markov chain Monte Carlo methods that justifies their use and the properties of the corresponding MCMCMLEs is both deep and involved, and is beyond the scope of this thesis. Hence the following treatment of MCMCML does not focus on why the method works, but rather on how it works. The interested reader can find numerous articles and textbooks in the literature devoted to both the how and the why of MCMCML, including [12], [18], and [31].

3.3.1 MCMCML: Background

Assume that the probability measure $\mathrm{P}_{\boldsymbol{\psi}}$ of the data $\boldsymbol{z}$ has a joint probability mass (or density) function $f_{\boldsymbol{\psi}}$ (with respect to a measure μ) that can be expressed as in the right hand side of (3.1), or as in the right hand side of (3.2) if there are also covariates observed on the lattice; for simplicity of presentation, assume without any loss of generality that no covariates have been observed over the MRF. Hence,

$$\begin{aligned} f_{\boldsymbol{\psi}}(\boldsymbol{z}) &= \frac{\exp\{\boldsymbol{T}'(\boldsymbol{z})\boldsymbol{\psi}\}}{\sum_{\boldsymbol{y}\in\Omega^n}\exp\{\boldsymbol{T}'(\boldsymbol{y})\boldsymbol{\psi}\}} \\ &= \frac{1}{c(\boldsymbol{\psi})}h_{\boldsymbol{\psi}}(\boldsymbol{z}), \end{aligned} \tag{3.10}$$

where $h_{\boldsymbol{\psi}}(\boldsymbol{z})$ is the known likelihood function, up to a proportionality constant, and $c(\boldsymbol{\psi})$ is the intractable normalizing constant. Now let $\boldsymbol{\phi}$ be an arbitrary, but fixed, point in the

parameter space $\boldsymbol{\Psi}$. Note that in theory $\boldsymbol{\phi}$ is arbitrary; in practice, it is desirable for $\boldsymbol{\phi}$ to be close to $\boldsymbol{\psi}$. Then observe the following, where integrals have been used instead of summations for generality (in the autologistic model, μ is taken as a counting measure):

$$\begin{aligned} c(\boldsymbol{\psi}) &= \int h_{\boldsymbol{\psi}}(\boldsymbol{z}) d\mu(\boldsymbol{z}) \\ &= \int h_{\boldsymbol{\psi}}(\boldsymbol{z}) \frac{1}{f_{\boldsymbol{\phi}}(\boldsymbol{z})} f_{\boldsymbol{\phi}}(\boldsymbol{z}) d\mu(\boldsymbol{z}) \\ &= \int h_{\boldsymbol{\psi}}(\boldsymbol{z}) \frac{1}{f_{\boldsymbol{\phi}}(\boldsymbol{z})} d\mathrm{P}_{\boldsymbol{\phi}}(\boldsymbol{z}) \\ &= \int h_{\boldsymbol{\psi}}(\boldsymbol{z}) \frac{c(\boldsymbol{\phi})}{h_{\boldsymbol{\phi}}(\boldsymbol{z})} d\mathrm{P}_{\boldsymbol{\phi}}(\boldsymbol{z}) \\ &= c(\boldsymbol{\phi}) \int \frac{h_{\boldsymbol{\psi}}(\boldsymbol{z})}{h_{\boldsymbol{\phi}}(\boldsymbol{z})} d\mathrm{P}_{\boldsymbol{\phi}}(\boldsymbol{z}) \\ &= c(\boldsymbol{\phi}) \mathrm{E}_{\boldsymbol{\phi}} \left\{ \frac{h_{\boldsymbol{\psi}}(\boldsymbol{Z})}{h_{\boldsymbol{\phi}}(\boldsymbol{Z})} \right\}. \end{aligned} \tag{3.11}$$

Hence, from (3.11), the following equation for the ratio of normalizing constants is obtained:

$$\frac{c(\boldsymbol{\psi})}{c(\boldsymbol{\phi})} = \mathrm{E}_{\boldsymbol{\phi}} \left\{ \frac{h_{\boldsymbol{\psi}}(\boldsymbol{Z})}{h_{\boldsymbol{\phi}}(\boldsymbol{Z})} \right\}. \tag{3.12}$$

This ratio of normalizing constants can then be estimated as a Monte Carlo average in the following manner, where $\boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots, \boldsymbol{Z}_m$ are realizations generated from $\mathrm{P}_{\boldsymbol{\phi}}$ using the Gibbs sampler (see section 3.3.2 below):

$$\frac{c(\boldsymbol{\psi})}{c(\boldsymbol{\phi})} \approx \frac{1}{m} \sum_{i=1}^{m} \frac{h_{\boldsymbol{\psi}}(\boldsymbol{Z}_i)}{h_{\boldsymbol{\phi}}(\boldsymbol{Z}_i)}. \tag{3.13}$$

While at first glance the above estimate of the ratio of normalizing constants may appear useless toward achieving an approximation to the MLE, it is in fact the cornerstone of MCMCML. To comprehend this crucial fact, note that $\operatorname{argmax}_{\boldsymbol{\psi} \in \boldsymbol{\Psi}} \log \{c(\boldsymbol{\phi}) \cdot l_{\boldsymbol{z}}(\boldsymbol{\psi})\} = \operatorname{argmax}_{\boldsymbol{\psi} \in \boldsymbol{\Psi}} \log \{l_{\boldsymbol{z}}(\boldsymbol{\psi})\}$, where $l_{\boldsymbol{z}}(\boldsymbol{\psi})$ denotes the log-likelihood. Hence, for purposes of op-

timization $l_{\boldsymbol{z}}(\boldsymbol{\psi})$ can be equivalently maximized as follows:

$$\begin{aligned} l_{\boldsymbol{z}}(\boldsymbol{\psi}) &\equiv \log\left\{c(\boldsymbol{\phi})f_{\boldsymbol{\psi}}(\boldsymbol{z})\right\} \\ &= \log\left\{c(\boldsymbol{\phi})\frac{h_{\boldsymbol{\psi}}(\boldsymbol{z})}{c(\boldsymbol{\psi})}\right\} \\ &= \log\left\{h_{\boldsymbol{\psi}}(\boldsymbol{z})\right\} + \log\left\{\frac{c(\boldsymbol{\phi})}{c(\boldsymbol{\psi})}\right\} \\ &= \log\left\{h_{\boldsymbol{\psi}}(\boldsymbol{z})\right\} - \log\left\{\frac{c(\boldsymbol{\psi})}{c(\boldsymbol{\phi})}\right\}. \end{aligned} \tag{3.14}$$

Consequently, an MCMC approximate log likelihood, $l_{m,\boldsymbol{z}_{obs}}(\boldsymbol{\psi})$, can be obtained by replacing $\frac{c(\boldsymbol{\psi})}{c(\boldsymbol{\phi})}$ with the right hand side of (3.13):

$$l_{m,\boldsymbol{z}_{obs}}(\boldsymbol{\psi}) = \log\left\{h_{\boldsymbol{\psi}}(\boldsymbol{z}_{obs})\right\} - \log\left\{\frac{1}{m}\sum_{i=1}^{m}\frac{h_{\boldsymbol{\psi}}(\boldsymbol{Z}_i)}{h_{\boldsymbol{\phi}}(\boldsymbol{Z}_i)}\right\}, \tag{3.15}$$

where the m subscript on the $l_{m,\boldsymbol{z}_{obs}}(\boldsymbol{\psi})$ indicates the amount of Monte Carlo effort and $\boldsymbol{z}_{obs}$ indicates the lattice realization of the MRF that was actually observed, rather than one of the realizations $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_m$ generated using the Gibbs sampler. The MCMCMLEs, $\hat{\boldsymbol{\psi}}_{MC}$, are then obtained by numerically maximizing (3.15) with respect to $\boldsymbol{\psi}$. The corresponding standard errors are obtained by calculating the MCMC approximate observed Fisher information matrix, which amounts to evaluating the inverse of the Hessian of $l_{m,\boldsymbol{z}_{obs}}(\boldsymbol{\psi})$ at $\hat{\boldsymbol{\psi}}_{MC}$ ([18]).

The Markov chain Monte Carlo maximum likelihood estimates are obtained through the clever use of an approximation of a ratio of normalizing constants, but what, if anything, can be said about the properties of such estimators? As it turns out, essentially every property the statistical modeler could possibly desire holds for the MCMCMLEs. In particular, Geyer and Thompson ([15]) proved that the MCMCMLEs converge almost surely to the true, but unknown, parameter values (i.e. the MCMCMLEs are strongly consistent) as a result of the Mean Ergodic Theorem. Additionally, Huffer and Wu ([23]) demonstrated in simulation studies that the MCMCMLEs are asymptotically normally distributed. Finally, Graham ([18])

verified via a simulation study that "both the Monte Carlo likelihood ratio test statistic and the Monte Carlo Wald test statistic have their usual asymptotic chi-squared distributions" ([23]). Hence, the method of MCMCML is a useful method for parameter estimation, with its only major drawback stemming from the computational expense associated with generating the realizations $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_m$ from $\mathrm{P}_{\boldsymbol{\phi}}$, and choice of an appropriate $\boldsymbol{\phi}$.

3.3.2 MCMCML: Gibbs Sampler

Section 3.3.1 elaborated on how a Markov chain of realizations $(\boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots, \boldsymbol{Z}_m)$ generated from $\mathrm{P}_{\boldsymbol{\phi}}$ could be used to form a Monte Carlo average that approximates the ratio of normalizing constants, which in turn leads to an approximation of the true MLE. It was assumed throughout section 3.3.1, however, that $\boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots, \boldsymbol{Z}_m$ were at hand. This section takes a step back and discusses how such realizations are generated using the Gibbs sampler, which was first described in the literature by Geman and Geman ([12]). Note that the Metropolis-Hastings algorithm can also be used to generate such realizations, but for the purposes of this thesis, only the Gibbs sampler is implemented, and, thus, it is the only method developed here in any detail. The following development of the Gibbs sampler represents a compilation of descriptions available in the literature, including [12], ([31]), and [23].

Suppose the lattice of interest, $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_n)'$, has probability measure $\mathrm{P}_{\boldsymbol{\phi}}$, where $\boldsymbol{\phi}$ is as defined in section 3.3.1, and a site conditional distribution is given by (3.3), or by (3.4) if there are covariates observed on the MRF, where a known $\boldsymbol{\phi}$ has been substituted for the unknown $\boldsymbol{\psi}$. For simplicity of presentation, assume without any loss of generality that covariates have not been observed over the MRF. Additionally, while the spatial Markovian assumption dictates that the conditional distribution for a site only depends on the nearest neighbors for that site, the presentation of the Gibbs sampler is more easily conveyed, as will be seen shortly, if the conditioning is expressed in terms of all other sites of the lattice, rather than in terms of only the nearest neighbors. Hence, the conditional distribution of the random

variable corresponding to site i is the following:

$$\begin{aligned} f(z_i|\{z_j : j \neq i\}; \boldsymbol{\phi}) &= f(z_i|z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_n; \boldsymbol{\phi}) \\ &= \frac{\exp\{\boldsymbol{t}'(z_i, \boldsymbol{z}_{-i})\boldsymbol{\phi}\}}{\sum_{y_i \in \Omega_i} \exp\{\boldsymbol{t}'(y_i, \boldsymbol{y}_{-i})\boldsymbol{\phi}\}}, \quad i = 1, \ldots, n, \end{aligned} \tag{3.16}$$

where $\boldsymbol{z}_{-i} = \{z_j : j \neq i\}$. Now let $\boldsymbol{Z}_t = (Z_{t1}, Z_{t2}, \ldots, Z_{tn})'$ be the t^{th} realization of the lattice within the Markov chain. Then the transition to the $(t+1)^{th}$ realization, $\boldsymbol{Z}_{t+1} = (Z_{(t+1)1}, Z_{(t+1)2}, \ldots, Z_{(t+1)n})'$, is accomplished through the following algorithm (i.e. the Gibbs sampler).

The Gibbs Sampler:

Given $\boldsymbol{z}_t = (z_{t1}, \ldots, z_{tn})'$, generate

1. $Z_{(t+1)1} \sim f(z_1|z_{t2}, \ldots, z_{tn}; \boldsymbol{\phi})$

2. $Z_{(t+1)2} \sim f(z_2|z_{(t+1)1}, z_{t3}, \ldots, z_{tn}; \boldsymbol{\phi})$

3. $Z_{(t+1)3} \sim f(z_2|z_{(t+1)1}, z_{(t+1)2}, z_{t4}, \ldots, z_{tn}; \boldsymbol{\phi})$

   $\vdots$

n. $Z_{(t+1)n} \sim f(z_2|z_{(t+1)1}, z_{(t+1)2}, \ldots, z_{(t+1)(n-1)}; \boldsymbol{\phi})$.

Note that "generate" in the above context means to randomly sample from the specified conditional distribution site by site. Hence, the realization for site 1 of the $(t+1)^{th}$ realization, $z_{(t+1)1}$, is obtained by randomly sampling a single value from the conditional distribution specified by $f(z_1|z_{t2}, \ldots, z_{tn}; \boldsymbol{\phi})$. Once all n steps have been carried out, the $(t+1)^{th}$ realization in the Markov chain is obtained. Each time the n steps listed above are completed, a single *sweep* of the Gibbs sampler has been performed. The starting, or initial, state for the Gibbs sampler, $\boldsymbol{Z}_0$, is arbitrary, but the observed lattice, $\boldsymbol{z}_{obs}$, or a lattice of zeros is often used as the initial state. Furthermore, as mentioned above in section 3.3.1, while the value used

for $\boldsymbol{\phi}$ is in theory arbitrary, in practice it is desirable that $\boldsymbol{\phi}$ be close to $\boldsymbol{\psi}$. Typically, a burn-in period, call it b, is specified where the first b sweeps of the Gibbs sampler are used to establish convergence of the chain to the desired MRF. These initial sweeps are not used in calculating the Monte Carlo average specified in (3.13). Additionally, a spacing value, call it δ, is often implemented such that only every δ^{th} sweep, after burn-in, is kept for use in computing the aforementioned Monte Carlo average. The objective of this spacing is to reduce the dependence between the kept realizations of the Gibbs sampler. Note that Geyer conveys that "in many cases $\delta = 1$ is optimal and in almost all cases the optimal δ is less than five" ([14]). Hence, if $\boldsymbol{Y}_j$ denotes the Markov chain lattice after the j^{th} sweep of the Gibbs sampler and $\boldsymbol{Z}_k$ denotes the lattice corresponding to the k^{th} sweep used in computing the Monte Carlo average in (3.13), then $\boldsymbol{Z}_k = \boldsymbol{Y}_{b+\delta k}$, $k = 1, \ldots, m$, where m is the desired number of realizations for computing the Monte Carlo average.

3.3.3 MCMCML: Application to the ALM

To demonstrate the use of MCMCML, as well as the Gibbs sampler, with a concrete example, consider the Ising model with a single covariate. In particular, the joint probability mass function, or likelihood, for the Ising model with a single covariate, $\boldsymbol{x} = (x_1, \ldots, x_n)'$, is the following:

$$f(\boldsymbol{z}|\boldsymbol{x};\boldsymbol{\psi}) = \frac{\exp\left\{\alpha\sum_{i=1}^{n} z_i + \frac{1}{2}\theta\sum_{i=1}^{n} z_i \sum_{j\in N_i} z_j + \beta\sum_{i=1}^{n} z_i x_i\right\}}{\sum_{\boldsymbol{y}\in\{0,1\}^n} \exp\left\{\alpha\sum_{i=1}^{n} y_i + \frac{1}{2}\theta\sum_{i=1}^{n} y_i \sum_{j\in N_i} y_j + \beta\sum_{i=1}^{n} y_i x_i\right\}}. \tag{3.17}$$

Now select a fixed point, $\boldsymbol{\phi}$, in the parameter space $\boldsymbol{\Psi}$. Typically, the pseudolikelihood estimates, $\widehat{\boldsymbol{\psi}}_{PL} = (\widehat{\alpha}_{PL}, \widehat{\theta}_{PL}, \widehat{\beta}_{PL})'$, are used for $\boldsymbol{\phi}$, i.e. $\boldsymbol{\phi} \equiv \widehat{\boldsymbol{\psi}}_{PL}$ ([18]). Then, by (3.12) and

(3.13), $\frac{c(\boldsymbol{\psi})}{c(\widehat{\boldsymbol{\psi}}_{PL})}$ is approximated as follows:

$$\frac{c(\boldsymbol{\psi})}{c(\widehat{\boldsymbol{\psi}}_{PL})} \approx \frac{1}{m}\sum_{i=1}^{m}\frac{h_{\boldsymbol{\psi}}(\boldsymbol{Z}_i)}{h_{\widehat{\boldsymbol{\psi}}_{PL}}(\boldsymbol{Z}_i)}, \tag{3.18}$$

where m is the number of Markov chain realizations, and $h_{\boldsymbol{\psi}}(\boldsymbol{Z}_i) = \exp\left\{\alpha\sum_{i=1}^{n} z_i + \frac{1}{2}\theta\sum_{i=1}^{n} z_i\sum_{j\in N_i} z_j + \beta\sum_{i=1}^{n} z_i x_i\right\}$, with $h_{\widehat{\boldsymbol{\psi}}_{PL}}(\boldsymbol{Z}_i)$ defined analogously, and $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_m$ are realizations generated from $\mathrm{P}_{\widehat{\boldsymbol{\psi}}_{PL}}$ using the Gibbs sampler (this is explained in detail below). Using (3.14) and (3.15), the MCMC approximate log likelihood, $l_{m,\boldsymbol{z}_{obs}}(\boldsymbol{\psi})$, is then given by the following equation:

$$\begin{aligned} l_{m,\boldsymbol{z}_{obs}}(\boldsymbol{\psi}) &= \log\left\{\exp\left\{\alpha\sum_{i=1}^{n} z_i^{obs} + \frac{1}{2}\theta\sum_{i=1}^{n} z_i^{obs}\sum_{j\in N_i} z_j^{obs} + \beta\sum_{i=1}^{n} z_i^{obs}x_i\right\}\right\} \\ &\quad -\log\left\{\frac{1}{m}\sum_{i=1}^{m}\frac{\exp\left\{\alpha\sum_{k=1}^{n} z_{ik} + \frac{1}{2}\theta\sum_{k=1}^{n} z_{ik}\sum_{j\in N_k} z_{ij} + \beta\sum_{k=1}^{n} z_{ik}x_k\right\}}{\exp\left\{\hat{\alpha}_{PL}\sum_{k=1}^{n} z_{ik} + \frac{1}{2}\hat{\theta}_{PL}\sum_{k=1}^{n} z_{ik}\sum_{j\in N_k} z_{ij} + \hat{\beta}_{PL}\sum_{k=1}^{n} z_{ik}x_k\right\}}\right\} \\ &= -\log\left\{\frac{1}{m}\sum_{i=1}^{m}\exp\left\{(\alpha-\hat{\alpha}_{PL})\sum_{k=1}^{n} z_{ik} + (\theta-\hat{\theta}_{PL})\frac{1}{2}\sum_{k=1}^{n} z_{ik}\sum_{j\in N_k} z_{ij} + (\beta-\hat{\beta}_{PL})\sum_{k=1}^{n} z_{ik}x_k\right\}\right\} \\ &\quad + \alpha\sum_{i=1}^{n} z_i^{obs} + \frac{1}{2}\theta\sum_{i=1}^{n} z_i^{obs}\sum_{j\in N_i} z_j^{obs} + \beta\sum_{i=1}^{n} z_i^{obs}x_i, \end{aligned} \tag{3.19}$$

where $\boldsymbol{z}_{obs} = (z_1^{obs}, \ldots, z_n^{obs})'$ and $\boldsymbol{z}_i = (z_{i1}, \ldots, z_{in})'$, $i = 1, \ldots, m$. The MCMCMLEs, $\widehat{\boldsymbol{\psi}}_{MC} = (\hat{\alpha}_{MC}, \hat{\theta}_{MC}, \hat{\beta}_{MC})'$, are obtained by numerically maximizing (3.19) with respect to $\boldsymbol{\psi} = (\alpha, \theta, \beta)'$. Finally, the standard errors corresponding to $\widehat{\boldsymbol{\psi}}_{MC}$ are obtained by evaluating the inverse of the Hessian of $l_{m,\boldsymbol{z}_{obs}}(\boldsymbol{\psi})$ at $\widehat{\boldsymbol{\psi}}_{MC}$, i.e. computing the observed Fisher information matrix.

To obtain the realizations $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_m$ generated from $\mathrm{P}_{\widehat{\boldsymbol{\psi}}_{PL}}$ using the Gibbs sampler, first note that the conditional distribution for the random variable corresponding to site i, given

the values of the random variables at all other locations of the lattice, is the following under the spatial Markovian assumption:

$$f(z_i|\{z_j \in N_i\}, x_i; \widehat{\boldsymbol{\psi}}_{PL}) = \frac{\exp\left\{\hat{\alpha}_{PL} z_i + \hat{\theta}_{PL} z_i \sum_{j \in N_i} z_j + \hat{\beta}_{PL} z_i x_i\right\}}{1 + \exp\left\{\hat{\alpha}_{PL} + \hat{\theta}_{PL} \sum_{j \in N_i} z_j + \hat{\beta}_{PL} x_i\right\}}. \tag{3.20}$$

As was done in section 3.3.2, the conditional distribution, given in (3.20), will be reexpressed with the conditioning being over all other sites of the lattice rather than just the neighboring sites, so as to simplify the following illustration of the Gibbs sampler. Hence, $f(z_i|z_1, \ldots z_{i-1}, z_{i+1}, \ldots, z_n, x_i; \widehat{\boldsymbol{\psi}}_{PL}) \equiv f(z_i|\{z_j \in N_i\}, x_i; \widehat{\boldsymbol{\psi}}_{PL})$. Let $\boldsymbol{Z}_0$, the initial state of the Markov chain, be a lattice of all zeros or the observed lattice $\boldsymbol{z}_{obs}$, then the $(t+1)^{th}$ lattice realization of the Markov chain, $\boldsymbol{Z}_{t+1} = (Z_{(t+1)1}, \ldots, Z_{(t+1)n})'$, given the t^{th} lattice realization of the Markov chain, $\boldsymbol{Z}_t = (Z_{t1}, \ldots, Z_{tn})'$, is transitioned to using the Gibbs sampler as follows.

1. (a) Calculate $r_1 \equiv P(Z_{(t+1)1} = 1|z_{t2}, \ldots, z_{tn}, x_1; \widehat{\boldsymbol{\psi}}_{PL})$, the conditional probability that the random variable corresponding to site 1 assumes a value of 1. Note that the conditional probability that the random variable corresponding to site 1 assumes a value of 0 is simply $1 - r_1$.

   (b) Randomly sample a single realization, u, from $U \sim \text{uniform}(0, 1)$.

   (c) If $u \leq r_1$, then set $z_{(t+1)1} = 1$, otherwise set $z_{(t+1)1} = 0$.

2. (a) Calculate $r_2 \equiv P(Z_{(t+1)2} = 1|z_{(t+1)1}, z_{t3}, \ldots, z_{tn}, x_2; \widehat{\boldsymbol{\psi}}_{PL})$, the conditional probability that the random variable corresponding to site 2 assumes a value of 1.

   (b) Randomly sample a single realization, u, from $U \sim \text{uniform}(0, 1)$.

   (c) If $u \leq r_2$, then set $z_{(t+1)2} = 1$, otherwise set $z_{(t+1)2} = 0$.

$\vdots$

n. (a) Calculate $r_n \equiv P(Z_{(t+1)n} = 1 | z_{(t+1)1}, z_{(t+1)2}, \ldots, z_{(t+1)(n-1)}, x_n; \widehat{\boldsymbol{\psi}}_{PL})$, the conditional probability that the random variable corresponding to site n assumes a value of 1.

(b) Randomly sample a single realization, u, from $U \sim \text{uniform}(0, 1)$.

(c) If $u \leq r_n$, then set $z_{(t+1)n} = 1$, otherwise set $z_{(t+1)n} = 0$.

The above algorithm is then performed until m realizations of the lattice have been obtained, after accounting for both the prespecified burn-in and spacing. This will leave the user with $\boldsymbol{Z_1}, \ldots, \boldsymbol{Z}_m$, or m realizations generated from $\mathrm{P}_{\widehat{\boldsymbol{\psi}}_{PL}}$, which are used in computing the MCMC approximate log likelihood (3.19) that is then numerically maximized to obtain the MCMCMLEs, $\widehat{\boldsymbol{\psi}}_{MC}$.

3.4 Maximum Generalized Pseudolikelihood Estimation

The method of generalized pseudolikelihood (GPL), developed by Huang and Ogata ([22]), is intended to serve as a compromise between PL and MCMCML, as previously discussed in Chapter 1. The essence of the method is to first divide the lattice of interest into n overlapping groups such that each site is centered in its own distinct group. Second, for each of these groups, a group joint likelihood function is constructed, where the corresponding normalizing constants are calculated via brute force, and then the product over all group joint likelihood functions is taken to produce the generalized pseudolikelihood function. Finally, the logarithm of this function is then numerically maximized to obtain the corresponding maximum generalized pseudolikelihood estimates (MGPLEs). The remainder of this section is broken into two subsections. In section 3.4.1 a rigorous treatment of GPL is provided that is essentially a compilation of the treatments of the method given in both [22] and [32], while in section 3.4.2 the method is demonstrated within the context of the Ising model, when generalized to incorporate a covariate.

Figure 3.1: "L-Shape" Group Structure

e	e	e	e	e	e	e
e	e	e	e	e	e	e
e	e	B_3	e	e	e	e
e	B_1	$Z_{(1)}$	B_5	e	e	e
e	B_2	Z_i	$Z_{(2)}$	B_7	e	e
e	e	B_4	B_6	e	e	e
e	e	e	e	e	e	e

An "L-shaped" group representing $g(i)$ on a 7×7 lattice subset of S, where $\boldsymbol{Z}_{g(i)} = \{Z_i, Z_{(1)}, Z_{(2)}\}$. Under a first-order neighborhood system, the 7 B sites represent the ***locations*** of the fixed boundary neighbors of $g(i)$, while the 39 e sites represent the ***locations*** of the external sites that are conditionally independent of the sites of $g(i)$ under a spatial Markovian assumption.

3.4.1 GPL: Background

Let $g(i)$ denote the *group* of sites "centered" at site i, $i = 1, \ldots, n$. Typically, each site's group consists of that site and its nearest neighbors, but in theory the modeler is free to implement any sort of group configuration imaginable and different sites can even have different group configurations. Additionally, let $\boldsymbol{Z}_{g(i)} = \{Z_k : k \in g(i)\}$ be the set of variables corresponding to sites contained within group i, and let $\boldsymbol{Z}^{g(i)} = \{Z_k : k \notin g(i)\}$ be the set of variables corresponding to all sites of the lattice that are not members of group i. For $k \in g(i)$, let $N_k^{g(i)}$ denote the neighbors of site k that are members of $g(i)$, or "internal" to $g(i)$, and let $N_k^{\partial g(i)}$ denote the neighbors of site k that are not members of $g(i)$, i.e. the sites of $N_k^{\partial g(i)}$ are "boundary" neighbors of the i^{th} group. Observe then that the Neighborhood set of site k, i.e. N_k, is $N_k = N_k^{g(i)} \cup N_k^{\partial g(i)}$, where $N_k^{g(i)} \cap N_k^{\partial g(i)} = \emptyset$. For example, in Figure 3.1 a three-site "L-shaped" group structure, which is used in Chapter 7 when modeling the fire occurrence data of Oregon and Washington state, is illustrated under a first-order neighborhood system. More specifically, $g(i) = \{i, (1), (2)\}$, where the parentheses are to

indicate that site (k) is not necessarily the same as site k. Furthermore, in Figure 3.1, $N_i^{g(i)} = \{(1),(2)\}$ and $N_i^{\partial g(i)} = \{B_2, B_4\}$, while $N_{(1)}^{g(i)} = \{i\}$ and $N_{(1)}^{\partial g(i)} = \{B_1, B_3, B_5\}$. As was indicated in Chapter 1, the need for the distinction between internal neighbors and boundary neighbors for the neighborhood sets of sites belonging to $g(i)$ is not expressed in the literature, but was discovered over the course of this thesis work. The relevance of this discovery is tied to the sufficient statistics for the group joint likelihood functions and is further addressed below. First, however, the formation of the group joint likelihoods must be discussed.

For each group, $g(i)$, the group joint likelihood, denoted $\mathrm{GL}_i(\boldsymbol{\psi})$, is constructed. The form of each group joint likelihood is essentially identical to the right hand side of (3.1), or (3.2) if covariates have also been observed over the MRF. For presentation ease, assume without any loss of generality that covariates have not been observed over the MRF. In particular, $\forall i = 1, \ldots, n$, the functional form of the i^{th} group joint likelihood is the following:

$$\mathrm{GL}_i(\boldsymbol{\psi}) \quad = \quad \left\{ \frac{\exp\{\boldsymbol{T}'\left(\boldsymbol{z}_{g(i)}, \boldsymbol{z}^{g(i)}\right) \cdot \boldsymbol{\psi}\}}{\sum_{\boldsymbol{y} \in \Omega^{|g(i)|}} \exp\{\boldsymbol{T}'\left(\boldsymbol{y}_{g(i)}, \boldsymbol{y}^{g(i)}\right) \cdot \boldsymbol{\psi}\}} \right\}^{\frac{1}{|g(i)|}}, \tag{3.21}$$

where $|g(i)|$ denotes the number of sites belonging to group i. Three things must be noted in considering the functional form of the group joint likelihoods. First, even though $\boldsymbol{z} = \boldsymbol{z}_{g(i)} \cup \boldsymbol{z}^{g(i)}$, the group sufficient statistics, $\boldsymbol{T}'\left(\boldsymbol{z}_{g(i)}, \boldsymbol{z}^{g(i)}\right)$, have been expressed as functions of $\boldsymbol{z}_{g(i)}$ and $\boldsymbol{z}^{g(i)}$ in (3.21), rather than as functions of $\boldsymbol{z}$, as in (3.1). This is done to emphasize that for each group joint likelihood $\boldsymbol{z}_{g(i)}$ and $\boldsymbol{z}^{g(i)}$ are viewed differently; in particular, while the $\boldsymbol{z}_{g(i)}$ represent the observed values of the *variables* corresponding to sites in group i, the $\boldsymbol{z}^{g(i)}$ represent *fixed* values corresponding to sites not in group i. Hence, the normalizing constant of the i^{th} group's joint likelihood (denominator of (3.21)) is formed by summing over all possible realizations of the variables corresponding to only the sites that are members of group i; for each such summand, $\boldsymbol{z}^{g(i)}$ remains fixed. Second, the sufficient statistics in (3.21) have a slightly different form than the sufficient statistics in (3.1). In particular, the sufficient

statistics for a group joint likelihood must distinguish a neighbor relation between two sites of the group and a neighbor relation between one site of the group and one boundary neighbor site of the group. For example, in Figure 3.1, the neighbor relation between sites i and (2) must be distinguished from the neighbor relation between sites i and B_3. This distinction is important because the neighbor relations between a group site and a boundary neighbor site must not be downweighted relative to the neighbor relations between two sites of a group in the joint likelihood. This crucial fact, which was alluded to above as the distinction between internal neighbors and boundary neighbors, is surprisingly void in the literature, but was discovered over the course of this thesis work. An example demonstrating how the sufficient statistics of the group joint likelihoods are altered when the internal neighbor relations of a group are distinguished from the boundary neighbor relations of that group is provided in section 3.4.2. Third and finally, while in section 1.2.4 it was assumed that $|g(i)|$ was the same for all i, this need not be the case and, consequently, the $|g(i)|^{th}$ root is taken to obtain the i^{th} group joint likelihood. If, however, $|g(i)| = C \ \forall i = 1, \ldots, n$, then taking the C^{th} root of each group joint likelihood is often omitted, as was done in section 1.2.4, since that constant, as will be seen shortly, is irrelevant.

The generalized pseudolikelihood function, denoted $\mathrm{GPL}(\boldsymbol{\psi})$, is formed by taking the product of the group joint likelihood functions over all sites:

$$\begin{aligned} \mathrm{GPL}(\boldsymbol{\psi}) &= \prod_{i=1}^{n} \mathrm{GL}_i(\boldsymbol{\psi}) \\ &= \prod_{i=1}^{n} \left\{ \frac{\exp\{\boldsymbol{T}'\left(\boldsymbol{z}_{g(i)}, \boldsymbol{z}^{g(i)}\right) \cdot \boldsymbol{\psi}\}}{\sum_{\boldsymbol{y} \in \Omega^{|g(i)|}} \exp\{\boldsymbol{T}'\left(\boldsymbol{y}_{g(i)}, \boldsymbol{y}^{g(i)}\right) \cdot \boldsymbol{\psi}\}} \right\}^{\frac{1}{|g(i)|}}. \end{aligned} \tag{3.22}$$

Note that the GPL function, just like the PL function, is not a true likelihood, but by multiplying joint likelihoods over small groups, it is in some sense "closer" to a true likelihood. Although already discussed in Chapter 1, it is worth mentioning again how GPL is a compromise between PL and MCMCML. If $g(i) = \{i\} \ \forall i = 1, \ldots, n$, then GPL reduces to PL since

(3.5) and (3.22) would be equivalent. Furthermore, if $g(i) = \{1, \ldots, n\}$ $\forall i = 1, \ldots, n$, then GPL is equivalent to maximum likelihood since (3.1) and (3.22) would be equivalent. To see this, note that if $g(i) = \{1, \ldots, n\}$ $\forall i = 1, \ldots, n$, then $|g(i)| = n$, $\boldsymbol{z}_{g(i)} = \boldsymbol{z}$, and $\boldsymbol{z}^{g(i)} = \emptyset$ $\forall i = 1, \ldots, n$, and (3.22) then becomes the following:

$$\begin{aligned}
\text{GPL}(\boldsymbol{\psi}) &= \prod_{i=1}^{n} \left\{ \frac{\exp\{\boldsymbol{T}'\left(\boldsymbol{z}_{g(i)}, \boldsymbol{z}^{g(i)}\right) \cdot \boldsymbol{\psi}\}}{\sum_{\boldsymbol{y} \in \Omega^{|g(i)|}} \exp\{\boldsymbol{T}'\left(\boldsymbol{y}_{g(i)}, \boldsymbol{y}^{g(i)}\right) \cdot \boldsymbol{\psi}\}} \right\}^{\frac{1}{|g(i)|}} \\
&= \prod_{i=1}^{n} \left\{ \frac{\exp\{\boldsymbol{T}'(\boldsymbol{z}) \cdot \boldsymbol{\psi}\}}{\sum_{\boldsymbol{y} \in \Omega^{n}} \exp\{\boldsymbol{T}'(\boldsymbol{y}) \cdot \boldsymbol{\psi}\}} \right\}^{\frac{1}{n}} \\
&= \left\{ \frac{\exp\{\boldsymbol{T}'(\boldsymbol{z}) \cdot \boldsymbol{\psi}\}}{\sum_{\boldsymbol{y} \in \Omega^{n}} \exp\{\boldsymbol{T}'(\boldsymbol{y}) \cdot \boldsymbol{\psi}\}} \right\}^{n\frac{1}{n}} \\
&= \frac{\exp\{\boldsymbol{T}'(\boldsymbol{z}) \cdot \boldsymbol{\psi}\}}{c(\boldsymbol{\psi})} \\
&= \text{L}(\boldsymbol{\psi}). \qquad (3.23)
\end{aligned}$$

Thus, for a fixed group structure over the entire lattice, if the corresponding group size is 1, then GPL is PL, if the corresponding group size is n, then GPL is maximum likelihood (which MCMCML approximates), and if the corresponding group size is strictly between 1 and n, then GPL is somewhere between PL and MCMCML.

As is the case with PL and MCMCML, the logarithm of the generalized pseudolikelihood function is typically used in practice and, hence, the log generalized pseudolikelihood function is the following:

$$\log\{\text{GPL}(\boldsymbol{\psi})\} = \log\left\{ \prod_{i=1}^{n} \left\{ \frac{\exp\{\boldsymbol{T}'\left(\boldsymbol{z}_{g(i)}, \boldsymbol{z}^{g(i)}\right) \cdot \boldsymbol{\psi}\}}{\sum_{\boldsymbol{y} \in \Omega^{|g(i)|}} \exp\{\boldsymbol{T}'\left(\boldsymbol{y}_{g(i)}, \boldsymbol{y}^{g(i)}\right) \cdot \boldsymbol{\psi}\}} \right\}^{\frac{1}{|g(i)|}} \right\}$$

$$= \sum_{i=1}^{n} \frac{1}{|g(i)|} \log\left\{ \frac{\exp\{\boldsymbol{T}'\left(\boldsymbol{z}_{g(i)}, \boldsymbol{z}^{g(i)}\right) \cdot \boldsymbol{\psi}\}}{\sum_{\boldsymbol{y} \in \Omega^{|g(i)|}} \exp\{\boldsymbol{T}'\left(\boldsymbol{y}_{g(i)}, \boldsymbol{y}^{g(i)}\right) \cdot \boldsymbol{\psi}\}} \right\}$$

$$= \sum_{i=1}^{n} \frac{1}{|g(i)|} \boldsymbol{T}' \left(\boldsymbol{z}_{g(i)}, \boldsymbol{z}^{g(i)} \right) \cdot \boldsymbol{\psi} - \sum_{i=1}^{n} \frac{1}{|g(i)|} \log \left\{ \sum_{\boldsymbol{y} \in \Omega^{|g(i)|}} \exp\{ \boldsymbol{T}' \left(\boldsymbol{y}_{g(i)}, \boldsymbol{y}^{g(i)} \right) \cdot \boldsymbol{\psi} \} \right\}. \tag{3.24}$$

The MGPLEs are then obtained by numerically maximizing (3.24) with respect to $\boldsymbol{\psi}$. Note that if $|g(i)| = C \ \forall i = 1, \ldots, n$, then the $\frac{1}{|g(i)|}$ terms become irrelevant in the maximization, which justifies their omission in the group joint likelihoods when the group sizes are all equivalent. The standard errors for the MGPLEs can be obtained by employing a resampling method, such as parametric bootstrapping. Finally, while GPL is a compromise between PL and MCMCML, basic asymptotic properties of the MGPLEs remain unestablished in the literature. As part of this dissertation, the strong consistency of the MGPLEs is proven in Chapter 4.

3.4.2 GPL: Application to ALM

To demonstrate the use of GPL with a concrete example, consider the Ising model with a single covariate, with probability mass function, or likelihood, as given in (3.17). If it is assumed that the group structure is the same for all n groups, then the group joint likelihood for group i, $i = 1, \ldots, n$, is the following:

$$\mathrm{GL}_i(\alpha, \theta, \beta) \quad =$$

$$\frac{\exp\left\{ \alpha \sum_{k \in g(i)} z_k + \theta \left\{ \frac{1}{2} \sum_{k \in g(i)} \sum_{j \in N_k^{g(i)}} z_k z_j + \sum_{k \in g(i)} \sum_{j \in N_k^{\partial g(i)}} z_k z_j \right\} + \beta \sum_{k \in g(i)} z_k x_k \right\}}{\sum_{\boldsymbol{y} \in \Omega^{|g(i)|}} \exp\left\{ \alpha \sum_{k \in g(i)} y_k + \theta \left\{ \frac{1}{2} \sum_{k \in g(i)} \sum_{j \in N_k^{g(i)}} y_k y_j + \sum_{k \in g(i)} \sum_{j \in N_k^{\partial g(i)}} y_k y_j \right\} + \beta \sum_{k \in g(i)} y_k x_k \right\}}, \tag{3.25}$$

where θ is the spatial dependence parameter multiplied by the two sums of neighboring site value products for both types of neighbor relations, and β is the covariate parameter. Hence, the vector of jointly sufficient statistics for the i^{th} group joint likelihood is $\boldsymbol{T}(\boldsymbol{z}_{g(i)}, \boldsymbol{z}^{g(i)}, \boldsymbol{x}^{g(i)}) = (T_1(\boldsymbol{z}_{g(i)}, \boldsymbol{z}^{g(i)}),\, T_2(\boldsymbol{z}_{g(i)},\, \boldsymbol{z}^{g(i)}), T_3(\boldsymbol{z}_{g(i)}, \boldsymbol{z}^{g(i)}, \boldsymbol{x}^{g(i)}))'$, where

$$T_1(\boldsymbol{z}_{g(i)}, \boldsymbol{z}^{g(i)}) = \sum_{k \in g(i)} z_k, \tag{3.26}$$

$$T_2(\boldsymbol{z}_{g(i)}, \boldsymbol{z}^{g(i)}) = \frac{1}{2} \sum_{k \in g(i)} \sum_{j \in N_k^{g(i)}} z_k z_j + \sum_{k \in g(i)} \sum_{j \in N_k^{\partial g(i)}} z_k z_j, \tag{3.27}$$

$$T_3(\boldsymbol{z}_{g(i)}, \boldsymbol{z}^{g(i)}, \boldsymbol{x}^{g(i)}) = \sum_{k \in g(i)} z_k x_k. \tag{3.28}$$

Note that $\boldsymbol{x}^{g(i)}$ denotes the vector of covariate values corresponding to the sites of group i. The analogous vector of jointly sufficient statistics for the true likelihood function, given by (3.17), is $\boldsymbol{T}(\boldsymbol{z}, \boldsymbol{x}) = (T_1(\boldsymbol{z}), T_2(\boldsymbol{z}), T_3(\boldsymbol{z}, \boldsymbol{x}))'$, where

$$T_1(\boldsymbol{z}) = \sum_{i=1}^{n} z_i, \tag{3.29}$$

$$T_2(\boldsymbol{z}) = \frac{1}{2} \sum_{i=1}^{n} \sum_{j \in N_i} z_i z_j, \tag{3.30}$$

$$T_3(\boldsymbol{z}, \boldsymbol{x}) = \sum_{i=1}^{n} z_i x_i. \tag{3.31}$$

In comparing the jointly sufficient statistics between the group joint likelihood and the true likelihood, it is evident that (3.26) and (3.29), as well as (3.28) and (3.31), are directly comparable. However, (3.27) and (3.30) are not directly comparable, and this is where the distinction between a neighbor relation involving two sites of the group and a neighbor relation involving one site of the group and one boundary neighbor site of the group is readily apparent. In particular, in (3.27), the purpose of the $\frac{1}{2}$ multiplier is to negate the double counting of each neighbor relation in the likelihood function since each $z_i z_j$ term appears twice in the sum. However, with a group joint likelihood function, a neighbor relation involving a site in the group and a boundary neighbor site of the group will not be double counted since the boundary

neighbor sites are treated as fixed. Thus, if those neighbor relation terms were multiplied by a $\frac{1}{2}$, such neighbor relations would be downweighted relative to the neighbor relations involving two sites within the group. Hence, the first portion of (3.27), $\frac{1}{2}\sum_{k\in g(i)}\sum_{j\in N_k^{g(i)}} z_k z_j$, deals with the terms corresponding to neighbor relations involving two sites of group i, while the second portion of (3.27), $\sum_{k\in g(i)}\sum_{j\in N_k^{\partial g(i)}} z_k z_j$, deals with the terms corresponding to neighbor relations involving one site of group i and one boundary neighbor site of group i.

The generalized pseudolikelihood function is then obtained by forming the product over all n group joint likelihoods given by (3.25). The logarithm of the GPL function would be the following:

$$\log\{\mathrm{GPL}(\alpha,\theta,\beta)\} = \sum_{i=1}^{n}\log\{\mathrm{GL}_i(\alpha,\theta,\beta)\}$$

$$= -\sum_{i=1}^{n}\log\left\{\sum_{\boldsymbol{y}\in\Omega^{|g(i)|}}\exp\left\{\alpha\sum_{k\in g(i)}y_k+\theta\left\{\frac{1}{2}\sum_{k\in g(i)}\sum_{j\in N_k^{g(i)}}y_k y_j+\sum_{k\in g(i)}\sum_{j\in N_k^{\partial g(i)}}y_k y_j\right\}+\beta\sum_{k\in g(i)}y_k x_k\right\}\right\}$$

$$+\sum_{i=1}^{n}\left\{\alpha\sum_{k\in g(i)}z_k+\theta\left\{\frac{1}{2}\sum_{k\in g(i)}\sum_{j\in N_k^{g(i)}}z_k z_j+\sum_{k\in g(i)}\sum_{j\in N_k^{\partial g(i)}}z_k z_j\right\}+\beta\sum_{k\in g(i)}z_k x_k\right\}. \qquad (3.32)$$

The MGPLEs, $\widehat{\boldsymbol{\psi}}_{GPL}=(\hat{\alpha}_{GPL},\hat{\theta}_{GPL},\hat{\beta}_{GPL})'$, are then obtained by numerically maximizing (3.32) with respect to $\boldsymbol{\psi}=(\alpha,\theta,\beta)'$. The MPLEs, $\widehat{\boldsymbol{\psi}}_{PL}=(\hat{\alpha}_{PL},\hat{\theta}_{PL},\hat{\beta}_{PL})'$, are generally used as starting values for the numerical optimization. The corresponding standard errors, as mentioned at the end of section 3.4.1, can be obtained using a resampling method such as parametric bootstrapping.

3.5 Maximum Block Generalized Pseudolikelihood Estimation

Similarly to GPL, the method of block generalized pseudolikelihood (BGPL), first presented in an unpublished paper by Lim et al. ([27]) and then developed in a Bayesian setting by Friel et al. ([11]), is intended to serve as a compromise between PL and MCMCML, as previously discussed in Chapter 1. The essence of the method is to first partition the lattice of interest into blocks of sites. Second, for each of these blocks, a block joint likelihood function is constructed, where all of the block normalizing constants are calculated via brute force. Third, the product of all block joint likelihood functions is formed to produce the block generalized pseudolikelihood function. Finally, the logarithm of this function is numerically maximized to obtain the corresponding maximum block generalized pseudolikelihood estimates (MBGPLEs).

The difference between GPL and BGPL was discussed in great detail in Chapter 1, but to reiterate, the *only* difference between the two methods is that GPL divides the lattice into *overlapping* groups, whereas BGPL *partitions* the lattice into (nonoverlapping) blocks perhaps reducing the level of dependence between blocks. Thus, while the following presentation of BGPL, which is divided into two subsections, will be nearly identical to the presentation of GPL given in section 3.4, it is given nonetheless for clarity and completeness. In section 3.5.1 a rigorous treatment of BGPL is provided that is essentially a compilation of the treatments of the method given in both [27] and [11], while in section 3.5.2 the method is demonstrated within the context of the Ising model, when generalized to incorporate a covariate.

3.5.1 BGPL: Background

The lattice of interest is first partitioned into $L \leq n$ blocks, where $b(l)$ denotes the sites "associated" with block l, $l = 1, \ldots, L$ (so $S = \bigcup_{l=1}^{L} b(l)$). Typically, the lattice is partitioned so that each block has an equal number of sites (if possible), where any "edge" sites not

Figure 3.2: 2×2 Block Structure

e	e	e	e	e	e	e
e	B_1	B_5	B_7	B_9	e	e
e	B_2	$Z_{(1)}$	$Z_{(3)}$	B_{10}	e	e
e	B_3	$Z_{(2)}$	$Z_{(4)}$	B_{11}	e	e
e	B_4	B_6	B_8	B_{12}	e	e
e	e	e	e	e	e	e
e	e	e	e	e	e	e

A 2×2 block representing $b(l)$ on a 7×7 lattice subset of S, where $\boldsymbol{Z}_{b(l)} = \{Z_{(1)}, Z_{(2)}, Z_{(3)}, Z_{(4)}\}$. Under a second-order neighborhood system, the 12 B sites represent the ***locations*** of the fixed boundary neighbors of $b(l)$, while the 33 e sites represent the ***locations*** of the external sites that are conditionally independent of the sites of $b(l)$ under a spatial Markovian assumption.

belonging to a block are then treated as fixed, but in theory the modeler is free to partition in such a way that each block has a different number of sites. Some practical blocking mechanisms include forming $b \times b$ blocks across the lattice for $b = 2, 3,$ or 4, or letting multiple adjacent columns (or rows) represent a block. Additionally, let $\boldsymbol{Z}_{b(l)} = \{Z_k : k \in b(l)\}$ be the set of variables corresponding to sites contained within block l, and let $\boldsymbol{Z}^{b(l)} = \{Z_k : k \notin b(l)\}$ be the set of variables corresponding to all sites of the lattice that are not members of block l. For $k \in b(l)$, let $N_k^{b(l)}$ denote the neighbors of site k that are members of $b(l)$, or "internal" to $b(l)$, and let $N_k^{\partial b(l)}$ denote the neighbors of site k that are not members of $b(l)$, i.e. the sites of $N_k^{\partial b(l)}$ are "boundary" neighbors of the l^{th} block. Observe that $N_k = N_k^{b(l)} \cup N_k^{\partial b(l)}$, where $N_k^{b(l)} \cap N_k^{\partial b(l)} = \emptyset$. For example, in Figure 3.2 a 2×2 block, under a second-order neighborhood system, is illustrated. More specifically, $b(l) = \{(1), (2), (3), (4)\}$, where the parentheses are to indicate that site (k) is not necessarily the same as site k. Furthermore, in Figure 3.2, $N_{(1)}^{b(l)} = \{(2), (3), (4)\}$ and $N_{(1)}^{\partial b(l)} = \{B_1, B_2, B_3, B_5, B_7\}$, while $N_{(4)}^{b(l)} = \{(1), (2), (3)\}$ and $N_{(4)}^{\partial b(l)} = \{B_6, B_8, B_{10}, B_{11}, B_{12}\}$. As was indicated in Chapter 1, and similarly discussed for

GPL, the need for the distinction between internal neighbors and boundary neighbors for the neighborhood sets of sites belonging to $b(l)$ is not expressed in the literature, but was discovered over the course of this thesis work. As with GPL, the relevance of this discovery is tied to the sufficient statistics for the block joint likelihood functions and is further addressed below. First, however, the formation of the block joint likelihoods is discussed.

For each block, $b(l)$, the block joint likelihood, denoted $\mathrm{BL}_l(\boldsymbol{\psi})$, is constructed. The form of each block joint likelihood is essentially identical to the right hand side of (3.1), or (3.2) if covariates have also been observed over the MRF. For presentation ease, again assume without any loss of generality that covariates have not been observed over the MRF. In particular, $\forall l = 1, \ldots, L$, the functional form of the l^{th} block joint likelihood is the following:

$$\mathrm{BL}_l(\boldsymbol{\psi}) \quad = \quad \frac{\exp\{\boldsymbol{T}'\left(\boldsymbol{z}_{b(l)}, \boldsymbol{z}^{b(l)}\right) \cdot \boldsymbol{\psi}\}}{\sum\limits_{\boldsymbol{y} \in \Omega^{|b(l)|}} \exp\{\boldsymbol{T}'\left(\boldsymbol{y}_{b(l)}, \boldsymbol{y}^{b(l)}\right) \cdot \boldsymbol{\psi}\}}, \tag{3.33}$$

where $|b(l)|$ denotes the number of sites belonging to block l. Two things must be noted in considering the functional form of the block joint likelihoods. First, even though $\boldsymbol{z} = \boldsymbol{z}_{b(l)} \cup \boldsymbol{z}^{b(l)}$, the block sufficient statistics, $\boldsymbol{T}'\left(\boldsymbol{z}_{b(l)}, \boldsymbol{z}^{b(l)}\right)$, have been expressed as functions of $\boldsymbol{z}_{b(l)}$ and $\boldsymbol{z}^{b(l)}$ in (3.33), rather than as functions of $\boldsymbol{z}$, as in (3.1). As with GPL, this is done to emphasize that for each block joint likelihood $\boldsymbol{z}_{b(l)}$ and $\boldsymbol{z}^{b(l)}$ are viewed differently; in particular, while $\boldsymbol{z}_{b(l)}$ represents the observed values of the *variables* corresponding to sites in block l, $\boldsymbol{z}^{b(l)}$ represents *fixed* values corresponding to sites not in block l. Hence, the normalizing constant of the l^{th} block's joint likelihood (denominator of (3.33)) is formed by summing over all possible realizations of the variables corresponding to only the sites that are members of block l; for each such summand, $\boldsymbol{z}^{b(l)}$ remains fixed. Second and finally, the sufficient statistics in (3.33) have a slightly different form than the sufficient statistics in (3.1). In particular, the sufficient statistics for a block joint likelihood, as with a group joint likelihood, must distinguish a neighbor relation between two sites of the block and a neighbor relation between one site of the block and one boundary neighbor site of the

block. For example, in Figure 3.2, the neighbor relation between sites (1) and (2) must be distinguished from the neighbor relation between sites (1) and B_5. This distinction is important because the neighbor relations between a block site and a boundary neighbor site must not be downweighted relative to the neighbor relations between two sites of a block in the joint likelihood. As with GPL, this crucial fact is surprisingly void in the literature, but was discovered over the course of this thesis work. An example demonstrating how the sufficient statistics of the block joint likelihoods are altered when the internal neighbor relations of a block are distinguished from the boundary neighbor relations of that block is provided in section 3.5.2.

The block generalized pseudolikelihood function, denoted BGPL($\boldsymbol{\psi}$), is formed by taking the product of the block joint likelihood functions over all blocks:

$$\begin{aligned} \text{BGPL}(\boldsymbol{\psi}) &= \prod_{l=1}^{L} \text{BL}_l(\boldsymbol{\psi}) \\ &= \prod_{l=1}^{L} \frac{\exp\{\boldsymbol{T}'\left(\boldsymbol{z}_{b(l)}, \boldsymbol{z}^{b(l)}\right) \cdot \boldsymbol{\psi}\}}{\sum_{\boldsymbol{y} \in \Omega^{|b(l)|}} \exp\{\boldsymbol{T}'\left(\boldsymbol{y}_{b(l)}, \boldsymbol{y}^{b(l)}\right) \cdot \boldsymbol{\psi}\}}. \end{aligned} \tag{3.34}$$

As with the GPL function, note that the BGPL function, just like the PL function, is not a true likelihood, but by multiplying joint likelihoods over small blocks, it is in some sense "closer" to a true likelihood. Although already discussed in Chapter 1, it is worth mentioning again how BGPL is a compromise between PL and MCMCML. If $b(l) = \{l\}\ \forall l = 1, \ldots, L$ (so $L = n$), then BGPL reduces to PL since (3.5) and (3.34) would be equivalent. Furthermore, if $b(l) = \{1, \ldots, n\}$ (so $L = 1$ block), then BGPL is equivalent to maximum likelihood since (3.1) and (3.34) would be equivalent. To see this, note that if $b(l) = \{1, \ldots, n\}$, then $|b(l)| = n$, $\boldsymbol{z}_{b(l)} = \boldsymbol{z}$, and $\boldsymbol{z}^{b(l)} = \emptyset$, and (3.34) then becomes the following:

$$\text{BGPL}(\boldsymbol{\psi}) = \prod_{l=1}^{L} \frac{\exp\{\boldsymbol{T}'\left(\boldsymbol{z}_{b(l)}, \boldsymbol{z}^{b(l)}\right) \cdot \boldsymbol{\psi}\}}{\sum_{\boldsymbol{y} \in \Omega^{|b(l)|}} \exp\{\boldsymbol{T}'\left(\boldsymbol{y}_{b(l)}, \boldsymbol{y}^{b(l)}\right) \cdot \boldsymbol{\psi}\}}$$

$$
\begin{aligned}
&= \frac{\exp\{\boldsymbol{T}'(\boldsymbol{z})\cdot\boldsymbol{\psi}\}}{\sum_{\boldsymbol{y}\in\Omega^n}\exp\{\boldsymbol{T}'(\boldsymbol{y})\cdot\boldsymbol{\psi}\}} \\
&= \frac{\exp\{\boldsymbol{T}'(\boldsymbol{z})\cdot\boldsymbol{\psi}\}}{c(\boldsymbol{\psi})} \\
&= \mathrm{L}(\boldsymbol{\psi}). \qquad (3.35)
\end{aligned}
$$

Thus, for a fixed block structure over the entire lattice, if the corresponding block size is 1, then BGPL is PL; if the corresponding block size is n, then BGPL is maximum likelihood (which MCMCML approximates); and if the corresponding block size is strictly between 1 and n, then BGPL is somewhere between PL and MCMCML.

As is the case with PL, GPL, and MCMCML, the logarithm of the block generalized pseudolikelihood function is typically used in practice and, hence, the log block generalized pseudolikelihood function is the following:

$$
\begin{aligned}
\log\{\mathrm{BGPL}(\boldsymbol{\psi})\} &= \log\left\{\prod_{l=1}^{L}\frac{\exp\{\boldsymbol{T}'\left(\boldsymbol{z}_{b(l)},\boldsymbol{z}^{b(l)}\right)\cdot\boldsymbol{\psi}\}}{\sum_{\boldsymbol{y}\in\Omega^{|b(l)|}}\exp\{\boldsymbol{T}'\left(\boldsymbol{y}_{b(l)},\boldsymbol{y}^{b(l)}\right)\cdot\boldsymbol{\psi}\}}\right\} \\
&= \sum_{l=1}^{L}\log\left\{\frac{\exp\{\boldsymbol{T}'\left(\boldsymbol{z}_{b(l)},\boldsymbol{z}^{b(l)}\right)\cdot\boldsymbol{\psi}\}}{\sum_{\boldsymbol{y}\in\Omega^{|b(l)|}}\exp\{\boldsymbol{T}'\left(\boldsymbol{y}_{b(l)},\boldsymbol{y}^{b(l)}\right)\cdot\boldsymbol{\psi}\}}\right\} \\
&= \sum_{l=1}^{L}\boldsymbol{T}'\left(\boldsymbol{z}_{b(l)},\boldsymbol{z}^{b(l)}\right)\cdot\boldsymbol{\psi} \\
&\quad -\sum_{l=1}^{L}\log\left\{\sum_{\boldsymbol{y}\in\Omega^{|b(l)|}}\exp\{\boldsymbol{T}'\left(\boldsymbol{y}_{b(l)},\boldsymbol{y}^{b(l)}\right)\cdot\boldsymbol{\psi}\}\right\}. \qquad (3.36)
\end{aligned}
$$

The MBGPLEs are then obtained by numerically maximizing (3.36) with respect to $\boldsymbol{\psi}$. As with PL and GPL, the standard errors of the MBGPLEs can be obtained by employing a resampling method, such as parametric bootstrapping. Finally, while BGPL is a compromise between PL and MCMCML, basic asymptotic properties of the MBGPLEs, as with the MGPLEs, remain unestablished in the literature. As part of this dissertation, the strong consistency of the MBGPLEs is proven in Chapter 4.

3.5.2 BGPL: Application to ALM

To demonstrate the use of BGPL with a concrete example, again consider the Ising model with a single covariate, with probability mass function, or likelihood, as given in (3.17). If it is assumed that the block size is the same for all L blocks, then the block joint likelihood for block l, $l = 1, \ldots, L$, is the following:

$$\mathrm{BL}_l(\alpha, \theta, \beta) \quad =$$

$$\frac{\exp\left\{\alpha\sum_{k\in b(l)} z_k + \theta\left\{\frac{1}{2}\sum_{k\in b(l)}\sum_{j\in N_k^{b(l)}} z_k z_j + \sum_{k\in b(l)}\sum_{j\in N_k^{\partial b(l)}} z_k z_j\right\} + \beta\sum_{k\in b(l)} z_k x_k\right\}}{\sum_{\boldsymbol{y}\in\Omega^{|b(l)|}}\exp\left\{\alpha\sum_{k\in b(l)} y_k + \theta\left\{\frac{1}{2}\sum_{k\in b(l)}\sum_{j\in N_k^{b(l)}} y_k y_j + \sum_{k\in b(l)}\sum_{j\in N_k^{\partial b(l)}} y_k y_j\right\} + \beta\sum_{k\in b(l)} y_k x_k\right\}}. \quad (3.37)$$

Hence, the vector of jointly sufficient statistics for the l^{th} block joint likelihood is $\boldsymbol{T}(\boldsymbol{z}_{b(l)}, \boldsymbol{z}^{b(l)}, \boldsymbol{x}^{b(l)}) = (T_1(\boldsymbol{z}_{b(l)}, \boldsymbol{z}^{b(l)}),\ T_2(\boldsymbol{z}_{b(l)},\ \boldsymbol{z}^{b(l)}), T_3(\boldsymbol{z}_{b(l)}, \boldsymbol{z}^{b(l)}, \boldsymbol{x}^{b(l)}))'$, where

$$T_1(\boldsymbol{z}_{b(l)}, \boldsymbol{z}^{b(l)}) \quad = \quad \sum_{k\in b(l)} z_k, \quad (3.38)$$

$$T_2(\boldsymbol{z}_{b(l)}, \boldsymbol{z}^{b(l)}) \quad = \quad \frac{1}{2}\sum_{k\in b(l)}\sum_{j\in N_k^{b(l)}} z_k z_j + \sum_{k\in b(l)}\sum_{j\in N_k^{\partial b(l)}} z_k z_j, \quad (3.39)$$

$$T_3(\boldsymbol{z}_{b(l)}, \boldsymbol{z}^{b(l)}, \boldsymbol{x}^{b(l)}) \quad = \quad \sum_{k\in b(l)} z_k x_k. \quad (3.40)$$

Note that $\boldsymbol{x}^{b(l)}$ denotes the vector of covariate values corresponding to the sites of block l. The analogous vector of jointly sufficient statistics for the likelihood function, given by (3.17), is the same as in section 3.4.2, with its three elements given in (3.29), (3.30), and (3.31). In comparing the sufficient statistics between the block joint likelihood and the true likelihood, it is evident that (3.38) and (3.29), as well as (3.40) and (3.31), are directly comparable. However, (3.39) and (3.30) are not directly comparable, and this is where the distinction

between a neighbor relation involving two sites of the block and a neighbor relation involving one site of the block and one boundary neighbor site of the block is readily apparent. In particular, in (3.39), recall that the purpose of the $\frac{1}{2}$ multiplier is to negate the double counting of each neighbor relation in the likelihood function since each $z_i z_j$ term appears twice in the sum. However, with a block joint likelihood function, as with a group joint likelihood function, a neighbor relation involving a site in the block and a boundary neighbor site of the block will not be double counted since the boundary neighbor sites are treated as fixed. Thus, if those neighbor relation terms were multiplied by $\frac{1}{2}$, such neighbor relations would be downweighted relative to the neighbor relations involving two sites within the block. Hence, the first portion of (3.39), $\frac{1}{2}\sum_{k\in b(l)}\sum_{j\in N_k^{b(l)}} z_k z_j$, deals with the terms corresponding to neighbor relations involving two sites of block l, while the second portion of (3.39), $\sum_{k\in b(l)}\sum_{j\in N_k^{\partial b(l)}} z_k z_j$, deals with the terms corresponding to neighbor relations involving one site of block l and one boundary neighbor site of block l.

The block generalized pseudolikelihood function is then obtained by forming the product of all L block joint likelihoods given by (3.37). The logarithm of the BGPL function would be the following:

$$\log\{\mathrm{BGPL}(\alpha,\theta,\beta)\} = \sum_{l=1}^{L}\log\{\mathrm{BL}_l(\alpha,\theta,\beta)\}$$

$$= -\sum_{l=1}^{L}\log\left\{\sum_{\boldsymbol{y}\in\Omega^{|b(l)|}}\exp\left\{\alpha\sum_{k\in b(l)}y_k+\theta\left\{\frac{1}{2}\sum_{k\in b(l)}\sum_{j\in N_k^{b(l)}}y_k y_j+\sum_{k\in b(l)}\sum_{j\in N_k^{\partial b(l)}}y_k y_j\right\}+\beta\sum_{k\in b(l)}y_k x_k\right\}\right\}$$

$$+\sum_{l=1}^{L}\left\{\alpha\sum_{k\in b(l)}z_k+\theta\left\{\frac{1}{2}\sum_{k\in b(l)}\sum_{j\in N_k^{b(l)}}z_k z_j+\sum_{k\in b(l)}\sum_{j\in N_k^{\partial b(l)}}z_k z_j\right\}+\beta\sum_{k\in b(l)}z_k x_k\right\}. \qquad (3.41)$$

The MBGPLEs, $\widehat{\boldsymbol{\psi}}_{BGPL} = (\hat{\alpha}_{BGPL}, \hat{\theta}_{BGPL}, \hat{\beta}_{BGPL})'$, are then obtained by numerically maximizing (3.41) with respect to $\boldsymbol{\psi} = (\alpha, \theta, \beta)'$. The MPLEs, $\widehat{\boldsymbol{\psi}}_{PL} = (\hat{\alpha}_{PL}, \hat{\theta}_{PL}, \hat{\beta}_{PL})'$, are generally used as starting values for the numerical optimization. The corresponding standard errors, as mentioned at the end of section 3.5.1, can be obtained using a resampling method such as parametric bootstrapping.

Chapter 4

Strong Consistency of the MGPLEs and MBGPLEs

4.1 Introduction

An important statistical question is whether or not the estimators obtained from both generalized pseudolikelihood and block generalized pseudolikelihood are strongly consistent. To date, however, this property has not been proven in the literature for either of these newer estimation methods. Hence, the proofs given in this chapter establishing the strong consistency of the MGPLEs and the MBGPLEs constitute the most substantial novel theoretical result of this dissertation. Note that these proofs are generalizations of Geman and Graffigne's ([13]) proof of the MPLE's strong consistency.

The remainder of this chapter is divided into four sections. In section 4.2, necessary notation, in addition to that already established in Chapter 3, is developed for use in the proof of the MGPLE's strong consistency. The theorem stating the strong consistency of the MGPLEs is also given near the end of section 4.2. In section 4.3, the five lemmas that will be used

to succinctly prove the strong consistency of the MGPLEs are sequentially presented, along with their corresponding proofs. In section 4.4, the concise proof of the MGPLE's strong consistency is given. Finally, in section 4.5, after developing the essential notation, the theorem conveying the MBGPLE's strong consistency is stated and the corresponding proof is discussed. More specifically, as this proof is entirely analogous and nearly identical in structure to that of the MGPLE's, only the few subtle adjustments that must be made to the proof of the MGPLE's strong consistency are presented in this final section.

4.2 Notation

Let the sequence of Markov random fields, $\mathrm{L}_1(\boldsymbol{\psi}_0), \mathrm{L}_2(\boldsymbol{\psi}_0), \ldots$, be a sequence of random vector distributions defined on expanding regular graphs (i.e. lattices) that share a common unknown parameter vector $\boldsymbol{\psi}_0 \in \boldsymbol{\Psi} \subseteq \mathbb{R}^p$. Let $\boldsymbol{z}(1), \boldsymbol{z}(2), \ldots$ be a sequence of realizations from $\mathrm{L}_1(\boldsymbol{\psi}_0), \mathrm{L}_2(\boldsymbol{\psi}_0), \ldots$, where $\boldsymbol{Z}(m)$ is the vector of random variables corresponding to the sites of $\mathrm{L}_m(\boldsymbol{\psi}_0)$. Note that the functional form of $\mathrm{L}_m(\boldsymbol{\psi}_0)$, $m = 1, 2, \ldots$, is given by (3.1), or by (3.2) if covariates have also been observed over the MRF; for ease of presentation, assume without any loss of generality that covariates have not been observed over the MRF. Let $\widehat{\boldsymbol{\psi}}_m$ denote the maximum generalized pseudolikelihood estimate of $\boldsymbol{\psi}_0$ obtained from the m^{th} sample in the sequence, $\boldsymbol{z}(m)$. The purpose of this section is to prove that as m tends toward infinity, $\widehat{\boldsymbol{\psi}}_m \to \boldsymbol{\psi}_0$ with probability one. The proof establishing this result is a generalization of the proof given by Geman and Graffigne ([13]) that establishes the strong consistency of the MPLEs. Consequently, the proof demonstrating the strong consistency of the MGPLEs parallels that of the MPLEs, though the notation has been altered to account for differences in the GPL method and to align with this dissertation. Before the proof of the MGPLE's strong consistency can be presented, however, some additional notation must first be introduced and some clarifying remarks made.

Let D now be a d-dimensional infinite square lattice (non-square lattices are a straightforward generalization) and let $S_m \subset D$ be a d-dimensional hypercube with sides of length m. As before, let i denote site i of S_m, $i = 1, \ldots, n_m$, where $n_m = m^d$. Let $S_1 \subseteq S_2 \subseteq \ldots$, and $\bigcup_{m=1}^{\infty} S_m = D$, so that $\boldsymbol{Z}(1) \subseteq \boldsymbol{Z}(2) \subseteq \ldots$, and $\bigcup_{m=1}^{\infty} \boldsymbol{Z}(m) = \boldsymbol{Z}$, where $\boldsymbol{Z}(m) = \{Z_i : i \in S_m\}$ and $\boldsymbol{Z} = \{Z_i : i \in D\}$. Further note that the sequence of distributions corresponding to $\boldsymbol{Z}(1), \boldsymbol{Z}(2), \ldots$, i.e. $\mathrm{L}_1(\boldsymbol{\psi}_0), \mathrm{L}_2(\boldsymbol{\psi}_0), \ldots$, are actually conditional distributions where the conditioning is on the complement of $\boldsymbol{Z}(m)$, $\boldsymbol{Z}(m') = \{Z_i : i \in D \backslash S_m\}$, $m = 1, 2, \ldots$. Let $N = \{N_i\}_{i \in D}$ and $G = \{g(i)\}_{i \in D}$ denote the translation invariant neighborhood system and group system, respectively, on D and its corresponding sublattices, $S_1, S_2, \ldots$. Recall that $\forall i \in D$, $g(i)$ denotes the group of sites, "centered" at site i, that are used in forming the group joint likelihood corresponding to the i^{th} site; hence, site i will be referred to as the "central" site of group i, $g(i)$. In this context, translation invariant means there is a fixed neighborhood structure and group structure over D, i.e. all groups are the same size, regardless of their "central" site's location on D. Note, however, that depending on the value of m, it is possible that groups whose "central" site is on or near the edge may have an incomplete group structure and/or an incomplete neighborhood structure. It is assumed that both a neighborhood set and a group set are finite, i.e. $\exists R_1, R_2 < \infty \ni$ if $j \in N_i$ then $|j - i| \leq R_1$ and if $j \in g(i)$ then $|j - i| \leq R_2$. Let $g(i)^c = \{k : k \in D \backslash g(i)\}$ be the sites of D that are not members of the i^{th} group and let $\partial g(i) = \{j : j \in N_k \cap g(i)^c, k \in g(i)\}$ be the boundary neighbor sites of group i, i.e. the sites that are neighbors of at least one site of group i, but are not themselves members of group i. Let $|g(i)|$ denote the number of sites in group i and let $|\partial g(i)|$ denote the number of boundary sites for group i, both of which are constant for all $i \in D$ under the translation invariant group structure assumption, i.e. for all $i = 1, 2, \ldots$, $|g(i)| = g$ and $|\partial g(i)| = \partial g$. Subgraphs of (D, N, G), consisting of sites S_m, are denoted by (S_m, N, G). Each subgraph (S_m, N, G), $m = 1, 2, \ldots$, has corresponding MRF distribution $\mathrm{L}_m(\boldsymbol{\psi}_0)$, where each variable of $\boldsymbol{Z}(m)$ has finite support set, Ω, and $\boldsymbol{Z}(m)$ has joint support set $\boldsymbol{\Omega}^{n_m} = \Omega \times \ldots \times \Omega$.

Now fix m and let $\boldsymbol{z} \in \boldsymbol{\Omega}^{n_m}$ be a realization from the MRF distribution $\mathrm{L}_m(\boldsymbol{\psi}_0)$. Then, as in Chapter 3, let $\boldsymbol{Z}_{g(i)} = \{Z_k : k \in g(i)\}$ denote the vector of variables corresponding to the sites of group i. Additionally, let $N_k^{g(i)} = \{j : j \in N_k \cap g(i), k \in g(i)\}$ denote the neighbors of site k, where k is a member of group i, that are also members of group i, and let $N_k^{\partial g(i)} = \{j : j \in N_k \cap \partial g(i), k \in g(i)\}$ denote the neighbors of site k, where k is a member of group i, that are boundary neighbors of group i. Furthermore, let $\boldsymbol{Z}^{g(i)} = \{Z_j : j \in N_k^{\partial g(i)}, k \in g(i)\}$ denote the vector of variables corresponding to the boundary neighbor sites of group i. Note that the group and/or neighborhood sets for edge sites of S_m will necessarily be incomplete in the sense that portions of these sets will belong to some $S_{m'}$, where $m' > m$. The conditional probabilities (i.e. group joint likelihood functions) $\mathrm{GL}_i^m(\boldsymbol{Z}_{g(i)} = \boldsymbol{z}_{g(i)} | \boldsymbol{Z}^{g(i)} = \boldsymbol{z}^{g(i)}; \boldsymbol{\psi}_0)$, for each group with "central" site $i \in S_m$, $\boldsymbol{z} \in \boldsymbol{\Omega}^{n_m}$, will be referred to as the "*local characteristics*" of $\mathrm{L}_m(\boldsymbol{\psi}_0)$.

The MRF distributions $\mathrm{L}_1(\boldsymbol{\psi}_0), \mathrm{L}_2(\boldsymbol{\psi}_0), \ldots$ are connected by the assumption that the local characteristics, which depend upon $\boldsymbol{\psi}_0$, are the same for all i and m on the interior S_m^0 of S_m, where $S_m^0 \equiv \{i \in S_m : \partial g(i) \subseteq S_m\}$; note that $\partial g(i) \subseteq S_m \Rightarrow g(i) \subseteq S_m$. In other words, the interior of S_m contains all of the sites of S_m whose corresponding group members and boundary neighbors are all contained within S_m. Consequently, $\forall m$, and $\forall i \in S_m^0$, $|g(i)| = g$ and $|\partial g(i)| = \partial g$. Hence, the conditional probability distribution form, under N and G, for interior sites of S_m does not depend on i or m:

$$\begin{aligned} \mathrm{GL}_i^m(\boldsymbol{Z}_{g(i)} = \boldsymbol{z}_{g(i)} | \boldsymbol{Z}^{g(i)} = \boldsymbol{z}^{g(i)}; \boldsymbol{\psi}_0) &= \mathrm{GL}(\boldsymbol{Z}_{g(i)} = \boldsymbol{z}_{g(i)} | \boldsymbol{Z}^{g(i)} = \boldsymbol{z}^{g(i)}; \boldsymbol{\psi}_0) \\ &= \frac{\exp\{\boldsymbol{T}'\left(\boldsymbol{z}_{g(i)}, \boldsymbol{z}^{g(i)}\right) \cdot \boldsymbol{\psi}_0\}}{\sum_{\boldsymbol{y} \in \boldsymbol{\Omega}^g} \exp\{\boldsymbol{T}'\left(\boldsymbol{y}_{g(i)}, \boldsymbol{y}^{g(i)}\right) \cdot \boldsymbol{\psi}_0\}}, \end{aligned} \tag{4.1}$$

for all m, and for all $i \in S_m^0$, $\boldsymbol{z}_{g(i)} \in \boldsymbol{\Omega}^g$, and $\boldsymbol{z}^{g(i)} \in \boldsymbol{\Omega}^{\partial g}$. Since interest is limited to local characteristics at interior sites, the sub/superscripts i and m will henceforth be dropped when writing the conditional probabilities $\mathrm{GL}(\cdot|\cdot)$. Thus, the corresponding generalized pseudolike-

lihood function of $\boldsymbol{\psi} \in \boldsymbol{\Psi}$, given a sample, $\boldsymbol{Z} = \boldsymbol{z}$, from $\mathrm{L}_m(\boldsymbol{\psi}_0)$, is

$$\begin{aligned} \mathrm{GPL}_m(\boldsymbol{\psi}|\boldsymbol{z}) &= \prod_{i \in S_m^0} \mathrm{GL}(\boldsymbol{z}_{g(i)}|\boldsymbol{z}^{g(i)}; \boldsymbol{\psi}) \\ &= \prod_{i \in S_m^0} \frac{\exp\{\boldsymbol{T}'\left(\boldsymbol{z}_{g(i)}, \boldsymbol{z}^{g(i)}\right) \cdot \boldsymbol{\psi}\}}{\sum_{\boldsymbol{y} \in \boldsymbol{\Omega}^g} \exp\{\boldsymbol{T}'\left(\boldsymbol{y}_{g(i)}, \boldsymbol{y}^{g(i)}\right) \cdot \boldsymbol{\psi}\}}. \end{aligned} \tag{4.2}$$

The generalized pseudolikelihood estimator of $\boldsymbol{\psi}_0$ is the set, $M_m(\boldsymbol{z})$, of $\boldsymbol{\psi}$ that maximizes $\mathrm{GPL}_m(\boldsymbol{\psi}|\boldsymbol{z})$:

$$M_m(\boldsymbol{z}) = \left\{ \boldsymbol{\psi} \in \boldsymbol{\Psi} : \mathrm{GPL}_m(\boldsymbol{\psi}|\boldsymbol{z}) = \sup_{\boldsymbol{\phi} \in \boldsymbol{\Psi}} \mathrm{GPL}_m(\boldsymbol{\phi}|\boldsymbol{z}) \right\}. \tag{4.3}$$

Throughout the remainder of this section identifiability of the MGPLE is assumed, where it is defined as follows:

Definition 4.1 (Identifiability – GPL)**.** *$\boldsymbol{\psi}_0 \in \boldsymbol{\Psi}$ is identifiable if $\boldsymbol{\psi} \neq \boldsymbol{\psi}_0 \Rightarrow \exists\ \boldsymbol{z}^{g(i)}, \boldsymbol{z}_{g(i)}$ such that $GL(\boldsymbol{z}_{g(i)}|\boldsymbol{z}^{g(i)}; \boldsymbol{\psi}) \neq GL(\boldsymbol{z}_{g(i)}|\boldsymbol{z}^{g(i)}; \boldsymbol{\psi}_0)$.*

Theorem 4.1 (Strong Consistency of Generalized Pseudolikelihood)**.** *For each $m = 1, 2, \ldots,$ let $\boldsymbol{z}(m)$ be a sample from the Markov random field $L_m(\boldsymbol{\psi}_0)$, with local characteristics* (4.1)*. If $\boldsymbol{\psi}_0$ is identifiable, then*

(a) $P(\log GPL_m(\boldsymbol{\psi}|\boldsymbol{z}(m))$ is strictly concave for all m sufficiently large$) = 1$;

(b) $P(M_m(\boldsymbol{z}(m))$ is a singleton for all m sufficiently large$) = 1$;

(c) $P\left(\sup_{\boldsymbol{\psi} \in M_m(\boldsymbol{z}(m))} |\boldsymbol{\psi} - \boldsymbol{\psi}_0| \to 0\right) = 1$, as $m \to \infty$.

It is worth emphasizing for Theorem 4.1 that since the number of sites of the lattice goes to infinity as $m \to \infty$, the asymptotics are when the number of sites of the lattice is large. The proof of Theorem 4.1 uses five lemmas, all of which are stated and proven in section 4.3. The proofs of the first two lemmas, however, require the following version of the strong law of large numbers (SLLN) ([13]).

Theorem 4.2 (SLLN)**.** *For each $m = 1, 2, \ldots,$ let $I_1(m), I_2(m), \ldots, I_{k_m}(m)$ be random variables and $\boldsymbol{Y}(m)$ be a random vector. Assume*

1. $\liminf_{m\to\infty} \frac{k_m}{m} > 0.$
2. *$I_1(m), I_2(m), \ldots, I_{k_m}(m)$ are conditionally independent, given $\boldsymbol{Y}(m)$.*
3. $|I_l(m)| \leq B < \infty \quad \forall \;\; l, m.$

Then as $m \to \infty$

$$\left| \frac{1}{k_m} \sum_{l=1}^{k_m} \{I_l(m) - E[I_l(m)|\boldsymbol{Y}(m)]\} \right| \to 0 \qquad a.s.$$

A proof of Theorem 4.2 is given in [13].

Additionally, the proofs of the final two lemmas, as well as the proof of Theorem 4.1, require a standard result from analysis. The following statement of that result is from [34], where it is labeled Theorem 9.3.3.

Theorem 4.3. *Let $f : M \to \mathbb{R}$ be continuous and M compact. Then* $\sup\{f(x)\}$ *and* $\inf\{f(x)\}$ *are both finite, and there are points in M where these values are assumed.*

A proof of 4.3 is also given in [34].

4.3 Lemmas

The five lemmas, and their corresponding proofs, are sequentially presented below and make use of the following notation. Let $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ denote arbitrary realizations of $\boldsymbol{\Omega}^g$ and $\boldsymbol{\Omega}^{\partial g}$, respectively. In other words, $\boldsymbol{\alpha}$ is a realization of the random vector corresponding to the sites of a group, while $\boldsymbol{\beta}$ is a realization of the random vector corresponding to the boundary neighbor sites of a group. Let $\eta_m = |S_m^0|$ denote the number of internal groups for the

m^{th} sublattice of D. Note that η_m is also equal to the number of internal sites for the m^{th} sublattice of D since each such site is also the "central" site of exactly one internal group of the m^{th} sublattice of D. Let $\eta_m(\boldsymbol{\beta}) = |\{i \in S_m^0 : \boldsymbol{Z}^{g(i)} = \boldsymbol{\beta}\}|$ denote the number of internal groups of S_m with boundary neighborhood configuration of type $\boldsymbol{\beta}$. Finally, let $\eta_m(\boldsymbol{\alpha}, \boldsymbol{\beta}) = |\{i \in S_m^0 : \boldsymbol{Z}_{g(i)} = \boldsymbol{\alpha}, \boldsymbol{Z}^{g(i)} = \boldsymbol{\beta}\}|$ denote the number of internal groups of S_m that have a group configuration of $\boldsymbol{\alpha}$ and a boundary neighborhood configuration of $\boldsymbol{\beta}$.

4.3.1 Lemma 1

The objective of this first lemma is to establish that the number of internal groups with boundary neighborhood configuration $\boldsymbol{\beta}$ is increasing with the expanding sublattices S_m.

Lemma 4.1. $\liminf_{m\to\infty} \frac{\eta_m(\boldsymbol{\beta})}{\eta_m} > 0$ *a.s.,* $\forall\ \boldsymbol{\beta}$.

Proof. For any $i \in D$, let

$$\Delta_i = \{r : \exists t \in [g(i) \cup \partial g(i)], r \in N_t, r \notin [g(i) \cup \partial g(i)]\}$$

denote the set of sites that border group i and its boundary neighbors, $g(i) \cup \partial g(i)$, according to the neighborhood and group structures, N and G. In other words, Δ_i consists of sites that are neighbors of group i or the boundary neighbors of group i (or both), but are neither a member of group i nor a member of the boundary neighbors of group i. For example, in Figure 4.1, $\Delta_i = \{O_j : j = 1, \ldots, 12\}$ under a first-order neighborhood system where $g(i) = \{i \cup N_i\}$, the 5 "X"-sites shown. Now let $|\Delta_i|$ denote the number of border sites for group i. Then for each m, choose groups $g(i_1), g(i_2), \ldots, g(i_{k_m})$, where $i_1, i_2, \ldots, i_{k_m} \in S_m$ are the "central" sites of these groups, such that

1. $\liminf_{m\to\infty} \frac{k_m}{\eta_m} > 0$,
   (This ensures that the number of chosen groups from S_m, k_m, increases "enough" as m grows.)

Figure 4.1: GPL Border Sites

e	e	e	O_6	e	e	e
e	e	O_4	B	O_8	e	e
e	O_2	B	X	B	O_{10}	e
O_1	B	X	X_i	X	B	O_{12}
e	O_3	B	X	B	O_{11}	e
e	e	O_5	B	O_9	e	e
e	e	e	O_7	e	e	e

For a 7x7 lattice subset of S_m, under a first-order neighborhood system with $g(i) = \{i \cup N_i\}$, the X sites represent the members of group i (X_i is the central site of the group), the B sites represent the boundary neighbors of the group, the O sites represent the border sites of the group (i.e. the sites belonging to Δ_i), and the e sites represent the external sites of the group.

2. $\Delta_{i_l} \subseteq S_m,\ \forall\ l = 1, \ldots, k_m,$

   (This ensures that the groups chosen, $g(i_1), \ldots, g(i_{k_m})$, are "internal enough" within S_m such that each $\Delta_{i_l}, l = 1, \ldots, k_m$, consists entirely of sites belonging to S_m. Notice that $\Delta_{i_l} \subseteq S_m \Rightarrow \{g(i_l) \cup \partial g(i_l)\} \subseteq S_m$.)

3. If $l \neq j$, then $\{g(i_l) \cup \partial g(i_l)\} \cap \Delta_{i_j} = \emptyset$. (This ensures that for each chosen group, that group, as well as its boundary neighbors, do not overlap with the border sites of any other such group and its corresponding boundary neighbors. Note that this condition does permit $\Delta_{i_l} \cap \Delta_{i_j} \neq \emptyset$, i.e. it permits distinct groups to share border sites.)

Since the neighborhood and group structures are both fixed and finite, the above three conditions can be accomplished, for example, if S_m is regularly partitioned into large hypercubes, with sizes independent of m, and big enough to accommodate $g(i) \cup \partial g(i) \cup \Delta_i$, for an arbitrary $i \in S_m^0$.

Now fix $\boldsymbol{\beta}$ and let $\boldsymbol{Y}(m) = \left\{ Z_j(m) : j \in \bigcup_{l=1}^{k_m} \{\Delta_{i_l} \cup g(i_l)\} \right\}$ and $I_{i_l}(m) = 1_{\boldsymbol{Z}^{g(i_l)}(m) = \boldsymbol{\beta}}$. In

other words, $\boldsymbol{Y}(m)$ is the vector of variables corresponding to the sites of group i_l and the sites bordering $g(i_l) \cup \partial g(i_l)$, $l = 1, \ldots, k_m$, and $I_{i_l}(m)$ is an indicator variable that is equal to one when the vector of variables corresponding to the boundary neighbors of group i_l have configuration (i.e. value) $\boldsymbol{\beta}$, and zero otherwise. Clearly, $|I_{i_l}(m)| \leq 1 < \infty \ \forall \ l, m$. By an appropriate spatial Markovian property, $I_{i_1}(m), I_{i_2}(m), \ldots, I_{i_{k_m}}(m)$ are conditionally independent, given $\boldsymbol{Y}(m)$. Hence, by Theorem 4.2 (SLLN),

$$\left| \frac{1}{k_m} \sum_{l=1}^{k_m} \{ I_{i_l}(m) - \mathrm{E}[I_{i_l}(m) | \boldsymbol{Y}(m)] \} \right| \quad \rightarrow \quad 0 \text{ a.s.} \tag{4.4}$$

Furthermore, by the same spatial Markovian property,

$$\begin{aligned} \mathrm{E}[I_{i_l}(m) | \boldsymbol{Y}(m)] &= \sum_{\boldsymbol{z}^{g(i_l)}(m) \in \boldsymbol{\Omega}^{\partial g}} I_{i_l}(m) \mathrm{GL}(\boldsymbol{Z}^{g(i_l)}(m) = \boldsymbol{z}^{g(i_l)}(m) | \boldsymbol{Y}(m); \boldsymbol{\psi}_0) \\ &= \sum_{\boldsymbol{z}^{g(i_l)}(m) \in \boldsymbol{\Omega}^{\partial g}} I_{i_l}(m) \mathrm{GL}(\boldsymbol{Z}^{g(i_l)}(m) = \boldsymbol{z}^{g(i_l)}(m) | \{ Z_j(m) : j \in \Delta_{i_l} \cup g(i_l) \}; \boldsymbol{\psi}_0) \\ &= \mathrm{GL}(\boldsymbol{Z}^{g(i_l)}(m) = \boldsymbol{\beta} | \{ Z_j(m) : j \in \Delta_{i_l} \cup g(i_l) \}; \boldsymbol{\psi}_0), \end{aligned} \tag{4.5}$$

which can have only a finite number of possible values, corresponding to the $\boldsymbol{\Omega}^{|\Delta_{i_l}| + g}$ configurations of group i_l and its border sites, all of which are positive.[1] Hence, for some $\epsilon > 0$,

$$\frac{1}{k_m} \sum_{l=1}^{k_m} \mathrm{E}[I_{i_l}(m) | \boldsymbol{Y}(m)] \quad > \quad \epsilon, \quad \forall m,$$

and, thus, by (4.4)

$$\liminf_{m \to \infty} \frac{1}{k_m} \sum_{l=1}^{k_m} I_{i_l}(m) \quad > \quad \epsilon \quad \text{ a.s.} \tag{4.6}$$

Then, since $\eta_m \geq k_m$, it follows that $\eta_m(\boldsymbol{\beta}) \geq \sum_{l=1}^{k_m} I_{i_l}(m)$, and, therefore, by (4.6), $\liminf_{m \to \infty} \frac{\eta_m(\boldsymbol{\beta})}{k_m} >$

[1] The local characteristics (4.1) determine the conditional probabilities (4.5) as well. Thus, the conditional distribution given in 4.5 is the same for all m.

ϵ a.s. Finally, since by design $\liminf_{m\to\infty} \frac{k_m}{\eta_m} > 0$, the desired result is obtained:

$$\begin{aligned} 0 &< \liminf_{m\to\infty} \frac{k_m}{\eta_m} \cdot \liminf_{m\to\infty} \frac{\eta_m(\boldsymbol{\beta})}{k_m} \quad \text{a.s.} \\ &\leq \liminf_{m\to\infty} \frac{k_m}{\eta_m} \cdot \frac{\eta_m(\boldsymbol{\beta})}{k_m} \\ &= \liminf_{m\to\infty} \frac{\eta_m(\boldsymbol{\beta})}{\eta_m}. \end{aligned}$$

□

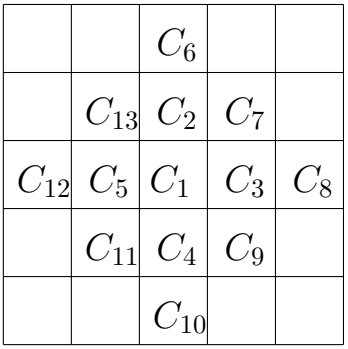
4.3.2 Lemma 2

The objective of this second lemma is to establish that among the groups with boundary neighborhood configuration $\boldsymbol{\beta}$, the proportion that have group configuration $\boldsymbol{\alpha}$ is essentially equivalent to the conditional probability specified by the local characteristics (4.1).

Lemma 4.2. $\lim_{m\to\infty} \frac{\eta_m(\boldsymbol{\alpha},\boldsymbol{\beta})}{\eta_m(\boldsymbol{\beta})} = GL(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0)$ *a.s.* $\forall$ $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$.

Proof. Let $C = \{c_l : l = 1, \ldots, k\}$ be a coloring of (D, N, G), i.e. $c_1, \ldots, c_k$ partition D such that if $i, j \in c_l$, then $j \notin \{g(i) \cup \partial g(i)\}$. In other words, among all of the sites of a group and its corresponding boundary neighbors, no two sites share the same color; hence, the neighborhood structure, as well as the group structure, should guide the coloring scheme over D. To help illustrate this visually, consider the group coloring given in Figure 4.2, where a first-order isotropic neighborhood system and a cross-shaped group structure have been assumed. Under such a neighborhood and group structure, 13 different colors $(c_1, \ldots, c_{13})$ are needed to color the sites of a group and its corresponding boundary neighbors; note that the "central" site of the group has color c_1 in Figure 4.2. Since (D, N, G) is a regular lattice (i.e. graph), it can be assumed that C is chosen in such a way that $\liminf_{m\to\infty} \frac{|S_m^0 \cap c_l|}{\eta_m} > 0$. Thus, this assumption is really asserting that only as many colors as are absolutely needed are used, and then because the neighborhood and group structures are finite, the regular lattice structure ensures that the

Figure 4.2: Coloring Schematic Example for a Group and its Boundary Neighbors



A coloring schematic for a group under a first-order neighborhood system and a cross-shaped group structure. For such neighborhood and group structures, 13 different colors are needed to color the 13 sites making up a group and its boundary neighbors. The "central" site of the group illustrated above has color c_1.

number of internal sites for the m^{th} sublattice of color c_l , $|S_m^0 \cap c_l|$, $l = 1, \ldots, k$, is increasing with the expanding sublattices, S_m.

For each color, c_l, define $\eta_m(\boldsymbol{\beta}; c_l) \equiv |\{i \in S_m^0 \cap c_l : \boldsymbol{Z}^{g(i)} = \boldsymbol{\beta}\}|$; i.e. $\eta_m(\boldsymbol{\beta}; c_l)$ is the number of internal sites of S_m that are of color c_l and are the "central" site of a group whose corresponding boundary neighbors have configuration $\boldsymbol{\beta}$. Additionally, define $\eta_m(\boldsymbol{\alpha}, \boldsymbol{\beta}; c_l) \equiv |\{i \in S_m^0 \cap c_l : \boldsymbol{Z}_{g(i)} = \boldsymbol{\alpha}, \boldsymbol{Z}^{g(i)} = \boldsymbol{\beta}\}|$; i.e. $\eta_m(\boldsymbol{\alpha}, \boldsymbol{\beta}; c_l)$ is the number of internal sites of S_m that are of color c_l and are the "central" site of a group whose configuration is $\boldsymbol{\alpha}$ and whose corresponding boundary neighbors have configuration $\boldsymbol{\beta}$. Now fix $l \in \{1, \ldots, k\}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\beta}$, and let $I_i(m) = 1_{\boldsymbol{Z}_{g(i)}=\boldsymbol{\alpha}, \boldsymbol{Z}^{g(i)}=\boldsymbol{\beta}}$ for each $i \in S_m^0 \cap c_l$. Thus, $I_i(m)$ is an indicator variable that is equal to 1 if site i is an internal site of S_m that is of color c_l and is a "central" site of a group with configuration $\boldsymbol{\alpha}$ and boundary neighborhood configuration $\boldsymbol{\beta}$. Let $\partial_{c_l}(m) = \{j : j \in \partial g(i), i \in (S_m^0 \cap c_l)\}$ and let $\boldsymbol{Y}(m) = \{Z_{i'}(m) : i' \in \partial_{c_l}(m)\}$. In other words, $\partial_{c_l}(m)$ is the set of all sites of S_m that are a boundary neighbor of a group whose defining, or "central," site is in S_m^0 and of color c_l, and $\boldsymbol{Y}(m)$ is the vector of variables corresponding to all such sites. Ultimately, as with Lemma 4.1, the successful completion of this proof of Lemma 4.2 is contingent upon being able to appeal to Theorem 4.2, which hinges on satisfying three

assumptions. To that end, since it is clear that $\liminf_{m\to\infty} \frac{\eta_m}{m} > 0$ and since it is assumed that $\liminf_{m\to\infty} \frac{|S_m^0 \cap c_l|}{\eta_m} > 0$, then $\liminf_{m\to\infty} \frac{|S_m^0 \cap c_l|}{m} \geq \liminf_{m\to\infty} \frac{\eta_m}{m} \liminf_{m\to\infty} \frac{|S_m^0 \cap c_l|}{\eta_m} > 0$, which satisfies the first assumption for $k_m = |S_m^0 \cap c_l|$. Furthermore, the coloring scheme, C, dictates that any two sites of color c_l cannot belong to the same group or corresponding set of boundary neighbors, and, thus, by a spatial Markovian property, $I_{i_1}(m), I_{i_2}(m), \ldots, I_{i_{|S_m^0 \cap c_l|}}(m)$ are conditionally independent, given $\boldsymbol{Y}(m)$, which satisfies the second assumption. Finally, because $I_i(m)$ is an indicator variable, $|I_i(m)| \leq 1 < \infty$ for all i, m, which satisfies the third assumption. Thus, by Theorem 4.2,

$$\left| \frac{1}{|S_m^0 \cap c_l|} \sum_{i \in S_m^0 \cap c_l} \{I_i(m) - \mathrm{E}[I_i(m)|\boldsymbol{Y}(m)]\} \right| \quad \to \quad 0 \text{ a.s.} \tag{4.7}$$

Then, again by the same spatial Markovian property, for some fixed $l \in \{1, \ldots, k\}$, and fixed site i:

$$\begin{aligned}
\mathrm{E}[I_i(m)|\boldsymbol{Y}(m)] &= \sum_{I_i(m)\in\{0,1\}} I_i(m)\mathrm{GL}(\boldsymbol{Z}_{g(i)}(m) = \boldsymbol{z}_{g(i)}(m)|\boldsymbol{Y}(m); \boldsymbol{\psi}_0) \\
&= \sum_{I_i(m)\in\{0,1\}} I_i(m)\mathrm{GL}(\boldsymbol{Z}_{g(i)}(m) = \boldsymbol{z}_{g(i)}(m)|\boldsymbol{Z}^{g(i)}(m) = \boldsymbol{z}^{g(i)}(m); \boldsymbol{\psi}_0) \\
&= \mathrm{GL}(\boldsymbol{Z}_{g(i)}(m) = \boldsymbol{\alpha}|\boldsymbol{Z}^{g(i)}(m) = \boldsymbol{\beta}; \boldsymbol{\psi}_0) \\
&= \mathrm{GL}(\boldsymbol{Z}_{g(i)}(m) = \boldsymbol{\alpha}|\boldsymbol{Z}^{g(i)}(m) = \boldsymbol{\beta}; \boldsymbol{\psi}_0) \cdot 1_{\boldsymbol{Z}^{g(i)}(m)=\boldsymbol{\beta}} \\
&= \mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta}; \boldsymbol{\psi}_0) \cdot 1_{\boldsymbol{Z}^{g(i)}(m)=\boldsymbol{\beta}}.
\end{aligned} \tag{4.8}$$

Note that the $1_{\boldsymbol{Z}^{g(i)}(m)=\boldsymbol{\beta}}$ piece in (4.8) appears redundant and unnecessary, but expressing the right hand side of (4.8) in such a manner will be useful below. Now observe that

$$\sum_{i \in S_m^0 \cap c_l} I_i(m) \quad = \quad \eta_m(\boldsymbol{\alpha}, \boldsymbol{\beta}; c_l) \tag{4.9}$$

and

$$\sum_{i \in S_m^0 \cap c_l} 1_{\boldsymbol{Z}^{g(i)}(m)=\boldsymbol{\beta}} \quad = \quad \eta_m(\boldsymbol{\beta}; c_l). \tag{4.10}$$

Hence, the left hand side of 4.7 can be reformulated as follows:

$$\left| \frac{1}{|S_m^0 \cap c_l|} \sum_{i \in S_m^0 \cap c_l} \{I_i(m) - \mathrm{E}[I_i(m)|\boldsymbol{Y}(m)]\} \right| \quad = \quad \frac{1}{|S_m^0 \cap c_l|} \left| \sum_{i \in S_m^0 \cap c_l} I_i(m) - \sum_{i \in S_m^0 \cap c_l} \mathrm{E}[I_i(m)|\boldsymbol{Y}(m)] \right|$$

$$\begin{aligned}
&= \frac{1}{|S_m^0 \cap c_l|} \left| \eta_m(\boldsymbol{\alpha}, \boldsymbol{\beta}; c_l) - \sum_{i \in S_m^0 \cap c_l} \mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta}; \boldsymbol{\psi}_0) \cdot 1_{\boldsymbol{Z}^{g(i)}(m)=\boldsymbol{\beta}} \right|, \quad \text{by (4.8) and (4.9)} \\
&= \frac{1}{|S_m^0 \cap c_l|} \left| \eta_m(\boldsymbol{\alpha}, \boldsymbol{\beta}; c_l) - \mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta}; \boldsymbol{\psi}_0) \sum_{i \in S_m^0 \cap c_l} 1_{\boldsymbol{Z}^{g(i)}(m)=\boldsymbol{\beta}} \right| \\
&= \frac{1}{|S_m^0 \cap c_l|} \left| \eta_m(\boldsymbol{\alpha}, \boldsymbol{\beta}; c_l) - \mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta}; \boldsymbol{\psi}_0) \eta_m(\boldsymbol{\beta}; c_l) \right|, \quad \text{by (4.10).}
\end{aligned} \tag{4.11}$$

Therefore, by (4.7) and (4.11),

$$\frac{1}{|S_m^0 \cap c_l|} \left| \eta_m(\boldsymbol{\alpha}, \boldsymbol{\beta}; c_l) - \mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta}; \boldsymbol{\psi}_0) \eta_m(\boldsymbol{\beta}; c_l) \right| \quad \to \quad 0 \text{ a.s.} \tag{4.12}$$

Noting that for all m, $\sum_{l=1}^{k} \frac{\eta_m(\boldsymbol{\beta}; c_l)}{\eta_m(\boldsymbol{\beta})} = 1$ and $\sum_{l=1}^{k} \frac{\eta_m(\boldsymbol{\alpha}, \boldsymbol{\beta}; c_l)}{\eta_m(\boldsymbol{\beta})} = \frac{\eta_m(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\eta_m(\boldsymbol{\beta})}$, since $C = \{c_l : l = 1, \ldots, k\}$ partitions $L_m(\boldsymbol{\psi}_0)$, observe the following:

$$\left| \frac{\eta_m(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\eta_m(\boldsymbol{\beta})} - \mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta}; \boldsymbol{\psi}_0) \right|$$

$$= \left| \sum_{l=1}^{k} \frac{\eta_m(\boldsymbol{\alpha}, \boldsymbol{\beta}; c_l)}{\eta_m(\boldsymbol{\beta})} - \mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta}; \boldsymbol{\psi}_0) \sum_{l=1}^{k} \frac{\eta_m(\boldsymbol{\beta}; c_l)}{\eta_m(\boldsymbol{\beta})} \right|$$

$$
\begin{aligned}
&= \left| \sum_{l=1}^{k} \frac{1}{\eta_m(\boldsymbol{\beta})} \left\{ \eta_m(\boldsymbol{\alpha}, \boldsymbol{\beta}; c_l) - \mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta}; \boldsymbol{\psi}_0) \eta_m(\boldsymbol{\beta}; c_l) \right\} \right| \\
&\leq \sum_{l=1}^{k} \frac{1}{\eta_m(\boldsymbol{\beta})} \left| \eta_m(\boldsymbol{\alpha}, \boldsymbol{\beta}; c_l) - \mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta}; \boldsymbol{\psi}_0) \eta_m(\boldsymbol{\beta}; c_l) \right| \qquad \text{(Triangle Inequality)} \\
&= \sum_{l=1}^{k} \frac{\eta_m}{\eta_m(\boldsymbol{\beta})} \frac{|S_m^0 \cap c_l|}{\eta_m} \frac{1}{|S_m^0 \cap c_l|} \left| \eta_m(\boldsymbol{\alpha}, \boldsymbol{\beta}; c_l) - \mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta}; \boldsymbol{\psi}_0) \eta_m(\boldsymbol{\beta}; c_l) \right|. \qquad (4.13)
\end{aligned}
$$

Finally, since $0 < \liminf_{m\to\infty} \frac{|S_m^0 \cap c_l|}{\eta_m} < 1$ for all l, and since (by Lemma 4.1) $\liminf_{m\to\infty} \frac{\eta_m(\boldsymbol{\beta})}{\eta_m} > 0$ a.s. $\Rightarrow 1 < \liminf_{m\to\infty} \frac{\eta_m}{\eta_m(\boldsymbol{\beta})} < \infty$ a.s., then appealing to (4.12) yields that

$$
\sum_{l=1}^{k} \frac{\eta_m}{\eta_m(\boldsymbol{\beta})} \frac{|S_m^0 \cap c_l|}{\eta_m} \frac{1}{|S_m^0 \cap c_l|} \left| \eta_m(\boldsymbol{\alpha}, \boldsymbol{\beta}; c_l) - \mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta}; \boldsymbol{\psi}_0) \eta_m(\boldsymbol{\beta}; c_l) \right| \quad \to \quad 0 \text{ a.s.} \quad (4.14)
$$

Therefore, by (4.13) and (4.14), $\forall\ \boldsymbol{\alpha}$ and $\boldsymbol{\beta}$,

$$
\lim_{m\to\infty} \frac{\eta_m(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\eta_m(\boldsymbol{\beta})} \quad = \quad \mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta}; \boldsymbol{\psi}_0) \quad \text{a.s.}
$$

□

4.3.3 Lemma 3

The objective of this third lemma is to establish that the logarithm of the GPL function, when shifted by a constant and rescaled by a constant, is almost surely strictly concave.

Lemma 4.3. *Let*

$$
\begin{aligned}
F_m(\boldsymbol{\psi}) &= \frac{1}{\eta_m} \left\{ \log GPL_m(\boldsymbol{Z}(m); \boldsymbol{\psi}) - \log GPL_m(\boldsymbol{Z}(m); \boldsymbol{\psi}_0) \right\} \qquad (4.15) \\
&= \sum_{\boldsymbol{\beta}} \frac{\eta_m(\boldsymbol{\beta})}{\eta_m} \sum_{\boldsymbol{\alpha}} \frac{\eta_m(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\eta_m(\boldsymbol{\beta})} \log \frac{GL(\boldsymbol{\alpha}|\boldsymbol{\beta}; \boldsymbol{\psi})}{GL(\boldsymbol{\alpha}|\boldsymbol{\beta}; \boldsymbol{\psi}_0)}. \qquad (4.16)
\end{aligned}
$$

Then $P(F_m(\cdot)$ *is strictly concave for* m *sufficiently large*$) = 1$.

Before proving Lemma 4.3, two observations should be made. First, both $\log \mathrm{GPL}_m(\boldsymbol{Z}(m);\boldsymbol{\psi}_0)$ and $\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0)$, for all $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, are constants since $\boldsymbol{\psi}_0$ is the true, but unknown, value of the parameter vector. Thus, $F_m(\boldsymbol{\psi})$ is just the logarithm of the generalized pseudolikelihood function shifted by a constant and rescaled by $\frac{1}{\eta_m}$. Second, (4.16) can be relatively easily obtained from (4.15), as demonstrated below:

$$\begin{aligned}
&\frac{1}{\eta_m}\left\{\log \mathrm{GPL}_m(\boldsymbol{Z}(m);\boldsymbol{\psi}) - \log \mathrm{GPL}_m(\boldsymbol{Z}(m);\boldsymbol{\psi}_0)\right\} \\
= \;& \frac{1}{\eta_m}\left\{\log \prod_{i\in S_m^0} \mathrm{GL}(\boldsymbol{z}_{g(i)}(m)|\boldsymbol{z}^{g(i)}(m);\boldsymbol{\psi}) - \log \prod_{i\in S_m^0} \mathrm{GL}(\boldsymbol{z}_{g(i)}(m)|\boldsymbol{z}^{g(i)}(m);\boldsymbol{\psi}_0)\right\} \\
= \;& \frac{1}{\eta_m}\left\{\log \prod_{\boldsymbol{\beta}}\prod_{\boldsymbol{\alpha}} \{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi})\}^{\eta_m(\boldsymbol{\alpha},\boldsymbol{\beta})} - \log \prod_{\boldsymbol{\beta}}\prod_{\boldsymbol{\alpha}} \{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0)\}^{\eta_m(\boldsymbol{\alpha},\boldsymbol{\beta})}\right\}
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\eta_m}\log \prod_{\boldsymbol{\beta}}\prod_{\boldsymbol{\alpha}} \frac{\{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi})\}^{\eta_m(\boldsymbol{\alpha},\boldsymbol{\beta})}}{\{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0)\}^{\eta_m(\boldsymbol{\alpha},\boldsymbol{\beta})}} \\
&= \frac{1}{\eta_m}\sum_{\boldsymbol{\beta}}\sum_{\boldsymbol{\alpha}} \eta_m(\boldsymbol{\alpha},\boldsymbol{\beta}) \log \frac{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi})}{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0)} \\
&= \sum_{\boldsymbol{\beta}} \frac{\eta_m(\boldsymbol{\beta})}{\eta_m} \sum_{\boldsymbol{\alpha}} \frac{\eta_m(\boldsymbol{\alpha},\boldsymbol{\beta})}{\eta_m(\boldsymbol{\beta})} \log \frac{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi})}{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0)}. \qquad (4.17)
\end{aligned}$$

Proof. Let $\boldsymbol{H}(F_m(\boldsymbol{\psi}))$ be the Hessian of $F_m(\boldsymbol{\psi})$ and let $\boldsymbol{\phi} \in \mathbb{R}^p$. The objective will be to derive a form of $\boldsymbol{\phi}'\boldsymbol{H}(F_m(\boldsymbol{\psi}))\boldsymbol{\phi}$ from which the concavity of $F_m(\boldsymbol{\psi})$ can relatively easily be demonstrated. Obtaining such a form, while routine, is tedious; therefore, to ease readability, the following shorthand references will be implemented, where the tilde will eventually be needed within complicated expressions to distinguish between multiple summations over all possible realizations of $\boldsymbol{\alpha}$.

$$\tilde{\boldsymbol{T}} \equiv \boldsymbol{T}(\tilde{\boldsymbol{\alpha}},\boldsymbol{\beta}), \qquad (4.18)$$

$$\boldsymbol{T} \equiv \boldsymbol{T}(\boldsymbol{\alpha},\boldsymbol{\beta}), \qquad (4.19)$$

$$\mathrm{E}[\cdot|\boldsymbol{\beta}] \equiv \mathrm{E}_{\boldsymbol{\psi}}[\cdot|\boldsymbol{\beta}], \tag{4.20}$$

where $\mathrm{E}_{\boldsymbol{\psi}}[\cdot|\boldsymbol{\beta}]$ is the expectation on $\boldsymbol{\Omega}^g$ with respect to $\mathrm{GL}(\cdot|\boldsymbol{\beta};\boldsymbol{\psi})$.

The first step in determining a "nice" form for the desired Hessian is to calculate the partial derivatives of $\log \frac{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi})}{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0)}$ with respect to ψ_i, $1 \le i \le p$:

$$\frac{\partial}{\partial\psi_i}\left\{\log\frac{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi})}{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0)}\right\}$$

$$\begin{aligned}
&= \frac{\partial}{\partial\psi_i}\left\{\log \mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi})\right\} \qquad (\text{since } \mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0) \text{ is constant w.r.t } \boldsymbol{\psi})\\
&= \frac{\partial}{\partial\psi_i}\left\{\log\left(\frac{\exp\{\boldsymbol{T}'\boldsymbol{\psi}\}}{\sum_{\boldsymbol{\alpha}\in\boldsymbol{\Omega}^g}\exp\{\boldsymbol{T}'\boldsymbol{\psi}\}}\right)\right\} \qquad (\text{definition of } \mathrm{GL}(\,\cdot\,|\cdot;\boldsymbol{\psi}))\\
&= \frac{\partial}{\partial\psi_i}\left\{\boldsymbol{T}'\boldsymbol{\psi} - \log\left(\sum_{\boldsymbol{\alpha}\in\boldsymbol{\Omega}^g}\exp\{\boldsymbol{T}'\boldsymbol{\psi}\}\right)\right\}\\
&= T_i - \frac{\sum_{\boldsymbol{\alpha}\in\boldsymbol{\Omega}^g} T_i\exp\{\boldsymbol{T}'\boldsymbol{\psi}\}}{\sum_{\boldsymbol{\alpha}\in\boldsymbol{\Omega}^g}\exp\{\boldsymbol{T}'\boldsymbol{\psi}\}}.
\end{aligned} \tag{4.21}$$

Continuing along this path, the next step is to calculate the second-order partial derivatives of $\log \frac{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi})}{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0)}$, first with respect to ψ_i, and then with respect to ψ_j, $1 \le i, j \le p$:

$$\begin{aligned}
\frac{\partial^2}{\partial\psi_i\partial\psi_j}\left\{\log\frac{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi})}{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0)}\right\} &= \frac{\partial}{\partial\psi_j}\left\{\frac{\partial}{\partial\psi_i}\left\{\log\frac{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi})}{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0)}\right\}\right\}\\
&= \frac{\partial}{\partial\psi_j}\left\{T_i - \frac{\sum_{\boldsymbol{\alpha}\in\boldsymbol{\Omega}^g} T_i\exp\{\boldsymbol{T}'\boldsymbol{\psi}\}}{\sum_{\boldsymbol{\alpha}\in\boldsymbol{\Omega}^g}\exp\{\boldsymbol{T}'\boldsymbol{\psi}\}}\right\}, \quad \text{by (4.21)}
\end{aligned}$$

$$= \frac{-\left(\sum_{\boldsymbol{\alpha}\in\boldsymbol{\Omega}^g} T_iT_j\exp\{\boldsymbol{T}'\boldsymbol{\psi}\}\right)\left(\sum_{\boldsymbol{\alpha}\in\boldsymbol{\Omega}^g}\exp\{\boldsymbol{T}'\boldsymbol{\psi}\}\right)}{\left(\sum_{\boldsymbol{\alpha}\in\boldsymbol{\Omega}^g}\exp\{\boldsymbol{T}'\boldsymbol{\psi}\}\right)^2}$$

$$+ \frac{\left(\sum_{\boldsymbol{\alpha}\in\boldsymbol{\Omega}^g} T_j\exp\{\boldsymbol{T}'\boldsymbol{\psi}\}\right)\left(\sum_{\boldsymbol{\alpha}\in\boldsymbol{\Omega}^g} T_i\exp\{\boldsymbol{T}'\boldsymbol{\psi}\}\right)}{\left(\sum_{\boldsymbol{\alpha}\in\boldsymbol{\Omega}^g}\exp\{\boldsymbol{T}'\boldsymbol{\psi}\}\right)^2}$$

$$\begin{aligned} &= -\mathrm{E}[T_iT_j|\boldsymbol{\beta}] + \mathrm{E}[T_i|\boldsymbol{\beta}]\mathrm{E}[T_j|\boldsymbol{\beta}] \\ &= -\mathrm{Cov}(T_i, T_j|\boldsymbol{\beta}). \end{aligned} \tag{4.22}$$

Thus, from (4.22), it is easily seen that the Hessian of $\log\frac{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi})}{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0)}$ is the following:

$$\begin{aligned} \mathrm{H}\left(\log\frac{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi})}{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0)}\right) &= -\mathrm{Var}(\boldsymbol{T}|\boldsymbol{\beta}) \\ &= -\mathrm{E}\left\{(\boldsymbol{T}-\mathrm{E}[\tilde{\boldsymbol{T}}|\boldsymbol{\beta}])(\boldsymbol{T}-\mathrm{E}[\tilde{\boldsymbol{T}}|\boldsymbol{\beta}])'\right\}, \end{aligned} \tag{4.23}$$

where the outer expectation is taken relative to all group configurations $\boldsymbol{\alpha}$ and the inner expectation is taken relative to all group configurations $\tilde{\boldsymbol{\alpha}}$. Then,

$$\begin{aligned} \boldsymbol{\phi}'\mathrm{H}\left(\log\frac{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi})}{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0)}\right)\boldsymbol{\phi} &= -\boldsymbol{\phi}'\mathrm{E}\left\{(\boldsymbol{T}-\mathrm{E}[\tilde{\boldsymbol{T}}|\boldsymbol{\beta}])(\boldsymbol{T}-\mathrm{E}[\tilde{\boldsymbol{T}}|\boldsymbol{\beta}])'\right\}\boldsymbol{\phi} \\ &= -\mathrm{E}\left\{\boldsymbol{\phi}'(\boldsymbol{T}-\mathrm{E}[\tilde{\boldsymbol{T}}|\boldsymbol{\beta}])(\boldsymbol{T}-\mathrm{E}[\tilde{\boldsymbol{T}}|\boldsymbol{\beta}])'\boldsymbol{\phi}\right\} \\ &= -\mathrm{E}\left\{\left(\boldsymbol{\phi}'(\boldsymbol{T}-\mathrm{E}[\tilde{\boldsymbol{T}}|\boldsymbol{\beta}])\right)^2\right\}. \end{aligned} \tag{4.24}$$

Note that $|\Omega| < \infty$ means that the number of summands in (4.17) is finite, which means that the Hessian of (4.17) can be expressed as the sum (over $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$) of the Hessians of

$\log \frac{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi})}{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0)}$. Hence, observe the following:

$$\boldsymbol{\phi}' H(F_m(\boldsymbol{\psi}))\boldsymbol{\phi}$$

$$\begin{aligned}
&= \boldsymbol{\phi}' H\left(\sum_{\boldsymbol{\beta}} \frac{\eta_m(\boldsymbol{\beta})}{\eta_m} \sum_{\boldsymbol{\alpha}} \frac{\eta_m(\boldsymbol{\alpha},\boldsymbol{\beta})}{\eta_m(\boldsymbol{\beta})} \log \frac{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi})}{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0)}\right)\boldsymbol{\phi} \\
&= \boldsymbol{\phi}' \sum_{\boldsymbol{\beta}} \frac{\eta_m(\boldsymbol{\beta})}{\eta_m} \sum_{\boldsymbol{\alpha}} \frac{\eta_m(\boldsymbol{\alpha},\boldsymbol{\beta})}{\eta_m(\boldsymbol{\beta})} H\left(\log \frac{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi})}{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0)}\right)\boldsymbol{\phi} \\
&= \sum_{\boldsymbol{\beta}} \frac{\eta_m(\boldsymbol{\beta})}{\eta_m} \sum_{\boldsymbol{\alpha}} \frac{\eta_m(\boldsymbol{\alpha},\boldsymbol{\beta})}{\eta_m(\boldsymbol{\beta})} \boldsymbol{\phi}' H\left(\log \frac{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi})}{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0)}\right)\boldsymbol{\phi} \\
&= -\sum_{\boldsymbol{\beta}} \frac{\eta_m(\boldsymbol{\beta})}{\eta_m} \sum_{\boldsymbol{\alpha}} \frac{\eta_m(\boldsymbol{\alpha},\boldsymbol{\beta})}{\eta_m(\boldsymbol{\beta})} \mathrm{E}\left\{\left(\boldsymbol{\phi}'(\boldsymbol{T} - \mathrm{E}[\tilde{\boldsymbol{T}}|\boldsymbol{\beta}])\right)^2\right\} \qquad \text{(by (4.24))}. \qquad (4.25)
\end{aligned}$$

Now note that $\mathrm{E}\left\{\left(\boldsymbol{\phi}'(\boldsymbol{T} - \mathrm{E}[\tilde{\boldsymbol{T}}|\boldsymbol{\beta}])\right)^2\right\}$ is independent of $\boldsymbol{\alpha}$ in the right hand side of (4.25) since the outer expectation sums over all $\boldsymbol{\alpha} \in \boldsymbol{\Omega}^g$. Hence, $\mathrm{E}\left\{\left(\boldsymbol{\phi}'(\boldsymbol{T} - \mathrm{E}[\tilde{\boldsymbol{T}}|\boldsymbol{\beta}])\right)^2\right\}$ can be pulled out in front of the sum, $\sum_{\boldsymbol{\alpha}} \frac{\eta_m(\boldsymbol{\alpha},\boldsymbol{\beta})}{\eta_m(\boldsymbol{\beta})}$, in the far right hand side of (4.25), which after additionally noting that $\sum_{\boldsymbol{\alpha}} \frac{\eta_m(\boldsymbol{\alpha},\boldsymbol{\beta})}{\eta_m(\boldsymbol{\beta})} = 1$, yields the following:

$$\begin{aligned}
\boldsymbol{\phi}' H(F_m(\boldsymbol{\psi}))\boldsymbol{\phi} &= -\sum_{\boldsymbol{\beta}} \frac{\eta_m(\boldsymbol{\beta})}{\eta_m} \mathrm{E}\left\{\left(\boldsymbol{\phi}'(\boldsymbol{T} - \mathrm{E}[\tilde{\boldsymbol{T}}|\boldsymbol{\beta}])\right)^2\right\} \\
&= -\sum_{\boldsymbol{\beta}} \frac{\eta_m(\boldsymbol{\beta})}{\eta_m} \cdot \frac{\sum_{\boldsymbol{\alpha}\in\boldsymbol{\Omega}^g} [\boldsymbol{\phi}'(\boldsymbol{T} - \mathrm{E}(\tilde{\boldsymbol{T}}|\boldsymbol{\beta}))]^2 \exp\{\boldsymbol{T}'\boldsymbol{\psi}\}}{\sum_{\boldsymbol{\alpha}\in\boldsymbol{\Omega}^g} \exp\{\boldsymbol{T}'\boldsymbol{\psi}\}}. \qquad (4.26)
\end{aligned}$$

Finally, as can clearly be seen from the right hand side of (4.26), $\boldsymbol{\phi}' H(F_m(\boldsymbol{\psi}))\boldsymbol{\phi} \leq 0 \; \forall \boldsymbol{\phi}$, and, therefore, $F_m(\boldsymbol{\psi})$ is concave. What remains to be shown is that for all m sufficiently large, $F_m(\boldsymbol{\psi})$ is almost surely *strictly* concave. The approach to demonstrating this will be to use proof by contradiction.

Lemma 4.1 conveys that for all m sufficiently large, $\inf_{\boldsymbol{\beta}} \frac{\eta_m(\boldsymbol{\beta})}{\eta_m} > 0$ with probability one. Suppose $F_m(\boldsymbol{\psi})$ is not almost surely strictly concave. In other words, assume there exists $\boldsymbol{\psi}$ and $\boldsymbol{\phi}$ such that $\boldsymbol{\phi}' H(F_m(\boldsymbol{\psi}))\boldsymbol{\phi} = 0$. If $\boldsymbol{\phi} = 0$, then $\boldsymbol{\phi}' H(F_m(\boldsymbol{\psi}))\boldsymbol{\phi} = 0$ trivially, so assume $\boldsymbol{\phi} \neq 0$. Consequently, by (4.26), for these values of $\boldsymbol{\psi}$ and $\boldsymbol{\phi}$, $\boldsymbol{\phi}'\boldsymbol{T} = \boldsymbol{\phi}'\mathrm{E}[\tilde{\boldsymbol{T}}|\boldsymbol{\beta}]$, or rather by (4.18), (4.19), and (4.20), $\boldsymbol{\phi}'\boldsymbol{T}(\boldsymbol{\alpha},\boldsymbol{\beta}) = \boldsymbol{\phi}'\mathrm{E}_{\boldsymbol{\psi}}[\boldsymbol{T}(\tilde{\boldsymbol{\alpha}},\boldsymbol{\beta})|\boldsymbol{\beta}]$ $\forall$ $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. This implies that for every $\boldsymbol{\beta}$, $\boldsymbol{\phi}'\boldsymbol{T}(\boldsymbol{\alpha},\boldsymbol{\beta})$ is independent (i.e. not a function) of $\boldsymbol{\alpha}$ since $\boldsymbol{\phi}'\mathrm{E}_{\boldsymbol{\psi}}[\boldsymbol{T}(\tilde{\boldsymbol{\alpha}},\boldsymbol{\beta})|\boldsymbol{\beta}]$ is independent of $\boldsymbol{\alpha}$. Such a conclusion implies

$$\begin{aligned}
\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}+\boldsymbol{\phi}) &= \frac{\exp\{(\boldsymbol{\psi}+\boldsymbol{\phi})'\boldsymbol{T}(\boldsymbol{\alpha},\boldsymbol{\beta})\}}{\sum_{\boldsymbol{\alpha}\in\boldsymbol{\Omega}^g}\exp\{(\boldsymbol{\psi}+\boldsymbol{\phi})'\boldsymbol{T}(\boldsymbol{\alpha},\boldsymbol{\beta})\}} \\
&= \frac{\exp\{\boldsymbol{\psi}'\boldsymbol{T}(\boldsymbol{\alpha},\boldsymbol{\beta})\}\exp\{\boldsymbol{\phi}'\boldsymbol{T}(\boldsymbol{\alpha},\boldsymbol{\beta})\}}{\sum_{\boldsymbol{\alpha}\in\boldsymbol{\Omega}^g}\exp\{\boldsymbol{\psi}'\boldsymbol{T}(\boldsymbol{\alpha},\boldsymbol{\beta})\}\exp\{\boldsymbol{\phi}'\boldsymbol{T}(\boldsymbol{\alpha},\boldsymbol{\beta})\}} \\
&= \frac{\exp\{\boldsymbol{\psi}'\boldsymbol{T}(\boldsymbol{\alpha},\boldsymbol{\beta})\}\exp\{\boldsymbol{\phi}'\boldsymbol{T}(\boldsymbol{\alpha},\boldsymbol{\beta})\}}{\exp\{\boldsymbol{\phi}'\boldsymbol{T}(\boldsymbol{\alpha},\boldsymbol{\beta})\}\sum_{\boldsymbol{\alpha}\in\boldsymbol{\Omega}^g}\exp\{\boldsymbol{\psi}'\boldsymbol{T}(\boldsymbol{\alpha},\boldsymbol{\beta})\}}, \quad (\text{since } \boldsymbol{\phi}'\boldsymbol{T}(\boldsymbol{\alpha},\boldsymbol{\beta}) \perp \boldsymbol{\alpha}) \\
&= \frac{\exp\{\boldsymbol{\psi}'\boldsymbol{T}(\boldsymbol{\alpha},\boldsymbol{\beta})\}}{\sum_{\boldsymbol{\alpha}\in\boldsymbol{\Omega}^g}\exp\{\boldsymbol{\psi}'\boldsymbol{T}(\boldsymbol{\alpha},\boldsymbol{\beta})\}} \\
&= \mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}), \qquad (4.27)
\end{aligned}$$

which contradicts the identifiability assumption (Definition 4.1) since $\boldsymbol{\phi} \neq 0$. Therefore, $F_m(\boldsymbol{\psi})$ is strictly concave whenever $\inf_{\boldsymbol{\beta}} \frac{\eta_m(\boldsymbol{\beta})}{\eta_m} > 0$, and thus $F_m(\boldsymbol{\psi})$ is a.s. strictly concave. □

4.3.4 Lemma 4

Using a function that approximates $F_m(\boldsymbol{\psi})$ from Lemma 4.3, the objective of the fourth lemma is to establish that this approximating function is uniquely maximized, with probability one, when it is evaluated at $\boldsymbol{\psi}_0$.

Lemma 4.4. *Let*

$$G_m(\boldsymbol{\psi}) = \sum_{\boldsymbol{\beta}} \frac{\eta_m(\boldsymbol{\beta})}{\eta_m} \sum_{\boldsymbol{\alpha}} GL(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0) \log \frac{GL(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi})}{GL(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0)}. \qquad (4.28)$$

1. *With probability one,* $\forall \epsilon > 0$, $\exists \delta > 0$ *such that*

$$\limsup_{m\to\infty} \sup_{|\boldsymbol{\psi}-\boldsymbol{\psi}_0|\leq\epsilon} \sup_{\boldsymbol{\phi}\in\mathbb{R}^p,|\boldsymbol{\phi}|=1} \boldsymbol{\phi}' H(G_m(\boldsymbol{\psi}))\boldsymbol{\phi} \;<\; -\delta,$$

   where $H(G_m(\boldsymbol{\psi}))$ *is the Hessian of* $G_m(\boldsymbol{\psi})$ *with respect to* $\boldsymbol{\psi}$.

2. $G_m(\boldsymbol{\psi}) \leq 0$ *for all* $\boldsymbol{\psi}, m$.

3. $G_m(\boldsymbol{\psi}_0) = 0$ *for all* m.

Before proving Lemma 4.4, note the following two things. First, motivated by Lemma 4.2, $G_m(\boldsymbol{\psi})$ is just an estimate of $F_m(\boldsymbol{\psi})$ in which $\frac{\eta_m(\boldsymbol{\alpha},\boldsymbol{\beta})}{\eta(\boldsymbol{\beta})}$ is now approximated by $\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0)$. Second, Lemma 4.4 establishes that $G_m(\boldsymbol{\psi}_0)$ is, with probability one, the *unique* maximum of $G_m(\boldsymbol{\psi})$.

Proof. By an essentially identical argument that was used to establish the strict concavity of $F_m(\boldsymbol{\psi})$ in the proof of Lemma 4.3, the strict concavity of $G_m(\boldsymbol{\psi})$ is also obtained whenever $\inf_{\boldsymbol{\beta}} \frac{\eta_m(\boldsymbol{\beta})}{\eta_m} > 0$. The only difference in the argument is that before (4.26), instead of noting that $\sum_{\boldsymbol{\alpha}} \frac{\eta_m(\boldsymbol{\alpha},\boldsymbol{\beta})}{\eta_m(\boldsymbol{\beta})} = 1$, we must note that $\sum_{\boldsymbol{\alpha}} \mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0) = 1$. Now, by Lemma 4.1, with probability one, there exists $\varsigma > 0$ such that $\inf_{\boldsymbol{\beta}} \frac{\eta_m(\boldsymbol{\beta})}{\eta_m} \geq \varsigma$ for all m sufficiently large. Furthermore, in light of (4.26), it is clear that $\boldsymbol{\phi}' H(G_m(\boldsymbol{\psi}))\boldsymbol{\phi}$ is differentiable in $\boldsymbol{\phi}, \boldsymbol{\psi}$, and the finite collection of variables $\frac{\eta_m(\boldsymbol{\beta})}{\eta_m}$; therefore, $\boldsymbol{\phi}' H(G_m(\boldsymbol{\psi}))\boldsymbol{\phi}$ is also jointly continuous in $\boldsymbol{\phi}, \boldsymbol{\psi}$, and the finite collection of variables $\frac{\eta_m(\boldsymbol{\beta})}{\eta_m}$. Then, since continuous functions over a compact domain necessarily achieve their maximum at a point in that domain (by Theorem 4.3), we know $\boldsymbol{\phi}' H(G_m(\boldsymbol{\psi}))\boldsymbol{\phi}$ must achieve its maximum on the compact set $|\boldsymbol{\phi}| = 1$, $|\boldsymbol{\psi} - \boldsymbol{\psi}_0| \leq \epsilon$, and $\frac{\eta_m(\boldsymbol{\beta})}{\eta_m} \in [\varsigma, 1]$ for all $\boldsymbol{\beta}$. Part 1 of Lemma 4.4 is then obtained by appealing to the strict concavity of $G_m(\boldsymbol{\psi})$.

As for part 2, first recall Jensen's inequality which asserts that for a random variable X and

a convex function $g(\cdot)$, $g(\mathrm{E}(X)) \leq \mathrm{E}[g(X)]$, provided that the expectations exist. Similarly, a simple corollary of Jensen's inequality asserts that for a random variable X and a concave function $g(\cdot)$, $g(\mathrm{E}(X)) \geq \mathrm{E}[g(X)]$, provided that the expectations exist. Now consider the following, noting that the logarithm is a concave function and that the specified expectations clearly exist:

$$\sum_{\boldsymbol{\alpha}} \mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0) \log\left\{\frac{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi})}{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0)}\right\}$$

$$\begin{aligned}
&= \mathrm{E}_{\boldsymbol{\psi}_0}\left[\log\left\{\frac{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi})}{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0)}\right\}\right] \\
&\leq \log\left\{\mathrm{E}_{\boldsymbol{\psi}_0}\left[\frac{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi})}{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0)}\right]\right\} \qquad \text{(by the Corollary to Jensen's Inequality)} \\
&= \log\left\{\sum_{\boldsymbol{\alpha}} \mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0)\left(\frac{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi})}{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0)}\right)\right\} \\
&= \log\left\{\sum_{\boldsymbol{\alpha}} \mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi})\right\} \\
&= \log\{1\} \\
&= 0 \qquad (4.29)
\end{aligned}$$

Hence, since the inequality given by (4.29) holds for all $\boldsymbol{\psi}$ and m, part 2 of Lemma 4.4 is easily obtained:

$$\begin{aligned}
G_m(\boldsymbol{\psi}) &= \sum_{\boldsymbol{\beta}} \frac{\eta_m(\boldsymbol{\beta})}{\eta_m} \sum_{\boldsymbol{\alpha}} \mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0) \log \frac{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi})}{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0)} \\
&\leq \sum_{\boldsymbol{\beta}} \frac{\eta_m(\boldsymbol{\beta})}{\eta_m} \cdot (0) \\
&= 0. \qquad (4.30)
\end{aligned}$$

Finally, part three of Lemma 4.4 is easily verified by evaluating $G_m(\cdot)$ at $\boldsymbol{\psi}_0$:

$$\begin{aligned} G_m(\boldsymbol{\psi}_0) &= \sum_{\boldsymbol{\beta}} \frac{\eta_m(\boldsymbol{\beta})}{\eta_m} \sum_{\boldsymbol{\alpha}} \mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0) \log \frac{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0)}{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0)} \\ &= \sum_{\boldsymbol{\beta}} \frac{\eta_m(\boldsymbol{\beta})}{\eta_m} \sum_{\boldsymbol{\alpha}} \mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0) \log\{1\} \\ &= 0. \end{aligned}$$

Therefore, parts two and three have established, when coupled with part one, that $G_m(\boldsymbol{\psi}_0)$ is almost surely the unique maximum of $G_m(\cdot)$. □

4.3.5 Lemma 5

The objective of this fifth and final lemma is to establish that the difference between $F_m(\boldsymbol{\psi})$, from Lemma 4.3, and its approximation, $G_m(\boldsymbol{\psi})$, from Lemma 4.4, is almost surely zero.

Lemma 4.5. *For all $\epsilon > 0$,*

$$\limsup_{m\to\infty} \sup_{|\boldsymbol{\psi}-\boldsymbol{\psi}_0|\leq\epsilon} |F_m(\boldsymbol{\psi}) - G_m(\boldsymbol{\psi})| = 0 \ \textit{a.s.}$$

Proof. First, observe the following:

$$\limsup_{m\to\infty} \sup_{|\boldsymbol{\psi}-\boldsymbol{\psi}_0|\leq\epsilon} |F_m(\boldsymbol{\psi}) - G_m(\boldsymbol{\psi})|$$

$$= \limsup_{m\to\infty} \sup_{|\boldsymbol{\psi}-\boldsymbol{\psi}_0|\leq\epsilon} \left| \sum_{\boldsymbol{\beta}} \frac{\eta_m(\boldsymbol{\beta})}{\eta_m} \sum_{\boldsymbol{\alpha}} \left(\frac{\eta_m(\boldsymbol{\alpha},\boldsymbol{\beta})}{\eta_m(\boldsymbol{\beta})} - \mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0) \right) \log \frac{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi})}{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0)} \right|$$

$$\leq \limsup_{m\to\infty} \sup_{|\boldsymbol{\psi}-\boldsymbol{\psi}_0|\leq\epsilon} \sum_{\boldsymbol{\beta}} \frac{\eta_m(\boldsymbol{\beta})}{\eta_m} \sum_{\boldsymbol{\alpha}} \left| \frac{\eta_m(\boldsymbol{\alpha},\boldsymbol{\beta})}{\eta_m(\boldsymbol{\beta})} - \mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0) \right| \left| \log \frac{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi})}{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0)} \right| \ \text{(Triangle Ineq.)}$$

$$\begin{aligned}
&\leq \limsup_{m\to\infty} \sup_{\boldsymbol{\alpha},|\boldsymbol{\psi}-\boldsymbol{\psi}_0|\leq\epsilon} \sum_{\boldsymbol{\beta}} \frac{\eta_m(\boldsymbol{\beta})}{\eta_m} |\boldsymbol{\Omega}^g| \left| \frac{\eta_m(\boldsymbol{\alpha},\boldsymbol{\beta})}{\eta_m(\boldsymbol{\beta})} - \mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0) \right| \left| \log \frac{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi})}{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0)} \right| \\
&\leq \limsup_{m\to\infty} \sup_{\boldsymbol{\alpha},\boldsymbol{\beta},|\boldsymbol{\psi}-\boldsymbol{\psi}_0|\leq\epsilon} |\boldsymbol{\Omega}^{g+\partial g}| \left| \frac{\eta_m(\boldsymbol{\alpha},\boldsymbol{\beta})}{\eta_m(\boldsymbol{\beta})} - \mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0) \right| \left| \log \frac{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi})}{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0)} \right| \\
&= |\boldsymbol{\Omega}^{g+\partial g}| \limsup_{m\to\infty} \sup_{\boldsymbol{\alpha},\boldsymbol{\beta},|\boldsymbol{\psi}-\boldsymbol{\psi}_0|\leq\epsilon} \left| \frac{\eta_m(\boldsymbol{\alpha},\boldsymbol{\beta})}{\eta_m(\boldsymbol{\beta})} - \mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0) \right| \left| \log \frac{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi})}{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0)} \right| \\
\leq\ & |\boldsymbol{\Omega}^{g+\partial g}| \sup_{\boldsymbol{\alpha},\boldsymbol{\beta},|\boldsymbol{\psi}-\boldsymbol{\psi}_0|\leq\epsilon} \left| \log \frac{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi})}{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0)} \right| \limsup_{m\to\infty} \sup_{\boldsymbol{\alpha},\boldsymbol{\beta}} \left| \frac{\eta_m(\boldsymbol{\alpha},\boldsymbol{\beta})}{\eta_m(\boldsymbol{\beta})} - \mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0) \right| . \quad (4.31)
\end{aligned}$$

Now by Lemma 4.2, we know that $\limsup_{m\to\infty} \sup_{\boldsymbol{\alpha},\boldsymbol{\beta}} \left| \frac{\eta_m(\boldsymbol{\alpha},\boldsymbol{\beta})}{\eta_m(\boldsymbol{\beta})} - \mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0) \right| = 0$ a.s. Furthermore, since $\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}) \neq 0$ for any $\boldsymbol{\alpha},\boldsymbol{\beta},\ \boldsymbol{\psi} \in \mathbb{R}^p$ and is continuous in $\boldsymbol{\psi}$ for each of the finitely many possibilities of $\boldsymbol{\alpha} \in \boldsymbol{\Omega}^g$ and $\boldsymbol{\beta} \in \boldsymbol{\Omega}^{\partial g}$, and since $\{\boldsymbol{\psi} : |\boldsymbol{\psi}-\boldsymbol{\psi}_0| \leq \epsilon\}$ is a compact set, $\sup_{\boldsymbol{\alpha},\boldsymbol{\beta},|\boldsymbol{\psi}-\boldsymbol{\psi}_0|\leq\epsilon} \left| \log \frac{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi})}{\mathrm{GL}(\boldsymbol{\alpha}|\boldsymbol{\beta};\boldsymbol{\psi}_0)} \right|$ is finite (by Theorem 4.3). Therefore, from the inequality given by (4.31), we get the desired result:

$$\limsup_{m\to\infty} \sup_{|\boldsymbol{\psi}-\boldsymbol{\psi}_0|\leq\epsilon} |F_m(\boldsymbol{\psi}) - G_m(\boldsymbol{\psi})| = 0 \text{ a.s.}$$

□

4.4 Proof of MGPLE's Strong Consistency

With the above five lemmas, a short proof of Theorem 4.1 can now be given.

Proof. By Lemma 4.3, we know that $F_m(\cdot)$ is almost surely strictly concave, and therefore part (a) of Theorem 4.1 is satisfied for $F_m(\boldsymbol{\psi})$. Now fix $\epsilon > 0$. By Lemma 4.4, we can conclude the following:

$$\liminf_{m\to\infty} \inf_{|\boldsymbol{\psi}_0-\boldsymbol{\psi}|=\epsilon} |G_m(\boldsymbol{\psi}) - G_m(\boldsymbol{\psi}_0)| > 0 \text{ a.s.} \quad (4.32)$$

Then, because Lemma 4.5 asserts that $F_m(\cdot)$ is uniformly approximated by $G_m(\cdot)$, the result given by (4.32) for $G_m(\cdot)$ also holds for $F_m(\cdot)$:

$$\liminf_{m\to\infty} \inf_{|\boldsymbol{\psi}_0-\boldsymbol{\psi}|=\epsilon} |F_m(\boldsymbol{\psi}) - F_m(\boldsymbol{\psi}_0)| \;\; > \;\; 0 \text{ a.s.} \tag{4.33}$$

Thus, whenever $\boldsymbol{\psi} \neq \boldsymbol{\psi}_0$, with probability one $F_m(\boldsymbol{\psi}) \neq F_m(\boldsymbol{\psi}_0)$. Now, because $F_m(\cdot)$ is almost surely strictly concave (by Lemma 4.3), we then know by Theorem 4.3 that $F_m(\cdot)$, a continuous function of $\boldsymbol{\psi}$, almost surely achieves its maximum, uniquely, in the compact set $\{\boldsymbol{\psi} : |\boldsymbol{\psi} - \boldsymbol{\psi}_0| < \epsilon\}$. Furthermore, since ϵ is arbitrary, $F_m(\boldsymbol{\psi})$ almost surely obtains its *unique* maximum when $\boldsymbol{\psi} = \boldsymbol{\psi}_0$; therefore, parts (b) and (c) of Theorem 4.1 are also satisfied for $F_m(\boldsymbol{\psi})$. Finally, since by (4.15) we know that $\log\{\text{GPL}_m(\boldsymbol{\psi}|\boldsymbol{Z}(m))\} = \eta_m F_m(\boldsymbol{\psi}) + \log\{\text{GPL}_m(\boldsymbol{\psi}_0; \boldsymbol{Z}(m))\}$, where $\log\{\text{GPL}_m(\boldsymbol{\psi}_0; \boldsymbol{Z}(m))\}$ is just a constant, parts (a), (b), and (c) of Theorem 4.1 are also satisfied for $\log\{\text{GPL}_m(\boldsymbol{\psi}|\boldsymbol{Z}(m))\}$. □

4.5 Strong Consistency of MBGPLEs

The theorem establishing the strong consistency of the MBGPLEs is completely analogous to Theorem 4.1, but before it can be formally stated, some of the notation introduced in section 4.2 must be reformulated under the BGPL paradigm. To that end, let $\widehat{\boldsymbol{\psi}}_m$ now denote the maximum block generalized pseudolikelihood estimate of $\boldsymbol{\psi}_0$ obtained from the m^{th} sample in the sequence, $\boldsymbol{z}(m)$. It will be assumed that D has been partitioned into a countably infinite number of equally and finitely sized blocks, denoted $b(l), l = 1, 2, \ldots$. Let $N = \{N_i\}_{i\in D}$ and $B = \{b(l)\}_{l\in\mathbb{Z}^+}$ denote the translation invariant neighborhood system and block system, respectively, on D and its corresponding sublattices, $S_1, S_2, \ldots$. In this context, translation invariant means there is a fixed neighborhood structure and block structure over D. It is assumed that the neighborhood set is finite, i.e. $\exists R_1 < \infty \ni$ if $j \in N_i$ then $|j - i| \leq R_1$.

Let $b(l)^c = \{k : k \in D \backslash b(l)\}$ be the sites of D that are not members of the l^{th} block and let

$\partial b(l) = \{j : j \in N_k \cap b(l)^c, k \in b(l)\}$ be the boundary neighbor sites of block l, i.e. the sites that are neighbors of at least one site of block l, but are not themselves members of block l. Let $|b(l)|$ denote the number of sites in block l and let $|\partial b(l)|$ denote the number of boundary sites for block l, both of which are constant for all $i \in D$ and $l \in \mathbb{Z}^+$ under the translation invariant block structure assumption. In other words, for all $l = 1, 2, \ldots$, $|b(l)| = b$ and $|\partial b(l)| = \partial b$. Subgraphs of (D, N, B), consisting of sites S_m, are denoted by (S_m, N, B). Each subgraph (S_m, N, B), $m = 1, 2, \ldots$, has corresponding MRF distribution $\mathrm{L}_m(\boldsymbol{\psi}_0)$, where each variable of $\boldsymbol{Z}(m)$ has finite support set, Ω, and $\boldsymbol{Z}(m)$ has joint support set $\boldsymbol{\Omega}^{n_m} = \Omega \times \ldots \times \Omega$.

Now fix m and let $\boldsymbol{z} \in \boldsymbol{\Omega}^{n_m}$ be a realization from the MRF distribution $\mathrm{L}_m(\boldsymbol{\psi}_0)$. Then, as in Chapter 3, let $\boldsymbol{Z}_{b(l)} = \{Z_k : k \in b(l)\}$ denote the vector of variables corresponding to the sites of block l. Additionally, let $N_k^{b(l)} = \{j : j \in N_k \cap b(l), k \in b(l)\}$ denote the neighbors of site k, where k is a member of block l, that are also members of block l, and let $N_k^{\partial b(l)} = \{j : j \in N_k \cap \partial b(l), k \in b(l)\}$ denote the neighbors of site k, where k is a member of block l, that are boundary neighbors of block l. Furthermore, let $\boldsymbol{Z}^{b(l)} = \{Z_j : j \in N_k^{\partial b(l)}, k \in b(l)\}$ denote the vector of variables corresponding to the boundary neighbor sites of block l. Note that the block and/or neighborhood sets located on the edge of S_m will necessarily be incomplete in the sense that portions of these sets will belong to some $S_{m'}$, where $m' > m$. The way S_m^0 is defined, further below, explains how such blocks are handled. The conditional probabilities (i.e. block joint likelihood functions) $\mathrm{BL}_l^m(\boldsymbol{Z}_{b(l)} = \boldsymbol{z}_{b(l)} | \boldsymbol{Z}^{b(l)} = \boldsymbol{z}^{b(l)}; \boldsymbol{\psi}_0)$, for each $b(l) \ni b(l) \cap S_m \neq \emptyset$, and each $\boldsymbol{z} \in \boldsymbol{\Omega}^{n_m}$, will be referred to as the "*local characteristics*" of $\mathrm{L}_m(\boldsymbol{\psi}_0)$.

The MRF distributions $\mathrm{L}_1(\boldsymbol{\psi}_0), \mathrm{L}_2(\boldsymbol{\psi}_0), \ldots$ are connected by the assumption that the local characteristics, which depend upon $\boldsymbol{\psi}_0$, are independent of l and m, for all l such that $b(l) \subset S_m^0$, the interior of S_m, where $S_m^0 \equiv \{i \in S_m : i \in b(l) \Rightarrow \partial b(l) \subseteq S_m\}$; note that $\partial b(l) \subseteq S_m \Rightarrow b(l) \subseteq S_m$. In other words, the interior of S_m contains all of the sites of S_m that belong to a block whose corresponding members and boundary neighbors are all contained within S_m. Consequently, $\forall m$, and $\forall l$ such that $b(l) \subset S_m^0$, $|b(l)| = b$ and $|\partial b(l)| = \partial b$. Hence, the

conditional probability distribution, under N and B, for interior sites of S_m does not depend on l or m:

$$\begin{aligned} \mathrm{BL}_l^m(\boldsymbol{Z}_{b(l)} = \boldsymbol{z}_{b(l)} | \boldsymbol{Z}^{b(l)} = \boldsymbol{z}^{b(l)}; \boldsymbol{\psi}_0) &= \mathrm{BL}(\boldsymbol{Z}_{b(l)} = \boldsymbol{z}_{b(l)} | \boldsymbol{Z}^{b(l)} = \boldsymbol{z}^{b(l)}; \boldsymbol{\psi}_0) \\ &= \frac{\exp\{\boldsymbol{T}'\left(\boldsymbol{z}_{b(l)}, \boldsymbol{z}^{b(l)}\right) \cdot \boldsymbol{\psi}_0\}}{\sum_{\boldsymbol{y} \in \boldsymbol{\Omega}^b} \exp\{\boldsymbol{T}'\left(\boldsymbol{y}_{b(l)}, \boldsymbol{y}^{b(l)}\right) \cdot \boldsymbol{\psi}_0\}}, \end{aligned} \tag{4.34}$$

for all m, $l \ni b(l) \subset S_m^0$, $\boldsymbol{z}_{b(l)} \in \boldsymbol{\Omega}^b$, and $\boldsymbol{z}^{b(l)} \in \boldsymbol{\Omega}^{\partial b}$. Since interest is limited to local characteristics at interior sites, the sub/superscripts l and m will henceforth be dropped when writing the conditional probabilities $\mathrm{BL}(\cdot|\cdot)$. Thus, the corresponding block generalized pseudolikelihood function of $\boldsymbol{\psi} \in \boldsymbol{\Psi}$, given a sample, $\boldsymbol{Z} = \boldsymbol{z}$, from $\mathrm{L}_m(\boldsymbol{\psi}_0)$, is

$$\begin{aligned} \mathrm{BGPL}_m(\boldsymbol{\psi}|\boldsymbol{z}) &= \prod_{l:b(l)\subset S_m^0} \mathrm{BL}(\boldsymbol{z}_{b(l)} | \boldsymbol{z}^{b(l)}; \boldsymbol{\psi}) \\ &= \prod_{l:b(l)\subset S_m^0} \frac{\exp\{\boldsymbol{T}'\left(\boldsymbol{z}_{b(l)}, \boldsymbol{z}^{b(l)}\right) \cdot \boldsymbol{\psi}\}}{\sum_{\boldsymbol{y} \in \boldsymbol{\Omega}^b} \exp\{\boldsymbol{T}'\left(\boldsymbol{y}_{b(l)}, \boldsymbol{y}^{b(l)}\right) \cdot \boldsymbol{\psi}\}}. \end{aligned} \tag{4.35}$$

The block generalized pseudolikelihood estimate of $\boldsymbol{\psi}_0$ is the set, $M_m(\boldsymbol{z})$, of $\boldsymbol{\psi}$ that maximizes $\mathrm{BGPL}_m(\boldsymbol{\psi}|\boldsymbol{z})$:

$$M_m(\boldsymbol{z}) = \left\{ \boldsymbol{\psi} \in \boldsymbol{\Psi} : \mathrm{BGPL}_m(\boldsymbol{\psi}|\boldsymbol{z}) = \sup_{\boldsymbol{\phi} \in \boldsymbol{\Psi}} \mathrm{BGPL}_m(\boldsymbol{\phi}|\boldsymbol{z}) \right\}. \tag{4.36}$$

As with GPL, identifiability is assumed, where it is defined as follows for BGPL.

Definition 4.2 (Identifiability – BGPL)**.** *$\boldsymbol{\psi}_0 \in \boldsymbol{\Psi}$ is identifiable if $\boldsymbol{\psi} \neq \boldsymbol{\psi}_0 \Rightarrow \exists\ \boldsymbol{z}^{b(l)}, \boldsymbol{z}_{b(l)}$ such that $BL(\boldsymbol{z}_{b(l)}|\boldsymbol{z}^{b(l)}; \boldsymbol{\psi}) \neq BL(\boldsymbol{z}_{b(l)}|\boldsymbol{z}^{b(l)}; \boldsymbol{\psi}_0)$.*

With the notation reformulated for BGPL, the strong consistency of the MBGPLEs can now be formally stated.

Theorem 4.4 (Strong Consistency of Block Generalized Pseudolikelihood)**.** *For each* $m = 1, 2, \ldots$, *let* $\boldsymbol{z}(m)$ *be a sample from the Markov random field* $L_m(\boldsymbol{\psi}_0)$*, with local characteristics* (4.34)*. If* $\boldsymbol{\psi}_0$ *is identifiable, then*

(a) $P(\log BGPL_m(\boldsymbol{\psi}|\boldsymbol{z}(m))$ *is strictly concave for all* m *sufficiently large*$) = 1$*;*

(b) $P(M_m(\boldsymbol{z}(m))$ *is a singleton for all* m *sufficiently large*$) = 1$*;*

(c) $P\left(\sup_{\boldsymbol{\psi} \in M_m(\boldsymbol{z}(m))} |\boldsymbol{\psi} - \boldsymbol{\psi}_0| \to 0\right) = 1$*, as* $m \to \infty$*.*

As was the case for Theorem 4.1, it is worth emphasizing for Theorem 4.4 that since the number of sites of the lattice goes to infinity as $m \to \infty$, the asymptotics are when the number of sites of the lattice is large. A proof of Theorem 4.4, with the notation established at the beginning of this section, would be almost perfectly analogous to the proof of Theorem 4.1, and will, therefore, not be provided in this dissertation. The only subtle differences in a proof of the MBGPLE's strong consistency, relative to that of the MGPLE's, would be in the proofs of the analogous versions of Lemmas 4.1 and 4.2. In particular, in proving the analogous version of Lemma 4.1 for BGPL, rather than selecting groups $g(i_1), g(i_2), \ldots, g(i_{k_m})$, where $i_1, i_2, \ldots, i_{k_m} \in S_m$ were the "central" sites of these groups, we would select blocks $b(l_1), b(l_2), \ldots, b(l_{k_m}) \subset S_m$. Lastly, in proving the analogous version of Lemma 4.2 for BGPL, the coloring of (D,N,B), $C = \{c_q : q = 1, \ldots, k\}$, still partitions D, but under a different schematic. If $i, j \in b(l)$, then $i, j \in c_q$, and if $i \in \{b(l) \cap c_q\}$ and $j \in \partial b(l)$, then $j \in c_r$, $r \neq q$. In other words, every site of a block is of the same color, but if another block has one or more sites in the set of boundary neighbors of the first block, then the second block's sites are necessarily of a different color. Note that the neighborhood structure and the block structure still guide the coloring scheme. As a result of this alteration to the coloring schematic, several quantities that would be defined and used throughout the remainder of a proof of the analogous version of Lemma 4.2 are in terms of blocks, rather than the "central" sites identifying groups. For example, in the proof of Lemma 4.2 for GPL, $\eta_m(\boldsymbol{\beta}; c_q) \equiv |\{i \in S_m^0 \cap c_q : \boldsymbol{Z}^{g(i)} = \boldsymbol{\beta}\}|$, but in an analogous proof for BGPL, $\eta_m(\boldsymbol{\beta}; c_q) \equiv |\{b(l) \subset S_m^0 \cap c_q : \boldsymbol{Z}^{b(l)} = \boldsymbol{\beta}\}|$. In other words,

$\eta_m(\boldsymbol{\beta}; c_q)$ is now the number of blocks of S_m^0 that are of color c_q and have boundary neighbor configuration $\boldsymbol{\beta}$. Do note, however, that this modification to an analogous proof of Lemma 4.2 is, as with an analogous proof of Lemma 4.1, not a substantial one. More specifically, while the $\eta_m(\boldsymbol{\beta}; c_q)$ from the GPL proof is a count of the number of "central" sites satisfying the specified conditions, it can in fact, as was discussed in the proof of Lemma 4.2, also be viewed as a count of the number of groups that satisfy the specified conditions. Therefore, when considered from this perspective, the few modifications to the proof of Theorem 4.1 that are needed to then prove Theorem 4.4 are truly minor.

Chapter 5

Autologistic Model Simulation Study

5.1 Introduction – Three Questions of Interest

Although numerous simulation studies involving the autologistic model have been carried out in the literature, as discussed in detail in section 1.4, a simulation study simultaneously comparing the estimation methods discussed in Chapter 3 (MPL, MCMCML, MGPL, MBGPL) has never been conducted. Hence, the large scale simulation study presented in this chapter comparing these four methods via an examination of the precision and accuracy associated with their corresponding estimators is the first of its kind. The overarching objective of this simulation study is to answer three questions that have either been ignored or unconvincingly addressed by the current simulation studies appearing in the literature. In particular, these questions are:

(1) Which method, if either, of GPL and BGPL is "better." More specifically, is there even

a noticeable difference between the estimates obtained from these two newer methods?

(2) Do the resulting MCMCMLEs noticeably differ when the MGPLE or MBGPLE are used as the reference point rather than the MPLE?

(3) How do the MPLEs, MCMCMLEs, MGPLEs, and MBGPLEs compare under different types/strengths of spatial correlation? More specifically, do GPL and BGPL achieve the intended compromise between PL and MCMCML and are there situations where some methods are preferable to the others?

To address the above three questions, multiple cases, corresponding to various neighborhood structures, both with and without a covariate term present in the model, were considered in the simulation study. The specifics of each case will be discussed below in section 5.2. Furthermore, to limit the extent to which the answers to the above three questions might be obscured by secondary components of the simulation study, three lattices sizes (26×26, 50×50, and 74×74), multiple spatial dependence strengths, and, when applicable, multiple covariate strengths and types were considered for each case of the simulation study. The specific values used for both the spatial dependence and covariate (when applicable) strengths, as well as the particular types of covariates considered (when applicable), varied from case to case and will be explained in detail in section 5.2. The three lattice sizes employed throughout the simulation study were chosen in tandem with the block sizes used for BGPL in the simulation study. The reason for such consideration is the fact that the blocks of BGPL partition the lattice and, thus, it is possible to select lattice sizes that do not result in partially incomplete blocks near the edges of the lattice. Note that group sizes for GPL were not taken into consideration when choosing the lattice sizes for the simulation study since groups unavoidably overlap and, therefore, regardless of the lattice size selected, incomplete groups are impossible to avoid. Now because blocks of size 5×5 and larger were found, a priori, to be prohibitively expensive to employ (because repeated generation of the 2^{25} realizations necessary to compute block normalizing constants was computationally expensive), block sizes of 2×2, 3×3, and 4×4

were settled upon for use in the simulation study. Then, since the least common multiple of 2, 3, and 4 is 12, we decided to use lattices of dimension 24×24, 48×48, and 72×72. However, because we also wanted to employ a "guard region" edge adjustment procedure for all of the estimation methods (to simplify the coding), which recall means that the outer "layer" of the lattice is treated as fixed, we finally settled on using lattices of size 26×26, 50×50, and 74×74.

The entire simulation study was carried out in R, and with the exception of using a handful of built-in R functions, including *glm* for obtaining the maximum pseudolikelihood estimates and *optim* for carrying out numerical optimization, most code used was written by the author for the simulation study. This entailed, among other things, writing a function that constructed the neighborhood matrices for an arbitrary lattice, a function that generated lattices for the simulation study via the Gibbs sampler, functions that constructed the generalized and block generalized pseudolikelihood functions, and functions that generated Monte Carlo samples and constructed the subsequent Monte Carlo approximate negative log likelihood function. Finally, note that as a result of personal communications between November 2010 and January 2011 with Dr. Jonathan Graham, a function that more efficiently computes the individual group and block normalizing constants within the aforementioned generalized and block generalized pseudolikelihood R functions was written by Dr. Graham and subsequently implemented in this simulation study.

The remainder of this chapter is divided into 5 sections. Section 5.2 systematically explains each of the seven cases of the simulation study, including how each case was implemented within the framework of the study. Sections 5.3, 5.4, and 5.5 present the answers to questions 1, 2, and 3, respectively, that were obtained from the simulation study. Finally, section 5.6 summarizes the key results of the simulation study and briefly discusses a few potentially lingering questions.

5.2 Cases of Simulation Study

Three neighborhood structures were considered within the simulation study, including a first-order isotropic structure, a second-order isotropic structure, and, finally, a first-order two-way anisotropic structure where the anisotropy corresponds to north/south and east/west gradients. The autologistic models associated with these three neighborhood structures, both with and without a covariate term, comprise the first six cases of the simulation study. The first three cases, which do not have a covariate term, are primarily concerned with comparing the performances of the estimation methods as the strength of the spatial dependence increases from weak to strong. The next three cases, which do have a covariate term, are primarily concerned with comparing the performances of the estimation methods for various types of covariates under both a moderately weak and a moderately strong covariate effect. The seventh and final case of the simulation study addresses estimation of the Ising model (i.e. first-order isotropic autologistic model) as the parameters near a critical value ($\theta \approx 1.76$ ($\alpha = -2\theta$)) associated with phase transition.

5.2.1 First-Order Isotropy without a Covariate – FI2p

The first case considered in this simulation study corresponds to the simplest autologistic model, known as the Ising model, which has no covariate parameters and follows an isotropic first-order neighborhood system. Thus, the Ising model has only two parameters, an intercept parameter and a spatial dependence parameter. Recall that the form of the Ising model is the following:

$$\mathrm{P}(z_i|\{z_j : j \in N_i\}) = \frac{\exp\{\alpha z_i + \theta \sum_{j\in N_i} z_i z_j\}}{1+\exp\{\alpha + \theta \sum_{j\in N_i} z_j\}}, \quad i = 1,\ldots,n, \quad z_i = 0, 1. \tag{5.1}$$

Eleven different parameter vectors, (α, θ), were considered for this case of the simulation study. In particular, α was held fixed at -1.0 for all eleven scenarios while $\theta = 0.0, 0.1, \ldots, 1.0$. These

values of θ correspond to no spatial dependence ($\theta = 0.0$) all the way up to extremely strong spatial dependence ($\theta = 1.0$). Values of $\theta > 1.0$, holding $\alpha = -1.0$ fixed, were not considered in the simulation as the number of sites that assume a value of one or zero in the resulting lattices would be disproportionate, which compromises the efficiency of the autologistic model ([2]). For example, when $\alpha = -1.0$ and $\theta = 1.0$, we would expect roughly 93% of the lattice sites to have value 1, and such a high percentage leads to a substantial decline in the performance of all four estimation methods, as will be seen in the forthcoming analysis. Hence, no value of θ was taken to be greater than 1.0. When $\alpha = -1$ and $\theta = 0.5$ ($\alpha = -2\theta$), the proportion of 1's is expected to be 50%.

Now with respect to the estimation methods, as was mentioned at the beginning of this chapter, three different block sizes were implemented under BGPL for the simulation study, including 2×2, 3×3, and 4×4. For GPL, two group sizes were considered, including a group size of 5, corresponding to a cross shape, and a group size of 9, corresponding to a 3×3 grid. Finally, for MCMCML, six different reference points were employed, including the MPLE, the MGPLE for both group sizes, and the MBGPLE for all three block sizes.

For this case, the simulation procedure for each of the 11 parameter vectors was the following:

(1) Using the Gibbs sampler, 500 80×80 Markov chain Monte Carlo (MCMC) lattice realizations (i.e. samples) were generated from the Ising model with true parameter vector (α_0, θ_0). A torus edge adjustment was used within the Gibbs sampler and each of the 500 realizations was obtained from an independent Markov chain in which a burn-in of 100 full lattice sweeps was employed to allow for convergence.

(2) For each of the above 500 lattice realizations, the central 26×26 lattice was extracted. These 500 26×26 samples will henceforth be referred to as the "original" samples.

(3) For the i^{th} "original" sample, $i = 1, \ldots, 500$, the pseudolikelihood function was constructed and maximized with respect to (α, θ) to produce the maximum pseudolikeli-

hood estimates (MPLEs) $(\hat{\alpha}_{PL_i}, \hat{\theta}_{PL_i})$. The mean, standard deviation, bias, and mean absolute error (MAE) of the parameter estimates were estimated using the 500 MPLEs as follows:

$$\bar{\hat{\alpha}}_{PL} = \frac{1}{500}\sum_{i=1}^{500}\hat{\alpha}_{PL_i}, \quad \bar{\hat{\theta}}_{PL} = \frac{1}{500}\sum_{i=1}^{500}\hat{\theta}_{PL_i},$$

$$\widehat{\mathrm{SD}}(\hat{\alpha}_{PL}) = \sqrt{\frac{1}{500}\sum_{i=1}^{500}(\hat{\alpha}_{PL_i} - \bar{\hat{\alpha}}_{PL})^2}, \quad \widehat{\mathrm{SD}}(\hat{\theta}_{PL}) = \sqrt{\frac{1}{500}\sum_{i=1}^{500}(\hat{\theta}_{PL_i} - \bar{\hat{\theta}}_{PL})^2},$$

$$\widehat{\mathrm{Bias}}(\bar{\hat{\alpha}}_{PL}) = (\bar{\hat{\alpha}}_{PL} - \alpha_0), \quad \widehat{\mathrm{Bias}}(\bar{\hat{\theta}}_{PL}) = (\bar{\hat{\theta}}_{PL} - \theta_0),$$

$$\widehat{\mathrm{MAE}}(\hat{\alpha}_{PL}) = \frac{1}{500}\sum_{i=1}^{500}\left|\hat{\alpha}_{PL_i} - \alpha_0\right|, \quad \widehat{\mathrm{MAE}}(\hat{\theta}_{PL}) = \frac{1}{500}\sum_{i=1}^{500}\left|\hat{\theta}_{PL_i} - \theta_0\right|.$$

(4) For the i^{th} "original" sample, $i = 1, \ldots, 500$, the generalized pseudolikelihood function under a group size of 5 (gs5) was constructed and numerically maximized with respect to (α, θ) to produce the maximum generalized pseudolikelihood estimates (MGPLEs) $(\hat{\alpha}_{gs5_i}, \hat{\theta}_{gs5_i})$. The mean, standard deviation, bias, and mean absolute error (MAE) of the parameter estimates were estimated in a manner analogous to the pseudolikelihood case above, yielding the sample means $(\bar{\hat{\alpha}}_{gs5}, \bar{\hat{\theta}}_{gs5})$, sample standard deviations $(\widehat{\mathrm{SD}}(\hat{\alpha}_{gs5}), \widehat{\mathrm{SD}}(\hat{\theta}_{gs5}))$, sample biases of the sample means $(\widehat{\mathrm{Bias}}(\bar{\hat{\alpha}}_{gs5}), \widehat{\mathrm{Bias}}(\bar{\hat{\theta}}_{gs5}))$, and sample mean absolute errors $(\widehat{\mathrm{MAE}}(\hat{\alpha}_{gs5}), \widehat{\mathrm{MAE}}(\hat{\theta}_{gs5}))$.

(5) Step (4) was repeated but for a group size of 9 (gs9).

(6) For the i^{th} "original" sample, $i = 1, \ldots, 500$, the block generalized pseudolikelihood function under a block size of 2×2 (bs2) was constructed and numerically maximized with respect to (α, θ) to produce the maximum block generalized pseudolikelihood estimates (MBGPLEs) $(\hat{\alpha}_{bs2_i}, \hat{\theta}_{bs2_i})$. The mean, standard deviation, bias, and mean absolute error (MAE) of the parameter estimates were estimated in a manner analogous to the pseudolikelihood case above, yielding the sample means $(\bar{\hat{\alpha}}_{bs2}, \bar{\hat{\theta}}_{bs2})$, sample standard deviations $(\widehat{\mathrm{SD}}(\hat{\alpha}_{bs2}), \widehat{\mathrm{SD}}(\hat{\theta}_{bs2}))$, sample biases of the sample means $(\widehat{\mathrm{Bias}}(\bar{\hat{\alpha}}_{bs2}), \widehat{\mathrm{Bias}}(\bar{\hat{\theta}}_{bs2}))$, and sample mean absolute errors $(\widehat{\mathrm{MAE}}(\hat{\alpha}_{bs2}), \widehat{\mathrm{MAE}}(\hat{\theta}_{bs2}))$.

(7) Step (6) was repeated but for a block size of 3×3 (bs3).

(8) Step (6) was repeated but for a block size of 4×4 (bs4).

(9) For the i^{th} "original" sample, $i = 1, \ldots, 500$, the Gibbs sampler was used to generate 550 Monte Carlo samples from an Ising distribution with parameter vector $\boldsymbol{\phi} = (\hat{\alpha}_{PL_i}, \hat{\theta}_{PL_i})$. A burn-in period of 100 full sweeps of the lattice was implemented for convergence and after each subsequent full sweep of the lattice, the resultant realization was retained as one Monte Carlo sample.

(10) For $i = 1, \ldots, 500$, the i^{th} "original" sample and the corresponding 550 Monte Carlo samples (from step (9)) were then used to construct the Monte Carlo approximate negative log likelihood function, which was then numerically maximized to obtain the corresponding MCMCMLEs $(\hat{\alpha}_{MC_i}, \hat{\theta}_{MC_i})$. The mean, standard deviation, bias, and mean absolute error of the parameter estimates were then computed in a manner analogous to the pseudolikelihood case above, yielding the sample means $(\bar{\hat{\alpha}}_{MC}, \bar{\hat{\theta}}_{MC})$, sample standard deviations $(\widehat{\mathrm{SD}}(\hat{\alpha}_{MC}), \widehat{\mathrm{SD}}(\hat{\theta}_{MC}))$, sample biases of the sample means $(\widehat{\mathrm{Bias}}(\bar{\hat{\alpha}}_{MC}), \widehat{\mathrm{Bias}}(\bar{\hat{\theta}}_{MC}))$, and sample mean absolute errors $(\widehat{\mathrm{MAE}}(\hat{\alpha}_{MC}), \widehat{\mathrm{MAE}}(\hat{\theta}_{MC}))$.

(11) Steps (9) and (10) were repeated but with $\boldsymbol{\phi} = (\hat{\alpha}_{gs5_i}, \hat{\theta}_{gs5_i})$ in step (9).

(12) Steps (9) and (10) were repeated but with $\boldsymbol{\phi} = (\hat{\alpha}_{gs9_i}, \hat{\theta}_{gs9_i})$ in step (9).

(13) Steps (9) and (10) were repeated but with $\boldsymbol{\phi} = (\hat{\alpha}_{bs2_i}, \hat{\theta}_{bs2_i})$ in step (9).

(14) Steps (9) and (10) were repeated but with $\boldsymbol{\phi} = (\hat{\alpha}_{bs3_i}, \hat{\theta}_{bs3_i})$ in step (9).

(15) Steps (9) and (10) were repeated but with $\boldsymbol{\phi} = (\hat{\alpha}_{bs4_i}, \hat{\theta}_{bs4_i})$ in step (9).

(16) For each of the 500 lattices from step (1) the central 50×50 lattice was extracted. These 500 lattice samples will henceforth be referred to as the "original" samples.

(17) Steps (3) through (15) were repeated for the "original" samples from step (16).

(18) For each of the 500 lattices from step (1) the central 74×74 lattice was extracted. These 500 lattice samples will henceforth be referred to as the "original" samples.

(19) Steps (3) through (15) were repeated for the "original" samples from step (18).

5.2.2 First-Order Two-Way Anisotropy without a Covariate – FAhv3p

The second case considered in this simulation study corresponds to an autologistic model under a first-order two-way anisotropic neighborhood structure without a covariate term. This autologistic model has three parameters and has the following functional form:

$$\mathrm{P}(z_i|\{z_j : j \in N_i\}) = \frac{\exp\left\{\alpha z_i + \theta_v \sum_{j\in N_i^v} z_i z_j + \theta_h \sum_{j\in N_i^h} z_i z_j\right\}}{1 + \exp\left\{\alpha + \theta_v \sum_{j\in N_i^v} z_j + \theta_h \sum_{j\in N_i^h} z_j\right\}}, \tag{5.2}$$

$i = 1, \ldots, n$, where θ_v and θ_h correspond to the north/south and east/west gradients, respectively. For this case of the simulation study, 18 different parameter vectors, $(\alpha, \theta_h, \theta_v)$, were considered. More specifically, three sets of six parameter vectors were employed. For the first set of six parameter vectors, α was held fixed at -1.0 and θ_h was held fixed at 0.2 while $\theta_v = 0.1, 0.4, \ldots, 1.6$. These values for θ_h and θ_v correspond to a relatively weak east/west spatial dependence ($\theta_h = 0.2$) and a north/south spatial dependence that increases from weak ($\theta_v = 0.1$) to extremely strong ($\theta_v = 1.6$). Since we would expect roughly 87% of lattice sites to have a value of 1 when $(\alpha, \theta_h, \theta_v) = (-1.0, 0.2, 1.6)$, no values of $\theta_v > 1.6$ were considered within this first set. For the second set of six parameter vectors, α was held fixed at -1.0 and θ_h was held fixed at 0.5 while $\theta_v = 0.4, 0.6, \ldots, 1.4$. These values for θ_h and θ_v correspond to a relatively moderate east/west spatial dependence strength ($\theta_h = 0.5$) and a north/south spatial dependence that increases from moderately strong ($\theta_v = 0.4$) to extremely strong ($\theta_v = 1.4$). Since we would expect roughly 90% of lattice sites to have a value of 1 when $(\alpha, \theta_h, \theta_v) = (-1.0, 0.5, 1.4)$, no values of $\theta_v > 1.4$ were considered within this second set. For the third and final set of six parameter vectors, α was again held fixed

at -1.0 and θ_h was held fixed at 0.7 while $\theta_v = 0.6, 0.7, \ldots, 1.1$. These values for θ_h and θ_v correspond to a relatively strong east/west spatial dependence strength ($\theta_h = 0.7$) and a north/south spatial dependence that increases from strong ($\theta_v = 0.6$) to extremely strong ($\theta_v = 1.1$). Since we would expect roughly 88% of lattice sites to have a value of 1 when $(\alpha, \theta_h, \theta_v) = (-1.0, 0.7, 1.1)$, no values of $\theta_v > 1.1$ were considered within this final set. The simulation procedure for this case of the study is analogous to the procedure for the first case, except that only groups of size 9 were considered for GPL and only the MPLEs were used as the reference point for MCMCML. The reasons for these exceptions will become apparent in the forthcoming sections.

5.2.3 Second-Order Isotropy without a Covariate – SI2p

The third case considered in this simulation study corresponds to an autologistic model under a second-order isotropic neighborhood structure without a covariate term. This autologistic model has two parameters, (α, θ), and its functional form is identical to equation (5.1), where the neighborhood sets now consist of 8 sites rather than 4 sites. Ten different parameter vectors were considered for this case. In particular, α was held fixed at -1.0 while $\theta = 0.00, 0.05, \ldots, 0.45$. These values of θ correspond to no spatial dependence ($\theta = 0.00$) all the way up to extremely strong spatial dependence ($\theta = 0.45$). Since we would expect roughly 90% of lattice sites to have value 1 when $(\alpha, \theta) = (-1.0, 0.45)$, no values of $\theta > 0.45$ were considered for this case. The simulation procedure for this case of the study is also analogous to the procedure for the first case, but, as with the second case, only groups of size 9 were considered for GPL and only the MPLEs were used as the reference point for MCMCML.

5.2.4 First-Order Isotropy with a Covariate – FI3p

The fourth case considered in this simulation study corresponds to an autologistic model under a first-order isotropic neighborhood structure with a covariate term. This autologistic model has three parameters, (α, θ, β), and its functional form is the following:

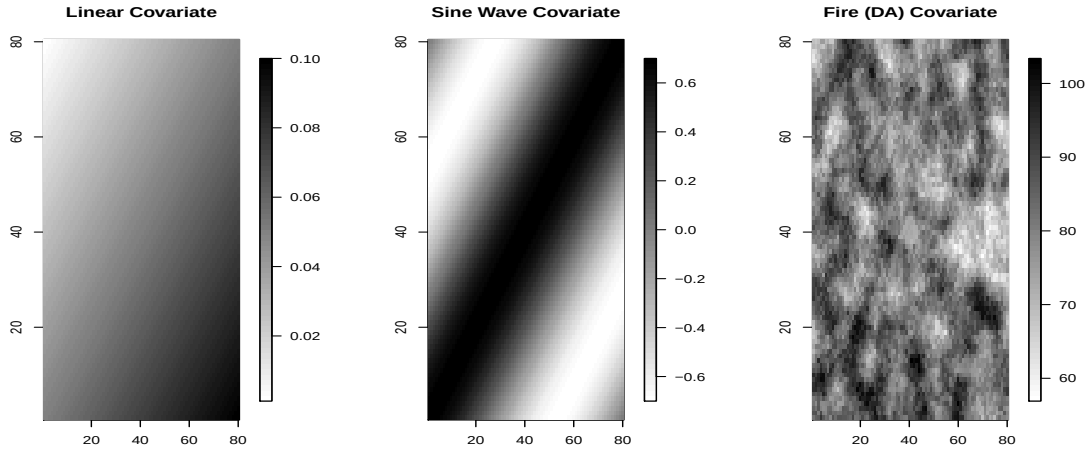
$$\mathrm{P}(z_i|\{z_j : j \in N_i\}, x_i) = \frac{\exp\{\alpha z_i + \theta \sum_{j \in N_i} z_i z_j + \beta z_i x_i\}}{1 + \exp\{\alpha + \theta \sum_{j \in N_i} z_j + \beta x_i\}}, \quad i = 1, \ldots, n. \tag{5.3}$$

Three different covariate types were considered for this case, including a diagonal linear covariate, a diagonal sine wave covariate, and, finally, a covariate corresponding to the departure from average (DA) variable of the Oregon and Washington fire data set. Note that this fire covariate was generated in R, through the *grf* package, using the estimated covariogram obtained from an arbitrary cross-sectional lattice of DA values from the aforementioned data set. For an $80x80$ lattice, the following functions were used to generate the linear and sine wave covariate lattices, where k denotes the row and l denotes the column of the lattice:

$$\begin{aligned} x_{kl} &= 0.05 * (k + 81 - l)/80 \\ x_{kl} &= 0.7 * \sin\{0.06 * (k + 81 - l) + 3\}. \end{aligned}$$

Figure 5.1 displays these three covariate types. For the linear covariate type, two parameter vectors were considered. In both, α was held fixed at -1.0 while $(\theta, \beta) = (0.5, 0.5)$ and $(0.2, 15.0)$. These values of β correspond to a relatively weak covariate effect (0.5) and a relatively strong covariate effect (15.0). The values of θ selected were such that the resultant lattices would yield a 50/50 split of zeros and ones. For the sine wave covariate type, two parameter vectors were also considered. In both, α was again held fixed at -1.0, but $(\theta, \beta) = (0.5, 0.3)$ and $(0.4, 3.0)$. These values of β also correspond to a relatively weak covariate effect (0.3) and a relatively strong covariate effect (3.0). The values of θ selected were again such that the resultant lattices would yield a 50/50 split of zeros and ones. Finally, for the fire covariate type, only one parameter vector was used: $(-8.6, 0.8, 0.065)$. This particular

Figure 5.1: Covariate Types



parameter vector was selected after maximum pseudolikelihood estimation was used to model the cross-sectional response lattice corresponding to the cross-sectional DA lattice used to create the fire covariate lattice given in Figure 5.1. Since an objective of this dissertation is to model the Oregon and Washington fire data using the four estimation methods of Chapter 3, we wanted to get an idea via simulation how these methods would perform on a similar lattice in which only approximately 5% of the sites have a value of one. The simulation procedure for this case of the study is also analogous, for all three covariate types, to the procedure for the first case, but only the MPLEs were used as the reference point for MCMCML.

5.2.5 First-Order Two-Way Anisotropy with a Covariate – FAhv4p

The fifth case considered in this simulation study corresponds to an autologistic model under a first-order two-way anisotropic neighborhood structure with a covariate term. This autologistic model has four parameters, $(\alpha, \theta_h, \theta_v, \beta)$, and has the following functional form:

$$\mathrm{P}(z_i|\{z_j : j \in N_i\}, x_i) = \frac{\exp\left\{\alpha z_i + \theta_v \sum_{j\in N_i^v} z_i z_j + \theta_h \sum_{j\in N_i^h} z_i z_j + \beta z_i x_i\right\}}{1 + \exp\left\{\alpha + \theta_v \sum_{j\in N_i^v} z_j + \theta_h \sum_{j\in N_i^h} z_j + \beta x_i\right\}}, \quad (5.4)$$

$i = 1, \ldots, n$, where θ_v and θ_h correspond to the north/south and east/west gradients, respectively. For this case, both the diagonal linear and diagonal sine wave covariates were again considered. For the linear covariate type, two parameter vectors were considered. In both, α was held fixed at -1.0 while $(\theta_h, \theta_v, \beta) = (0.3, 0.1, 14.0)$ and $(0.7, 0.3, 0.5)$. These values of β correspond to a relatively weak covariate effect (0.5) and a relatively strong covariate effect (14.0). For the sine wave covariate type, two parameter vectors were also considered. In both, α was again held fixed at -1.0, but $(\theta_h, \theta_v, \beta) = (0.4, 0.2, 3.0)$ and $(0.7, 0.3, 0.2)$. These values of β also correspond to a relatively weak covariate effect (0.2) and a relatively strong covariate effect (3.0). For both covariate types, the values of θ_h and θ_v selected were again such that the resultant lattices would yield a roughly 50/50 split of zeros and ones. The simulation procedure, for both covariate types, is also analogous to the procedure for the first case, but, as with the second and third cases, only groups of size 9 were considered for GPL and only the MPLEs were used as the reference point for MCMCML.

5.2.6 Second-Order Isotropy with a Covariate – SI3p

The sixth case considered in this simulation study corresponds to an autologistic model under a second-order isotropic neighborhood structure with a covariate term. This autologistic model has three parameters, (α, θ, β), and its functional form is identical to equation (5.3), where the neighborhood sets now consist of 8 sites rather than 4 sites. For this case, both the diagonal linear and diagonal sine wave covariates were again considered. For the linear covariate type, two parameter vectors were considered. In both, α was held fixed at -2.0 while $(\theta, \beta) = (0.3, 17.0)$ and $(0.5, 0.5)$. These values of β correspond to a relatively weak covariate effect (0.5) and a relatively strong covariate effect (17.0). For the sine wave covariate type, two parameter vectors were also considered. In both, α was again held fixed at -2.0, while $(\theta, \beta) = (0.45, 2.5)$ and $(0.5, 0.2)$. These values of β also correspond to a relatively weak covariate effect (0.2) and a relatively strong covariate effect (2.5). For both covariate types, the values of θ selected were again such that the resultant lattices would yield a roughly 50/50

split of zeros and ones. The simulation procedure for this case, for both covariate types, is also analogous to the procedure for the first case, but, as with the second, third, and fifth cases, only groups of size 9 were considered for GPL and only the MPLEs were used as the reference point for MCMCML.

5.2.7 The Ising Model and Phase Transition

The seventh and final case of this simulation study revisits the Ising model from case one (see equation 5.1). The objective of this case, however, is to explore the performance of the estimation methods as the parameter vector nears a critical value associated with phase transition ($\theta \approx 1.76$ ($\alpha = -2\theta$)) for this model. Eight different parameter vectors were considered for this case, all of which maintain the relationship $\alpha = -2\theta$: $(\alpha, \beta) = (-0.4, 0.2), (-0.8, 0.4), \ldots,$ $(-3.2, 1.6)$, where the last of these is approaching this critical value. The simulation procedure for this case is also analogous to the procedure for the first case, but only groups of size 9 were considered for GPL and only three types of estimates were used as the reference point for MCMCML, including the MPLEs, the MGPLEs for a group size of 9, and the MBGPLEs for a block size of 4×4.

5.3 GPL vs. BGPL – Question 1

To address the first question of interest, which seeks to determine which method, if either, between GPL or BGPL is "better," comparisons of precision, bias, and accuracy (using MAE), will be made between the estimates from both methods. Before we present the findings of the simulation study with respect to this first question, recall that the precision of an estimation procedure is a measure of its variability, while the accuracy of an estimation procedure is often assessed though an estimate of its bias in conjunction with its variance. Due to its resistance to outliers, the measure of accuracy used throughout this simulation study was the

mean absolute error (MAE), defined as:

$$MAE(\widehat{\boldsymbol{\theta}}) \quad = \quad \frac{1}{n}\sum_{i=1}^{n}|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0|,$$

where $\widehat{\boldsymbol{\theta}}$ is an estimator of the true parameter $\boldsymbol{\theta}_0$ and n is the number of samples generated in a given simulation.

Although clear differences between the estimates obtained from GPL and BGPL were discovered over the course of this simulation study, as will be demonstrated in the forthcoming discussion, two phenomena that are present in the estimates obtained from both procedures should first be addressed. In particular, the further a lattice deviates from a 50/50 split of zeros and ones, 1) the more variable and 2) the more biased both the MGPLEs and MBGPLEs generally become. For instance, consider Figure 5.2, which contains plots of the standard errors and of the estimated biases for the MGPLEs of α and θ, under a group size of 9, from Case 1. With respect to the first phenomenon, plots (a) and (c) clearly demonstrate that as the spatial dependence increases, especially above 0.5, the corresponding standard errors of $\hat{\alpha}$ and $\hat{\theta}$ increase, but to a lesser extent for larger lattices. Observe, however, that in both plots (a) and (c), it is evident that, as would be expected, the standard errors decrease as the lattice size increases. Finally, with respect to the second phenomenon, plots (b) and (d) clearly demonstrate that the estimated biases for $\hat{\alpha}$ and $\hat{\theta}$ move further away from 0 as the spatial dependence increases, but the magnitude of this deviation decreases as the lattice size increases. Note that these phenomena were analogously observed in every applicable case of the simulation study, not just in Case 1, and for every group size and block size considered, not just a group size of 9. It is worth mentioning here that these phenomena are consistent with the notion that the numerical efficiency of the autologistic model is optimal when the lattice does not contain a disproportionate number of ones or zeros ([2]).

While the above phenomena apply to both estimation methods, several observations can be made within each of the two estimation methods; i.e. important differences between the

MGPLEs for groups of size 5 and 9, as well as important differences between the MBGPLEs for the three different block sizes, exist. To this end, first consider Table 5.1 in which the average estimates and corresponding standard errors of the MGPLEs, under all three lattice sizes, are given for the parameters of Case 1 (α and θ). In particular, observe that the standard errors of both $\hat{\alpha}$ and $\hat{\theta}$ for a group size of 9 are uniformly less than or equal to the corresponding standard errors for a group size of 5. Furthermore, although there is no clear trend in the estimated biases for $\hat{\alpha}$, the estimated biases of $\hat{\theta}$ for a group size of 9 tend to be larger than the corresponding estimated biases for a group size of 5, especially for large $\boldsymbol{\theta}$. Hence, there appears to be some evidence of a variance/bias trade off among the MGPLEs as the group size is increased. Such evidence exists in every applicable case of the simulation study, not just in Case 1. Now consider Table 5.2 in which the average estimates and corresponding standard errors of the MBGPLEs, under all three lattice sizes, are given for the parameters of Case 1 (α and θ). In particular, note that the standard errors of both $\hat{\alpha}$ and $\hat{\theta}$ for a block size of 4×4 tend to be smaller, though not uniformly so, than the corresponding standard errors for either a block size of 3×3 or 2×2. Additionally, and perhaps surprisingly, the standard errors of both $\hat{\alpha}$ and $\hat{\theta}$ for a block size of 3×3 tend to be larger than the corresponding standard errors for a block size of 2×2. Furthermore, there is no clear trend in the estimated biases of $\hat{\alpha}$ and $\hat{\theta}$ among the three different block sizes. Hence, unlike the MGPLEs, there does not appear to be any evidence of a variance/bias trade off among the MBGPLEs as the block size is increased. Such findings are evident in every applicable case of the simulation study, not just in Case 1.

Having addressed phenomena present in both estimation methods, and having addressed differences within each method, a comparison of the two estimation procedures is now made. Since such a comparison is more easily conveyed visually, multiple plots within figures will be both given and discussed below. Although plots of the standard errors and estimated biases could be given for various cases of the simulation study to point out the major differences between the two methods, plots of the MAE are more revealing and will, therefore, be the

primary visual tool to compare GPL and BGPL. Specifically, plots of the relative MAE will be employed, rather than just plots of the MAE, since the broad range of parameter vectors considered in many cases of the simulation study makes it difficult to discern what are in fact clear differences in MAE between the two estimation procedures. To calculate relative MAE, the MAEs from each set of estimates are divided by their corresponding MAE from an arbitrarily selected set of estimates. To be precise, for all plots and figures used in this section, the arbitrarily selected set of estimates are those corresponding to the MGPLEs under a group size of 9. Thus, for each case of the simulation study, the MAEs for every set of estimates for a particular case are divided by the corresponding MAEs from the group size 9 set of estimates for that case. Note that relative MAE values greater than 1 indicate that the corresponding estimates performed worse than the group size 9 estimates, while relative MAE values less than 1 indicate that the corresponding estimates performed better than the group size 9 estimates; the relative MAE values for the group size 9 estimates will obviously be 1. For clarity, a relative MAE value of 1.10 will mean that the MAE of the corresponding estimator is 1.10 times as large as the MAE for the group size 9 estimator. The following comparison of GPL and BGPL is broken into two parts. The first part of the comparison addresses the cases of the simulation study that do not include a covariate term, while the second part of the comparison addresses the cases of the simulation study that do include a covariate term.

In all four cases of the simulation study that do not include a covariate term, there is clear evidence that GPL is superior to BGPL. To illustrate this point, consider Figure 5.3, which contains plots of the relative MAE values, for both $\hat{\alpha}$ and $\hat{\theta}$, from Case 1 of the simulation study. [For the interested reader, Tables A.1 and A.2 in section A.1 of Appendix A provide the actual MAE values for Case 1.] Now, in plots (a) and (b), which correspond to 26×26 lattices, no single set among the five sets of relative MAEs tends to be better than the others, although the relative MAEs for the group size 5 estimates tend to be the worst. Hence, these first two plots suggest that GPL, at least for a group size of 9, and BGPL are relatively

competitive. However, in plots (c) and (d), which correspond to 50×50 lattices, the relative MAEs for both sets of MGPLEs tend to be better than the relative MAEs for all three sets of MBGPLEs. Furthermore, in plots (e) and (f), which correspond to 74×74 lattices, the relative MAEs for both sets of MGPLEs are uniformly better than the relative MAEs for all three sets of MBGPLEs. Hence, these six plots clearly suggest that while BGPL is competitive with GPL for relatively small lattices, BGPL is inferior to GPL for larger lattices. Finally, as was alluded to above, such results were consistent across the other three cases of the simulation study that did not include a covariate term. In fact, the inferiority of BGPL is most extreme in set 3 of Case 3, where for the north/south gradient parameter, θ_v, the relative MAE for all three sets of MBGPLEs reaches values as large as 1.15. For the interested reader, a table of the average estimates and corresponding standard errors, as well as a figure of the relative MAEs, for set 3 of Case 3 of the simulation study are given in section A.1 (see Table A.3 and Figure A.1.

Before moving on to a comparison of GPL and BGPL for the three cases of the simulation study that do include a covariate term, a few additional observations should be pointed out from Figure 5.3. First, notice that in all six plots, the relative MAEs for the group size 5 estimates are *always* larger than those of the group size 9 estimates. Such results were evident in every applicable case of the simulation study, not just in Case 1. Hence, this suggests that increased group size does in fact lead to an improvement in performance among the MGPLEs; additionally, it suggests that within the variance/bias trade off mentioned earlier between the two group sizes, variability is the more dominant contributor to the MAE. Finally, in all six plots of Figure 5.3, and in particular plots (c)-(f), the relative MAEs for the 4×4 block size estimates tend to be the smallest of the three block sizes, but again somewhat surprisingly, the relative MAEs for the 3×3 block size estimates tend to be larger than those of the 2×2 block size estimates. Such results were again evident in every applicable case of the simulation study, not just in Case 1. Hence, despite the fact that the 4×4 block size estimates tended to perform the best of the MBGPLEs, there is no clear evidence from this simulation study

that increased block size necessarily leads to increased performance among the MBGPLEs. While we would like to explore the effect of increasing block (and group) size, it becomes prohibitively expensive once the block size becomes 5×5.

Finally, in the three cases of the simulation study that do include a covariate term, there is clear evidence that GPL still tends to be superior to BGPL in estimating the spatial parameters and the intercept parameter, but now BGPL tends to be superior to GPL in estimating the covariate parameter. To illustrate this point, consider Table 5.3, which provides the 74×74 lattice size relative MAE values for $\hat{\alpha}$, $\hat{\theta}_h$, $\hat{\theta}_v$, and $\hat{\beta}$ from Case 5, for the diagonal sine wave covariate type, of the simulation study. From this table it is clear that for both spatial parameters, θ_h and θ_v, the MGPLEs uniformly outperform the MBGPLEs, under all three block sizes. Furthermore, it is evident from the table that the MGPLEs tend to do better than the MBGPLEs in estimating α, the intercept parameter. However, it is also clear from the table that for the diagonal sine wave parameter, β, the MBGPLEs uniformly outperform the MGPLEs. Such results, for all covariate types and all lattice sizes, were typical among the three cases of the simulation study incorporating a covariate parameter into the model.

Table 5.1: Sample Means and Standard Errors of α and θ when the Model is FI2p and the Method of Estimation is GPL for Group Sizes of 5 (gs5) and 9 (gs9)

		26x26		50x50		74x74	
Model	Estimate	gs5	gs9	gs5	gs9	gs5	gs9
	$\bar{\hat{\alpha}}$	-1.0150	-1.0139	-1.0045	-1.0046	-0.9990	-0.9992
Ising	$\hat{SD}(\hat{\alpha})$	0.1904	0.1877	0.0901	0.0895	0.0612	0.0609
$(-1.0, 0.0)$	$\bar{\hat{\theta}}$	0.0021	0.0013	0.0010	0.0009	-0.0005	-0.0005
	$\hat{SD}(\hat{\theta})$	0.1612	0.1587	0.0721	0.0716	0.0471	0.0468
	$\bar{\hat{\alpha}}$	-1.0018	-1.0015	-1.0054	-1.0054	-1.0041	-1.0042
Ising	$\hat{SD}(\hat{\alpha})$	0.1938	0.1914	0.0983	0.0972	0.0616	0.0615
$(-1.0, 0.1)$	$\bar{\hat{\theta}}$	0.0993	0.0989	0.1021	0.1020	0.1022	0.1022
	$\hat{SD}(\hat{\theta})$	0.1421	0.1406	0.0731	0.0723	0.0445	0.0445
	$\bar{\hat{\alpha}}$	-0.9874	-0.9866	-0.9935	-0.9936	-0.9964	-0.9965
Ising	$\hat{SD}(\hat{\alpha})$	0.2009	0.1979	0.0991	0.0986	0.0659	0.0654
$(-1.0, 0.2)$	$\bar{\hat{\theta}}$	0.1931	0.1923	0.1977	0.1977	0.1986	0.1984
	$\hat{SD}(\hat{\theta})$	0.1332	0.1314	0.0674	0.0671	0.0446	0.0444
	$\bar{\hat{\alpha}}$	-1.0067	-1.0034	-1.0027	-1.0018	-1.0009	-1.0009
Ising	$\hat{SD}(\hat{\alpha})$	0.2210	0.2162	0.1055	0.1045	0.0670	0.0667
$(-1.0, 0.3)$	$\bar{\hat{\theta}}$	0.2983	0.2967	0.2982	0.2976	0.2986	0.2986
	$\hat{SD}(\hat{\theta})$	0.1351	0.1320	0.0648	0.0641	0.0416	0.0414
	$\bar{\hat{\alpha}}$	-0.9879	-0.9857	-1.0031	-1.0029	-1.0033	-1.0033
Ising	$\hat{SD}(\hat{\alpha})$	0.2165	0.2126	0.1094	0.1078	0.0726	0.0722
$(-1.0, 0.4)$	$\bar{\hat{\theta}}$	0.3916	0.3901	0.4020	0.4018	0.4013	0.4014
	$\hat{SD}(\hat{\theta})$	0.1174	0.1157	0.0585	0.0578	0.0391	0.0389
	$\bar{\hat{\alpha}}$	-0.9791	-0.9740	-0.9952	-0.9941	-0.9994	-0.9993
Ising	$\hat{SD}(\hat{\alpha})$	0.2494	0.2440	0.1172	0.1151	0.0795	0.0784
$(-1.0, 0.5)$	$\bar{\hat{\theta}}$	0.4873	0.4847	0.4968	0.4962	0.4992	0.4992
	$\hat{SD}(\hat{\theta})$	0.1204	0.1176	0.0554	0.0544	0.0378	0.0372
	$\bar{\hat{\alpha}}$	-0.9639	-0.9595	-1.0037	-1.0013	-1.0008	-0.9998
Ising	$\hat{SD}(\hat{\alpha})$	0.2936	0.2874	0.1446	0.1428	0.0929	0.0919
$(-1.0, 0.6)$	$\bar{\hat{\theta}}$	0.5863	0.5842	0.6016	0.6006	0.6005	0.6001
	$\hat{SD}(\hat{\theta})$	0.1243	0.1212	0.0615	0.0606	0.0391	0.0386
	$\bar{\hat{\alpha}}$	-0.9476	-0.9415	-0.9970	-0.9941	-0.9988	-0.9977
Ising	$\hat{SD}(\hat{\alpha})$	0.3712	0.3619	0.1787	0.1750	0.1153	0.1144
$(-1.0, 0.7)$	$\bar{\hat{\theta}}$	0.6833	0.6809	0.6989	0.6978	0.6995	0.6991
	$\hat{SD}(\hat{\theta})$	0.1366	0.1332	0.0670	0.0657	0.0435	0.0431
	$\bar{\hat{\alpha}}$	-0.9099	-0.9001	-0.9796	-0.9764	-0.9932	-0.9914
Ising	$\hat{SD}(\hat{\alpha})$	0.5109	0.4939	0.2512	0.2445	0.1626	0.1608
$(-1.0, 0.8)$	$\bar{\hat{\theta}}$	0.7745	0.7710	0.7936	0.7925	0.7979	0.7973
	$\hat{SD}(\hat{\theta})$	0.1692	0.1635	0.0839	0.0817	0.0545	0.0540
	$\bar{\hat{\alpha}}$	-0.8620	-0.8504	-0.9787	-0.9737	-0.9755	-0.9708
Ising	$\hat{SD}(\hat{\alpha})$	0.8181	0.7854	0.3576	0.3513	0.2448	0.2416
$(-1.0, 0.9)$	$\bar{\hat{\theta}}$	0.8639	0.8602	0.8951	0.8936	0.8931	0.8917
	$\hat{SD}(\hat{\theta})$	0.2497	0.2403	0.1103	0.1084	0.0747	0.0739
	$\bar{\hat{\alpha}}$	-0.7536	-0.7235	-0.9505	-0.9432	-0.9834	-0.9754
Ising	$\hat{SD}(\hat{\alpha})$	1.2098	1.1514	0.5140	0.5042	0.3452	0.3403
$(-1.0, 1.0)$	$\bar{\hat{\theta}}$	0.9362	0.9271	0.9882	0.9862	0.9967	0.9944
	$\hat{SD}(\hat{\theta})$	0.3452	0.3286	0.1495	0.1468	0.1009	0.0996

Table 5.2: Sample Means and Standard Errors of α and θ when the Model is FI2p and the Method of Estimation is BGPL for Block Sizes of 2x2 (bs2), 3x3 (bs3), and 4x4 (bs4)

		26x26			50x50			74x74		
Model	Estimate	bs2	bs3	bs4	bs2	bs3	bs4	bs2	bs3	bs4
	$\bar{\hat{\alpha}}$	-1.0140	-1.0097	-1.0127	-1.0036	-1.0044	-1.0044	-0.9992	-0.9990	-0.9994
Ising	$\hat{\text{SD}}(\hat{\alpha})$	0.1815	0.1819	0.1810	0.0910	0.0925	0.0924	0.0629	0.0623	0.0622
$(-1.0, 0.0)$	$\bar{\hat{\theta}}$	0.0005	-0.0034	-0.0007	0.0000	0.0008	0.0008	-0.0007	-0.0009	-0.0005
	$\hat{\text{SD}}(\hat{\theta})$	0.1597	0.1599	0.1589	0.0742	0.0740	0.0752	0.0492	0.0489	0.0486
	$\bar{\hat{\alpha}}$	-1.0050	-1.0029	-1.0052	-1.0045	-1.0040	-1.0061	-1.0039	-1.0037	-1.0044
Ising	$\hat{\text{SD}}(\hat{\alpha})$	0.1834	0.1857	0.1854	0.0973	0.0968	0.0983	0.0630	0.0644	0.0630
$(-1.0, 0.1)$	$\bar{\hat{\theta}}$	0.1020	0.1001	0.1018	0.1014	0.1010	0.1027	0.1019	0.1017	0.1023
	$\hat{\text{SD}}(\hat{\theta})$	0.1377	0.1417	0.1400	0.0738	0.0735	0.0746	0.0459	0.0472	0.0462
	$\bar{\hat{\alpha}}$	-0.9877	-0.9880	-0.9896	-0.9940	-0.9953	-0.9934	-0.9973	-0.9983	-0.9964
Ising	$\hat{\text{SD}}(\hat{\alpha})$	0.1999	0.2013	0.1992	0.0975	0.1001	0.0993	0.0683	0.0687	0.0695
$(-1.0, 0.2)$	$\bar{\hat{\theta}}$	0.1928	0.1932	0.1942	0.1978	0.1987	0.1974	0.1988	0.1996	0.1981
	$\hat{\text{SD}}(\hat{\theta})$	0.1356	0.1367	0.1351	0.0669	0.0694	0.0680	0.0468	0.0469	0.0475
	$\bar{\hat{\alpha}}$	-0.9978	-1.0067	-0.9976	-0.9997	-1.0020	-0.9997	-1.0003	-1.0009	-1.0008
Ising	$\hat{\text{SD}}(\hat{\alpha})$	0.2147	0.2218	0.2111	0.1057	0.1075	0.1044	0.0685	0.0688	0.0677
$(-1.0, 0.3)$	$\bar{\hat{\theta}}$	0.2940	0.2997	0.2939	0.2966	0.2981	0.2966	0.2981	0.2985	0.2985
	$\hat{\text{SD}}(\hat{\theta})$	0.1326	0.1374	0.1287	0.0654	0.0661	0.0641	0.0431	0.0429	0.0426
	$\bar{\hat{\alpha}}$	-0.9970	-0.9866	-0.9887	-1.0042	-1.0009	-1.0024	-1.0040	-1.0021	-1.0039
Ising	$\hat{\text{SD}}(\hat{\alpha})$	0.2104	0.2093	0.2034	0.1101	0.1101	0.1070	0.0746	0.0747	0.0750
$(-1.0, 0.4)$	$\bar{\hat{\theta}}$	0.3957	0.3899	0.3911	0.4025	0.4007	0.4015	0.4017	0.4007	0.4017
	$\hat{\text{SD}}(\hat{\theta})$	0.1171	0.1154	0.1138	0.0593	0.0595	0.0578	0.0401	0.0402	0.0405
	$\bar{\hat{\alpha}}$	-0.9794	-0.9764	-0.9805	-0.9933	-0.9935	-0.9952	-1.0015	-1.0008	-1.0002
Ising	$\hat{\text{SD}}(\hat{\alpha})$	0.2483	0.2447	0.2413	0.1173	0.1189	0.1179	0.0812	0.0818	0.0816
$(-1.0, 0.5)$	$\bar{\hat{\theta}}$	0.4873	0.4858	0.4877	0.4960	0.4961	0.4968	0.5003	0.4999	0.4997
	$\hat{\text{SD}}(\hat{\theta})$	0.1201	0.1177	0.1166	0.0556	0.0563	0.0561	0.0385	0.0389	0.0388
	$\bar{\hat{\alpha}}$	-0.9744	-0.9629	-0.9741	-1.0064	-1.0015	-1.0024	-1.0018	-1.0012	-1.0011
Ising	$\hat{\text{SD}}(\hat{\alpha})$	0.2907	0.2948	0.2927	0.1486	0.1469	0.1467	0.0965	0.0963	0.0937
$(-1.0, 0.6)$	$\bar{\hat{\theta}}$	0.5898	0.5852	0.5898	0.6029	0.6009	0.6013	0.6008	0.6007	0.6006
	$\hat{\text{SD}}(\hat{\theta})$	0.1230	0.1246	0.1232	0.0633	0.0625	0.0623	0.0407	0.0406	0.0395
	$\bar{\hat{\alpha}}$	-0.9587	-0.9494	-0.9512	-0.9957	-0.9941	-0.9919	-0.9982	-0.9981	-0.9959
Ising	$\hat{\text{SD}}(\hat{\alpha})$	0.3637	0.3566	0.3578	0.1774	0.1787	0.1760	0.1201	0.1187	0.1179
$(-1.0, 0.7)$	$\bar{\hat{\theta}}$	0.6866	0.6833	0.6839	0.6982	0.6977	0.6969	0.6993	0.6994	0.6986
	$\hat{\text{SD}}(\hat{\theta})$	0.1343	0.1312	0.1323	0.0670	0.0673	0.0662	0.0453	0.0448	0.0446
	$\bar{\hat{\alpha}}$	-0.9253	-0.9029	-0.9167	-0.9837	-0.9744	-0.9866	-0.9983	-0.9909	-0.9965
Ising	$\hat{\text{SD}}(\hat{\alpha})$	0.5027	0.4912	0.4875	0.2541	0.2521	0.2506	0.1697	0.1724	0.1682
$(-1.0, 0.8)$	$\bar{\hat{\theta}}$	0.7788	0.7714	0.7762	0.7951	0.7921	0.7960	0.7996	0.7972	0.7989
	$\hat{\text{SD}}(\hat{\theta})$	0.1666	0.1627	0.1614	0.0855	0.0846	0.0844	0.0571	0.0582	0.0567
	$\bar{\hat{\alpha}}$	-0.8920	-0.8916	-0.8704	-0.9865	-0.9810	-0.9717	-0.9716	-0.9723	-0.9696
Ising	$\hat{\text{SD}}(\hat{\alpha})$	0.8080	0.7756	0.7680	0.3677	0.3713	0.3673	0.2471	0.2573	0.2487
$(-1.0, 0.9)$	$\bar{\hat{\theta}}$	0.8725	0.8722	0.8660	0.8976	0.8959	0.8930	0.8920	0.8922	0.8913
	$\hat{\text{SD}}(\hat{\theta})$	0.2483	0.2409	0.2356	0.1132	0.1149	0.1127	0.0756	0.0789	0.0759
	$\bar{\hat{\alpha}}$	-0.7386	-0.7358	-0.7255	-0.9539	-0.9478	-0.9521	-0.9807	-0.9715	-0.9760
Ising	$\hat{\text{SD}}(\hat{\alpha})$	1.2029	1.1541	1.1495	0.5297	0.5304	0.5202	0.3518	0.3529	0.3454
$(-1.0, 1.0)$	$\bar{\hat{\theta}}$	0.9309	0.9300	0.9271	0.9897	0.9879	0.9891	0.9958	0.9932	0.9945
	$\hat{\text{SD}}(\hat{\theta})$	0.3433	0.3299	0.3295	0.1540	0.1543	0.1515	0.1030	0.1030	0.1011

Figure 5.2: Standard Errors and Estimated Biases of α and θ when the Model is FI2p and the Method of Estimation is GPL for a Group Size of 9 (gs9)

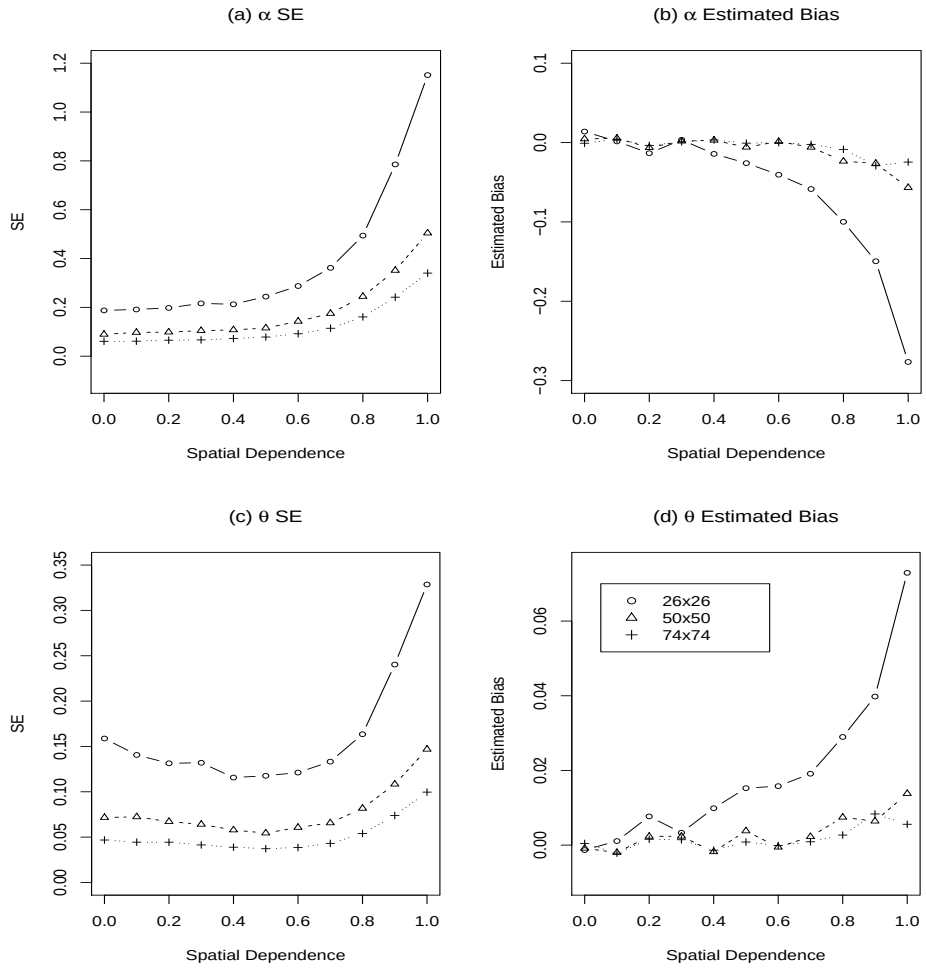


Figure 5.3: Relative MAEs of α and θ when the Model is FI2p and the Methods of Estimation are GPL for Group Sizes of 5 (gs5) and 9 (gs9) and BGPL for Block Sizes of 2x2 (bs2), 3x3 (bs3), and 4x4 (bs4)

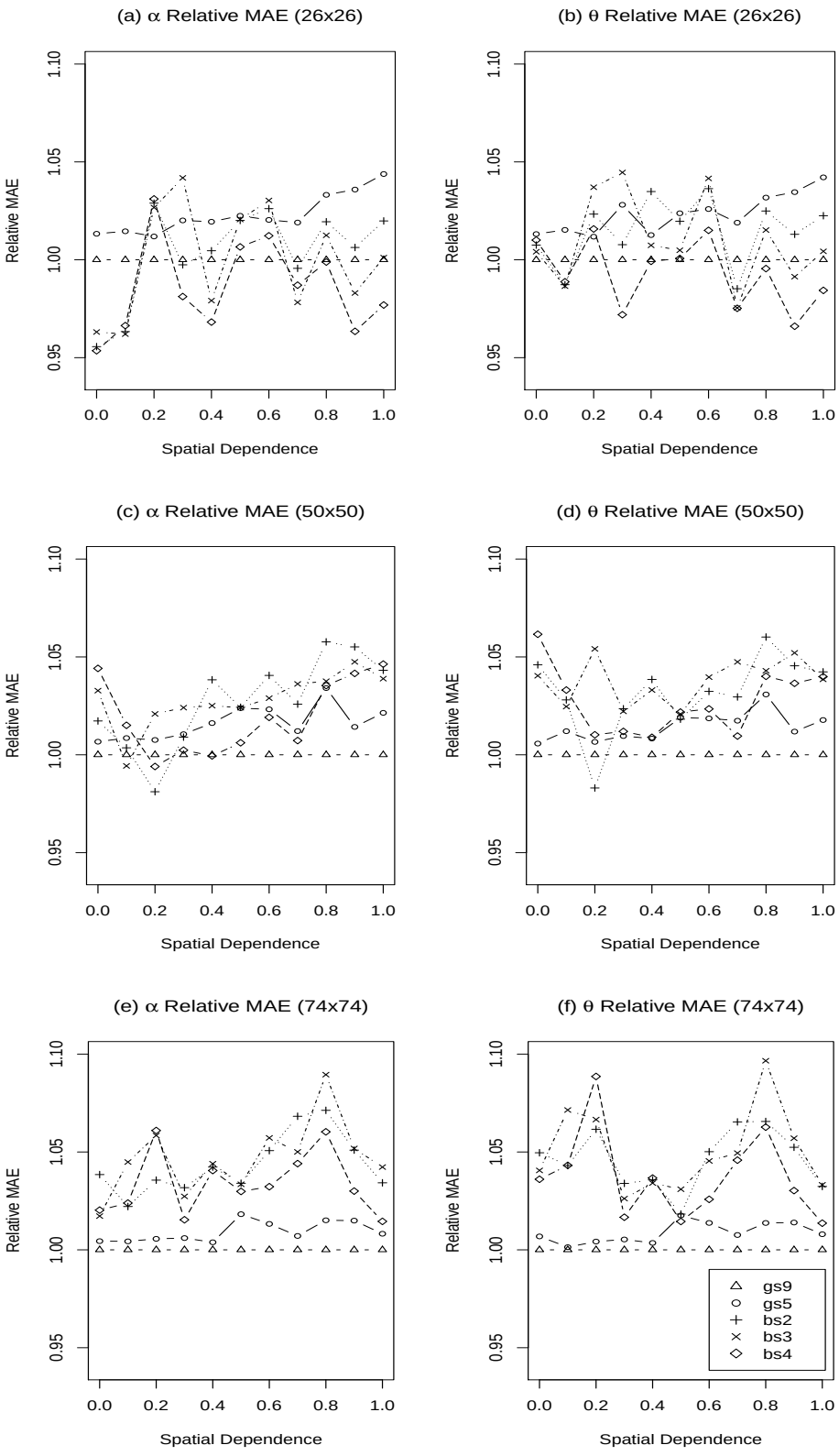


Table 5.3: Relative MAEs of α, θ_h, θ_v, and β (Sine Wave Covariate) when the Model is FAhv4p and the Methods of Estimation are GPL for a Group Size of 9 (gs9) and BGPL for Block Sizes of 2x2 (bs2), 3x3 (bs3), and 4x4 (bs4)

		74x74			
Model	Estimate	gs9	bs2	bs3	bs4
FAhv4p SineWave $(-1.0, 0.4, 0.2, 3.0)$	Rel. MAE($\hat{\alpha}$)	1	1.0004	0.9896	1.0031
	Rel. MAE($\hat{\theta}_h$)	1	1.0136	1.0299	1.0283
	Rel. MAE($\hat{\theta}_v$)	1	1.0458	1.0364	1.0485
	Rel. MAE($\hat{\beta}$)	1	0.9860	0.9847	0.9907
FAhv4p SineWave $(-1.0, 0.7, 0.3, 0.2)$	Rel. MAE($\hat{\alpha}$)	1	1.0221	1.0282	1.0207
	Rel. MAE($\hat{\theta}_h$)	1	1.0336	1.0327	1.0334
	Rel. MAE($\hat{\theta}_v$)	1	1.0549	1.0752	1.0490
	Rel. MAE($\hat{\beta}$)	1	0.9680	0.9689	0.9679

5.4 Reference Points for MCMCML – Question 2

The second question of interest for this simulation study addresses the performance of the MCMCMLEs under various reference point values. In particular, while use of the MPLEs as the reference point value for the Gibbs sampler is typically sufficient, it is of interest to know whether or not use of the MGPLEs and/or MBGPLEs as the reference point value will improve the performance of the MCMCMLEs, especially in situations where the difficulties of MCMCML are well-established. Such situations include autologistic models under extremely strong spatial correlation, as well as autologistic models near phase transition. As with question 1 in section 5.3, answers to this second question are pursued through comparisons of the precision, bias, and MAE of the MCMCMLEs obtained from all of the various reference point values. Based on the conclusions that can ultimately be drawn, such analysis is separated below into two parts. The first part briefly addresses the cases of the simulation study, for this particular question, that pertain to autologistic models under extremely strong spatial correlation, while the second part thoroughly addresses the one case of the simulation study that pertains to Ising models near phase transition.

For Case 1 of the simulation study, the MCMCMLEs were obtained using the MPLEs, the MGPLEs, under both group sizes, and the MBGPLES, under all three block sizes, as

the reference point values. The differences in the sample means, standard errors, and MAEs between the MCMCMLEs obtained under these six different sets of reference points were, for both α and θ, virtually nonexistent. In fact, considering the MCMCMLEs with the MPLEs as the reference point as the baseline case, all sets of relative MAEs, across all 11 parameter vectors and all three lattice sizes, were within 0.01 of 1 with no consistent patterns. In other words, the other reference point sets offered, at best, an MAE that was 0.99 times that of the baseline MAE and, at worst, an MAE that was 1.01 times that of the baseline MAE. The interested reader can find all such corresponding relative MAE plots in Figure A.2 of section A.2, where, for example, MCMCML(PL) denotes the MCMCMLEs with the MPLEs as the reference point. With such clear evidence that the MGPLEs and MBGPLEs are not better than the MPLEs as reference points when implementing MCMCML to estimate the parameters of an autologistic model under extremely strong spatial correlation, analogous comparisons were not conducted among the other applicable cases of the simulation study.

For Case 7 of the simulation study, which recall explores the Ising model as the parameter vector nears one of the critical values associated with phase transition ($\theta = 1.76$, $\alpha = -2\theta$), the results were not nearly as mundane as those for Case 1. Note that since the results from Case 1 indicated that there were no differences in the MCMCMLEs obtained under the six reference point sets, and because section 5.4 revealed 1) that the group size 9 MGPLEs were uniformly better than the group size 5 MGPLEs, and 2) that the block size 4×4 MBGPLEs tended to be better than both the block size 3×3 and 2×2 MBGPLEs, only three reference point sets were considered. In particular, the MPLEs, the group size 9 MGPLEs, and the block size 4×4 MBGPLEs were used as the three reference point sets for obtaining the MCMCMLEs. The sample means and corresponding standard errors of these three sets of MCMCMLEs are given below in Table 5.4, where, for example, MCMC(PL) indicates MCMCML with MPLE reference points. From the table it is evident that, with the exception of the two models closest to phase transition ($\theta = 1.4$ and $\theta = 1.6$), there are negligible differences between these three sets of MCMCMLEs. However, for these last two models there are differences

between the three sets of MCMCMLEs. The relative MAE plots given in Figure 5.4, where the MCMCMLEs with MPLE reference points serve as the baseline set, more clearly reveal these differences. From these six plots, there is some indication that when the group size 9 MGPLEs and the block size 4×4 MBGPLEs are used as reference points, they tend to produce better MCMCMLEs near phase transition. For instance, in plots (c) and (d) the MAEs corresponding to the MCMCMLEs with MGPLE and MBGPLE reference points are, for $\alpha = -3.2$ and $\theta = 1.6$, roughly 0.85 to 0.90 times the corresponding MAE for the MCMCMLEs with MPLE reference points. At the surface, such observations would seem to suggest that when dealing with models near phase transition, the MGPLEs and MBGPLEs might be better reference points than the MPLEs when employing MCMCML. A deeper analysis, however, reveals that such observations must be interpreted cautiously.

The above cautionary warning stems from the fact that there are vast discrepancies in the number of estimates used in obtaining the sample means, standard errors, and MAEs given in Table 5.4 and plotted in Figure 5.4. This is a result of the fact that for parameter values near phase transition, the Gibbs sampler used within the MCMCML algorithm frequently fails to converge, and, therefore, fails to produce an estimate. To quantify such convergence failures visually, consider plots (a), (c), and (e) of Figure 5.5, which display the proportion of runs (out of 500) that the MCMCML algorithm actually converged for the three different lattice sizes. From these plots, when $\theta = 1.6$, it can be seen that in lattices of size 74×74, for example, only roughly 45%, 55%, and 80% of the time did the MCMCML algorithm converge when the MPLEs, MBGPLEs, and MGPLEs, respectively, were used as the reference points. Furthermore, many of the estimates that were produced were extremely, i.e. orders of magnitude, unusual, and, thus, the $1.5 \times IQR$ rule was used to filter out such abberant values. Plots (b), (d), and (f) in Figure 5.4 display the proportion of runs (out of 500) for the three different lattice sizes that the MCMCML algorithm actually converged *and* the resultant estimate was not an outlier. From these plots, when $\theta = 1.6$, it can be seen that in lattices of size 74×74, for example, only roughly 35%, 50%, and 70% of the time did the MCMCML

algorithm converge and produce non-outlying estimates when the MPLEs, MBGPLEs, and MGPLEs, respectively, were used as the reference points. This perhaps suggests that, in a return on computational investment sense, the MGPLEs, and to a lesser extent the MBGPLEs, are better reference points than the MPLEs are for MCMCML. However, in a relative performance sense, measured through numerical quantities like the MAE, it is difficult to draw such firm conclusions. More specifically, because direct comparisons can only be made for the runs in which all three sets of reference points produced "reasonable" MCMCMLEs, the sample means, standard errors, and MAEs from the simulation study for the models closer to phase transition are based on substantially fewer runs/trials. For instance, when $\theta = 1.6$, at most 35% of the 500 runs are used in calculating the summary statistics. Therefore, caution should be exercised in inferring from Table 5.4 and Figure 5.4 that the MGPLEs and MBGPLEs tend to be better, in a relative error sense, reference points when employing MCMCML to estimate model parameters near phase transition. Finally, note that it may be because there was such a limited number of usable trials for these more extreme models that the standard errors in Table 5.4 surprisingly seem to decrease as the model nears phase transition. Perhaps these estimates were usable precisely because they fell within some relatively narrow window of the parameter space.

At this point, it would appear that not much can be said in way of an answer to question 2. However, if we realize that the underlying objective in answering question 2 is to ultimately determine if there is a better procedure for implementing MCMCML when the model is near phase transition, then a heuristic approach presents itself. More specifically, consider Table 5.5, which displays, for $\theta = 1.6$, the proportion of runs that either failed to converge or produced an outlying estimate under one set of reference points, but produced "reasonable" estimates according to one of the other sets of reference points. For example, the 0.6000 value in the MCMCML_{gs9} row and MCMCML_{PL} column indicates that 60% of the runs that either failed to converge or produced an outlying observation with the MPLEs as the reference point, produced "reasonable" estimates with the MGPLEs as the reference point.

While it is evident from the table that the MGPLEs have the highest, and the MPLEs the lowest, percentages as an alternative reference point when either one of the other two reference point options has failed, it is more important to note that all three sets of reference points, when used after one of the others has failed, had a decent chance of producing reasonable estimates. To elaborate on this point, consider Table 5.6, which displays the proportion of trials (out of 500) in which at least one of the three reference points produced a "reasonable" MCMCMLE. From this table we see that, for example, when $\alpha = -3.2$ and $\theta = 1.6$, 84.8% of the runs produced a usable estimate from at least one of the three sets of reference points. Compare this with the 35%, 50%, and 70% obtained individually for the MPLEs, MBGPLEs, and MGPLEs, respectively. Thus, for models near phase transition, parameter estimation using MCMCML can be improved, in the sense that there is a greater chance of obtaining a "reasonable" estimate, if all three reference point options are utilized. The interested reader should note that tables analogous to Tables 5.5 and 5.6, which are based only on whether the MCMCML algorithm converged, and not additionally on whether the resultant estimate was an outlier, are given in section A.2 and suggest the same heuristic indicated above (see Tables A.4 and A.5). It is worth mentioning that this lack of convergence could also be overcome using a suite of parameter vectors, but having these three computationally simpler estimation methods provides three informed estimates of the parameter vector.

Table 5.4: Sample Means and Standard Errors of α and θ when the Model is an Ising Model Near Phase Transition and the Method of Estimation is MCMCML Under the Following Reference Points: MPLE (PL), MGPLE for a Group Size of 9 (gs9), and MBGPLE for a Block Size of 4x4 (bs4)

		26x26			50x50			74x74		
Model	Estimate	MCMC (PL)	MCMC (gs9)	MCMC (bs4)	MCMC (PL)	MCMC (gs9)	MCMC (bs4)	MCMC (PL)	MCMC (gs9)	MCMC (bs4)
	$\hat{\alpha}$	-0.3990	-0.4002	-0.3991	-0.4019	-0.4012	-0.4012	-0.4020	-0.4019	-0.4018
Ising	$\hat{SD}(\hat{\alpha})$	0.2350	0.2348	0.2366	0.1160	0.1162	0.1160	0.0838	0.0838	0.0837
$(-0.4, 0.2)$	$\bar{\hat{\theta}}$	0.1985	0.1990	0.1987	0.1998	0.1994	0.1995	0.2002	0.2002	0.2001
	$\hat{SD}(\hat{\theta})$	0.1091	0.1089	0.1099	0.0549	0.0551	0.0549	0.0395	0.0395	0.0395
	$\hat{\alpha}$	-0.7816	-0.7816	-0.7825	-0.7978	-0.7982	-0.7973	-0.8034	-0.8030	-0.8032
Ising	$\hat{SD}(\hat{\alpha})$	0.2244	0.2233	0.2245	0.1097	0.1099	0.1100	0.0757	0.0760	0.0756
$(-0.8, 0.4)$	$\bar{\hat{\theta}}$	0.3924	0.3924	0.3929	0.3997	0.3999	0.3994	0.4020	0.4018	0.4019
	$\hat{SD}(\hat{\theta})$	0.1067	0.1062	0.1068	0.0521	0.0521	0.0522	0.0364	0.0367	0.0365
	$\hat{\alpha}$	-1.1704	-1.1711	-1.1711	-1.1940	-1.1942	-1.1937	-1.1973	-1.1976	-1.1973
Ising	$\hat{SD}(\hat{\alpha})$	0.2236	0.2229	0.2241	0.1124	0.1116	0.1125	0.0785	0.0786	0.0784
$(-1.2, 0.6)$	$\bar{\hat{\theta}}$	0.5852	0.5856	0.5855	0.5973	0.5973	0.5972	0.5995	0.5996	0.5996
	$\hat{SD}(\hat{\theta})$	0.1070	0.1065	0.1074	0.0535	0.0531	0.0534	0.0371	0.0372	0.0372
	$\hat{\alpha}$	-1.5590	-1.5601	-1.5588	-1.5909	-1.5910	-1.5908	-1.6008	-1.6009	-1.6008
Ising	$\hat{SD}(\hat{\alpha})$	0.2203	0.2209	0.2216	0.1080	0.1079	0.1078	0.0716	0.0716	0.0719
$(-1.6, 0.8)$	$\bar{\hat{\theta}}$	0.7808	0.7813	0.7808	0.7962	0.7963	0.7962	0.8007	0.8007	0.8007
	$\hat{SD}(\hat{\theta})$	0.1034	0.1032	0.1033	0.0507	0.0508	0.0509	0.0342	0.0342	0.0343
	$\hat{\alpha}$	-1.9667	-1.9672	-1.9677	-1.9943	-1.9941	-1.9952	-1.9945	-1.9939	-1.9943
Ising	$\hat{SD}(\hat{\alpha})$	0.1994	0.2001	0.2004	0.0945	0.0943	0.0942	0.0648	0.0639	0.0645
$(-2.0, 1.0)$	$\bar{\hat{\theta}}$	0.9841	0.9845	0.9847	0.9978	0.9978	0.9982	0.9976	0.9973	0.9975
	$\hat{SD}(\hat{\theta})$	0.0964	0.0963	0.0962	0.0458	0.0458	0.0456	0.0314	0.0309	0.0312
	$\hat{\alpha}$	-2.3435	-2.3475	-2.3425	-2.3916	-2.3906	-2.3906	-2.4002	-2.4007	-2.4009
Ising	$\hat{SD}(\hat{\alpha})$	0.1735	0.1731	0.1740	0.0941	0.0908	0.0917	0.0585	0.0570	0.0573
$(-2.4, 1.2)$	$\bar{\hat{\theta}}$	1.1739	1.1751	1.1733	1.1959	1.1955	1.1953	1.2000	1.2001	1.2002
	$\hat{SD}(\hat{\theta})$	0.0824	0.0819	0.0815	0.0446	0.0434	0.0431	0.0282	0.0275	0.0275
	$\hat{\alpha}$	-2.7344	-2.7345	-2.7323	-2.7876	-2.7903	-2.7876	-2.7946	-2.7940	-2.7948
Ising	$\hat{SD}(\hat{\alpha})$	0.1644	0.1536	0.1518	0.0829	0.0821	0.0816	0.0595	0.0560	0.0576
$(-2.8, 1.4)$	$\bar{\hat{\theta}}$	1.3672	1.3668	1.3661	1.3934	1.3949	1.3934	1.3974	1.3969	1.3971
	$\hat{SD}(\hat{\theta})$	0.0803	0.0760	0.0750	0.0406	0.0407	0.0399	0.0293	0.0275	0.0282
	$\hat{\alpha}$	-3.0642	-3.0567	-3.0548	-3.1667	-3.1763	-3.1578	-3.1753	-3.1861	-3.1801
Ising	$\hat{SD}(\hat{\alpha})$	0.1397	0.1247	0.1268	0.0953	0.0781	0.0726	0.0455	0.0480	0.0474
$(-3.2, 1.6)$	$\bar{\hat{\theta}}$	1.5258	1.5225	1.5214	1.5792	1.5809	1.5747	1.5855	1.5879	1.5852
	$\hat{SD}(\hat{\theta})$	0.0739	0.0652	0.0664	0.0476	0.0407	0.0379	0.0242	0.0278	0.0249

Table 5.5: Proportion of Trials Converging/Non-Outlying when the Model is an Ising Model Near Phase Transition with $(\alpha, \theta) = (-3.2, 1.6)$ and the Method of Estimation is MCMCML Under the Following Reference Points: MPLE (MCMCML$_{PL}$), MGPLE for a Group Size of 9 (MCMCML$_{gs9}$), and MBGPLE for a Block Size of 4x4 (MCMCML$_{bs4}$)

			Failed To Converge or was an Outlier For		
Lattice Size		Method	MCMCML$_{PL}$	MCMCML$_{gs9}$	MCMCML$_{bs4}$
26×26	But Converged or was not an Outlier For	MCMCML$_{PL}$	0	0.2549	0.3070
		MCMCML$_{gs9}$	0.6000	0	0.5658
		MCMCML$_{bs4}$	0.4456	0.3529	0
50×50	But Converged or was not an Outlier For	MCMCML$_{PL}$	0	0.2887	0.2987
		MCMCML$_{gs9}$	0.6854	0	0.6623
		MCMCML$_{bs4}$	0.4953	0.4507	0
74×74	But Converged or was not an Outlier For	MCMCML$_{PL}$	0	0.2573	0.3308
		MCMCML$_{gs9}$	0.5968	0	0.6353
		MCMCML$_{bs4}$	0.4349	0.4327	0

Table 5.6: Proportion of MCMCML Trials Converging as Non-Outliers for at Least One of the Three Reference Point Sets (MPLE, MGPLE for a Group Size of 9, or MBGPLE for a Block Size of 4x4) when the Model is an Ising Model Near Phase Transition

Model	26×26	50×50	74×74
Ising (-0.4,0.2)	0.994	0.986	0.996
Ising (-0.8,0.4)	0.998	0.984	0.986
Ising (-1.2,0.6)	0.990	0.990	0.988
Ising (-1.6,0.8)	0.996	0.986	0.990
Ising (-2.0,1.0)	0.994	0.992	0.990
Ising (-2.4,1.2)	0.990	0.996	0.994
Ising (-2.8,1.4)	0.982	0.998	0.992
Ising (-3.2,1.6)	0.832	0.884	0.848

Figure 5.4: Relative MAEs of α and θ when the Model is an Ising Model Near Phase Transition and the Method of Estimation is MCMCML Under the Following Reference Points: MPLE (MCMCML(PL)), MGPLE for a Group Size of 9 (MCMCML(gs9)), and MBGPLE for a Block Size of 4x4 (MCMCML(bs4))

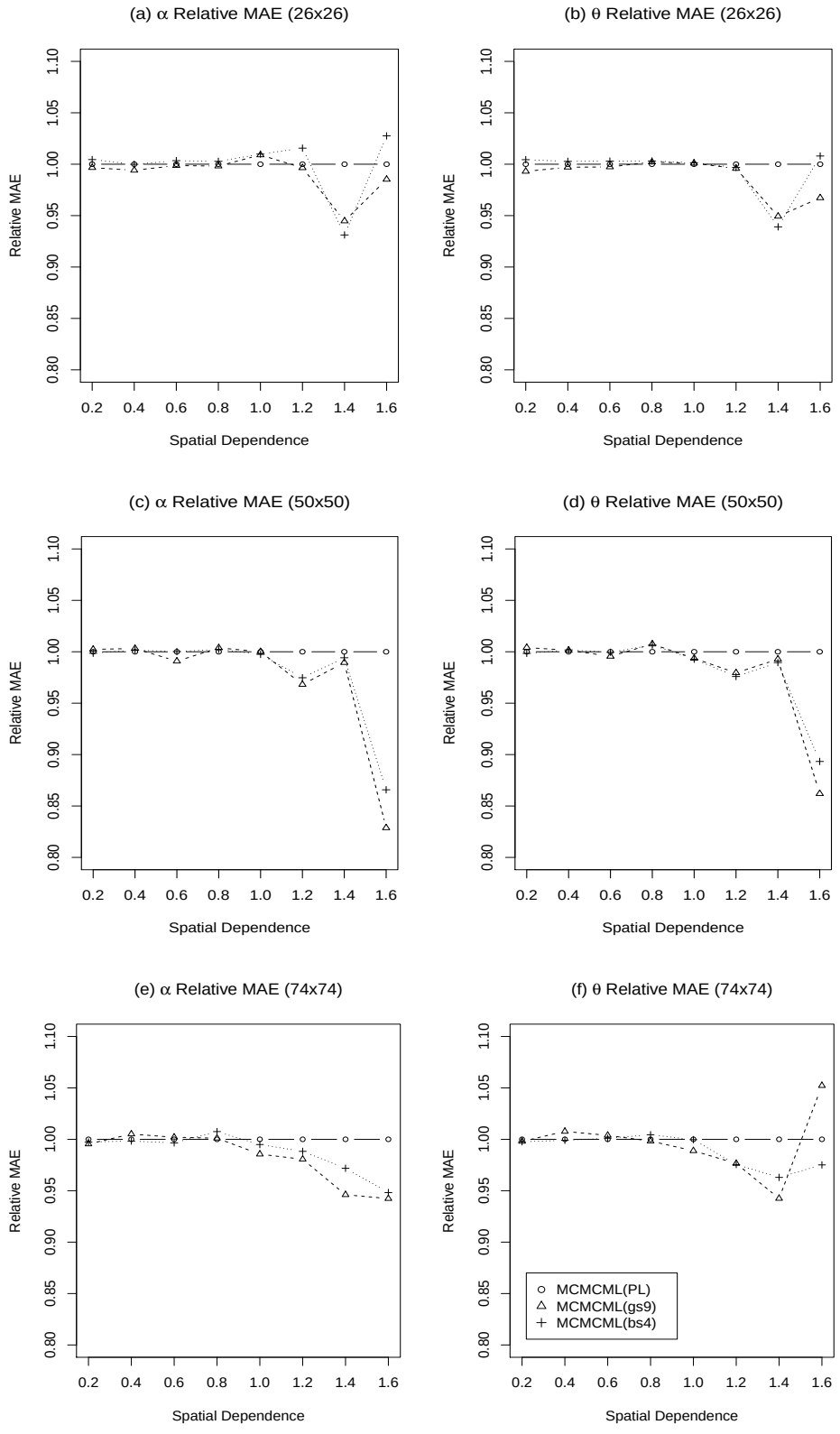
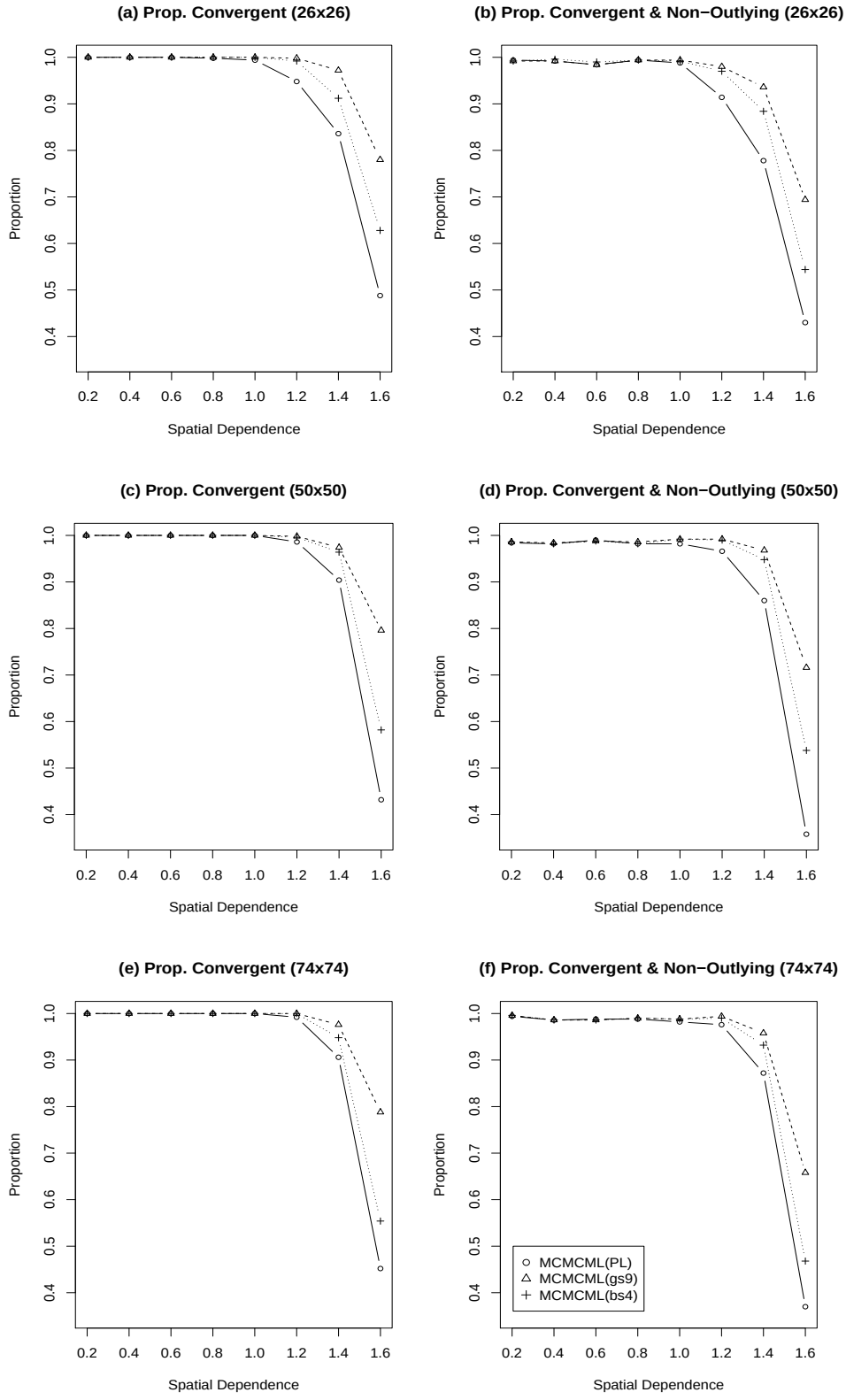


Figure 5.5: Proportion of Convergent Trials ((a), (c), (e)) and Proportion of Convergent and Non-Outlying Trials ((b), (d), (f)) when the Model is an Ising Model Near Phase Transition and the Method of Estimation is MCMCML Under the Following Reference Points: MPLE (MCMCML(PL)), MGPLE for a Group Size of 9 (MCMCML(gs9)), and MBGPLE for a Block Size of 4x4 (MCMCML(bs4))



5.5 PL vs. MCMCML vs. GPL vs. BGPL – Question 3

The third and final question of interest for this simulation study addresses the relative performances of the four different estimation methods. In particular, do the MGPLEs and/or MBGPLEs achieve the intended compromise between PL and MCMCML, and are there situations where one of the four methods is preferable to the other three? As with the first two questions covered in sections 5.3 and 5.4, answers to this third question are pursued through comparisons of the precision, bias, and MAE of the estimates from all four methods. Based on the conclusions that can ultimately be drawn, such analysis is separated below into three parts. The first part addresses, in detail, the initial three cases of the simulation study that deal with purely spatial autologistic models, i.e. models that contain only an intercept and one or more spatial parameters. The second part briefly addresses Cases 4 through 6 of the simulation study which deal with autologistic models with an intercept, one or more spatial parameters, and a covariate parameter. Finally, the third part thoroughly addresses the one case of the simulation study that pertains to Ising models near phase transition.

Before addressing each part in turn, however, note that since the block size 4×4 MBGPLEs tended to be the best of the three block sizes considered, and since any statements below for these MBGPLEs also typically hold in an analogous manner for the MBGPLEs corresponding to the two smaller block sizes, only the block size 4×4 MBGPLEs are formally compared below with the MPLEs, MGPLEs, and MCMCMLEs. Similarly, since the group size 9 MGPLEs were uniformly the best of the two group sizes considered, and since any statements below for these MGPLEs also typically hold in an analogous manner for the group size 5 MGPLEs, only the group size 9 MGPLEs are formally compared below with the MPLEs, MBGPLEs, and MCMCMLEs. Additionally, for the first two parts, the MCMCMLEs were obtained using the MPLEs as the reference point, but for the third part, the MCMCMLEs were obtained using a variety of reference points. The procedure for obtaining these "hybrid" MCMCMLEs, as they will be referenced here, will be discussed below at the appropriate time.

The first three cases of the simulation study revealed several trends in the relative performances of the four estimation methods. Consider Table 5.7, which provides the sample means and corresponding standard errors for the estimates obtained from each method for Case 2, set 3. From this table it is clear that for all four methods and for all three parameters, as the strength of the north/south spatial dependence increases (i.e. θ_v increases), both the standard errors and estimated biases also tend to increase. Furthermore, as would be expected, for all four methods and for all three parameters, as the lattice size increases, the estimated biases and corresponding standard errors tend to decrease. Such trends were present in all cases of the simulation study where the strength of the spatial dependence was incrementally increased, i.e. Cases 1 through 3. As for trends explaining how the methods differed, consider the relative (to PL) MAE plots for α, θ_h, and θ_v, corresponding to set 3 of Case 2, given in Figure 5.6. From this figure, the most obvious trend is that BGPL not only tends to fail in achieving the intended compromise between PL and MCMCML, but it is in fact often noticeably worse than PL. Among the 4 sets of MAEs corresponding to each of the spatial parameters, θ_h and θ_v, the MAEs for the MBGPLEs are almost uniformly the worst, with MAEs, in some instances, more than 1.10 times those of the MPLEs. As for GPL, in the instances where the differences in PL, GPL, and MCMCML were not negligible, GPL not only appears to achieve the intended compromise between PL and MCMCML, but at times, it is also competitive, or even better than, MCMCML. This tendency is more evident in the two larger lattice sizes, and more evident for $\hat{\alpha}$ and $\widehat{\theta_v}$, the correlation direction whose strength was sequentially increased. As these trends were evident across each of the first three cases of the simulation study, it appears that for purely spatial autologistic models, GPL tends to achieve its intended purpose while BGPL tends to fail to achieve its intended purpose.

Before moving on to the results of Cases 4 through 6, we should pause for a moment and note the implications of consolidating the knowledge we have just obtained with the knowledge we have from the previous simulation studies discussed in section 1.4. Such implications, I believe, are clearly synopsized in Figure 5.7, which gives the relative (to PL) MAE plots for

α and θ, corresponding to Case 1, for 74×74 lattices. In particular, when the strength of the spatial dependence is weak ($0 \leq \theta \leq 0.4$ in Figure 5.7), PL appears to be the practical choice for estimating the parameters of the autologistic model. However, when the strength of spatial dependence is moderate to strong ($0.4 \leq \theta \leq 0.9$ in Figure 5.7), either MCMCML or GPL should be used instead of PL. Note that while it appears that MCMCML is not really any better than GPL in these situations, it is possible that this is merely an artifact of needing to assuage the Gibbs sampler, in some form, to improve the estimates, such as, for example, using more sweeps or a longer burn-in. Finally, in situations where the strength of the spatial dependence is extremely strong ($\theta \geq 1.0$ in Figure 5.7), GPL is probably the best choice, or perhaps even just PL. Note that in such situations MCMCML tends to struggle. In fact, for the strongest spatial dependency parameter scenarios of Case 3, the MCMCMLEs had MAEs that were nearly 1.10 times the MAEs of the corresponding MPLEs. I suspect that the difficulties of MCMCML in these situations may be attributable to the Gibbs sampler, since its inherent use of randomness may lead it to produce "bad" realizations with increased probability. In any event, when the spatial dependence is extremely strong, the performance of any of these estimation methods will be less than desirable.

While several trends were evident in the purely spatial cases of the simulation study, for the cases of the simulation study incorporating a covariate parameter into the autologistic model (Cases 4 through 6), no such trends were clearly evident. More specifically, there were instances when GPL did achieve its intended compromise, and there where instances when it did not. An analogous statement also applies to BGPL in these cases. Perhaps the only trend in the results from these three cases was that the differences in the performances of the estimates between the four methods tended to be negligible. I suspect that this is largely attributable to the fact that the few parameter vectors considered in these cases were selected so that the proportion of ones/zeros across a lattice was roughly 0.50/0.50. As was inferred from the first three cases of the simulation study, differences in the methods are typically more apparent in cases of stronger spatial dependence, where the proportion of ones across

the lattice dominates the proportion of zeros. Hence, in the more "balanced" scenarios of Cases 4 through 6, the relative results of the estimates from the four methods are less consistent. In order for more general conclusions to be made, a more exhaustive simulation study involving covariate parameters will have to be conducted. Such a study will need to explore numerous parameter vectors in which the strength of the spatial dependence, as well as the strength of the covariate effect, are sequentially increased.

Before moving on to the results of the final case of the simulation study, it is of value to briefly discuss the results of the fire covariate scenario of Case 4. In particular, as it is the only scenario in Cases 4 through 6 in which the lattice exhibits very strong spatial dependence, and as the fire ignition data from Oregon/Washington, from which this covariate was constructed, are modeled in Chapter 7, it is of interest to note how the four estimation methods performed, both in general and relative to each other. To that end, while all four methods did produce average estimates that were "reasonable" (see Table A.6 in section A.3), especially as the lattice size was increased, there were differences in their relative performances. Table 5.8 contains the relative (to PL) MAEs for the four estimation methods. From this table, it is evident that GPL is only better than BGPL at estimating the spatial parameter for the largest lattice size, but BGPL is better than GPL, especially for the smallest lattice sizes, at estimating both the intercept and fire covariate parameters. Finally, it is also evident from the table that MCMCML uniformly produces the best estimates for the two larger lattice sizes, although PL is arguably competitive with MCMCML. Hence, to some extent, both GPL and BGPL fail to achieve the intended compromise between PL and MCMCML in this more extreme scenario.

For the final part of this section, we consider Case 7 of the simulation study in which the Ising model is considered as its parameter vector approaches one of the critical values associated with phase transition. It is in this case that the most pronounced trends and differences are found between the four estimation methods. Before presenting these results, however, the set of MCMCMLEs that were ultimately used for the following comparisons should be described.

In particular, since the number of available estimates in a set of MCMCMLEs obtained from any one type of reference point (MPLEs, MGPLEs, or MBGPLEs), as discussed in section 5.4, becomes extremely limited as the parameter vector nears phase transition, a hybrid set of estimates, related to the heuristic suggested at the end of section 5.4, was employed. For each of the 500 trials, the MCMCMLE obtained from the MPLE reference point was automatically in this hybrid set if 1) the Gibbs sampler converged and produced an estimate and 2) that estimate was not deemed an outlier, with respect to the MCMCMLEs with MPLE reference points, by the 1.5×IQR rule. If an estimate was not produced or it was deemed an outlier, then the corresponding MCMCMLE obtained from the MGPLE reference point was included in the hybrid set if it satisfied conditions 1) and 2) above, where an outlier is now determined with respect to the MCMCMLEs with MGPLE reference points. If an estimate was still either not produced or deemed an outlier, then the corresponding MCMCMLE obtained from the MBGPLE reference point was included in the hybrid set if it satisfied conditions 1) and 2) above, where an outlier is now determined with respect to the MCMCMLEs with MBGPLE reference points. If an estimate was still either not produced or deemed an outlier, then the hybrid set failed to have an estimate for that given trial. In other words, if an estimate satisfied (1) and (2) for at least one of the three reference values, it was included in the hybrid set. Table 5.9 displays the proportion (out of 500) of trials, for each parameter vector, used to compare the four estimation methods (PL, GPL, BGPL, and "hybrid"-MCMCML). From the table, we see that in all but the $(\alpha, \beta) = (-3.2, 1.6)$ case, at least 97% of the trials were used in the comparison. Furthermore, in the extreme case, $(\alpha, \beta) = (-3.2, 1.6)$, well over 80% of trials were used in the comparison, which is far larger than the roughly 35%, 50%, and 70% that would have been used if only one set of MCMCMLEs had been used rather than the hybrid set. Therefore, the clear advantage in this hybrid approach to MCMCML is that it allows for a clearer comparison of the four estimation methods since the overwhelming majority of trials are included in the evaluation.

With the description of how the MCMCMLEs were ultimately obtained for this comparison,

the aforementioned trends and differences can finally be presented. To that end, the table of sample means and corresponding standard errors for the estimates from each method are given in Table A.7 in section A.3 of Appendix A, but the clear message conveyed by this table is most succinctly expressed by the relative (to PL) MAE plots in Figure 5.8. From these plots it is clear that in the scenarios corresponding to the parameter vectors that are the furthest from the critical value (i.e. $\theta \leq 0.6$), PL appears to do as well as the other three estimation methods and thus is preferred due to its simplicity of use. However, as the parameter vector approaches the critical value ($\theta \geq 0.8$), GPL, BGPL, and MCMCML are all better than PL, and the amount of improvement increases as θ increases. In fact, for $\theta \geq 1.4$, all three methods have MAEs that are at most 0.80 of the corresponding MAEs for the MPLEs. Furthermore, both GPL and BGPL have achieved the intended compromise between PL and MCMCML, with the MGPLEs tending to perform slightly better than the MBGPLEs. Finally, the MCMCMLEs are almost uniformly best among the four sets of estimates. Hence, it appears that it is near phase transition that both GPL and BGPL not only achieve their intended compromise, but show their most substantial improvement relative to PL.

Table 5.7: Sample Means and Standard Errors of α, θ_h, and θ_v when the Model is FAhv3p and the Methods of Estimation are PL, GPL for a Group Size of 9 (gs9), BGPL for a Block Size of 4x4 (bs4), and MCMCML (26×26, 50×50, and 74×74 Lattices)

		26x26				50x50			
Model	Estimate	PL	gs9	bs4	MCMCML	PL	gs9	bs4	MCMCML
	$\hat{\alpha}$	-0.9924	-0.9767	-0.9856	-0.9715	-1.0079	-1.0035	-1.0040	-1.0037
	$\hat{\mathrm{SD}}(\hat{\alpha})$	0.3326	0.3236	0.3232	0.3257	0.1597	0.1541	0.1580	0.1545
FAhv3p	$\bar{\hat{\theta}}_h$	0.7086	0.6998	0.7117	0.6996	0.7056	0.7022	0.7106	0.7032
$(-1.0, 0.7, 0.6)$	$\hat{\mathrm{SD}}(\hat{\theta}_h)$	0.1864	0.1890	0.1963	0.1836	0.0912	0.0904	0.0944	0.0904
	$\bar{\hat{\theta}}_v$	0.5930	0.5907	0.5840	0.5865	0.6029	0.6025	0.5938	0.6020
	$\hat{\mathrm{SD}}(\hat{\theta}_v)$	0.1931	0.1914	0.2114	0.1901	0.0904	0.0905	0.0973	0.0895
	$\hat{\alpha}$	-0.9829	-0.9520	-0.9776	-0.9660	-0.9907	-0.9913	-0.9927	-0.9875
	$\hat{\mathrm{SD}}(\hat{\alpha})$	0.3483	0.3425	0.3430	0.3399	0.1718	0.1679	0.1721	0.1706
FAhv3p	$\bar{\hat{\theta}}_h$	0.6814	0.6691	0.6824	0.6757	0.6977	0.6980	0.6999	0.6978
$(-1.0, 0.7, 0.7)$	$\hat{\mathrm{SD}}(\hat{\theta}_h)$	0.1921	0.1943	0.2048	0.1900	0.0927	0.0934	0.0974	0.0942
	$\bar{\hat{\theta}}_v$	0.7080	0.6978	0.7033	0.7023	0.6957	0.6957	0.6947	0.6932
	$\hat{\mathrm{SD}}(\hat{\theta}_v)$	0.1927	0.1895	0.2144	0.1931	0.0963	0.0949	0.1069	0.0949
	$\hat{\alpha}$	-0.9839	-0.9547	-0.9821	-0.9525	-1.0025	-0.9975	-1.0021	-0.9938
	$\hat{\mathrm{SD}}(\hat{\alpha})$	0.4371	0.4312	0.4265	0.4377	0.2027	0.2013	0.1989	0.2004
FAhv3p	$\bar{\hat{\theta}}_h$	0.6894	0.6833	0.6879	0.6841	0.6987	0.6981	0.6959	0.6979
$(-1.0, 0.7, 0.8)$	$\hat{\mathrm{SD}}(\hat{\theta}_h)$	0.2259	0.2271	0.2404	0.2317	0.1080	0.1088	0.1129	0.1061
	$\bar{\hat{\theta}}_v$	0.8068	0.7942	0.8085	0.7914	0.8053	0.8019	0.8086	0.7998
	$\hat{\mathrm{SD}}(\hat{\theta}_v)$	0.2157	0.2119	0.2344	0.2169	0.1023	0.1012	0.1106	0.1027
	$\hat{\alpha}$	-0.9705	-0.9147	-0.9653	-0.9282	-0.9922	-0.9792	-0.9932	-0.9735
	$\hat{\mathrm{SD}}(\hat{\alpha})$	0.5177	0.5177	0.5084	0.5112	0.2554	0.2394	0.2545	0.2430
FAhv3p	$\bar{\hat{\theta}}_h$	0.6941	0.6743	0.6889	0.6806	0.6979	0.6935	0.6887	0.6910
$(-1.0, 0.7, 0.9)$	$\hat{\mathrm{SD}}(\hat{\theta}_h)$	0.2469	0.2526	0.2607	0.2487	0.1181	0.1153	0.1225	0.1179
	$\bar{\hat{\theta}}_v$	0.8957	0.8811	0.8994	0.8832	0.8997	0.8954	0.9111	0.8943
	$\hat{\mathrm{SD}}(\hat{\theta}_v)$	0.2426	0.2427	0.2624	0.2435	0.1221	0.1177	0.1298	0.1198
	$\hat{\alpha}$	-0.9365	-0.8860	-0.9331	-0.8740	-0.9905	-0.9741	-0.9972	-0.9758
	$\hat{\mathrm{SD}}(\hat{\alpha})$	0.6315	0.6223	0.6307	0.6413	0.2996	0.2841	0.2973	0.2846
FAhv3p	$\bar{\hat{\theta}}_h$	0.6827	0.6645	0.6683	0.6721	0.7019	0.6951	0.6926	0.6987
$(-1.0, 0.7, 1.0)$	$\hat{\mathrm{SD}}(\hat{\theta}_h)$	0.2808	0.2790	0.2959	0.2831	0.1480	0.1457	0.1547	0.1510
	$\bar{\hat{\theta}}_v$	0.9912	0.9792	1.0075	0.9629	0.9943	0.9911	1.0102	0.9888
	$\hat{\mathrm{SD}}(\hat{\theta}_v)$	0.2996	0.3043	0.3263	0.3120	0.1432	0.1409	0.1499	0.1408
	$\hat{\alpha}$	-0.8578	-0.7929	-0.8532	-0.8024	-0.9743	-0.9516	-0.9783	-0.9547
	$\hat{\mathrm{SD}}(\hat{\alpha})$	0.7355	0.7299	0.7294	0.7601	0.3697	0.3564	0.3678	0.3501
FAhv3p	$\bar{\hat{\theta}}_h$	0.6466	0.6344	0.6317	0.6251	0.6876	0.6798	0.6753	0.6793
$(-1.0, 0.7, 1.1)$	$\hat{\mathrm{SD}}(\hat{\theta}_h)$	0.3629	0.3550	0.3723	0.3731	0.1652	0.1602	0.1692	0.1628
	$\bar{\hat{\theta}}_v$	1.0825	1.0589	1.0991	1.0738	1.0998	1.0941	1.1170	1.0966
	$\hat{\mathrm{SD}}(\hat{\theta}_v)$	0.3310	0.3368	0.3439	0.3471	0.1610	0.1621	0.1664	0.1568

Table 5.7 (continued): Sample Means and Standard Errors of α, θ_h, and θ_v when the Model is FAhv3p and the Methods of Estimation are PL, GPL for a Group Size of 9 (gs9), BGPL for a Block Size of 4x4 (bs4), and MCMCML (26×26, 50×50, and 74×74 Lattices)

		74x74			
Model	Estimate	PL	gs9	bs4	MCMCML
FAhv3p $(-1.0, 0.7, 0.6)$	$\hat{\bar{\alpha}}$	-1.0017	-1.0005	-0.9998	-0.9993
	$\hat{SD}(\hat{\alpha})$	0.0984	0.0967	0.0978	0.0959
	$\hat{\bar{\theta}}_h$	0.7028	0.7028	0.7070	0.7026
	$\hat{SD}(\hat{\theta}_h)$	0.0583	0.0582	0.0603	0.0588
	$\hat{\bar{\theta}}_v$	0.5991	0.5983	0.5925	0.5974
	$\hat{SD}(\hat{\theta}_v)$	0.0608	0.0605	0.0651	0.0608
FAhv3p $(-1.0, 0.7, 0.7)$	$\hat{\bar{\alpha}}$	-0.9956	-0.9958	-0.9949	-0.9947
	$\hat{SD}(\hat{\alpha})$	0.1137	0.1096	0.1126	0.1096
	$\hat{\bar{\theta}}_h$	0.6983	0.6992	0.6989	0.6982
	$\hat{SD}(\hat{\theta}_h)$	0.0642	0.0638	0.0684	0.0640
	$\hat{\bar{\theta}}_v$	0.6984	0.6978	0.6974	0.6979
	$\hat{SD}(\hat{\theta}_v)$	0.0666	0.0645	0.0736	0.0643
FAhv3p $(-1.0, 0.7, 0.8)$	$\hat{\bar{\alpha}}$	-1.0018	-0.9972	-1.0037	-0.9944
	$\hat{SD}(\hat{\alpha})$	0.1413	0.1339	0.1407	0.1363
	$\hat{\bar{\theta}}_h$	0.6976	0.6974	0.6944	0.6964
	$\hat{SD}(\hat{\theta}_h)$	0.0728	0.0725	0.0759	0.0728
	$\hat{\bar{\theta}}_v$	0.8047	0.8016	0.8099	0.8005
	$\hat{SD}(\hat{\theta}_v)$	0.0758	0.0727	0.0809	0.0731
FAhv3p $(-1.0, 0.7, 0.9)$	$\hat{\bar{\alpha}}$	-0.9925	-0.9861	-0.9956	-0.9873
	$\hat{SD}(\hat{\alpha})$	0.1692	0.1648	0.1674	0.1619
	$\hat{\bar{\theta}}_h$	0.6993	0.6963	0.6936	0.6966
	$\hat{SD}(\hat{\theta}_h)$	0.0801	0.0799	0.0817	0.0785
	$\hat{\bar{\theta}}_v$	0.8964	0.8954	0.9054	0.8958
	$\hat{SD}(\hat{\theta}_v)$	0.0795	0.0789	0.0845	0.0797
FAhv3p $(-1.0, 0.7, 1.0)$	$\hat{\bar{\alpha}}$	-0.9985	-0.9877	-1.0028	-0.9875
	$\hat{SD}(\hat{\alpha})$	0.2028	0.1908	0.2008	0.1959
	$\hat{\bar{\theta}}_h$	0.7020	0.6977	0.6937	0.6978
	$\hat{SD}(\hat{\theta}_h)$	0.0970	0.0942	0.0993	0.0955
	$\hat{\bar{\theta}}_v$	0.9979	0.9953	1.0105	0.9954
	$\hat{SD}(\hat{\theta}_v)$	0.0919	0.0895	0.0929	0.0906
FAhv3p $(-1.0, 0.7, 1.1)$	$\hat{\bar{\alpha}}$	-0.9864	-0.9764	-0.9900	-0.9756
	$\hat{SD}(\hat{\alpha})$	0.2403	0.2329	0.2397	0.2344
	$\hat{\bar{\theta}}_h$	0.6909	0.6878	0.6834	0.6861
	$\hat{SD}(\hat{\theta}_h)$	0.1078	0.1055	0.1085	0.1071
	$\hat{\bar{\theta}}_v$	1.1029	1.1004	1.1143	1.1014
	$\hat{SD}(\hat{\theta}_v)$	0.1071	0.1042	0.1096	0.1046

Table 5.8: Relative MAEs of α, θ, and β (Fire Covariate) when the Model is FI3p and the Methods of Estimation are PL, GPL for a Group Size of 9 (gs9), BGPL for a Block Size of 4x4 (bs4), and MCMCML

		26x26			
Model	Estimate	PL	gs9	bs4	MCMCML
FI3p	Rel MAE($\hat{\alpha}$)	1.0000	1.0836	0.9846	1.0102
(-8.6,0.8,0.065)	Rel MAE($\hat{\theta}$)	1.0000	1.0245	0.9405	0.9443
	Rel MAE($\hat{\beta}$)	1.0000	1.0757	0.9836	1.0089
		50x50			
Model	Estimate	PL	gs9	bs4	MCMCML
FI3p	Rel MAE($\hat{\alpha}$)	1.0000	1.0269	1.0003	0.9970
(-8.6,0.8,0.065)	Rel MAE($\hat{\theta}$)	1.0000	0.9816	0.9772	0.9675
	Rel MAE($\hat{\beta}$)	1.0000	1.0278	1.0017	0.9993
		74x74			
Model	Estimate	PL	gs9	bs4	MCMCML
FI3p	Rel MAE($\hat{\alpha}$)	1.0000	1.0201	1.0063	0.9901
(-8.6,0.8,0.065)	Rel MAE($\hat{\theta}$)	1.0000	0.9489	0.9880	0.9343
	Rel MAE($\hat{\beta}$)	1.0000	1.0147	1.0005	0.9920

Table 5.9: Proportion of Trials (out of 500) Used in Comparisons of MPLEs, MGPLEs, MBGPLEs, and MCMCMLEs when the Model is an Ising Model Near Phase Transition

Model	26×26	50×50	74×74
Ising (-0.4,0.2)	0.994	0.986	0.996
Ising (-0.8,0.4)	0.996	0.984	0.986
Ising (-1.2,0.6)	0.990	0.990	0.988
Ising (-1.6,0.8)	0.992	0.984	0.990
Ising (-2.0,1.0)	0.990	0.990	0.990
Ising (-2.4,1.2)	0.986	0.994	0.994
Ising (-2.8,1.4)	0.976	0.998	0.988
Ising (-3.2,1.6)	0.828	0.882	0.848

Figure 5.6: Relative MAEs of α, θ_h, and θ_v when the Model is FAhv3p and the Methods of Estimation are PL, GPL for a Group Size of 9 (gs9), BGPL for a Block Size of 4x4 (bs4), and MCMCML

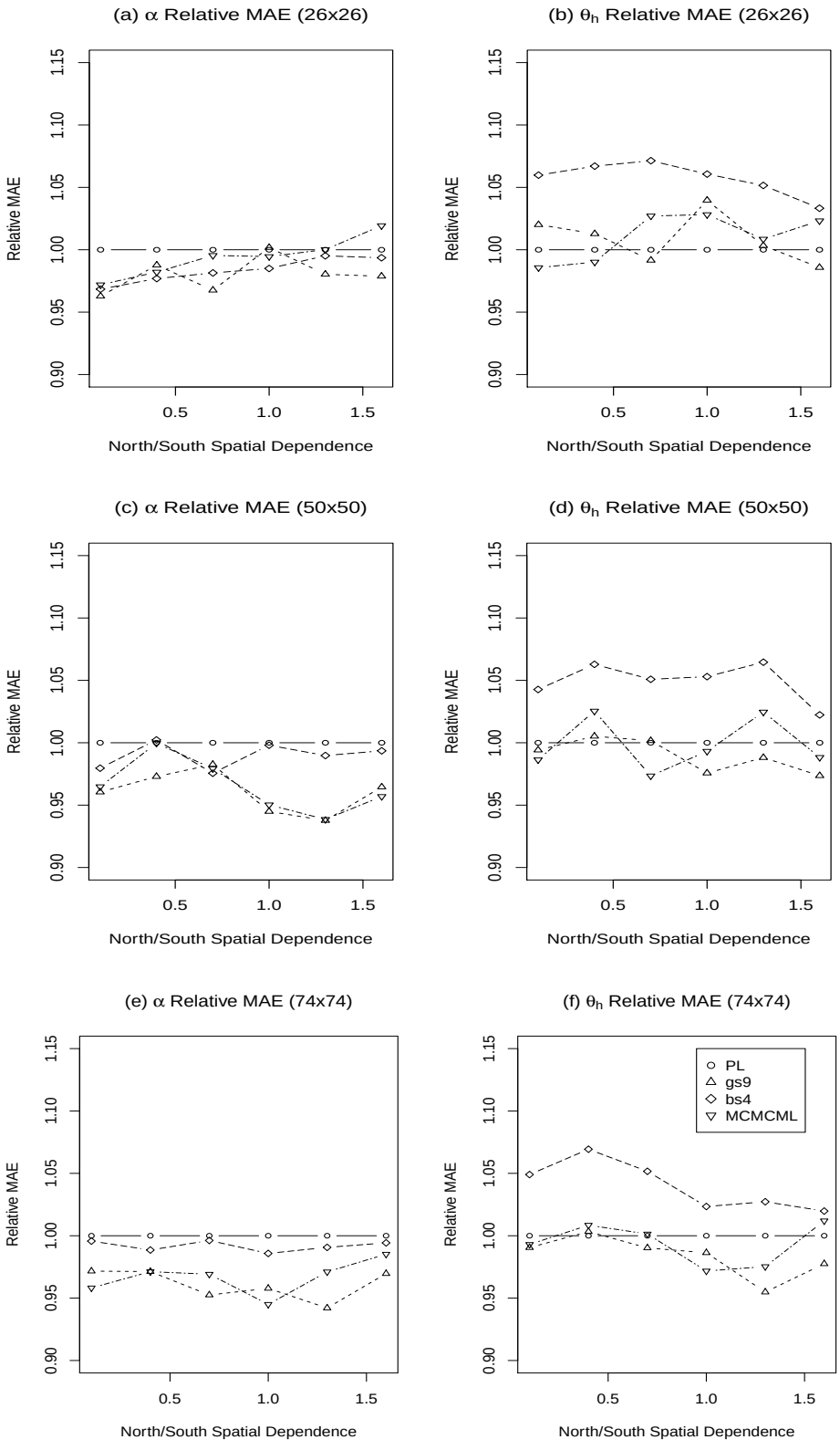


Figure 5.6 (continued): Relative MAEs of α, θ_h, and θ_v when the Model is FAhv3p and the Methods of Estimation are PL, GPL for a Group Size of 9 (gs9), BGPL for a Block Size of 4x4 (bs4), and MCMCML

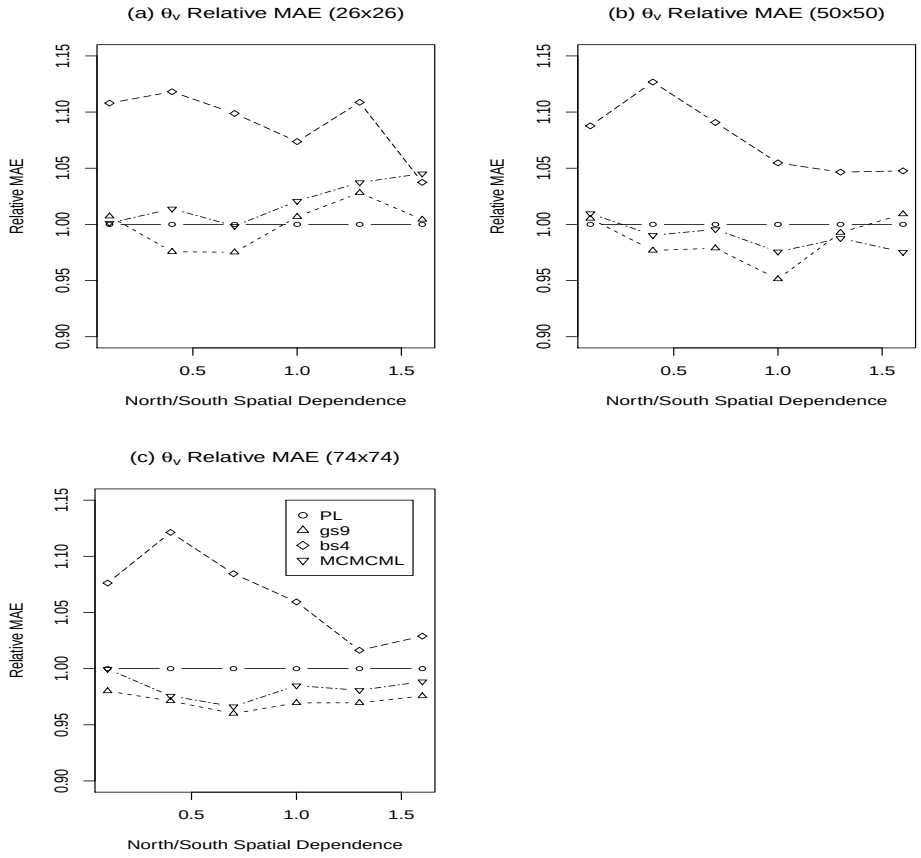
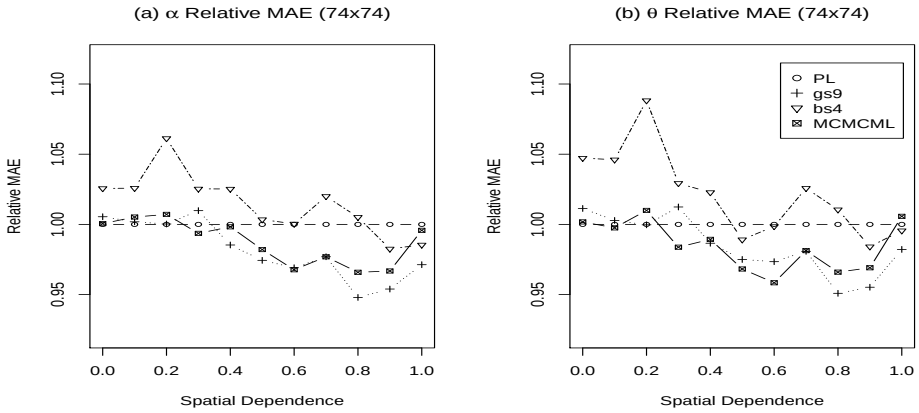


Figure 5.7: Relative MAEs of α and θ when the Model is FI2p and the Methods of Estimation are PL, GPL for a Group Size of 9 (gs9), BGPL for a Block Size of 4x4 (bs4), and MCMCML

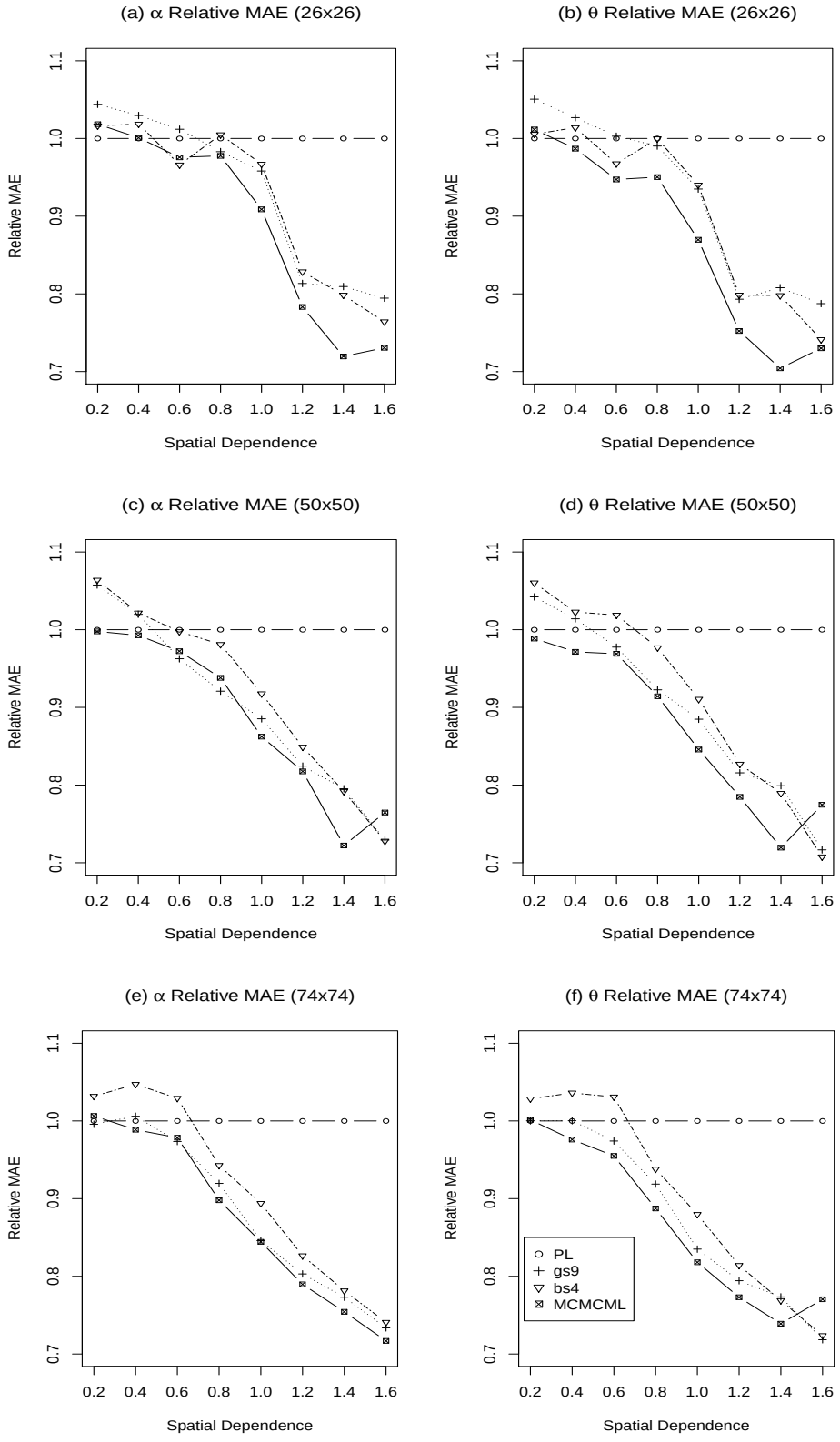


5.6 Concluding Remarks

With respect to the first question posed at the beginning of this chapter, direct comparisons of GPL and BGPL through the seven cases of this simulation study clearly revealed that GPL is better than BGPL for purely spatial autologistic models with an intercept term. However, for autologistic models that incorporate a covariate term, this simulation study also revealed that BGPL tended to outperform GPL with respect to estimating the covariate parameter (at least for the three cases considered), but not the spatial parameters or the intercept parameter. Hence, we are left with a lingering question: Does this caveat hold for autologistic models with more than one covariate parameter? In other words, will BGPL continue to outperform GPL in estimating all of the covariate parameters, but none of the spatial parameters or intercept parameter? Such a question was not explored in this dissertation, but could be the focus of further simulation studies.

As for the second question posed at the beginning of this chapter, direct comparisons of the MCMCMLEs produced under various reference points for autologistic models whose parameters are notoriously difficult to estimate revealed a wide spectrum of conclusions. For models

Figure 5.8: Relative MAEs of α and θ when the Model is an Ising Model Near Phase Transition and the Methods of Estimation are PL, GPL for a Group Size of 9 (gs9), BGPL for a Block Size of 4x4 (bs4), and MCMCML



corresponding to extremely strong spatial correlation, virtually no differences were found to exist between the various MCMCMLEs. However, for models near phase transition, inconsistencies were found to exist in the MCMCMLEs produced from different reference points. Analysis of these inconsistencies suggested, without convincing evidence, that perhaps the MGPLEs, and to a lesser extent the MBGPLEs, served as better reference points than the corresponding MPLEs. While no conclusive results to this end were obtained, further analysis did suggest a heuristic approach to improving MCMCML for use with autologistic models near phase transition. In particular, if all three reference points are separately implemented, then there is a reasonably high probability of obtaining a useful parameter estimate. Thus, this heuristic procedure offers an approach to obtaining, with relatively high probability, reasonable MCMCMLEs in situations were the method is known to struggle. Such a procedure is certainly an improvement over using MCMCML with only the MPLEs as the reference point.

Finally, with respect to the last question posed at the beginning of this chapter, comparisons of the relative performances of the four estimation methods revealed some interesting results. In particular, for the purely spatial autologistic model simulations, BGPL generally did not achieve the intended compromise between PL and MCMCML and it was in fact often the worst of the four methods. On the other hand, GPL either achieved the intended compromise between PL and MCMCML, or its performance was competitive with MCMCML and PL when these two methods were negligibly different. In the simulations where a covariate term was included in the autologistic model, neither GPL nor BGPL consistently succeeded or failed in achieving the intended compromise. Furthermore, in these cases the four methods generally produced estimates that were negligibly different when precision, bias, and accuracy were compared. Ultimately, these cases revealed that a more extensive simulation study involving covariates is needed to more conclusively establish the relative performances of the four methods. Finally, for the simulations dealing with the Ising model as its parameter vector nears the critical value associated with phase transition, the most pronounced results were obtained. Both GPL and BGPL achieved the intended compromise between PL and

MCMCML, with GPL consistently performing slightly better than BGPL. Additionally, it was also clearly evident that the closer the Ising model moves toward phase transition, the greater the improvement GPL, BGPL, and MCMCML all have relative to PL. Such results for the MGPLEs of Ising model parameters near phase transition are consistent with the results obtained from the simulation studies conducted by Huang and Ogata ([22]) and Sherman et al. ([32]).

Chapter 6

Spatio-Temporal Autologistic Model

6.1 Introduction

The preceding chapters have primarily focused on binary Markov random fields corresponding to, or observed at, a single instance in time. Often, however, such binary MRFs are repeatedly observed at a discrete number of equally spaced time points. In these situations, the sites of such a space-time MRF may exhibit both spatial and temporal dependencies; i.e. we expect sites that are relatively close together in both space and time to have more similar response values than sites that are relatively far apart in space and/or time. Furthermore, although it is assumed otherwise for this dissertation, it is even likely that the strength and/or type of spatial dependency may depend on the temporal location of the MRF, and vice versa. In other words, the effects of space and time on the binary response variables of the MRF may interact.

Fortunately, the purely spatial autologistic model has an intuitive extension into the space-time domain, but this spatio-temporal autologistic model still possesses an intractable normalizing constant that makes exact maximum likelihood methods either impractical or impossible.

Thus, the primary objective of this chapter is to present PL, MCMCML, GPL, and BGPL under the space-time paradigm, rather than the purely spatial paradigm of Chapter 3, so that their implementation with the spatio-temporal autologistic model is tractable. It should be noted that while (1) PL's use in the space-time domain is already established ([37]) and (2) MCMCML's initial theoretical development naturally permits its use in the space-time domain ([15]), the development of GPL and BGPL in the space-time domain, as presented in this chapter, is novel. More specifically, both methods were originally proposed in only the purely spatial domain, as presented in sections 3.4 and 3.5, and as far as I am aware, their use in the (three-dimensional) space-time domain, to date, has been neither developed nor implemented in the literature.

Let us pause for a moment here and note several things. First, both in section 1.5 and the preceding paragraphs of this chapter, the term space-time MRF has been used as if its meaning were intuitively understood. This is because the definition given for a MRF in Chapter 2 (Definition 2.3), under a purely spatial domain paradigm, also holds for a space-time MRF, except that now the neighborhood structure is three-dimensional, rather than two-dimensional. More will be said about such spatio-temporal neighborhood structures further below. Second, as discussed in section 1.5, we will assume, based on the data set being modeled in Chapter 7, that (1) the type of temporal dependency exhibited by the space-time MRF adheres to a non-absorbing state and (2) the temporal neighborhood structure of the space-time MRF is autoregressive (with past-lags only) in nature. Recall that the second assumption stipulates that only values of the response variable observed in the past can be temporally conditioned on to help model the current value of the response variable, and, thus, future values of the response variable cannot be temporally conditioned on to help model the current value of the response variable. Third and finally, we will assume that the parameters used in modeling the space-time MRF are separable and space and time invariant. Recall that while the separability of space and time means that their respective effects on the response variable do not interact, the invariance of space and time means that their respective effects

are the same across all spatial locations and all time steps (i.e. the process is stationary in space and time).

As the notation used throughout this chapter is predominantly in accordance with the notation of Chapter 3, in general only the notation that is unique to this chapter, or altered from Chapter 3, will be formally established. Furthermore, when considering space-time MRFs, edge effects exist not only in space, but also in time. The spatial edge effects can still be handled as in previous chapters of this dissertation, and the resulting temporal edge effects can be handled in a manner analogous to the spatial edge effects. For instance, a guard-region can be employed in which all of the observations from the first t (where the value of t depends on the temporal neighborhood structure) time steps are treated as fixed, or a toroidal temporal edge adjustment can be employed in which the initial time steps have the later time steps as temporal neighbors. For ease of presentation, edge effects (both spatial and temporal) will be ignored while presenting/developing the space-time extensions of the four estimation methods within this chapter.

The remainder of this chapter is divided into six sections. In section 6.2, the two-dimensional spatial MRF of Chapter 3 is generalized to a three-dimensional space-time MRF. Additionally, the spatio-temporal autologistic model is formally presented. In sections 6.3-6.6, the respective methods of PL, MCMCML, GPL, and BGPL are presented/developed and then exemplified with the spatio-temporal autologistic model. Finally, in section 6.7, the results of a small-scale simulation study evaluating and comparing the performances of these methods in the space-time domain for the autologistic model are presented.

6.2 The Spatio-Temporal Autologistic Model

As in Chapter 3, let $D = \{(i, j, t, \ldots) : i = 1, 2, \ldots; j = 1, 2, \ldots; t = 1, 2, \ldots; \ldots\}$ be an infinite dimensional square lattice, where $(i, j, t, \ldots)$ denotes site $(i, j, t, \ldots)$ on the lattice.

Unlike in Chapter 3, however, S will now denote a finite three-dimensional (rather than two-dimensional) subset of D. To facilitate a more tractable development, S will be referred to as a three-dimensional array, rather than lattice, with its three dimensions corresponding to the number of rows, columns, and lattices, respectively. Hence, $S = \{(i,j,t) : i = 1,\ldots,m_r; j = 1,\ldots,m_c; t = 1,\ldots,m_T\} \subset D$, where m_r is the number of rows, m_c is the number of columns, and m_T is the number of lattices. Let $S_t = \{(i,j,t) : i = 1,\ldots,m_r; j = 1,\ldots,m_c\}$ denote the t^{th} lattice of S (or the lattice of S at time t). For notational simplicity, we will again let $n \equiv m_r \times m_c$, and we will numerically label sites of S_t from top to bottom within its columns, and from left to right across its columns, $\forall\ t = 1,\ldots,m_T$. Hence, $S_t = \{(i,t) : i = 1,\ldots,n\}$, where (i,t) denotes site i of the t^{th} lattice of the array.

Let $\boldsymbol{Z} = (\boldsymbol{Z}_1', \boldsymbol{Z}_2', \ldots, \boldsymbol{Z}_{m_T}')'$ again be a vector of discrete random variables on S, with joint support $\boldsymbol{\Omega}^{n \cdot m_T}$, where $\boldsymbol{Z}$ is distributed according to a Markov random field (MRF) and $\boldsymbol{Z}_t' = (Z_{1,t}, Z_{2,t}, \ldots, Z_{n,t})\ \forall\ t = 1,\ldots,m_T$. The joint density (i.e. likelihood), or MRF distribution, of $\boldsymbol{Z}$ has the same functional form as that of (3.1), if no covariates are in the model, and of (3.2) if covariates are in the model. In particular, if covariates are in the model, then the likelihood function has the following general form:

$$\begin{aligned} \mathrm{L}(\boldsymbol{\psi}) &= \mathrm{P}(\boldsymbol{Z} = \boldsymbol{z} | \boldsymbol{X}; \boldsymbol{\psi}) \\ &= \frac{\exp\{\boldsymbol{T}'(\boldsymbol{z}; \boldsymbol{X}) \cdot \boldsymbol{\psi}\}}{\sum_{\boldsymbol{y} \in \boldsymbol{\Omega}^{n \cdot m_T}} \exp\{\boldsymbol{T}'(\boldsymbol{y}; \boldsymbol{X}) \cdot \boldsymbol{\psi}\}} \\ &= \frac{\exp\{\boldsymbol{T}'(\boldsymbol{z}; \boldsymbol{X}) \cdot \boldsymbol{\psi}\}}{c(\boldsymbol{\psi})}, \end{aligned} \tag{6.1}$$

where $\boldsymbol{\psi} = (\alpha, \theta_1 \ldots, \theta_{p-1}, \beta_1, \ldots, \beta_k, \tau_1, \ldots, \tau_q)'$ is the vector of parameters, with parameter space $\boldsymbol{\Psi}$, $\boldsymbol{T}(\boldsymbol{z}) = (T_1(\boldsymbol{z}), \ldots, T_{p+q+k}(\boldsymbol{z}))'$ is the corresponding vector of jointly sufficient statistics (determined by the particular realization $\boldsymbol{z}$), and $c(\boldsymbol{\psi})$ is the intractable normalizing constant. Recall that α is the spatial trend parameter, $\theta_1, \ldots, \theta_{p-1}$ are the spatial dependence parameters, and $\beta_1, \ldots, \beta_k$ are the covariate parameters. The $\tau_1, \ldots, \tau_q$ terms are used to denote the q temporal dependence parameters.

The corresponding conditional probability density for $Z_{i,t}$, given $\{Z_{j,t'} : (j,t') \neq (i,t)\}$, is analogous to (3.3), if no covariates are in the model, and to (3.4) if covariates are in the model. In particular, using spatial and temporal Markovian properties, the corresponding conditional probability density for (6.1), $\forall\ (i,t) \in S$, can be written:

$$\mathrm{P}(Z_{i,t} = z_{i,t} | \{z_{j,t'} : (j,t') \in N_{i,t} \cup \Gamma_{i,t}\}, \boldsymbol{x}^{i,t}; \boldsymbol{\psi}) \quad =$$

$$\frac{\exp\{\boldsymbol{t}'(z_{i,t}, \{z_{j,t'} : (j,t') \in N_{i,t} \cup \Gamma_{i,t}\}, \boldsymbol{x}^{i,t}) \cdot \boldsymbol{\psi}\}}{\sum_{y_{i,t} \in \Omega_{i,t}} \exp\{\boldsymbol{t}'(y_{i,t}, \{y_{j,t'} : (j,t') \in N_{i,t} \cup \Gamma_{i,t}\}, \boldsymbol{x}^{i,t}) \cdot \boldsymbol{\psi}\}}, \tag{6.2}$$

where $\boldsymbol{\psi}$ is the same parameter vector as in (6.1) and $\boldsymbol{t}(z_{i,t}, \{z_{j,t'} : (j,t') \in N_{i,t} \cup \Gamma_{i,t}\}, \boldsymbol{x}^{i,t}) = (t_1(z_{i,t}, \{z_{j,t'} : (j,t') \in N_{i,t} \cup \Gamma_{i,t}\}, \boldsymbol{x}^{i,t}), \ldots, t_{p+q+k}(z_{i,t}, \{z_{j,t'} : (j,t') \in N_{i,t} \cup \Gamma_{i,t}\}, \boldsymbol{x}^{i,t})'$. As in all previous chapters, $N_{i,t}$ denotes the set of "spatial" (within a lattice subset S_t of S) neighbors of site (i,t), but now $\Gamma_{i,t}$, as introduced in section 1.5, denotes the set of "temporal" (between lattice subsets S_t and $S_{t'}, S_{t''}, \ldots$ of S) neighbors of site (i,t). Technically, distinguishing between spatial versus temporal neighbors of a site is unnecessary since their respective effects can be separated within the MRF model, as in (6.2), when spatio-temporal anisotropy is exhibited across the array. Note that spatio-temporal anisotropy is analogous to the concept of spatial anisotropy and simply means that the effect of time on the response variable is different than the effect, or one or more of the effects under spatial anisotropy, of space. The notational distinction between spatial and temporal neighbors made above, and throughout the remainder of this dissertation, is employed only because it allows for a more tractable presentation. Finally, recall that $\boldsymbol{x}^{i,t} = (x_{1,i,t}, x_{2,i,t}, \ldots, x_{k,i,t})'$ denotes the vector of k covariate values for site (i,t).

As we want to model the fire ignitions data from Oregon and Washington in Chapter 7, the spatio-temporal autologistic model we will present below and use to exemplify each of the estimation methods throughout the remainder of the chapter, will be the model that is ultimately fit to these data in Chapter 7. In particular, for illustrative purposes, assume we

are working with a spatio-temporal autologistic model that has four parameters, including an intercept parameter, a first-order isotropic spatial parameter, a first-order autoregressive temporal parameter, and a single covariate parameter. Note that a first-order autoregressive temporal parameter means that site i at time t has only one temporal neighbor, which is site i at time $t-1$; i.e. $\Gamma_{i,t} = \{(i, t-1)\}\ \forall\ (i,t) \in S$. The likelihood function for such a spatio-temporal autologistic model has the following form:

$$L(\boldsymbol{\psi}) \;=\; \mathrm{P}(\boldsymbol{Z} = \boldsymbol{z}|\boldsymbol{X};\boldsymbol{\psi})$$

$$= \frac{\exp\left\{\alpha\sum_{t=1}^{m_T}\sum_{i=1}^{n} z_{i,t} + \frac{1}{2}\theta\sum_{t=1}^{m_T}\sum_{i=1}^{n} z_{i,t}\sum_{(j,t')\in N_{i,t}} z_{j,t'} + \beta\sum_{t=1}^{m_T}\sum_{i=1}^{n} z_{i,t}x_{i,t} + \tau\sum_{t=1}^{m_T}\sum_{i=1}^{n} z_{i,t}\sum_{(j,t')\in \Gamma_{i,t}} z_{j,t'}\right\}}{\sum\limits_{\boldsymbol{y}\in\boldsymbol{\Omega}^{n\cdot m_T}} \exp\left\{\alpha\sum_{t=1}^{m_T}\sum_{i=1}^{n} y_{i,t} + \frac{1}{2}\theta\sum_{t=1}^{m_T}\sum_{i=1}^{n} y_{i,t}\sum_{(j,t')\in N_{i,t}} y_{j,t'} + \beta\sum_{t=1}^{m_T}\sum_{i=1}^{n} y_{i,t}x_{i,t} + \tau\sum_{t=1}^{m_T}\sum_{i=1}^{n} y_{i,t}\sum_{(j,t')\in \Gamma_{i,t}} y_{j,t'}\right\}}, \quad (6.3)$$

for all $\boldsymbol{z} \in \boldsymbol{\Omega}^{n\cdot m_T} = \{0,1\}^{n\cdot m_T}$. The corresponding conditional probability form of such a spatio-temporal autologistic model, by the spatial and temporal Markovian properties, is then

$$\mathrm{P}(Z_{i,t} = z_{i,t}|\{z_{j,t'} : (j,t') \in N_{i,t}\cup\Gamma_{i,t}\}, \boldsymbol{x}^{i,t};\boldsymbol{\psi}) \;=$$

$$= \frac{\exp\left\{\alpha z_{i,t} + \theta z_{i,t}\sum\limits_{(j,t')\in N_{i,t}} z_{j,t'} + \beta z_{i,t}x_{i,t} + \tau z_{i,t}\sum\limits_{(j,t')\in\Gamma_{i,t}} z_{j,t'}\right\}}{1 + \exp\left\{\alpha + \theta\sum\limits_{(j,t')\in N_{i,t}} z_{j,t'} + \beta x_{i,t} + \tau\sum\limits_{(j,t')\in\Gamma_{i,t}} z_{j,t'}\right\}}, \quad (6.4)$$

for all $z_{i,t} \in \Omega = \{0,1\}$, where $\boldsymbol{x}^{i,t} = x_{i,t}$ since there is only one covariate term in the model in our case. Observe that (6.4) strongly resembles the two-way anisotropic spatial autologistic model given by (2.41), and, thus, θ and τ communicate spatio-temporal anisotropy across

the array in a manner analogous to the way that θ_v and θ_h communicate spatial anisotropy across the lattice. This emphasizes the point made previously that the distinction between temporal and spatial neighbors is really only necessary for clarity of presentation. Finally, note that the functional form of (6.4) still bears a strong resemblance to that of the standard logistic model that is commonly employed with independent binary random variables. In fact, if the sites of the array were spatially and temporally independent, i.e. $\theta = 0$ and $\tau = 0$, then (6.4) reduces to a simple intercept-only logistic model. In general, the functional form of spatio-temporal autologistic models, as with purely spatial autologistic models, is identical to that of the logistic model, except that it also conditions on neighboring values, both spatially and temporally, of the binary response variable.

6.3 Spatio-Temporal PL

Employing pseudolikelihood in the space-time domain is straightforward, and the development presented here is analogous to that suggested by Zhu et al. ([37]). For the purely spatial MRFs of Chapter 3, the PL function was defined as the product of all n site conditional distributions. Hence, in the space-time domain, the PL function is still formed by taking the product over all of the site conditional distributions, except now there are $n \cdot m_T$, rather than just n, of them. In other words, the pseudolikelihood function, assuming covariates are in the model, is defined as follows in the space-time domain:

$$\mathrm{PL}(\boldsymbol{\psi}|\boldsymbol{z}, \boldsymbol{X}) \;=\; \prod_{t=1}^{m_T}\prod_{i=1}^{n} \mathrm{P}(Z_{i,t} = z_{i,t}|\{z_{j,t'} : (j,t') \in N_{i,t} \cup \Gamma_{i,t}\}, \boldsymbol{x}^{i,t}; \boldsymbol{\psi}), \tag{6.5}$$

$\boldsymbol{z} \in \boldsymbol{\Omega}^{n \cdot m_T}, \boldsymbol{\psi} \in \boldsymbol{\Psi}$.

The logarithm of (6.5) is then numerically maximized with respect to $\boldsymbol{\psi}$ to obtain the corresponding maximum pseudolikelihood estimate, $\hat{\boldsymbol{\psi}}_{PL}$, of $\boldsymbol{\psi}$. As in the purely spatial case, the numerical optimization can be carried out using any statistical software that has a

generalized linear models package, such as R. However, the corresponding asymptotic standard errors obtained from such optimization are once again invalid as the observations over the array are not independent. As before, a resampling method such as parametric bootstrapping may be used to obtain valid standard errors for $\hat{\boldsymbol{\psi}}_{PL}$. As mentioned in section 3.2, this technique will be presented in Chapter 7 when PL, as well as the other three estimation methods discussed in this current chapter, are applied to the fire data from Oregon and Washington. The computational expense of employing PL to obtain parameter estimates in the space-time domain is still negligible, although the expense in obtaining their corresponding standard errors with this additional dimension can be relatively substantial if m_T is large. Finally, note that the strong consistency of the MPLEs still applies here in the space-time domain as a result of the general nature under which this property was proven by Geman and Graffigne ([13]).

To provide a concrete example of (6.5) for the autologistic model, once again consider the four parameter spatio-temporal autologistic model with conditional distribution given by (6.4). In this context, the logarithm of the PL function is the following:

$$\begin{aligned}\log\{\mathrm{PL}(\boldsymbol{\psi}|\boldsymbol{z},\boldsymbol{X})\} &= \sum_{t=1}^{m_T}\sum_{i=1}^{n}\alpha z_{i,t}+\theta z_{i,t}\sum_{(j,t')\in N_{i,t}} z_{j,t'}+\beta z_{i,t}x_{i,t}+\tau z_{i,t}\sum_{(j,t')\in\Gamma_{i,t}} z_{j,t'} \\ &\quad -\sum_{t=1}^{m_T}\sum_{i=1}^{n}\log\left\{1+\exp\left\{\alpha+\theta\sum_{(j,t')\in N_{i,t}} z_{j,t'}+\beta x_{i,t}+\tau\sum_{(j,t')\in\Gamma_{i,t}} z_{j,t'}\right\}\right\}. \qquad (6.6)\end{aligned}$$

The MPLEs, $\hat{\boldsymbol{\psi}}_{PL}=(\hat{\alpha}_{PL},\hat{\theta}_{PL},\hat{\beta}_{PL},\hat{\tau}_{PL})'$, are then obtained by numerically maximizing (6.6) with respect to $\boldsymbol{\psi}$, and the corresponding standard error can be obtained using a parametric bootstrap via the Gibbs sampler.

6.4 Spatio-Temporal MCMCML

Employing Markov chain Monte Carlo maximum likelihood in the space-time domain is also straightforward, and the development presented here is derived from Geyer and Thompson's ([15]) canonical paper on MCMCML. In particular, a ratio of normalizing constants is used, in a fashion analogous to that described in section 3.3, to obtain a MCMC approximate log likelihood. The general functional form of this MCMC approximate log likelihood is identical to (3.15) given in section 3.3, namely:

$$l_{m,\boldsymbol{z}_{obs}}(\boldsymbol{\psi}) \quad = \quad \log\left\{h_{\boldsymbol{\psi}}(\boldsymbol{z}_{obs})\right\} - \log\left\{\frac{1}{m}\sum_{r=1}^{m}\frac{h_{\boldsymbol{\psi}}(\boldsymbol{Z}_{(r)})}{h_{\boldsymbol{\phi}}(\boldsymbol{Z}_{(r)})}\right\}, \tag{6.7}$$

where $h_{\boldsymbol{\psi}}(\cdot)$ is now just the un-normalized spatio-temporal joint likelihood and $\boldsymbol{\phi}$ is the reference point for the Gibbs sampler. As in the purely spatial case, we expect that using the MPLEs for $\boldsymbol{\phi}$ should generally yield satisfactory results. Recall that the m subscript on the $l_{m,\boldsymbol{z}_{obs}}(\boldsymbol{\psi})$ indicates the amount of Monte Carlo effort, $\boldsymbol{z}_{obs}$ indicates the array realization of the MRF that was actually observed, and $\boldsymbol{Z}_{(r)}$ is the r^{th} array realization of the Markov chain generated using the Gibbs sampler, whose use in the space-time domain is addressed further below.

The MCMCMLEs, $\hat{\boldsymbol{\psi}}_{MC}$, of $\boldsymbol{\psi}$ are obtained by numerically maximizing (6.7) with respect to $\boldsymbol{\psi}$, while the corresponding standard errors are again obtained by evaluating the inverted Hessian of $l_{m,\boldsymbol{z}_{obs}}(\boldsymbol{\psi})$ at $\hat{\boldsymbol{\psi}}_{MC}$. The computational expense of employing MCMCML in the space-time domain is, as would be expected, substantial relative to PL, especially if m_T is large. This fact is once again the motivation for alternative methods, such as GPL and BGPL, which aim to strike a computational, as well as a performance-based, compromise between PL and MCMCML. Finally, note that, as with PL, the strong consistency of the MCMCMLEs still applies here in the space-time domain as a result of the general nature under which this property was proven by Geyer and Thompson ([15]).

The use of the Gibbs sampler in the space-time domain is entirely analogous to its use in the purely spatial domain, as presented in section 3.3.2. Some important things to note in the space-time domain, however, are (1) the conditional probability densities that are now employed are given by (6.2), assuming covariates are in the model, and (2) a sweep now constitutes a realization being generated for each of the $n \cdot m_T$ sites of the array. Finally, note that the MRF realizations generated through the Gibbs sampler are now $m_r \times m_c \times m_t$ arrays, rather than $m_r \times m_c$ lattices as in the purely spatial setting.

To provide a concrete example of (6.7) for the autologistic model, once again consider the four parameter spatio-temporal autologistic model with conditional distribution given by (6.4). In this context, with the MPLEs serving as the reference point, the MCMC approximate log likelihood function is the following:

$$l_{m,\boldsymbol{z}_{obs}}(\boldsymbol{\psi}) \quad =$$

$$-\log\left\{\frac{1}{m}\sum_{r=1}^{m}\frac{\exp\left\{\alpha\sum_{t=1}^{m_T}\sum_{i=1}^{n}z_{i,t}^{(r)}+\frac{1}{2}\theta\sum_{t=1}^{m_T}\sum_{i=1}^{n}z_{i,t}^{(r)}\sum_{(j,t')\in N_{i,t}}z_{j,t'}^{(r)}+\beta\sum_{t=1}^{m_T}\sum_{i=1}^{n}z_{i,t}^{(r)}x_{i,t}+\tau\sum_{t=1}^{m_T}\sum_{i=1}^{n}z_{i,t}^{(r)}\sum_{(j,t')\in\Gamma_{i,t}}z_{j,t'}^{(r)}\right\}}{\exp\left\{\hat{\alpha}_{PL}\sum_{t=1}^{m_T}\sum_{i=1}^{n}z_{i,t}^{(r)}+\frac{1}{2}\hat{\theta}_{PL}\sum_{t=1}^{m_T}\sum_{i=1}^{n}z_{i,t}^{(r)}\sum_{(j,t')\in N_{i,t}}z_{j,t'}^{(r)}+\hat{\beta}_{PL}\sum_{t=1}^{m_T}\sum_{i=1}^{n}z_{i,t}^{(r)}x_{i,t}+\hat{\tau}_{PL}\sum_{t=1}^{m_T}\sum_{i=1}^{n}z_{i,t}^{(r)}\sum_{(j,t')\in\Gamma_{i,t}}z_{j,t'}^{(r)}\right\}}\right\}$$

$$+\quad \alpha\sum_{t=1}^{m_T}\sum_{i=1}^{n}z_{i,t}^{obs}+\frac{1}{2}\theta\sum_{t=1}^{m_T}\sum_{i=1}^{n}z_{i,t}^{obs}\sum_{(j,t')\in N_{i,t}}z_{j,t'}^{obs}+\beta\sum_{t=1}^{m_T}\sum_{i=1}^{n}z_{i,t}^{obs}x_{i,t}+\tau\sum_{t=1}^{m_T}\sum_{i=1}^{n}z_{i,t}^{obs}\sum_{(j,t')\in\Gamma_{i,t}}z_{j,t'}^{obs}, \quad (6.8)$$

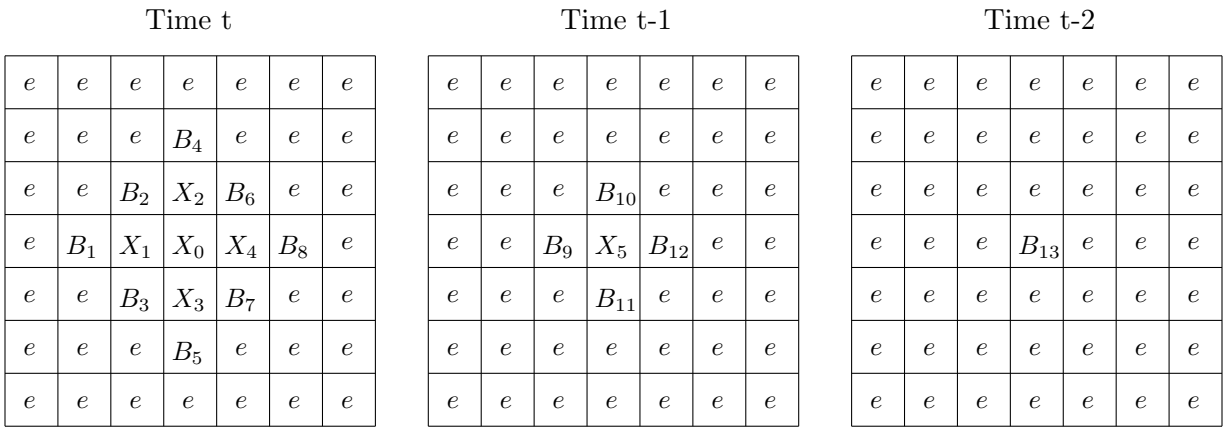
where $\boldsymbol{z}_{obs} = (\boldsymbol{z}'_{obs,1}, \ldots, \boldsymbol{z}'_{obs,m_T})'$ with $\boldsymbol{z}'_{obs,t} = (z_{1,t}^{obs}, \ldots, z_{n,t}^{obs})$, and $\boldsymbol{z}_{(r)} = (\boldsymbol{z}'_{(r)1}, \ldots, \boldsymbol{z}'_{(r)m_t})'$ with $\boldsymbol{z}'_{(r)t} = (z_{1,t}^{(r)}, \ldots, z_{n,t}^{(r)})$. The Gibbs sampler, with conditional distribution given by (6.4), is used to generate the m Markov chain space-time realizations $\boldsymbol{z}_{(r)}, r = 1, \ldots, m$. The MCMCMLEs, $\hat{\boldsymbol{\psi}}_{MC} = (\hat{\alpha}_{MC}, \hat{\theta}_{MC}, \hat{\beta}_{MC}, \hat{\tau}_{MC})'$, are obtained by numerically maximizing (6.8) with respect to $\boldsymbol{\psi} = (\alpha, \theta, \beta, \tau)'$. Finally, the standard errors corresponding to $\hat{\boldsymbol{\psi}}_{MC}$ are obtained by evaluating the inverse of the Hessian of (6.8) at $\hat{\boldsymbol{\psi}}_{MC}$, i.e. by computing the observed Fisher information matrix.

6.5 Spatio-Temporal GPL

The generalization of generalized pseudolikelihood for use in the space-time domain, while not terribly difficult, is novel. The critical component in making this generalization is establishing the group structure. In the purely spatial domain, there is a group corresponding to each site of the lattice and the group structure implemented (e.g. the cross structure) is the same for every group. These two characteristics of groups in the spatial domain will be analogously extended into the space-time domain. However, in the space-time domain, groups can now consist of not only multiple sites of the array from the same time point, but also of sites of the array at previous time points. Thus, if we let $g(i,t)$ denote the group of sites associated with site $(i,t) \in S$, then a reasonable group structure might consist of site (i,t), its spatial neighbors and its temporal neighbors, i.e. $g(i,t) = \{\{(i,t)\} \cup N_{i,t} \cup \Gamma_{i,t}\}$. For example, Figure 6.1 illustrates a possible group structure for the spatio-temporal autologistic model corresponding to (6.3), where $g(i,t)$ consists of site (i,t), its four nearest neighbors at time t, and site $(i,t-1)$ (i.e. site i at the previous time point, $t-1$). The set of boundary neighbors of a group consists of all sites that are not themselves members of the group, but are spatial and/or temporal neighbors of at least one site belonging to the group. For instance, in Figure 6.1, sites $B_1, \ldots, B_{13}$ are the boundary neighbors of the group. Notice, for example, that site B_{13} is only a temporal neighbor of group site X_5, site B_2 is only a spatial neighbor for group sites X_1 and X_2, and site B_9 is both a temporal neighbor of group site X_1 and a spatial neighbor of group site X_5.

With the general notion of a group in the space-time domain established, the corresponding group joint likelihood functions and generalized pseudolikelihood function, which are analogous to their counterparts (3.21) and (3.22) presented in section 3.4, can now be defined. In particular, $\forall\ (i,t) \in S$, the group joint likelihood function for the $(i,t)^{th}$ group, assuming no

Figure 6.1: Space-Time Domain Group Structure Example



e	e	e	e	e	e	e
e	e	e	B_4	e	e	e
e	e	B_2	X_2	B_6	e	e
e	B_1	X_1	X_0	X_4	B_8	e
e	e	B_3	X_3	B_7	e	e
e	e	e	B_5	e	e	e
e	e	e	e	e	e	e

e	e	e	e	e	e	e
e	e	e	e	e	e	e
e	e	e	B_{10}	e	e	e
e	e	B_9	X_5	B_{12}	e	e
e	e	e	B_{11}	e	e	e
e	e	e	e	e	e	e
e	e	e	e	e	e	e

e	e	e	e	e	e	e
e	e	e	e	e	e	e
e	e	e	e	e	e	e
e	e	e	B_{13}	e	e	e
e	e	e	e	e	e	e
e	e	e	e	e	e	e
e	e	e	e	e	e	e

A group, its boundary neighbors, and the corresponding external sites on a $7 \times 7 \times 3$ array subset of S under a first-order isotropic spatial structure and a first-order autoregressive temporal structure. Moving left to right, the three lattices correspond to times $t, t-1$, and $t-2$, respectively. Assume that sites X_1, X_2, X_3, X_4, and X_5 are the additional group members, as well as the spatial and temporal neighbors, corresponding to site X_0. Then the 13 B sites represent the fixed spatial and/or temporal boundary neighbors of the group and the 128 e sites represent the external sites of the group, i.e., the sites that are conditionally independent of the group sites under the spatial and temporal Markov assumptions.

covariates are in the model, still has the following general form:

$$\mathrm{GL}_{i,t}(\boldsymbol{\psi}) \;=\; \left\{ \frac{\exp\{\boldsymbol{T}'\left(\boldsymbol{z}_{g(i,t)}, \boldsymbol{z}^{g(i,t)}\right)\cdot\boldsymbol{\psi}\}}{\sum_{\boldsymbol{y}\in\boldsymbol{\Omega}^{|g(i,t)|}}\exp\{\boldsymbol{T}'\left(\boldsymbol{y}_{g(i,t)}, \boldsymbol{y}^{g(i,t)}\right)\cdot\boldsymbol{\psi}\}} \right\}^{\frac{1}{|g(i,t)|}}, \tag{6.9}$$

where $|g(i,t)|$ is the number of sites belonging to group (i,t). Recall from section 3.4.1 that the sufficient statistics for a group joint likelihood must distinguish between a (spatial) neighbor relation between two sites of the group and a (spatial) neighbor relation between one site of the group and one boundary neighbor site of the group. No such distinction is needed for temporal neighbor relationships since they are naturally ordered in time. More specifically, because the space-time paradigm implemented for this dissertation assumes that the temporal neighbors of any site, (i,t), of the array can only belong to an "earlier" lattice of the array, i.e. $(j,t') \in \Gamma_{i,t} \Rightarrow t' \leq (t-1)$, the temporal relations between two sites (in different time steps) of the group will not be "double-counted" relative to the temporal relations between a site (at one time step) of the group and a boundary neighbor site (at a different time step) of

the group. For example, in Figure 6.1 the temporal relation between sites X_0 and X_5 is not double-counted relative to the other temporal relations of the group, such as those between sites X_1 and B_9 or sites X_5 and B_{13}.

The generalized pseudolikelihood function, again assuming no covariates are in the model, is then just the product of all group joint likelihood functions from (6.9):

$$\text{GPL}(\boldsymbol{\psi}) \;=\; \prod_{t=1}^{m_T}\prod_{i=1}^{n}\left\{\frac{\exp\{\boldsymbol{T}'\left(\boldsymbol{z}_{g(i,t)},\boldsymbol{z}^{g(i,t)}\right)\cdot\boldsymbol{\psi}\}}{\sum_{\boldsymbol{y}\in\boldsymbol{\Omega}^{|g(i,t)|}}\exp\{\boldsymbol{T}'\left(\boldsymbol{y}_{g(i,t)},\boldsymbol{y}^{g(i,t)}\right)\cdot\boldsymbol{\psi}\}}\right\}^{\frac{1}{|g(i,t)|}}. \qquad (6.10)$$

Note that, just as in the purely spatial domain, if $g(i,t) = \{(i,t)\}\ \forall\ (i,t) \in S$, then GPL reduces to PL, and if $g(i,t) = S\ \forall\ (i,t) \in S$, then GPL is ultimately maximum likelihood. The MGPLEs of $\boldsymbol{\psi}$, $\hat{\boldsymbol{\psi}}_{GPL}$, are then obtained by numerically maximizing the logarithm of (6.10). As in the purely spatial domain, the standard errors of the MGPLEs can be obtained using a resampling method, such as a parametric bootstrap. The strong consistency of the MGPLEs in the space-time domain still applies since the proof given in Chapter 4 is general enough to accommodate more than two-dimensions.

Before demonstrating the extended GPL method within the context of the spatio-temporal autologistic model, we should briefly discuss a potentially computational issue with GPL in the space-time domain. In particular, just as was the case in the purely spatial domain, the group size in the space-time domain cannot be too large or the computational compromise GPL strives to achieve between PL and MCMCML will be lost. Thus, while GPL can straightforwardly be extended into the space-time domain, it is computationally difficult to use group sizes any larger than the biggest group sizes used in the purely spatial domain. Such limitations, however, can make choosing a group structure unsatisfactory since striking the appropriate balance between spatial and temporal relations can then be difficult. For example, in Figure 6.1, the given group structure has 5 sites from the same time step (spatial relations) in the group, but only one site from a different time step (temporal relations) in

the group. Maybe this is ultimately a satisfactory group structure, but perhaps this group structure actually overemphasizes the spatial relations relative to the temporal relations of the array. Thus, the issue of selecting a suitable group structure becomes more important in the space-time domain.

Finally, as an example of (6.9) and (6.10), again consider the four parameter spatio-temporal autologistic model conveyed by (6.3). In this context, the group likelihood functions have the following form for all $(i,t) \in S$:

$$\mathrm{GL}_{i,t}(\boldsymbol{\psi}) \quad = \quad \frac{\exp\left\{\alpha \sum_{(j,t')\in g(i,t)} z_{j,t'} + \theta \sum_{(j,t')\in g(i,t)} z_{j,t'} \left(\frac{1}{2} \sum_{(k,t'')\in N_{j,t'}^{g(i,t)}} z_{k,t''} + \sum_{(k,t'')\in N_{j,t'}^{\partial g(i,t)}} z_{k,t''} \right) + \beta \sum_{(j,t')\in g(i,t)} z_{j,t'} x_{j,t'} + \tau \sum_{(j,t')\in g(i,t)} z_{j,t'} \sum_{(k,t'')\in \Gamma_{j,t'}} z_{k,t''} \right\}}{\sum_{\boldsymbol{y}\in\boldsymbol{\Omega}^{|g(i,t)|}} \exp\left\{\alpha \sum_{(j,t')\in g(i,t)} y_{j,t'} + \theta \sum_{(j,t')\in g(i,t)} y_{j,t'} \left(\frac{1}{2} \sum_{(k,t'')\in N_{j,t'}^{g(i,t)}} y_{k,t''} + \sum_{(k,t'')\in N_{j,t'}^{\partial g(i,t)}} y_{k,t''} \right) + \beta \sum_{(j,t')\in g(i,t)} y_{j,t'} x_{j,t'} + \tau \sum_{(j,t')\in g(i,t)} y_{j,t'} \sum_{(k,t'')\in \Gamma_{j,t'}} y_{k,t''} \right\}}. \quad (6.11)$$

The corresponding GPL function is then just the product over all $n \cdot T_m$ group likelihood functions from (6.11). The logarithm of this function is then numerically maximized with respect to $\boldsymbol{\psi}$ to obtain the MGPLEs, $\hat{\boldsymbol{\psi}}_{GPL} = (\hat{\alpha}_{GPL}, \hat{\theta}_{GPL}, \hat{\beta}_{GPL}, \hat{\tau}_{GPL})'$, of $\boldsymbol{\psi}$, while a parametric bootstrap can be used to obtain the corresponding standard errors.

6.6 Spatio-Temporal BGPL

Just as with GPL, the generalization of block generalized pseudolikelihood for use in the space-time domain, while not terribly difficult, is novel. The critical component in making this generalization, as was analogously the case with GPL, is establishing the block structure.

Figure 6.2: Space-Time Domain Block Structure Example

Time t

e	e	e	e	e	e	e
e	B_1	B_6	B_8	B_{10}	B_{12}	e
e	B_2	X_1	X_4	X_7	B_{13}	e
e	B_3	X_2	X_5	X_8	B_{14}	e
e	B_4	X_3	X_6	X_9	B_{15}	e
e	B_5	B_7	B_9	B_{11}	B_{16}	e
e	e	e	e	e	e	e

Time t-1

e	e	e	e	e	e	e
e	B_{17}	B_{22}	B_{24}	B_{26}	B_{28}	e
e	B_{18}	X_{10}	X_{13}	X_{16}	B_{29}	e
e	B_{19}	X_{11}	X_{14}	X_{17}	B_{30}	e
e	B_{20}	X_{12}	X_{15}	X_{18}	B_{31}	e
e	B_{21}	B_{23}	B_{25}	B_{27}	B_{32}	e
e	e	e	e	e	e	e

Time t-2

e	e	e	e	e	e	e
e	e	e	e	e	e	e
e	e	B_{33}	B_{36}	B_{39}	e	e
e	e	B_{34}	B_{37}	B_{40}	e	e
e	e	B_{35}	B_{38}	B_{41}	e	e
e	e	e	e	e	e	e
e	e	e	e	e	e	e

A block, its boundary neighbors, and the corresponding external sites on a $7 \times 7 \times 3$ array subset of S under a second-order isotropic spatial structure and a first-order autoregressive temporal structure. Moving left to right, the three lattices correspond to times $t, t-1$, and $t-2$, respectively. Assume that sites $X_i, i = 1, \ldots 18$ are members of block l. Then the 41 B sites represent the fixed spatial and/or temporal boundary neighbors of the block and the 88 e sites represent the external sites of the block, i.e., the sites that are conditionally independent of the block sites under the spatial and temporal Markov assumptions.

Once again, the array must be partitioned into $L \leq n \cdot m_T$ (non overlapping) blocks, where $b(l)$ denotes the sites "contained" within the l^{th} block; so $S = \bigcup_{l=1}^{L} b(l)$. While it is not theoretically necessary that the L blocks all be the same size or shape, we will assume, as was done in the purely spatial domain, that they are all the same size and shape. Possible blocking mechanisms include forming $b \times b \times d$ blocks, where $1 \leq d \leq m_T$ and $1 \leq b \leq \min\{m_r, m_c\}$. For example, Figure 6.2 illustrates a possible block structure for the spatio-temporal autologistic model example corresponding to (6.3), where $b(l)$ is $3 \times 3 \times 2$. Analogous to GPL, the set of boundary neighbors of a block consists of all sites that are not themselves members of the block, but are spatial and/or temporal neighbors of at least one site belonging to the block. For instance, in Figure 6.2, sites $B_1, \ldots, B_{41}$ are the boundary neighbors of the block. Notice, for example, that site B_{31} is only a temporal neighbor of block site X_{10} and site B_{10} is only a spatial neighbor for block sites X_4 and X_7.

With the general notion of a block in the space-time domain established, the corresponding block joint likelihood functions and block generalized pseudolikelihood function, which are

analogous to their counterparts (3.33) and (3.34) presented in section 3.5, can now be defined. In particular, $\forall\ l,\ l = 1, \ldots, L$, the block joint likelihood function for the l^{th} block, assuming no covariates are in the model, still has the following general form:

$$\mathrm{BL}_l(\boldsymbol{\psi}) \quad = \quad \frac{\exp\{\boldsymbol{T}'\left(\boldsymbol{z}_{b(l)}, \boldsymbol{z}^{b(l)}\right) \cdot \boldsymbol{\psi}\}}{\sum_{\boldsymbol{y} \in \boldsymbol{\Omega}^{|b(l)|}} \exp\{\boldsymbol{T}'\left(\boldsymbol{y}_{b(l)}, \boldsymbol{y}^{b(l)}\right) \cdot \boldsymbol{\psi}\}}, \tag{6.12}$$

where $|b(l)|$ is the number of sites belonging to block l. Recall from section 3.5.1 that the sufficient statistics for a block joint likelihood must distinguish between a (spatial) neighbor relation between two sites of the block and a (spatial) neighbor relation between one site of the block and one boundary neighbor site of the block. As was the case for GPL, no such distinction is needed here for temporal neighbor relationships since they are naturally ordered in time. More specifically, as was discussed in section 6.5, because the space-time paradigm implemented for this dissertation assumes that the temporal neighbors of any site, (i, t), of the array can only belong to an "earlier" lattice of the array, i.e. $(j, t') \in \Gamma_{i,t} \Rightarrow t' \leq (t-1)$, the temporal relations between two sites of the block will not be "double-counted" relative to the temporal relations between a site of the block and a boundary neighbor site of the block. For example, in Figure 6.2 the temporal relation between sites X_1 and X_{10}, or sites X_5 and X_{14}, is not double-counted relative to the other temporal relations of the group, such as those between sites X_{10} and B_{33} or sites X_{14} and B_{37}.

The block generalized pseudolikelihood function, again assuming no covariates are in the model, is then just the product of all block joint likelihood functions from (6.12):

$$\mathrm{BGPL}(\boldsymbol{\psi}) \quad = \quad \prod_{l=1}^{L} \frac{\exp\{\boldsymbol{T}'\left(\boldsymbol{z}_{b(l)}, \boldsymbol{z}^{b(l)}\right) \cdot \boldsymbol{\psi}\}}{\sum_{\boldsymbol{y} \in \boldsymbol{\Omega}^{|b(l)|}} \exp\{\boldsymbol{T}'\left(\boldsymbol{y}_{b(l)}, \boldsymbol{y}^{b(l)}\right) \cdot \boldsymbol{\psi}\}} \tag{6.13}$$

Note that, just as in the purely spatial domain, if $|b(l)| = 1\ \forall\ l \in L$ (i.e. $L = n \cdot m_T$), then BGPL reduces to PL, and if $b(l) = S$ (i.e. $L = 1$), then BGPL is ultimately maximum likelihood. The MBGPLEs of $\boldsymbol{\psi}$, $\hat{\boldsymbol{\psi}}_{BGPL}$, are then obtained by numerically maximizing the

logarithm of (6.13). As in the purely spatial domain, the standard errors of the MBGPLEs can be obtained using a resampling method, such as a parametric bootstrap. As with the MGPLEs, the strong consistency of the MBGPLEs in the space-time domain still applies since the proof given in Chapter 4 is general enough to accommodate more than two-dimensions. Note that the computational issue discussed in section 6.5 in choosing a group structure also applies to choosing a block structure for BGPL in the space-time domain.

Finally, as an example of (6.12) and (6.13), again consider the four parameter spatio-temporal autologistic model conveyed by (6.3). In this context, the block likelihood functions have the following form for all $l = 1, \ldots, L$:

$$\mathrm{BL}_l(\boldsymbol{\psi}) \quad =$$

$$\frac{\exp\left\{\alpha \sum\limits_{(i,t)\in b(l)} z_{i,t} + \theta \sum\limits_{(i,t)\in b(l)} z_{i,t}\left(\frac{1}{2}\sum\limits_{(j,t')\in N_{i,t}^{b(l)}} z_{j,t'} + \sum\limits_{(j,t')\in N_{i,t}^{\partial b(l)}} z_{j,t'}\right) + \beta \sum\limits_{(i,t)\in b(l)} z_{i,t}x_{i,t} + \tau \sum\limits_{(i,t)\in b(l)} z_{i,t} \sum\limits_{(j,t')\in\Gamma_{i,t}} z_{j,t'}\right\}}{\sum\limits_{\boldsymbol{y}\in\boldsymbol{\Omega}^{|b(l)|}} \exp\left\{\alpha \sum\limits_{(i,t)\in b(l)} y_{i,t} + \theta \sum\limits_{(i,t)\in b(l)} y_{i,t}\left(\frac{1}{2}\sum\limits_{(j,t')\in N_{i,t}^{b(l)}} y_{j,t'} + \sum\limits_{(j,t')\in N_{i,t}^{\partial b(l)}} y_{j,t'}\right) + \beta \sum\limits_{(i,t)\in b(l)} y_{i,t}x_{i,t} + \tau \sum\limits_{(i,t)\in b(l)} y_{i,t} \sum\limits_{(j,t')\in\Gamma_{i,t}} y_{j,t'}\right\}}. \quad (6.14)$$

The corresponding BGPL function is then just the product over all L block likelihood functions from (6.14). The logarithm of this function is then numerically maximized with respect to $\boldsymbol{\psi}$ to obtain the MBGPLEs, $\hat{\boldsymbol{\psi}}_{BGPL} = (\hat{\alpha}_{BGPL}, \hat{\theta}_{BGPL}, \hat{\beta}_{BGPL}, \hat{\tau}_{BGPL})'$, of $\boldsymbol{\psi}$, while a parametric bootstrap can be used to obtain the corresponding standard errors.

6.7 Spatio-Temporal Autologistic Model Simulation Study

As discussed in section 1.5, a simulation study involving the spatio-temporal autologistic model does not currently exist in the literature. Hence, the small-scale simulation study of the spatio-temporal autologistic model carried out for this dissertation, which evaluates and compares the performances of the estimates obtained via PL, MCMCML, GPL, and BGPL, is the first of its kind. The two primary objectives of this simulation study were exploratory in nature. More specifically, we wanted to (1) evaluate, via measures of bias, precision, and accuracy, the utility of these four methods in estimating the parameters of the spatio-temporal autologistic model and (2) compare, via the relative mean absolute error (MAE), the performances of the estimates obtained from these four different methods.

6.7.1 Scope

As the computational expense of conducting a simulation study for the spatio-temporal autologistic model is substantial, the scope of the study that was conducted is necessarily limited. In particular, the only (spatial/temporal) neighborhood structure considered had a first-order isotropic spatial structure and an autoregressive lag-one temporal structure. In other words, each internal site of the array has 4 spatial neighbors and 1 temporal neighbor at the same location but previous time step. Furthermore, the only array size considered was $50 \times 50 \times 11$. Note that the 50×50 spatial dimensions were ultimately selected because they were the most moderate of the three lattice sizes used in the purely spatial simulation study of Chapter 5, while the temporal dimension of 11 was ultimately selected because that is the number of time steps available in the Oregon/Washington fire occurrence data set modeled in Chapter 7. Additionally, only a $3 \times 3 \times 1$ group structure, centered at the "defining" site of the group, and a $4 \times 4 \times 1$ block structure were considered. Note that these structures were ultimately used for two reasons. First, the 3×3 group structure, denoted by gs9 in Chapter 5, and the 4×4

block structure, denote by bs4 in Chapter 5, were the best of their respective structures in the purely spatial simulations of Chapter 5. Second, adding any "depth" to these structures in the space-time simulations, such as a $3 \times 3 \times 2$ group structure or a $4 \times 4 \times 2$ block structure, would have compromised the computational objective of GPL and BGPL in that both methods would then be computationally more expensive, i.e. take longer, than MCMCML.

In addition to the above stipulations, the first-order isotropic autoregressive lag-one spatio-temporal autologistic model (FIAR1) was considered both with and without a covariate term in the model. The functional form of the conditional spatio-temporal autologistic model, without a covariate term, is given by the following:

$$\begin{aligned} \mathrm{P}(Z_{i,t} = z_{i,t} | \{z_{j,t'} : (j,t') \in N_{i,t} \cup \Gamma_{i,t}\}; \boldsymbol{\psi}) \quad &= \\ = \quad \frac{\exp\left\{\alpha z_{i,t} + \theta z_{i,t} \sum\limits_{(j,t') \in N_{i,t}} z_{j,t'} + \tau z_{i,t} z_{i,t-1}\right\}}{1 + \exp\left\{\alpha + \theta \sum\limits_{(j,t') \in N_{i,t}} z_{j,t'} + \tau z_{i,t-1}\right\}}. & \end{aligned} \tag{6.15}$$

The functional form of the conditional spatio-temporal autologistic model, with a covariate term, is given by (6.4), where recall $\Gamma_{i,t} = \{(i, t-1)\} \; \forall \; (i,t) \in S$. For the model without a covariate term, two parameter vectors were considered for $(\alpha, \theta, \tau)'$. In both, α and θ were held fixed at -1.0 and 0.4, respectively, while $\tau = 0.5$ and 1.5. Note that in both of these vectors the strength of the spatial dependence is moderate, while the strength of the temporal dependence is relatively moderate for $\tau = 0.5$ and relatively strong for $\tau = 1.5$. In particular, for $(\alpha, \theta, \tau)' = (-1.0, 0.4, 0.5)$, approximately 52% of the sites of the array have value 1, while for $(\alpha, \theta, \tau)' = (-1.0, 0.4, 1.5)$, approximately 79% of the sites of the array have value 1. For the model with a covariate term, two covariate types were considered. The first covariate type corresponds to the departure from average (DA) variable of the Oregon and Washington fire occurrence data set. Note that this fire covariate, as in the analogous spatial

simulations of Chapter 5, was generated via the *grf* package in R by using the estimated covariograms obtained from each cross-sectional lattice of DA values from the aforementioned data set. Only the parameter vector $(\alpha, \theta, \tau, \beta)' = (-6.87, 1.20, 0.015, 0.023)$ was used for this covariate type. This particular parameter vector was selected after maximum pseudolikelihood estimation was used to model the array of data from Oregon/Washington. Since an objective of this dissertation is to model the Oregon and Washington fire data using the four estimation methods previously developed in this chapter, we wanted to get an idea via simulation how these methods would perform on a similar array in which approximately $< 1\%$ of the sites have a value of 1.

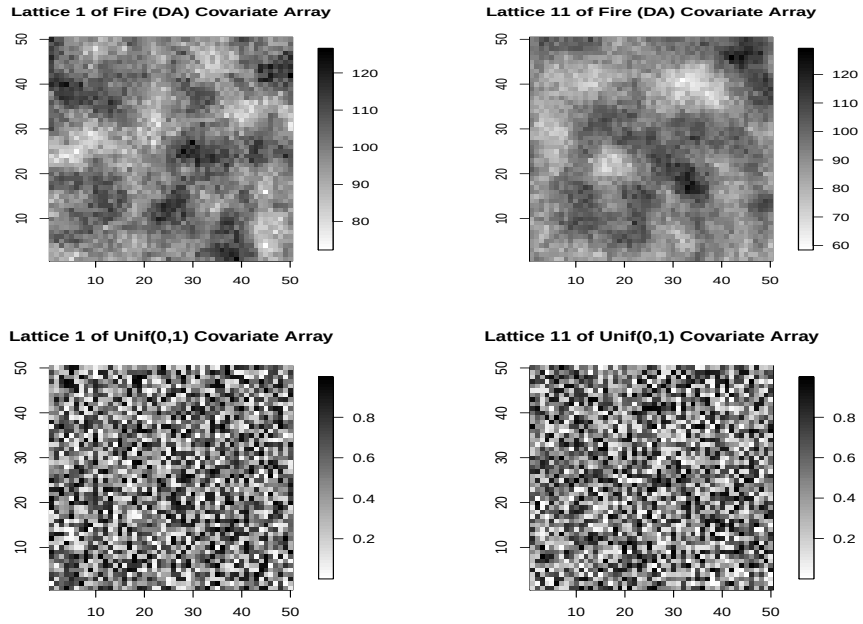
The second covariate corresponds to a Uniform(0,1) random variable in which the sites of the array were each randomly assigned a value from this distribution. Only the parameter vector $(\alpha, \theta, \tau, \beta)' = (-2.0, 0.80, 0.30, 0.50)$ was used for this covariate type. This particular parameter vector was selected for two reasons. First, we wanted the spatial dependence, temporal dependence, and uniform covariate to all have a relatively moderate effect on the binary response variable. Second, we wanted approximately 50% of the sites of the array to have a value of 1. Figure 6.3 displays, for both covariate types, the 1^{st} and 11^{th} covariate lattices (cross-sections) of the array.

6.7.2 Procedure

The simulation procedures used for this space-time simulation study was analogous to the procedures used and outlined in Chapter 5 for the purely spatial simulation study. To avoid any confusion, however, the simulation procedure used for the 2 parameter vectors of the first-order isotropic autoregressive lag-one spatio-temporal autologistic model, without a covariate term, was the following:

(1) Using the Gibbs sampler, 510 $56 \times 56 \times 11$ Markov chain Monte Carlo (MCMC) array re-

Figure 6.3: Cross-Sections 1 and 11 for Both Covariate Types



alizations (i.e. samples) were generated from the Ising model with true parameter vector $(\alpha_0, \theta_0, \tau_0)$. A torus spatial and temporal edge adjustment was used within the Gibbs sampler and each of the 510 realizations was obtained from an independent Markov chain in which a burn-in of 100 full array sweeps was employed to allow for convergence.

(2) For each of the above 510 array realizations, the "central" $50 \times 50 \times 11$ lattice was extracted. These 510 $50 \times 50 \times 11$ samples will henceforth be referred to as the "original" Markov chain samples.

(3) For the i^{th} "original" sample, $i = 1, \ldots, 510$, the pseudolikelihood function was constructed and maximized with respect to (α, θ, τ) to produce the maximum pseudolikelihood estimates (MPLEs) $(\hat{\alpha}_{PL_i}, \hat{\theta}_{PL_i}, \hat{\tau}_{PL_i})$. The mean, standard deviation, bias, and mean absolute error (MAE) of the parameter estimates were estimated using the 510

MPLEs as follows:

$$\bar{\hat{\alpha}}_{PL} = \frac{1}{510}\sum_{i=1}^{510}\hat{\alpha}_{PL_i}, \quad \bar{\hat{\theta}}_{PL} = \frac{1}{510}\sum_{i=1}^{510}\hat{\theta}_{PL_i}, \quad \bar{\hat{\tau}}_{PL} = \frac{1}{510}\sum_{i=1}^{510}\hat{\tau}_{PL_i},$$

$$\widehat{\mathrm{SD}}(\hat{\alpha}_{PL}) = \sqrt{\frac{1}{510}\sum_{i=1}^{510}(\hat{\alpha}_{PL_i} - \bar{\hat{\alpha}}_{PL})^2}, \quad \widehat{\mathrm{SD}}(\hat{\theta}_{PL}) = \sqrt{\frac{1}{510}\sum_{i=1}^{510}(\hat{\theta}_{PL_i} - \bar{\hat{\theta}}_{PL})^2},$$

$$\widehat{\mathrm{SD}}(\hat{\tau}_{PL}) = \sqrt{\frac{1}{510}\sum_{i=1}^{510}(\hat{\tau}_{PL_i} - \bar{\hat{\tau}}_{PL})^2},$$

$$\widehat{\mathrm{Bias}}(\bar{\hat{\alpha}}_{PL}) = (\bar{\hat{\alpha}}_{PL} - \alpha_0), \quad \widehat{\mathrm{Bias}}(\bar{\hat{\theta}}_{PL}) = (\bar{\hat{\theta}}_{PL} - \theta_0), \quad \widehat{\mathrm{Bias}}(\bar{\hat{\tau}}_{PL}) = (\bar{\hat{\tau}}_{PL} - \tau_0),$$

$$\widehat{\mathrm{MAE}}(\hat{\alpha}_{PL}) = \frac{1}{510}\sum_{i=1}^{510}\left|\hat{\alpha}_{PL_i} - \alpha_0\right|, \quad \widehat{\mathrm{MAE}}(\hat{\theta}_{PL}) = \frac{1}{510}\sum_{i=1}^{510}\left|\hat{\theta}_{PL_i} - \theta_0\right|,$$

$$\widehat{\mathrm{MAE}}(\hat{\tau}_{PL}) = \frac{1}{510}\sum_{i=1}^{510}\left|\hat{\tau}_{PL_i} - \tau_0\right|.$$

(4) For the i^{th} "original" sample, $i = 1, \ldots, 510$, the generalized pseudolikelihood function under a $3 \times 3 \times 1$ group size was constructed and numerically maximized with respect to (α, θ, τ) to produce the maximum generalized pseudolikelihood estimates (MGPLEs) $(\hat{\alpha}_{GPL_i}, \hat{\theta}_{GPL_i}, \hat{\tau}_{GPL_i})$. The mean, standard deviation, bias, and mean absolute error (MAE) of the parameter estimates were estimated in a manner analogous to the pseudolikelihood case above, yielding the sample means $(\bar{\hat{\alpha}}_{GPL}, \bar{\hat{\theta}}_{GPL}, \bar{\hat{\tau}}_{GPL})$, sample standard deviations $(\widehat{\mathrm{SD}}(\hat{\alpha}_{GPL}), \widehat{\mathrm{SD}}(\hat{\theta}_{GPL}), \widehat{\mathrm{SD}}(\hat{\tau}_{GPL}))$, sample biases of the sample means $(\widehat{\mathrm{Bias}}(\bar{\hat{\alpha}}_{GPL}), \widehat{\mathrm{Bias}}(\bar{\hat{\theta}}_{GPL})\widehat{\mathrm{Bias}}(\bar{\hat{\tau}}_{GPL}))$, and sample mean absolute errors $(\widehat{\mathrm{MAE}}(\hat{\alpha}_{GPL})$, $\widehat{\mathrm{MAE}}(\hat{\theta}_{GPL})$, $\widehat{\mathrm{MAE}}(\hat{\tau}_{GPL}))$.

(5) For the i^{th} "original" sample, $i = 1, \ldots, 510$, the block generalized pseudolikelihood function under a $4 \times 4 \times 1$ block size was constructed and numerically maximized with respect to (α, θ, τ) to produce the maximum block generalized pseudolikelihood estimates (MBGPLEs) $(\hat{\alpha}_{BGPL_i}, \hat{\theta}_{BGPL_i}, \hat{\tau}_{BGPL_i})$. The mean, standard deviation, bias, and mean absolute error (MAE) of the parameter estimates were estimated in a manner analogous

to the pseudolikelihood case above, yielding the sample means $(\bar{\hat{\alpha}}_{BGPL}, \bar{\hat{\theta}}_{BGPL}, \bar{\hat{\tau}}_{BGPL})$, sample standard deviations $(\widehat{\text{SD}}(\hat{\alpha}_{BGPL}), \widehat{\text{SD}}(\hat{\theta}_{BGPL}), \widehat{\text{SD}}(\hat{\tau}_{BGPL}))$, sample biases of the sample means $(\widehat{\text{Bias}}(\bar{\hat{\alpha}}_{BGPL}), \widehat{\text{Bias}}(\bar{\hat{\theta}}_{BGPL}), \widehat{\text{Bias}}(\bar{\hat{\tau}}_{BGPL}))$, and sample mean absolute errors $(\widehat{\text{MAE}}(\hat{\alpha}_{BGPL}), \widehat{\text{MAE}}(\hat{\theta}_{BGPL}), \widehat{\text{MAE}}(\hat{\tau}_{BGPL}))$.

(6) For the i^{th} "original" sample, $i = 1, \ldots, 510$, the Gibbs sampler was used to generate 550 Monte Carlo samples from a first-order isotropic autoregressive lag-one spatio-temporal autologistic distribution with parameter vector $\boldsymbol{\phi} = (\hat{\alpha}_{PL_i}, \hat{\theta}_{PL_i}, \hat{\tau}_{PL_i})$. A burn-in period of 100 full sweeps of the array was implemented for convergence and after each subsequent full sweep of the array, the resultant realization was retained as one Monte Carlo sample.

(7) For $i = 1, \ldots, 510$, the i^{th} "original" sample and the corresponding 550 Monte Carlo samples (from step (6)) were then used to construct the Monte Carlo approximate negative log likelihood function, which was then numerically maximized to obtain the corresponding MCMCMLEs $(\hat{\alpha}_{MC_i}, \hat{\theta}_{MC_i}, \hat{\tau}_{MC_i})$. The mean, standard deviation, bias, and mean absolute error of the parameter estimates were then computed in a manner analogous to the pseudolikelihood case above, yielding the sample means $(\bar{\hat{\alpha}}_{MC}, \bar{\hat{\theta}}_{MC}, \bar{\hat{\tau}}_{MC})$, sample standard deviations $(\widehat{\text{SD}}(\hat{\alpha}_{MC}), \widehat{\text{SD}}(\hat{\theta}_{MC}), \widehat{\text{SD}}(\hat{\tau}_{MC}))$, sample biases of the sample means $(\widehat{\text{Bias}}(\bar{\hat{\alpha}}_{MC}), \widehat{\text{Bias}}(\bar{\hat{\theta}}_{MC}), \widehat{\text{Bias}}(\bar{\hat{\tau}}_{MC}))$, and sample mean absolute errors $(\widehat{\text{MAE}}(\hat{\alpha}_{MC})$, $\widehat{\text{MAE}}(\hat{\theta}_{MC})$, $\widehat{\text{MAE}}(\hat{\tau}_{MC}))$.

The simulation procedure for the first-order isotropic autoregressive lag-one spatio-temporal autologistic model, with a covariate term, is analogous to the above procedure for both covariate types and, therefore, will not be provided.

6.7.3 Results

As with the purely spatial simulation study of Chapter 5, the performances of the four estimation methods in the space-time domain for this small-scale simulation study are assessed via evaluations and comparisons of their respective measures of bias, precision, and accuracy, via the MAE. Based on the conclusions that can ultimately be drawn, the analysis is separated below into two parts. The first part addresses the two scenarios considered in the study that do not incorporate a covariate term, while the second part addresses the two scenarios considered in the study that do incorporate a covariate term.

For the two scenarios, or rather parameter vectors, considered under a first-order isotropic autoregressive lag-one spatio-temporal autologistic model, without a covariate term, the sample means and corresponding standard errors of the estimates obtained from each of the four methods are given in Table 6.1. From this table, we highlight two observations of note. First, the sample means for all three parameters from all four estimation methods appear to be relatively close to the truth, relative to their corresponding standard errors. For example, for the $(\alpha, \theta, \tau)' = (-1.0, 0.4, 0.5)$ scenario, GPL yielded $(\bar{\widehat{\alpha}}, \bar{\widehat{\theta}}, \bar{\widehat{\tau}})' = (-0.9875, 0.3975, 0.4886)$ with $(\widehat{SD}(\widehat{\alpha}), \widehat{SD}(\widehat{\theta}), \widehat{SD}(\widehat{\tau}))' = (0.0410, 0.0190, 0.0263)$. This suggest that the extensions of PL, MCMCML, GPL, and BGPL into the space-time domain presented above in sections 6.3-6.6 are indeed useful in estimating the parameters of the spatio-temporal autologistic model. Second, for the stronger temporal dependence scenario ($\tau = 1.5$), both the biases and standard errors for all 4 methods are larger than for the smaller temporal dependence scenario ($\tau = 0.5$). For example, when using PL under the stronger temporal dependence scenario, the estimated bias in estimating α is nearly 4.5 times larger (0.0705 vs. 0.0157), while the corresponding standard error is nearly 2 times larger (0.0800 vs. 0.0415), than when using PL under the weaker temporal dependence scenario. This is not surprising since an increase in the strength of temporal dependence will tend to produce arrays that deviate further from a 50/50 split of zeros and ones, which we know from section 1.4 and Chapter 5 are associated

Table 6.1: Sample Means and Standard Errors of α, θ, and τ when the Model is FIAR1 (No Covariate) and the Methods of Estimation are PL, GPL, BGPL, and MCMCML

		50x50x11			
Model	Estimate	PL	GPL	BGPL	MCMCML
	$\bar{\hat{\alpha}}$	-0.9843	-0.9875	-0.9843	-0.9952
	$\widehat{\text{SD}}(\hat{\alpha})$	0.0415	0.0410	0.0428	0.0424
FIAR1	$\bar{\hat{\theta}}$	0.3958	0.3975	0.3963	0.3982
$(-1.0, 0.4, 0.5)$	$\widehat{\text{SD}}(\hat{\theta})$	0.0191	0.0190	0.0197	0.0190
	$\bar{\hat{\tau}}$	0.4890	0.4886	0.4869	0.4949
	$\widehat{\text{SD}}(\hat{\tau})$	0.0255	0.0263	0.0257	0.0277
	$\bar{\hat{\alpha}}$	-0.9295	-0.9600	-0.9427	-0.9737
	$\widehat{\text{SD}}(\hat{\alpha})$	0.0800	0.0859	0.0890	0.0913
FIAR1	$\bar{\hat{\theta}}$	0.3816	0.3916	0.3866	0.3926
$(-1.0, 0.4, 1.5)$	$\widehat{\text{SD}}(\hat{\theta})$	0.0256	0.0271	0.0284	0.0274
	$\bar{\hat{\tau}}$	1.4808	1.4815	1.4783	1.4885
	$\widehat{\text{SD}}(\hat{\tau})$	0.0373	0.0376	0.0377	0.0416

with more biased and more variable parameter estimates.

With the MPLEs serving as the baseline, the four sets of relative MAEs, for both scenarios, under a first-order isotropic autoregressive lag-one spatio-temporal autologistic model, without a covariate term, are given in Table 6.2. From this table, we infer the following. In particular, first note that for the weaker temporal dependence scenario ($\tau = 0.5$), where the array has a roughly 48/52 split of zeros and ones, the relative MAEs between the four methods, across all three parameters, do not appear to be substantially different. For example, the best and worst MAEs are only 0.9610 and 1.0548, respectively, times their corresponding PL MAEs. However, for the stronger temporal dependence scenario ($\tau = 1.5$), where the array has a roughly 21/79 split of zeros and ones, the relative MAEs for $\hat{\alpha}$ and $\hat{\theta}$ corresponding to GPL and MCMCML appear to be substantially less, i.e. better, than the corresponding MAEs for PL and BGPL. For example, the MAEs of $\hat{\alpha}$ for GPL and MCMCML are only approximately 0.862 and 0.855, respectively, times the MAE corresponding to PL. Hence, this suggests that, as in the purely spatial case, the further the proportion of ones on the array increases above

Table 6.2: Relative MAEs of α, θ, and τ when the Model is FIAR1 (No Covariate) and the Methods of Estimation are PL, GPL, BGPL, and MCMCML

		50x50x11			
Model	Estimate	PL	GPL	BGPL	MCMCML
FIAR1	Rel MAE($\hat{\alpha}$)	1	0.9610	1.0383	0.9699
$(-1.0, 0.4, 0.5)$	Rel MAE($\hat{\theta}$)	1	0.9739	1.0382	0.9815
	Rel MAE($\hat{\tau}$)	1	1.0548	1.0424	1.0053
FIAR1	Rel MAE($\hat{\alpha}$)	1	0.8623	0.9885	0.8554
$(-1.0, 0.4, 1.5)$	Rel MAE($\hat{\theta}$)	1	0.8826	0.9925	0.8754
	Rel MAE($\hat{\tau}$)	1	0.9941	1.0370	1.0152

0.50, the greater the improvement both GPL and MCMCML demonstrate relative to PL. In addition to this inference, it is also worth noting that, just as in the purely spatial simulations (without a covariate term) of Chapter 5, in situations where the array has a relatively high proportion of ones, GPL appears to not only achieve its compromise between PL and GPL, but it also appears to even be competitive with MCMCML. Consistent with the results of Chapter 5, this does not appear to be the case for BGPL, which is generally only competitive, at best, with PL.

For the two scenarios, or rather covariate types, considered under a first-order isotropic autoregressive lag-one spatio-temporal autologistic model, with a covariate term, the sample means and corresponding standard errors of the estimates obtained from each of the four methods are given in Table 6.3. The main thing to note from this table is that the sample means, for all but the temporal parameter of the fire covariate scenario, from all four estimation methods appear to be relatively close to the truth, relative to their corresponding standard errors. For example, for the $(\alpha, \theta, \tau, \beta)' = (-2.0, 0.8, 0.3, 0.5)$ scenario, BGPL yielded $(\bar{\hat{\alpha}}, \bar{\hat{\theta}}, \bar{\hat{\tau}}, \bar{\hat{\beta}})' = (-1.9966, 0.7997, 0.2745, 0.5198)$ with $(\widehat{SD}(\hat{\alpha}), \widehat{SD}(\hat{\theta}), \widehat{SD}(\hat{\tau}), \widehat{SD}(\hat{\beta}))' = (0.0473, 0.0174, 0.0270, 0.0470)$. Hence, as in the covariate-free scenarios above, this suggests that the extensions of PL, MCMCML, GPL, and BGPL into the space-time domain are indeed useful for estimating the parameters of the spatio-temporal autologistic model, even when a covariate term has been incorporated into the model. Note that the one inconsistency with

the above statement, as already alluded to, is the temporal component from the fire covariate scenario. In particular, the MCMCML estimate, for example, of $\tau = 0.015$ is roughly 2.7 times larger than the truth ($\hat{\tau} = 0.0408$), and the corresponding standard error of 0.5420 is relatively large. This "abnormality" is most likely attributable to at least one of the following two conjectures. First, perhaps the autoregressive lag-one temporal parameter for the fire occurrence data from Oregon/Washington is insignificant in the model in the presence of the spatial and covariate effects. Second, as $< 1\%$ of the array's sites have a value of 1, the array is extremely disproportionate and this too could be affecting the estimates of the temporal parameter. Note that in this extreme situation, 79 of the 510 runs of the simulation procedure produced estimates that were orders of magnitude from the truth; such estimates were removed from the analysis. Hence, the plausible insignificance of this temporal parameter in the model, along with the extremely disproportionate number of zero-valued sites on the array, could explain why all four methods preform relatively poorly in estimating the value of this parameter.

With the MPLEs serving as the baseline, the four sets of relative MAEs, for both covariate types, under a first-order isotropic autoregressive lag-one spatio-temporal autologistic model, with a covariate term, are given in Table 6.4. From this table, it is arguably evident that the relative MAEs between the four methods, across all four parameters, do not appear to be substantially different. In particular, while for the Uniform(0,1) covariate scenario the relative MAE values of 1.0696 and 0.9307 for $\hat{\beta}_{GPL}$ and $\hat{\theta}_{GPL}$, respectively, might be substantially different from the MAEs for the corresponding MPLEs, the differences in relative MAEs for the fire covariate scenario estimates seem negligible. For example, the best and worst MAEs for the fire covariate scenario are only 0.9907 and 1.0483, respectively, times their corresponding PL MAEs. Hence, PL appears to be a good choice for parameter estimation, especially in the extreme fire covariate scenario. We suspect that in the space-time domain, as in the purely spatial domain, GPL and MCMCML will demonstrate more substantial performance-based improvements relative to PL in situations where the proportion of ones in the array is

Table 6.3: Sample Means and Standard Errors of α, θ, τ, and β (Unif(0,1) or Fire (DA) Covariate) when the Model is FIAR1 (With a Covariate) and the Methods of Estimation are PL, GPL, BGPL, and MCMCML

		50x50x11			
Model	Estimate	PL	GPL	BGPL	MCMCML
	$\bar{\hat{\alpha}}$	-1.9844	-2.0005	-1.9966	-1.9937
Unif(0,1)	$\hat{\text{SD}}(\hat{\alpha})$	0.0488	0.0479	0.0473	0.0510
Covariate	$\bar{\hat{\theta}}$	0.7977	0.8004	0.7997	0.7984
	$\hat{\text{SD}}(\hat{\theta})$	0.0185	0.0172	0.0174	0.0182
	$\bar{\hat{\tau}}$	0.2763	0.2787	0.2745	0.2846
FIAR1	$\hat{\text{SD}}(\hat{\tau})$	0.0270	0.0271	0.0270	0.0327
$(-2.0, 0.8, 0.3, 0.5)$	$\bar{\hat{\beta}}$	0.5012	0.5206	0.5198	0.5007
	$\hat{\text{SD}}(\hat{\beta})$	0.0483	0.0477	0.0470	0.0472
	$\bar{\hat{\alpha}}$	-6.8741	-6.8822	-6.8807	-6.8697
Fire (DA)	$\hat{\text{SD}}(\hat{\alpha})$	0.6309	0.6364	0.6325	0.6309
Covariate	$\bar{\hat{\theta}}$	1.1614	1.1600	1.1605	1.1594
	$\hat{\text{SD}}(\hat{\theta})$	0.3151	0.3167	0.3226	0.3119
	$\bar{\hat{\tau}}$	0.0457	0.0296	0.0432	0.0408
FIAR1	$\hat{\text{SD}}(\hat{\tau})$	0.5330	0.5851	0.5331	0.5420
$(-6.87, 1.2, 0.015, 0.023)$	$\bar{\hat{\beta}}$	0.0230	0.0231	0.0231	0.0230
	$\hat{\text{SD}}(\hat{\beta})$	0.0070	0.0071	0.0070	0.0070

somewhere roughly between 0.2 and 0.8.

In conclusion, the results of this small-scale spatio-temporal autologistic model simulation study ultimately suggested two important outcomes. First, the extensions of PL, MCMCML, GPL, and BGPL into the space-time domain, as described in the preceding sections of this chapter, appear to be useful for estimating the parameters of the spatio-temporal autologistic model. Second, and finally, the relative performances of these four methods, under the limited scope of this space-time study, appeared to be fairly consistent with the results obtained from the large-scale purely spatial simulation study of Chapter 5. In particular, in situations of moderate to extreme spatial and/or temporal dependency, GPL and MCMCML (and occasionally BGPL) are the best methods, but in situations of weak spatial and/or temporal dependency, or aberrantly extreme spatial and/or temporal dependency, PL is no worse

Table 6.4: Relative MAEs of α, θ, τ, and β (Unif(0,1) or Fire (DA) Covariate) when the Model is FIAR1 (With a Covariate) and the Methods of Estimation are PL, GPL, BGPL, and MCMCML

		50x50x11			
Model	Estimate	PL	GPL	BGPL	MCMCML
Unif(0,1)	Rel MAE($\hat{\alpha}$)	1	1.0164	1.0124	1.0094
Covariate	Rel MAE($\hat{\theta}$)	1	0.9307	0.9319	0.9610
FIAR1	Rel MAE($\hat{\tau}$)	1	0.9544	1.0584	0.9621
$(-2.0, 0.8, 0.3, 0.5)$	Rel MAE($\hat{\beta}$)	1	1.0696	1.0483	0.9841
Fire (DA)	Rel MAE($\hat{\alpha}$)	1	1.0014	1.0011	0.9993
Covariate	Rel MAE($\hat{\theta}$)	1	1.0059	1.0294	0.9907
FIAR1	Rel MAE($\hat{\tau}$)	1	1.0483	0.9995	1.0169
$(-6.87, 1.2, 0.015, 0.023)$	Rel MAE($\hat{\beta}$)	1	1.0180	1.0048	1.0025

than the other three methods. Ultimately, however, a more extensive space-time simulation study must be conducted if such conjectures are to be validated. Furthermore, it would be of interest to explore in such a simulation study how adding "depth" to both the group and block structures would affect the performances of GPL and BGPL, particularly with respect to estimating any temporal parameters. For this to be accomplished, more efficient code needs to be developed to counteract the increased computational complexity necessary to do this.

Chapter 7

Spatio-Temporal Autologistic Model Application

7.1 Introduction

Throughout this dissertation, references to a wild fire occurrence data set from regions of Oregon and Washington state have been made. In fact, the essence of this data set was used in Chapter 1 to help motivate the need for the autologistic model. Furthermore, aspects of this data set were incorporated into the simulation studies presented in Chapters 5 and 6. In this chapter, a subset of these data will be modeled. The objectives of such modeling are three-fold. First, we want to demonstrate the use of the autologistic model on an actual, rather than simulated, data set. The methods of PL, MCMCML, GPL, and BGPL will all be employed to estimate the corresponding model parameters. Second, since in practice the truth is never known, we want to present an approach to model selection for situations when the autologistic model is being employed. Such an approach is a hybrid of the approaches suggested by Gumpertz et al. ([19]) and Wu and Huffer ([35]). Third, and finally, we want

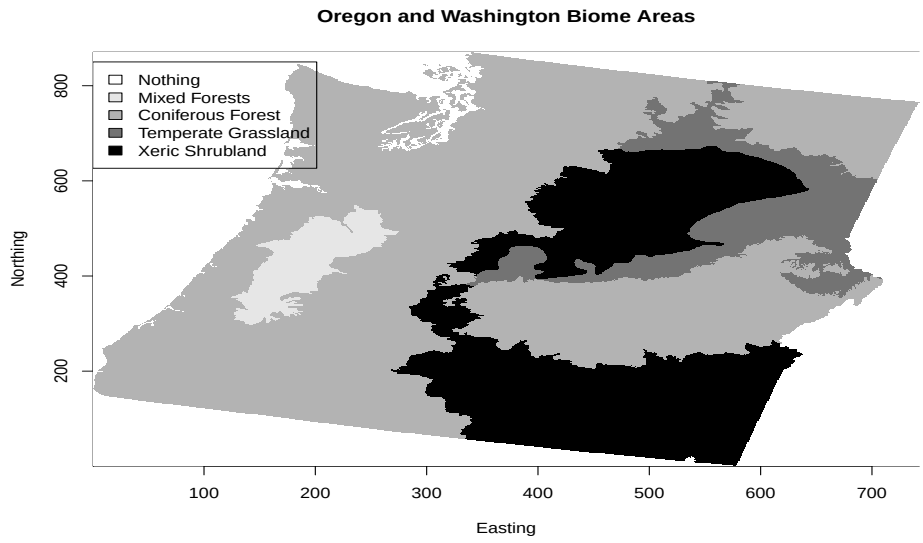
to better understand what affects the pattern (presence/absence) of wild fire occurrence in Oregon and Washington. More specifically, we want to know if accounting for the effects of space, time, and/or the Departure from Average covariate (see section 7.2) will provide us with a "better" understanding of the wild fire occurrence pattern in these two states.

The remainder of this chapter is divided into three sections. In section 7.2, a detailed description of the fire occurrence data from Oregon and Washington is provided. In section 7.3, the model fitting process is explained along with the aforementioned model selection procedure. Note that the parametric bootstrap procedure, which has been alluded to throughout this dissertation as a means for obtaining the standard errors of the MPLEs, MGPLEs, and MBGPLEs, is also presented in this section. Finally, in section 7.4, the results of the model fitting process are presented along with the subsequent interpretations.

7.2 Description of Data Set

The data for the analysis carried out in this chapter were graciously shared by Cindy Leary, a researcher at The University of Montana, during the course of her work with the Fire Sciences Laboratory in Missoula, Montana. The description of the data set given throughout this section is a summary of the more exhaustive description given by Leary et al. ([26]). The spatial domain of the data set consists of an 870 km by 740 km regular lattice comprising Oregon and Washington state. This regular lattice, as displayed in Figure 7.1, consists of four distinct terrestrial ecosystems including Temperate Broadleaf and Mixed Forests, Temperate Coniferous Forests, Temperate Grasslands, Savannas, and Shrublands, and finally, Deserts and Xeric Shrublands. The temporal domain of the data set covers the months between May and October for each of the four years from 2002 to 2005. More specifically, over the span of these six months, for each of the four years, data were collected on the lattice at 11 distinct and equally-spaced times; thus, the temporal lag is roughly two weeks. Thus, there are 4 data

Figure 7.1: Four Terrestrial Ecosystems of Oregon and Washington



arrays, each consisting of 11 cross-sectional readings of the lattice. It should be pointed out here that as annual wild fire occurrence values are believed to be unrelated, these 4 arrays are viewed as four distinct data sets rather than as one massive data set.

The response variable for this data set is fire ignition, recorded as the presence or absence of fire. More specifically, each 1 km^2 pixel (or site) of the lattice has a response value of 1 at a given time point if any location within that 1 km^2 pixel experiences fire ignition at some instance between the last time point and the current time point, and a value of 0 otherwise. These response values were constructed by first "merging state and federal fire occurrence records for Oregon and Washington into a single historical database" ([26]) and then converting these data into arrays of binary values.

The only covariate considered in this data set is the Departure from Average (DA) variable, which is a MODIS-derived (Moderate Resolution Imaging Spectroradiometer) fire potential metric. As was mentioned in Chapter 1, this covariate is essentially a surrogate for the relative change in vegetation across the sites of the lattice over time. In particular, let $EVI_{i,t}$ denote

the Enhanced Vegetation Index of pixel i at time t, where i denotes one of the 1 km^2 pixels of the lattice and t denotes one of the 11 readings collected annually. Additionally, let $EVI_{i,t}^{avg}$ denote the historical average Enhanced Vegetation Index of pixel i at time t. Then the DA value for site i at time t is given by the following:

$$DA_{i,t} = 100 + (EVI_{i,t} - EVI_{i,t}^{avg}) * 100, \tag{7.1}$$

where all relevant Enhanced Vegetation Index (EVI) values are obtained from MODIS satellite platform data, which are available through NASA's Land Processes Distributed Active Archive Center (LP DAAC). Leary et al. convey that "DA values generally range from 0 to 200 where values of 100 indicate that a site is equal to the historic average (EVI) for that time period and values closer to 200 indicate that a site appears much greener than usual" ([26]). It is believed that lower DA values are associated with an increased probability of fire occurrence as they indicate "dryer than usual" conditions.

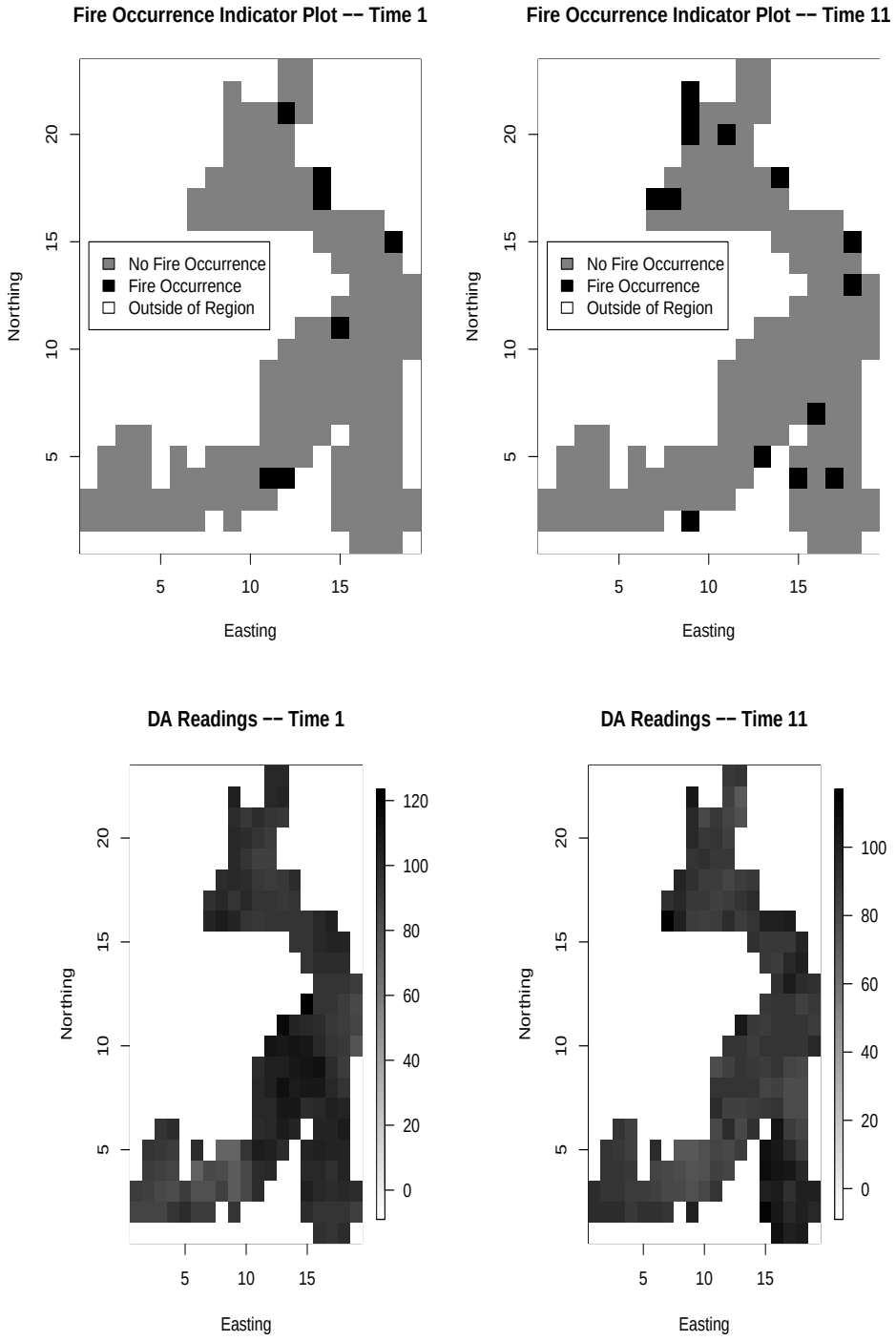
While it is theoretically possible to model this entire data set at a 1 km^2 resolution, practically speaking, this is not reasonable. More specifically, at such a resolution, the proportion of pixels with a response value of 1 is substantially less than 1%, which, as discussed in previous chapters of this dissertation, makes the parameter estimates for the autologistic model highly variable. Thus, a 20 km^2 resolution was utilized so as to produce a data set with a greater proportion of pixels having a response value of 1. Such a scale remains practical from the perspective of a forest manager seeking to acquire resoures to combat wild fires. Note that finer resolutions did not sufficiently increase the proportion of pixels with a response value of 1, while coarser resolutions reduced the number of sites to an unacceptable level. Under a 20 km^2 resolution (similar to a 1 km^2 resolution) a pixel, or site, has a response value of 1 at a given time point if any location within that 20 km^2 pixel experiences fire ignition at some instance between the last time point and the current time point, and a value of 0 otherwise. The DA value for a pixel under a 20 km^2 resolution is the median DA value of the 400 1 km^2 DA values enclosed in the 20 km^2 pixel, all 400 of which are obtained using (7.1).

For the purposes of this dissertation, only the temperate grasslands terrestrial ecosystem will be modeled since, among the four ecosystems, it had the greatest proportion of sites with a response value of 1. Furthermore, of the four years of data collected for this ecosystem, only the first year (2002) will be modeled since, among the four years, it had the greatest proportion of sites with a response value of 1. Note that this subset of the Oregon and Washington data set, under a 20 km^2 resolution, is contained within a 19×23 lattice, but only 170 of the pixels in this lattice make up the temperate grasslands ecosystem. Hence, the data array being modeled in this chapter consists of 170 sites observed at 11 time points. Fire occurrence indicator plots of this ecosystem for the first and last time points of 2002 are displayed in Figure 7.2, along with the corresponding DA plots. From these plots, it appears that fire occurrence is more prevalent at the end of the fire season (Time 11) than at the beginning (Time 1) of the fire season, and that the temperate grasslands ecosystem is dryer at the end of the fire season than it was at the beginning of the fire season. It is worth mentioning that intermediate weeks had even higher proportions of ignitions, ranging as high as $\frac{46}{170} \approx .271$ at Time 5. Note that of the $170 * 11 = 1870$ sites making up the data array for the temperate grasslands ecosystem, 237 of them, or $\frac{237}{1870} \approx 0.1267$, have a response value of 1.

7.3 Approach to Model Fitting and Model Selection

The process of model selection when employing the spatio-temporal autologistic model is not routine since the corresponding likelihood function is almost always intractable. Hence, the process suggested here is a hybrid of the processes suggested by Gumpertz et al. ([19]) and Wu and Huffer ([35]). In particular, once any number of possible models are initially considered, PL can be used to estimate the parameters for each of these initial models. As PL is extremely computationally cheap, especially relative to other methods of parameter estimation, obtaining the MPLEs for each of the possible models poses no real difficulties. Using the standard output that is associated with any standard statistical software package, such as R, each model can

Figure 7.2: Fire Occurrence and Departure from Average (DA) Readings for the Temperate Grasslands Ecosystem – Cross-Sections 1 and 11 of 2002



then be evaluated by checking the significance of each of the parameters in the respective model. Note that while the standard errors used to assess such parameter significance are usually incorrect, as discussed in previous chapters of this dissertation, they can still be used as a rough guide to determine a subset of "best" models from the initial set of models. Gumpertz et al. ([19]) even suggest employing an AIC-like measure (PL-AIC), using the pseudolikelihood function in lieu of the true likelihood function, to determine the aforementioned "best" subset of models for further consideration. The approach implemented for this application uses both PL and PL-AIC to determine the "best" subset of models for further examination. After employing the above process to reduce the number of viable models down to just a few potential models, the MGPLEs, MBGPLEs, and MCMCMLEs could then be obtained. Note that as these estimates are computationally more expensive to obtain than the MPLEs, it is best in terms of computational expense if relatively few models are considered at this stage of the fitting process.

For each model under consideration, and for each of the four sets of parameter estimates for that model, a Monte Carlo approach can then be used to determine the "best-fitting" model. This approach, based on a sum of absolute error (SAE), was first proposed by Wu and Huffer ([35]) and will be employed in this application as a further means for selecting the "best" model. To determine the SAE for a particular model and a particular set of parameter estimates, first let $\hat{\boldsymbol{\psi}}$ denote a set of parameter estimates for one of the models. Then with the unknown parameter vector set equal to $\hat{\boldsymbol{\psi}}$, the Gibbs sampler is employed to generate N samples, or realizations, of the model. Next, the average realization for each site is calculated using these N samples, i.e. let $\hat{p}_{i,t}$ be the average value of the N realizations for site i at time t. Note that $\hat{p}_{i,t}$ can also be thought of as the estimated probability of a fire occurrence at site i at time t. Finally, with $z_{i,t}$ denoting the observed fire occurrence value at site i at time t, the sum of absolute error is calculated as follows:

$$SAE = \sum_{\text{All Array Sites}} |\hat{p}_{i,t} - z_{i,t}|. \tag{7.2}$$

Observe that the SAE is ultimately just an aggregate measure of the fitted errors, or rather the differences between the unconditional predicted values and the observed data. The SAE for each model and for each set of parameter estimates is then obtained in a manner analogous to the above. In this way, the SAE can be used to compare models and estimation methods for a given set of data. Note that instead of using the SAE, Gumpertz et al. ([19]) suggest using cross-validation, but this was not considered here.

To determine the standard errors of any estimates obtained using PL, GPL, or BGPL, a parametric bootstrap can be employed. For ease of presentation, assume we want the standard errors for a set of MPLEs, $\hat{\boldsymbol{\psi}}_{PL}$. Then the parametric bootstrap procedure entails using the Gibbs sampler (see sections 3.3.2 and 3.3.3), with the unknown parameter vector set equal to $\hat{\boldsymbol{\psi}}_{PL}$, to generate K realizations of the data array. For each of these K realizations, PL is used to estimate the model parameters. The standard deviations of the MPLEs from these K sets of estimates are then calculated to obtain the standard errors of the original MPLEs, i.e. SE($\hat{\boldsymbol{\psi}}_{PL}$). The parametric bootstrap procedure for obtaining the standard errors of the MGPLEs, or MBGPLEs, is analogous to the above procedure for the MPLEs.

Before presenting the results of the model fitting process for the temperate grasslands ecosystem (see section 7.4), the type of spatial and temporal edge-adjustments implemented for this data set should be discussed. First, with respect to the spatial edges, since the temperate grasslands ecosystem is rather irregular in shape and since we have no reason to believe, for example, that fire occurrence on the northern edge of the region is similar to fire occurrence on the southern edge of the region, a toroidal spatial edge adjustment was not considered. Furthermore, since the temperate grasslands ecosystem consists of only 170 sites, we decided that implementing a guard-region spatial edge adjustment would be too expensive, from an information point of view. Hence, we have employed a weighted spatial edge adjustment procedure (see section 2.2.4) in modeling the temperate grasslands ecosystem data set. Finally, with respect to the temporal edge (recall from Chapter 6 that we are assuming a past-lags-only temporal Markov assumption), since we have no reason to believe that the fire occurrence at

the later time steps is similar to the fire occurrence at the initial time steps, a toroidal temporal edge adjustment was not considered. Furthermore, since we have absolutely no information on the temporal neighbors of sites from the first time step, it is impossible to employ a weighted temporal edge adjustment at this time point. Thus, we have employed a guard-region temporal edge adjustment procedure in modeling the temperate grasslands ecosystem data set, which means we are simply not modeling the first time point.

7.4 Results

In accordance with the process described above in section 7.3, the first step in modeling the temperate grasslands ecosystem data array was to determine an initial set of models to consider. For the purposes of this dissertation, only five such models were considered. The most complex of these models was a first-order isotropic autoregressive lag-one autologistic model with a term for the DA covariate. The conditional probability form of such a model, using the notation established in Chapter 6, is the following:

$$\mathrm{P}(Z_{i,t} = z_{i,t} | \{z_{j,t'} : (j,t') \in N_{i,t} \cup \Gamma_{i,t}\}, x_{i,t}; \boldsymbol{\psi}) \quad =$$

$$= \quad \frac{\exp\left\{\alpha z_{i,t} + \theta z_{i,t} \sum_{(j,t') \in N_{i,t}} z_{j,t'} + \beta z_{i,t} x_{i,t} + \tau z_{i,t} \sum_{(j,t') \in \Gamma_{i,t}} z_{j,t'}\right\}}{1 + \exp\left\{\alpha + \theta \sum_{(j,t') \in N_{i,t}} z_{j,t'} + \beta x_{i,t} + \tau \sum_{(j,t') \in \Gamma_{i,t}} z_{j,t'}\right\}}, \tag{7.3}$$

where in this case, $\Gamma_{i,t} = \{(i, t-1)\}$, $N_{i,t}$ consists of the four first-order isotropic spatial neighbors of site (i,t), and $x_{i,t}$ is the DA value for site i at time t. The other four models considered are all nested within the model given by (7.3), which we will henceforth refer to as Model 5. In particular, those four models consist of the following subset of parameters:

- Model 1: $\boldsymbol{\psi}(M1) = (\alpha, \beta)' \Rightarrow$ covariate effect only (no spatial or temporal effects);
- Model 2: $\boldsymbol{\psi}(M2) = (\alpha, \theta)' \Rightarrow$ spatial effect only (no covariate or temporal effects);
- Model 3: $\boldsymbol{\psi}(M3) = (\alpha, \theta, \beta)' \Rightarrow$ spatial and covariate effects (no temporal effect);
- Model 4: $\boldsymbol{\psi}(M4) = (\alpha, \beta, \tau)' \Rightarrow$ temporal and covariate effects (no spatial effect).

Note that more complex models than Model 5, such as models allowing for spatial anisotropy or higher-order spatial and/or temporal dependencies, were not considered due to the limited size of the data set.

The next step in the model fitting process was to use PL to obtain parameter estimates and PL-AIC values for Models 1 through 5. The resulting PL-AIC values for the five models are given in Table 7.1. From this table, it is evident that Model 5 is the "best" model, according to PL-AIC, and that Model 3 is a distant second "best." In terms of tests of parameter significance, for each model including a spatial effect parameter (Models 2, 3, and 5), the corresponding test of significance produced a p-value of less than 0.0001. Hence, fire occurrence in the temperate grasslands ecosystem is significantly spatially correlated at the 20-km resolution. Furthermore for each of the two models including a temporal effect parameter (Models 4 and 5), the corresponding test of significance produced a p-value greater than 0.1. Thus, there is no evidence of temporal correlation in fire occurrence in the temperate grasslands ecosystem. Finally, in each of the four models including a DA covariate parameter (Models 1, 3, 4, and 5), the corresponding test of significance produced a p-value less than 0.1. It is worth noting, however, that the aforementioned p-values for the DA covariate parameter were less than 0.002 for both models not including a temporal effect (Models 1 and 3). Therefore, it appears that fire occurrence in the temperate grasslands ecosystem is dependent on the DA covariate, but the strength of this relationship may depend on whether a temporal component is included. The above results indicate that any model considered further must certainly have a spatial dependence parameter and a DA covariate parameter.

Table 7.1: PL-AIC Values for the Five Fire Application Models

Model	M1	M2	M3	M4	M5
PL-AIC	1410.4	1322.1	1314.0	1325.7	1247.1

Of the 5 models considered, only Models 3 and 5 satisfy this condition, the models with the lowest PL-AIC values. Thus, even though the significance of the temporal effect parameter is questionable, both Models 3 and 5 were considered for further examination, which coincides with what the PL-AIC values in Table 7.1 would suggest. It is worth mentioning here that while PL-AIC was used as a component of this analysis, little is understood regarding its use.

Having isolated Models 3 and 5 as two potential models to consider further, the methods of GPL, BGPL, and MCMCML were all then used to obtain separate estimates for the parameters of both models. Before presenting the corresponding estimates, however, we must first discuss a couple of issues involved in employing these three additional estimation methods with the temperate grasslands ecosystem data array. First, for this application the group and block structures implemented for GPL and BGPL, respectively, were ultimately determined by the shape of and the relatively limited number of sites making up the temperate grasslands ecosystem. In particular, the group structure ultimately used to model these data was neither the cross-shape nor the 3×3 shape used in the simulation studies of Chapters 5 and 6 since such group structures produced relatively few groups with a complete set of sites within the region of interest. Instead, the three-site L-shaped group structure displayed in Figure 3.1 was employed as it yielded the best compromise between the largest group size and the greatest number of complete groups. Recognize that this does not imply that the spatial dependence only exists in particular orientations from a given site; all 3 sites within this L-shaped group have their complete neighborhoods. Also, note that there was no "depth" to this group structure, i.e. a group does not contain sites from different time steps. Similarly, a block size of $2 \times 2 \times 1$ was used for BGPL since it yielded the best compromise between the largest block

size and the greatest number of complete blocks. Second, and finally, the MCMCML algorithm did not converge with the MPLEs as the reference point (i.e. $\boldsymbol{\phi} = \hat{\boldsymbol{\psi}}_{PL}$) for the Gibbs sampler for either Model 3 or Model 5. Fortunately, the MCMCML algorithm did converge for Model 3 when the MGPLEs, and the MBGPLEs for that matter, were instead used as the reference point for the Gibbs sampler. However, the MCMCML algorithm still did not converge for Model 5 when the MGPLEs or MBGPLEs were instead used as the reference point for the Gibbs sampler. Hence, an additional approach was needed to get a reference point for the Gibbs sampler that would yield the MCMCMLEs for Model 5. The approach that we ultimately employed was essentially the approach suggested and demonstrated by Wu and Huffer ([35]). In particular, Wu and Huffer suggest adjusting the initial choice for $\boldsymbol{\phi}$ such that the observed sufficient statistics fall within the data cloud produced by the sufficient statistics resulting from the sweeps of the Gibbs sampler. Such an adjustment ultimately entailed using our intuition to guide a guess-and-check procedure, but eventually we were able to achieve convergence of the MCMCML algorithm and, therefore, obtain the MCMCMLEs for Model 5.

The parameter estimates obtained for Models 3 and 5, from all four estimation procedures, are given in Table 7.2, along with the corresponding standard errors. Note that the standard errors for the MPLEs, MGPLEs, and MBGPLEs were all obtained using a 500-sweep parametric bootstrap, after 100 burn-in sweeps, as described in section 7.3. From the table, several things are evident for Model 3. First, the standard errors for the MCMCMLEs are uniformly the smallest. Second, while all four methods produced comparable estimates for the three parameters, the PL estimate of θ is noticeably smaller than the corresponding estimates from the other three methods. Third, and finally, with the one exception being the BGPL estimate of β, the absolute value of a parameter estimate, across all four methods for all three parameters, is more than two standard errors from zero. Hence, appealing to the asymptotic normality of the MPLEs and MCMCMLEs, and possibly erroneously assuming asymptotic normality of the MGPLEs and MBGPLEs, this suggests that there is a consensus among the

four methods that each of the parameters in Model 3 is statistically significant. From Table 7.2, several things are also evident with respect to Model 5. First, the standard errors for the MPLEs and MCMCMLEs are both uniformly larger than the standard errors for the MGPLEs and MBGPLEs, particularly for τ. A histogram of the 500 PL parametric bootstrap estimates of τ was bimodal with approximately half of the estimates for τ being centered around a value of -14, and the other half being centered around a value of 0, which explains why the $\text{SE}(\hat{\tau})$ is so much larger than the corresponding standard errors for the other three methods. Second, with the exception of the PL estimate of τ and the MCMCML estimate of θ, all four methods produced comparable estimates of the 4 parameters in the model. Third, and finally, once again appealing to asymptotic normality, while GPL and BGPL both suggest that each of the four model parameters is significant, MCMCML suggests that the spatial and temporal effects are significant, and PL suggests that none of the effects are significant in the model. Such discrepancies are surely attributable to the inclusion of the temporal parameter in Model 5 since the four methods produced largely comparable results for Model 3. Perhaps the temporal variable is correlated with, and thus competing with, the spatial variable and/or the covariate, causing instability in the model estimates.

To help assess the adequacy of the fit for each of the four sets of parameters, for both Models 3 and 5, the SAE, as described in section 7.3, was computed with $N = 500$. The resulting SAEs are given in Table 7.3. Note that the SAEs in this table were computed using only the sites of the region that were not treated as fixed (i.e. all sites from time steps 2-11) so that the SAEs from Models 3 and 5 were comparable. From Table 7.3, it is clear that for Model 3 the MGPLEs provided the "best" fit with respect to SAE. Similarly, it is also clear from the table that for Model 5 the MPLEs clearly provided the "best" fit with respect to SAE, with the MCMCMLEs a close second. However, as the Model 5 fit of the MPLEs is questionable, for reasons discussed in the preceding paragraph, we would be hesitant to consider such estimates as reliable. Thus, we would argue that for Model 5 the MCMCMLEs provide the "best" fit in terms of SAE. As for which model provides the overall best fit for the temperate grasslands

Table 7.2: Parameter Estimates and Standard Errors for Fire Application Models M3 and M5 where the Methods of Estimation are PL, GPL for a three-site L-shaped Group Size, BGPL for a 2x2x1 Block Size, and MCMCML

	Model 3			
Method	PL	GPL	BGPL	MCMCML
$\hat{\alpha}$	-4.4430	-4.2081	-3.9041	-4.0969
SE($\hat{\alpha}$)	0.5568	0.6388	0.8652	0.5196
$\hat{\theta}$	0.6260	0.8158	0.8742	0.8931
SE($\hat{\theta}$)	0.1121	0.1053	0.1334	0.1015
$\hat{\beta}$	0.0229	0.0179	0.0140	0.0169
SE($\hat{\beta}$)	0.0062	0.0071	0.0095	0.0058
	Model 5			
Method	PL	GPL	BGPL	MCMCML
$\hat{\alpha}$	-5.3912	-4.8143	-4.9408	-5.0132
SE($\hat{\alpha}$)	1.2920	0.7796	1.0434	1.8522
$\hat{\theta}$	0.5735	0.6189	0.6420	1.2441
SE($\hat{\theta}$)	0.4133	0.1268	0.1518	0.5137
$\hat{\beta}$	0.0197	0.0231	0.0245	0.0141
SE($\hat{\beta}$)	0.0145	0.0088	0.0115	0.0165
$\hat{\tau}$	0.0150	1.6891	1.6520	2.1766
SE($\hat{\tau}$)	6.6738	0.2198	0.2532	0.6039

Table 7.3: SAEs for Fire Application Models M3 and M5 where the Methods of Estimation are PL, GPL for a three-site L-shaped Group Size, BGPL for a 2x2x1 Block Size, and MCMCML

	Model 3			
Method	PL	GPL	BGPL	MCMCML
SAE	371.67	357.22	361.45	378.80
	Model 5			
Method	PL	GPL	BGPL	MCMCML
SAE	264.90	355.57	357.76	281.92

ecosystem data array, it depends on the criterion used. Model 5, with the MCMCMLEs as the estimated values of the parameters, and Model 3, with the MGPLEs as the estimated values of the parameters, appear to be two of the better models, but without a true cross validation of these models, it is difficult to determine a "best" model or even whether any of these models are very useful.

While we have determined that, among the five models considered, Model 5 is potentially the best, we should take a moment and interpret the MCMCMLEs of Model 5. In particular, since the autologistic model is just an intuitive extension of the logistic model, its parameters are interpreted in the same manner as the parameters of a logistic model. Thus, if we let $Y_{i,t}$ denote one of the regressor variables in the autologistic model for site i at time t , then the *odds ratio* for an increase of one unit in $Y_{i,t}$, holding the values of the other regressor variables constant, is the following:

$$\text{odds ratio} \quad = \quad \frac{\text{odds of presence if } Y_{i,t} = (y_{i,t} + 1)}{\text{odds of presence if } Y_{i,t} = y_{i,t}}. \tag{7.4}$$

Therefore, $\hat{\theta}_{MCMC} = 1.2441$ produces an odds ratio of $e^{1.2441} \approx 3.47$. In other words, for each additional first-order isotropic neighbor experiencing a fire occurrence, the odds of a fire occurrence increase by approximately 247%. Similarly, $\hat{\beta}_{MCMC} = 0.0141$ produces an odds ratio of $e^{0.0141} \approx 1.0142$. Hence, a ten unit increase in Departure from Average increases the odds of fire occurrence by approximately 15.14% ($e^{0.0141 \times 10} \approx 1.1514$). Note that this

indicates that the dryer the region becomes over the year, the less likely it is to experience a fire occurrence, which is contrary to our intuition. We are not suggesting here, however, that our intuition be labeled as incorrect since alternative explanations exist which can rectify these inconsistencies. For example, perhaps the dryness of the region is related to weather patterns over the area, which somehow influence fire occurrence, or there may be "swamping" or competition effects with the space or time variables. In other words, perhaps there is a confounding variable not accounted for that explains the inconsistency between our analysis and long-held beliefs. Finally, $\hat{\tau}_{MCMC} = 2.1766$ produces an odds ratio of $e^{2.1766} \approx 8.8163$, which means that if the lag-one temporal neighbor experiences a fire occurrence, the odds of fire occurrence increase by approximately 782%. Therefore, in conclusion, the analysis carried out in this chapter suggests that space, time, and the Departure from Average fire potential metric are all helpful in explaining the spread of fire occurrence across the temperate grasslands ecosystem of Oregon and Washington state.

Chapter 8

Concluding Remarks

To conclude the presentation of the research conducted for this dissertation, the novel contributions made over the course of the previous seven chapters, as well as several problems that remain either completely uninvestigated or unsatisfactorily explored, are summarized and discussed in this eighth and final chapter. There are five novel contributions made in this thesis to the topic of Markov Random Field (MRF) modeling, and particularly to Binary MRF modeling, as highlighted in Chapter 1. First, the need for a distinction between types of spatial neighbor relations when using GPL and BGPL was recognized and justified in this dissertation. More specifically, the spatial neighbor relations between two sites of any group/block cannot be treated in the same fashion as the spatial neighbor relations between a site of that group/block and a boundary site of that group/block and the conditional distribution forms were modified to reflect this. Second, in Theorem 4.1 the strong consistency of the MGPLEs and MBGPLEs was established for any finite and discrete state-space MRF. The proof of this property that was given in Chapter 4, for both GPL and BGPL, was a generalization of the proof of the MPLE's strong consistency given by Geman and Graffigne ([13]). Third, the first simulation study of the autologistic model comparing the newer and subtly different methods of GPL and BGPL was conducted. This large-scale simulation study, which also in-

corporated the well-established methods of PL and MCMCML, was conducted entirely in the spatial domain and demonstrated that GPL is not only usually superior to BGPL, but is also frequently competitive with MCMCML. This simulation study was also the first such study to simultaneously incorporate covariates, spatial-anisotropy, and higher-order neighborhood systems. Fourth, extensions of GPL and BGPL into the space-time domain for non-absorbing state models were proposed and validated via a comparison with PL and MCMCML through a small-scale simulation study using the spatio-temporal autologistic model. Fifth and finally, the spatio-temporal autologistic model was applied to the fire occurrence data of the temperate grasslands ecosystem of Oregon and Washington state. The methods of PL, GPL, BGPL, and MCMCML were all used to estimate the corresponding model parameters and an SAE procedure ([35]) was employed for model selection.

While this dissertation is both extensive and thorough, there remain topics within these pages that require further investigation. The most pertinent of these topics, at least in our opinion, along with a corresponding brief description, are given in the following list:

(1) **Implementation of more efficient programming languages**: As one of the primary motivations behind the methods of GPL and BGPL is achieving a compromise between PL and MCMCML, in terms of computational expense, programming these methods in other languages may be worthwhile. More specifically, as R was the only language used to carry out parameter estimation in this dissertation, GPL and BGPL can more efficiently be carried out in another language, such as C, which would allow for the use of larger group/block sizes when employing GPL/BGPL.

(2) **Exploration of group/block structures that have "depth"**: For this dissertation, all group/block structures employed, either via simulation or application, involved sites from the current time step only and, therefore, sites from previous time steps were only conditioned upon in forming the group/block joint likelihoods. In other words, the groups/blocks had no spatial or temporal "depth." Incorporating such "depth" into the

group/block structure was theoretically developed in Chapter 6, but the utility of such structures for GPL and BGPL was never evaluated in this dissertation. Hence, it is of interest to know whether or not the additional "depth" in such group/block structures would produce "better" MGPLEs/MBGPLEs. Such a question could be addressed in a straightforward manner via a simulation study.

(3) **Incorporation of a space-time interaction component**: Throughout this dissertation the separability of space and time was assumed. In reality, however, such a simplifying assumption is likely invalid. In other words, the effect of space may depend on time, and vice versa. Thus, spatio-temporal autologistic models that incorporate a space-time interaction term need to be explored. The difficulty in such exploration should not be in the form of such an autologistic model, but in estimating its corresponding parameters. In particular, generalizing GPL and BGPL for use with such a model will require theoretical investigation along with a simulation study validating any proposed extensions.

(4) **Development and validation of a model selection criterion**: While the SAE approach of Wu and Huffer ([35]) was used in this dissertation as a model selection criterion, along with the PL-AIC measure suggested by Gumpertz et al. ([19]), the validity of either of these methods has never been formally explored. Note that a cross-validation approach to model selection has also been suggested by Gumpertz et al. ([19]), as mentioned in Chapter 7, but the validity of this approach also has never been formally explored. Hence, as modern computing has made implementing auto-models, such as the autologistic model, fairly practical with respect to computational expense, we suspect that the need for a model selection criterion that is both practical and reliable is only going to grow. Perhaps one of the aforementioned methods is such a method, or perhaps a new method needs to be developed, but either way the validity of such a procedure must be theoretically justified and practically evaluated via a rigorous simulation-based study.

(5) **Extensions of GPL and BGPL for spatio-temporal autologistic models under absorbing states**: The focus of this dissertation within the space-time domain was on non-absorbing state models. However, absorbing state models are also needed in the space-time domain for many applications, as briefly alluded to in Chapter 1. Recall that an absorbing state is a type of temporal dependency in which a site remains permanently in one state (say diseased) for the duration of the study, if or once it becomes diseased. Hence, in order to employ GPL and BGPL in such a domain, these methods would have to be adapted. We suspect that such an adaptation is not terribly difficult, but nonetheless one does not currently exist in the literature.

(6) **Extensions of the proofs of the strong consistency of the MGPLEs and MBGPLEs**: The proofs given for the MGPLE's and MBGPLE's strong consistency assumes a finite state-space and a fixed group/block structure. Extending these proofs so that such assumptions are not needed would be extremely useful for the following two reasons. First, the strong consistency of the MGPLEs and MBGPLEs would then apply to additional models such as the auto-Poisson model. Second, the statistical modeler could then confidently apply GPL and/or BGPL under varying group/block structures, which may be helpful in modeling small and awkwardly shaped regions like the temperate grasslands ecosystem modeled in Chapter 7.

In addition to the aforementioned topics, many other concepts from this dissertation deserve further investigation, but have been omitted from the above list to limit the length of this concluding chapter. Finally, note that as modern computing continues to advance, estimating the parameters of the autologistic model, as well as the parameters of other auto-models, will becomes less and less computationally expensive, which should allow for larger group/block sizes and, therefore, better compromises between PL and MCMCML.

Bibliography

[1] Julian Besag, *Nearest-neighbor Systems and the Auto-logistic Model for Binary Data*, Journal of the Royal Statistical Society. Series B **34** (1972), no. 1, 75–83.

[2] ———, *Spatial Interaction and the Statistical Analysis of Lattice Systems*, Journal of the Royal Statistical Society. Series B **36** (1974), no. 2, 192–236.

[3] ———, *Statistical Analysis of Non-Lattice Data*, Journal of the Royal Statistical Society. Series D (The Statistician) **24** (1975), no. 3, 179–195.

[4] ———, *Some Methods of Statistical Analysis for Spatial Data*, Bulletin of the International Statistical Institute **47** (1977), 77–92.

[5] Petruta C. Caragea and Mark S. Kaiser, *Autologistic Models with Interpretable Parameters*, Journal of Agricultural, Biological, and Environmental Statistics **14** (2009), no. 3, 281–300.

[6] J. Chadoeuf, D. Nandris, J.P. Geiger, M. Nicole, and J.C. Pierrat, *Spatio-Temporal Modeling of an Epidemic Using a Gibbs Process: Estimation and Testing*, Biometrics **48** (1992), no. 4, 1165–1175. Trans. Jonathan M. Graham.

[7] F. Comets, *On Consistency of a Class of Estimators of Exponential Families of Markov Random Fields on the Lattice*, Annals of Statistics **20** (1992), no. 1, 455–468.

[8] Noel A.C. Cressie, *Statistics For Spatial Data*, Revised, Wiley Series in Probability and Mathematical Statistics, Wiley, New York, 1993.

[9] R. L. Dobrushin, *The Description of a Random Field by Means of Conditional Probabilities and Conditions of its Regularity*, Theory of Probability and Its Applications **13** (1968), 197–224.

[10] H. Follmer, *A Covariance Estimate for Gibbs Measures*, Journal of Functional Analysis **46** (1968), 387–395.

[11] N. Friel, A.N. Pettitt, R. Reeves, and E. Wit, *Bayesian Inference in Hidden Markov Random Fields for Binary Data Defined on Large Lattices*, Journal of Computational and Graphical Statistics **18** (2009), no. 2, 243–261.

[12] Stuart Geman and Donald Geman, *Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images*, IEEE Transactions on Pattern Analysis and Machine Intelligence **PAM1-6** (1984), no. 6, 721–741.

[13] Stuart Geman and Christine Graffigne, *Markov Random Field Image Models and Their Applications to Computer Vision*, Proceedings of the international congress of mathematicians, 1986, pp. 1496–1517.

[14] Charles Geyer, *Markov Chain Monte Carlo Maximum Likelihood*, Interface proceedings, 1991, pp. 1–8.

[15] Charles J. Geyer and Elizabeth A. Thompson, *Constrained Monte Carlo Maximum Likelihood for Dependent Data*, Journal of the Royal Statistical Society. Series B **54** (1992), no. 3, 657–699.

[16] B. Gidas, *Consistency of Maximum Likelihood and Pseudo-likelihood Estimates for Gibbs Distributions*, Stochastic differential systems, stochastic control theory and applications, 1988.

[17] Jonathan Mills Graham, *Markov Chain Monte Carlo Inference Procedures for Discrete Spatial Lattice Models*, Ph.D. Thesis, 1995. Department of Statistics, North Carolina State University.

[18] ———, *Markov Chain Monte Carlo Methods for Modeling the Spatial Pattern of Disease Spread in Bell Pepper*, Proceedings of the 1996 kansas st. university conference on applied statistics in agriculture, 1997, pp. 91–108.

[19] Marcia L. Gumpertz, Jonathan M. Graham, and Jean B. Ristaino, *Autologistic Model of Spatial Pattern of Phytophthora Epidemic in Bell Pepper: Effects of Soil Variables on Disease Presence*, Journal of Agricultural, Biological, and Environmental Statistics **2** (1997), no. 2, 131–156.

[20] X. Guyon and H.R. Kunsch, *Asymptotic Comparison of Estimators in the Ising Model*, Stochastic models, statistical methods and algorithms in image analysis, 1992, pp. 177–198.

[21] Fangliang He, Julie Zhou, and Hongtu Zhu, *Autologistic Regression Model for the Distribution of Vegetation*, Journal of Agricultural, Biological, and Environmental Statistics **8** (2003), no. 2, 205–222.

[22] Fuchun Huang and Yosihiko Ogata, *Generalized Pseudo-likelihood Estimates for Markov Random Fields on Lattice*, Annals of the Institute of Statistical Mathematics **54** (2002), no. 1, 1–18.

[23] Fred W. Huffer and Hulin Wu, *Markov Chain Monte Carlo for Autologistic Regression Models with Applications to the Distributions of Plant Species*, Biometrics **54** (1998), no. 2, 509–524.

[24] J.L. Jensen and J. Møller, *Pseudolikelihood for Exponential Family Models of Spatial Point Processes*, Annals of Applied Probability **1** (1991), no. 3, 445–461.

[25] Elias Teixeira Krainski, Paulo Justiniano Ribeiro Junior, Renato Beozzo Bassanezi, and Luziane Franciscon, *Autologistic Model with an Application to the Citrus "Sudden Death" Disease*, Scientia Agricola **65** (2008), no. 5, 541–547.

[26] Cindy S. Leary, W. Matt Jolly, Jonathan M. Graham, and Patricia L. Andrews, *Spatial statistical methods for evaluating satellite-derived metrics of seasonal wildland fire potential*. Unpublished article: The University of Montana, Missoula, MT.

[27] Johan Lim, Haiyan Liu, and Michael Sherman, *Parameter Estimation in the Spatial Auto-Logistic Model with Varying Independent Subblocks*. Unpublished article: Texas A&M University, College Station, TX.

[28] David K. Pickard, *Inference for Discrete Markov Fields: The Simplest Nontrivial Case*, Journal of the American Statistical Association **82** (1987), no. 397, 90–96.

[29] Haiganoush K. Preisler, *Modelling Spatial Patterns of Trees Attacked by Bark-Beetles*, Journal of the Royal Statistical Society **42** (1993), no. 3, 501–514.

[30] R. Reeves and A.N. Pettitt, *Efficient Recursions for General Factorisable Models*, Biometrika **91** (2004), no. 3, 751–757.

[31] Christian P. Robert and George Casella, *Monte Carlo Statistical Methods*, Springer Texts in Statistics, Springer, New York, 1999.

[32] Michael Sherman, Tatiyana V. Apanasovich, and Raymond J. Carroll, *On Estimation in Binary Autologistic Spatial Models*, Journal of Statistical Computation and Simulation **76** (2006), no. 2, 167–179.

[33] David Strauss, *The Many Faces of Logistic Regression*, The American Statistician **46** (1992), no. 4, 321–327.

[34] Robert S. Strichartz, *The Way of Analysis*, Revised, Jones and Bartlett Books in Mathematics, Jones and Bartlett, Boston, 2000.

[35] Hulin Wu and Fred Huffer, *Modeling and Regressing Spatial Binary Data with Applications to the Distribution of Species*, Florida State University, 1993.

[36] ———, *Modeling the Distribution of Plant Species Using the Autologistic Regression Model*, Environmental and Ecological Statistics **4** (1997), 49–64.

[37] Jun Zhu, Hsin-Cheng Huang, and Jungpin Wu, *Modeling Spatial-Temporal Binary Data Using Markov Random Fields*, Journal of Agricultural, Biological, and Environmental Statistics **10** (2005), no. 2, 212–225.

Appendix A

Supplementary Tables and Figures for Chapter 5

A.1 Supplemental Items for Section 5.3

A.2 Supplemental Items for Section 5.4

A.3 Supplemental Items for Section 5.5

Table A.1: MAEs of α and θ when the Model is FI2p and the Method of Estimation is GPL for Group Sizes of 5 (gs5) and 9 (gs9)

		26x26		50x50		74x74	
Model	Estimate	gs5	gs9	gs5	gs9	gs5	gs9
FI2p	MAE($\hat{\alpha}$)	0.1542	0.1522	0.0720	0.0715	0.0491	0.0489
$(-1.0, 0.0)$	MAE($\hat{\theta}$)	0.1299	0.1282	0.0576	0.0572	0.0378	0.0376
FI2p	MAE($\hat{\alpha}$)	0.1551	0.1529	0.0783	0.0777	0.0500	0.0498
$(-1.0, 0.1)$	MAE($\hat{\theta}$)	0.1149	0.1132	0.0578	0.0571	0.0356	0.0356
FI2p	MAE($\hat{\alpha}$)	0.1597	0.1578	0.0803	0.0797	0.0527	0.0524
$(-1.0, 0.2)$	MAE($\hat{\theta}$)	0.1072	0.1059	0.0539	0.0535	0.0351	0.0349
FI2p	MAE($\hat{\alpha}$)	0.1787	0.1752	0.0850	0.0841	0.0535	0.0531
$(-1.0, 0.3)$	MAE($\hat{\theta}$)	0.1095	0.1065	0.0518	0.0513	0.0340	0.0338
FI2p	MAE($\hat{\alpha}$)	0.1698	0.1666	0.0869	0.0855	0.0580	0.0578
$(-1.0, 0.4)$	MAE($\hat{\theta}$)	0.0919	0.0907	0.0464	0.0461	0.0315	0.0313
FI2p	MAE($\hat{\alpha}$)	0.1981	0.1938	0.0932	0.0910	0.0633	0.0622
$(-1.0, 0.5)$	MAE($\hat{\theta}$)	0.0964	0.0941	0.0444	0.0435	0.0305	0.0299
FI2p	MAE($\hat{\alpha}$)	0.2372	0.2325	0.1176	0.1150	0.0743	0.0733
$(-1.0, 0.6)$	MAE($\hat{\theta}$)	0.1004	0.0979	0.0500	0.0490	0.0317	0.0313
FI2p	MAE($\hat{\alpha}$)	0.3012	0.2956	0.1406	0.1389	0.0929	0.0922
$(-1.0, 0.7)$	MAE($\hat{\theta}$)	0.1119	0.1098	0.0529	0.0520	0.0352	0.0349
FI2p	MAE($\hat{\alpha}$)	0.4102	0.3970	0.2008	0.1941	0.1277	0.1258
$(-1.0, 0.8)$	MAE($\hat{\theta}$)	0.1366	0.1324	0.0674	0.0654	0.0430	0.0425
FI2p	MAE($\hat{\alpha}$)	0.6552	0.6325	0.2839	0.2799	0.1952	0.1924
$(-1.0, 0.9)$	MAE($\hat{\theta}$)	0.2019	0.1952	0.0881	0.0871	0.0597	0.0589
FI2p	MAE($\hat{\alpha}$)	0.9572	0.9170	0.4096	0.4010	0.2709	0.2687
$(-1.0, 1.0)$	MAE($\hat{\theta}$)	0.2751	0.2640	0.1194	0.1173	0.0794	0.0787

Table A.2: MAEs of α and θ when the Model is FI2p and the Method of Estimation is BGPL for Block Sizes of 2x2 (bs2), 3x3 (bs3), and 4x4 (bs4)

		26x26			50x50			74x74		
Model	Estimate	bs2	bs3	bs4	bs2	bs3	bs4	bs2	bs3	bs4
FI2p	MAE($\hat{\alpha}$)	0.1454	0.1466	0.1451	0.0728	0.0739	0.0747	0.0507	0.0497	0.0498
$(-1.0, 0.0)$	MAE($\hat{\theta}$)	0.1292	0.1287	0.1295	0.0599	0.0596	0.0608	0.0394	0.0391	0.0389
FI2p	MAE($\hat{\alpha}$)	0.1473	0.1471	0.1477	0.0779	0.0772	0.0788	0.0509	0.0520	0.0510
$(-1.0, 0.1)$	MAE($\hat{\theta}$)	0.1117	0.1117	0.1119	0.0587	0.0585	0.0590	0.0371	0.0381	0.0371
FI2p	MAE($\hat{\alpha}$)	0.1624	0.1620	0.1627	0.0782	0.0814	0.0792	0.0543	0.0555	0.0556
$(-1.0, 0.2)$	MAE($\hat{\theta}$)	0.1084	0.1099	0.1076	0.0526	0.0564	0.0541	0.0371	0.0373	0.0380
FI2p	MAE($\hat{\alpha}$)	0.1748	0.1826	0.1719	0.0849	0.0862	0.0843	0.0548	0.0546	0.0540
$(-1.0, 0.3)$	MAE($\hat{\theta}$)	0.1073	0.1112	0.1035	0.0525	0.0524	0.0519	0.0350	0.0347	0.0344
FI2p	MAE($\hat{\alpha}$)	0.1673	0.1631	0.1613	0.0888	0.0877	0.0855	0.0602	0.0603	0.0601
$(-1.0, 0.4)$	MAE($\hat{\theta}$)	0.0939	0.0914	0.0907	0.0478	0.0476	0.0465	0.0325	0.0324	0.0325
FI2p	MAE($\hat{\alpha}$)	0.1977	0.1978	0.1951	0.0932	0.0932	0.0916	0.0643	0.0643	0.0641
$(-1.0, 0.5)$	MAE($\hat{\theta}$)	0.0960	0.0946	0.0942	0.0443	0.0444	0.0445	0.0305	0.0309	0.0304
FI2p	MAE($\hat{\alpha}$)	0.2386	0.2396	0.2354	0.1196	0.1183	0.1172	0.0770	0.0775	0.0757
$(-1.0, 0.6)$	MAE($\hat{\theta}$)	0.1014	0.1019	0.0993	0.0506	0.0510	0.0502	0.0328	0.0327	0.0321
FI2p	MAE($\hat{\alpha}$)	0.2943	0.2891	0.2917	0.1425	0.1440	0.1399	0.0985	0.0969	0.0963
$(-1.0, 0.7)$	MAE($\hat{\theta}$)	0.1082	0.1071	0.1071	0.0536	0.0545	0.0525	0.0372	0.0367	0.0365
FI2p	MAE($\hat{\alpha}$)	0.4047	0.4020	0.3966	0.2053	0.2014	0.2010	0.1347	0.1370	0.1333
$(-1.0, 0.8)$	MAE($\hat{\theta}$)	0.1357	0.1344	0.1318	0.0693	0.0682	0.0680	0.0452	0.0466	0.0451
FI2p	MAE($\hat{\alpha}$)	0.6365	0.6218	0.6094	0.2953	0.2932	0.2915	0.2022	0.2023	0.1981
$(-1.0, 0.9)$	MAE($\hat{\theta}$)	0.1977	0.1935	0.1885	0.0911	0.0916	0.0903	0.0620	0.0622	0.0607
FI2p	MAE($\hat{\alpha}$)	0.9352	0.9180	0.8958	0.4183	0.4166	0.4196	0.2779	0.2800	0.2726
$(-1.0, 1.0)$	MAE($\hat{\theta}$)	0.2699	0.2651	0.2599	0.1223	0.1218	0.1220	0.0813	0.0814	0.0798

Table A.3: Sample Means and Standard Errors of α, θ_h, and θ_v when the Model is FAhv3p and the Methods of Estimation are GPL for a Group Size of 9 (gs9) and BGPL for Block Sizes of 2x2 (bs2), 3x3 (bs3), and 4x4 (bs4) (26×26, 50×50, and 74×74 Lattices)

		26x26				50x50			
Model	Estimate	gs9	bs2	bs3	bs4	gs9	bs2	bs3	bs4
	$\hat{\bar{\alpha}}$	-0.9754	-0.9814	-0.9815	-0.9843	-1.0039	-1.0050	-1.0063	-1.0044
	$\hat{\mathrm{SD}}(\hat{\alpha})$	0.3247	0.3286	0.3297	0.3242	0.1541	0.1576	0.1603	0.1580
FAhv3p	$\hat{\bar{\theta}}_h$	0.6993	0.7091	0.7091	0.7111	0.7028	0.7095	0.7074	0.7110
$(-1.0, 0.7, 0.6)$	$\hat{\mathrm{SD}}(\hat{\theta}_h)$	0.1892	0.1955	0.1950	0.1965	0.0911	0.0943	0.0963	0.0949
	$\hat{\bar{\theta}}_v$	0.5904	0.5826	0.5828	0.5838	0.6023	0.5953	0.5990	0.5936
	$\hat{\mathrm{SD}}(\hat{\theta}_v)$	0.1914	0.2100	0.2127	0.2113	0.0905	0.0978	0.1006	0.0973
	$\hat{\bar{\alpha}}$	-0.9477	-0.9757	-0.9775	-0.9735	-0.9920	-0.9898	-0.9908	-0.9937
	$\hat{\mathrm{SD}}(\hat{\alpha})$	0.3484	0.3532	0.3520	0.3484	0.1684	0.1723	0.1755	0.1732
FAhv3p	$\hat{\bar{\theta}}_h$	0.6677	0.6802	0.6797	0.6812	0.6983	0.6992	0.6978	0.7002
$(-1.0, 0.7, 0.7)$	$\hat{\mathrm{SD}}(\hat{\theta}_h)$	0.1951	0.2056	0.2063	0.2053	0.0936	0.0987	0.0996	0.0976
	$\hat{\bar{\theta}}_v$	0.6962	0.7037	0.7072	0.7017	0.6959	0.6931	0.6958	0.6951
	$\hat{\mathrm{SD}}(\hat{\theta}_v)$	0.1908	0.2167	0.2174	0.2155	0.0949	0.1081	0.1074	0.1072
	$\hat{\bar{\alpha}}$	-0.9535	-0.9817	-0.9764	-0.9798	-0.9970	-1.0044	-1.0012	-1.0014
	$\hat{\mathrm{SD}}(\hat{\alpha})$	0.4420	0.4433	0.4442	0.4372	0.2015	0.2009	0.2019	0.1994
FAhv3p	$\hat{\bar{\theta}}_h$	0.6819	0.6806	0.6840	0.6860	0.6985	0.6944	0.6959	0.6962
$(-1.0, 0.7, 0.8)$	$\hat{\mathrm{SD}}(\hat{\theta}_h)$	0.2310	0.2448	0.2444	0.2441	0.1090	0.1160	0.1154	0.1130
	$\hat{\bar{\theta}}_v$	0.7952	0.8159	0.8090	0.8093	0.8013	0.8122	0.8083	0.8078
	$\hat{\mathrm{SD}}(\hat{\theta}_v)$	0.2136	0.2371	0.2352	0.2354	0.1021	0.1138	0.1146	0.1117
	$\hat{\bar{\alpha}}$	-0.9146	-0.9687	-0.9731	-0.9654	-0.9766	-0.9945	-0.9942	-0.9904
	$\hat{\mathrm{SD}}(\hat{\alpha})$	0.5192	0.5189	0.5158	0.5123	0.2427	0.2578	0.2570	0.2578
FAhv3p	$\hat{\bar{\theta}}_h$	0.6747	0.6860	0.6842	0.6887	0.6924	0.6878	0.6901	0.6878
$(-1.0, 0.7, 0.9)$	$\hat{\mathrm{SD}}(\hat{\theta}_h)$	0.2535	0.2622	0.2576	0.2617	0.1163	0.1249	0.1245	0.1234
	$\hat{\bar{\theta}}_v$	0.8808	0.9051	0.9106	0.9001	0.8947	0.9132	0.9107	0.9102
	$\hat{\mathrm{SD}}(\hat{\theta}_v)$	0.2423	0.2595	0.2615	0.2624	0.1184	0.1276	0.1296	0.1309
	$\hat{\bar{\alpha}}$	-0.8925	-0.9468	-0.9433	-0.9410	-0.9741	-0.9966	-0.9947	-0.9972
	$\hat{\mathrm{SD}}(\hat{\alpha})$	0.6251	0.6289	0.6309	0.6341	0.2841	0.2984	0.2969	0.2973
FAhv3p	$\hat{\bar{\theta}}_h$	0.6672	0.6704	0.6710	0.6720	0.6951	0.6906	0.6903	0.6926
$(-1.0, 0.7, 1.0)$	$\hat{\mathrm{SD}}(\hat{\theta}_h)$	0.2822	0.3053	0.2958	0.3013	0.1457	0.1525	0.1530	0.1547
	$\hat{\bar{\theta}}_v$	0.9811	1.0142	1.0111	1.0094	0.9911	1.0116	1.0107	1.0102
	$\hat{\mathrm{SD}}(\hat{\theta}_v)$	0.3049	0.3277	0.3302	0.3268	0.1409	0.1507	0.1522	0.1499
	$\hat{\bar{\alpha}}$	-0.7971	-0.8683	-0.8626	-0.8591	-0.9516	-0.9819	-0.9790	-0.9783
	$\hat{\mathrm{SD}}(\hat{\alpha})$	0.7327	0.7419	0.7481	0.7369	0.3564	0.3696	0.3667	0.3678
FAhv3p	$\hat{\bar{\theta}}_h$	0.6340	0.6254	0.6248	0.6321	0.6798	0.6725	0.6775	0.6753
$(-1.0, 0.7, 1.1)$	$\hat{\mathrm{SD}}(\hat{\theta}_h)$	0.3580	0.3778	0.3766	0.3775	0.1602	0.1670	0.1675	0.1692
	$\hat{\bar{\theta}}_v$	1.0623	1.1141	1.1129	1.1030	1.0941	1.1219	1.1149	1.1170
	$\hat{\mathrm{SD}}(\hat{\theta}_v)$	0.3368	0.3498	0.3632	0.3454	0.1621	0.1666	0.1692	0.1664

Table A.3 (continued): Sample Means and Standard Errors of α, θ_h, and θ_v when the Model is FAhv3p and the Methods of Estimation are GPL for a Group Size of 9 (gs9) and BGPL for Block Sizes of 2x2 (bs2), 3x3 (bs3), and 4x4 (bs4) (26×26, 50×50, and 74×74 Lattices)

		74x74			
Model	Estimate	gs9	bs2	bs3	bs4
	$\hat{\bar{\alpha}}$	-0.9999	-0.9984	-0.9982	-0.9992
	$\hat{\mathrm{SD}}(\hat{\alpha})$	0.0976	0.0995	0.0982	0.0985
FAhv3p	$\hat{\bar{\theta}}_h$	0.7025	0.7060	0.7057	0.7068
$(-1.0, 0.7, 0.6)$	$\hat{\mathrm{SD}}(\hat{\theta}_h)$	0.0585	0.0606	0.0613	0.0605
	$\hat{\bar{\theta}}_v$	0.5981	0.5922	0.5924	0.5923
	$\hat{\mathrm{SD}}(\hat{\theta}_v)$	0.0607	0.0644	0.0651	0.0652
	$\hat{\bar{\alpha}}$	-0.9958	-0.9947	-0.9959	-0.9949
	$\hat{\mathrm{SD}}(\hat{\alpha})$	0.1096	0.1146	0.1142	0.1126
FAhv3p	$\hat{\bar{\theta}}_h$	0.6992	0.6993	0.6988	0.6989
$(-1.0, 0.7, 0.7)$	$\hat{\mathrm{SD}}(\hat{\theta}_h)$	0.0638	0.0684	0.0689	0.0684
	$\hat{\bar{\theta}}_v$	0.6978	0.6968	0.6982	0.6974
	$\hat{\mathrm{SD}}(\hat{\theta}_v)$	0.0645	0.0738	0.0737	0.0736
	$\hat{\bar{\alpha}}$	-0.9972	-1.0043	-1.0036	-1.0037
	$\hat{\mathrm{SD}}(\hat{\alpha})$	0.1339	0.1405	0.1397	0.1407
FAhv3p	$\hat{\bar{\theta}}_h$	0.6974	0.6942	0.6949	0.6944
$(-1.0, 0.7, 0.8)$	$\hat{\mathrm{SD}}(\hat{\theta}_h)$	0.0725	0.0758	0.0760	0.0759
	$\hat{\bar{\theta}}_v$	0.8016	0.8108	0.8095	0.8099
	$\hat{\mathrm{SD}}(\hat{\theta}_v)$	0.0727	0.0807	0.0821	0.0809
	$\hat{\bar{\alpha}}$	-0.9851	-0.9970	-0.9960	-0.9946
	$\hat{\mathrm{SD}}(\hat{\alpha})$	0.1661	0.1693	0.1714	0.1687
FAhv3p	$\hat{\bar{\theta}}_h$	0.6959	0.6920	0.6930	0.6933
$(-1.0, 0.7, 0.9)$	$\hat{\mathrm{SD}}(\hat{\theta}_h)$	0.0802	0.0824	0.0827	0.0821
	$\hat{\bar{\theta}}_v$	0.8952	0.9081	0.9065	0.9052
	$\hat{\mathrm{SD}}(\hat{\theta}_v)$	0.0790	0.0811	0.0846	0.0846
	$\hat{\bar{\alpha}}$	-0.9864	-1.0020	-1.0016	-1.0016
	$\hat{\mathrm{SD}}(\hat{\alpha})$	0.1926	0.2034	0.2039	0.2025
FAhv3p	$\hat{\bar{\theta}}_h$	0.6972	0.6915	0.6920	0.6931
$(-1.0, 0.7, 1.0)$	$\hat{\mathrm{SD}}(\hat{\theta}_h)$	0.0949	0.0982	0.0985	0.1000
	$\hat{\bar{\theta}}_v$	0.9951	1.0120	1.0115	1.0103
	$\hat{\mathrm{SD}}(\hat{\theta}_v)$	0.0896	0.0910	0.0939	0.0929
	$\hat{\bar{\alpha}}$	-0.9735	-0.9883	-0.9882	-0.9872
	$\hat{\mathrm{SD}}(\hat{\alpha})$	0.2369	0.2437	0.2441	0.2434
FAhv3p	$\hat{\bar{\theta}}_h$	0.6872	0.6814	0.6809	0.6827
$(-1.0, 0.7, 1.1)$	$\hat{\mathrm{SD}}(\hat{\theta}_h)$	0.1057	0.1083	0.1080	0.1088
	$\hat{\bar{\theta}}_v$	1.0993	1.1151	1.1158	1.1133
	$\hat{\mathrm{SD}}(\hat{\theta}_v)$	0.1054	0.1090	0.1095	0.1105

Figure A.1: Relative MAEs of α, θ_h, and θ_v when the Model is FAhv3p and the Methods of Estimation are GPL for a Group Size of 9 (gs9) and BGPL for Block Sizes of 2x2 (bs2), 3x3 (bs3), and 4x4 (bs4)

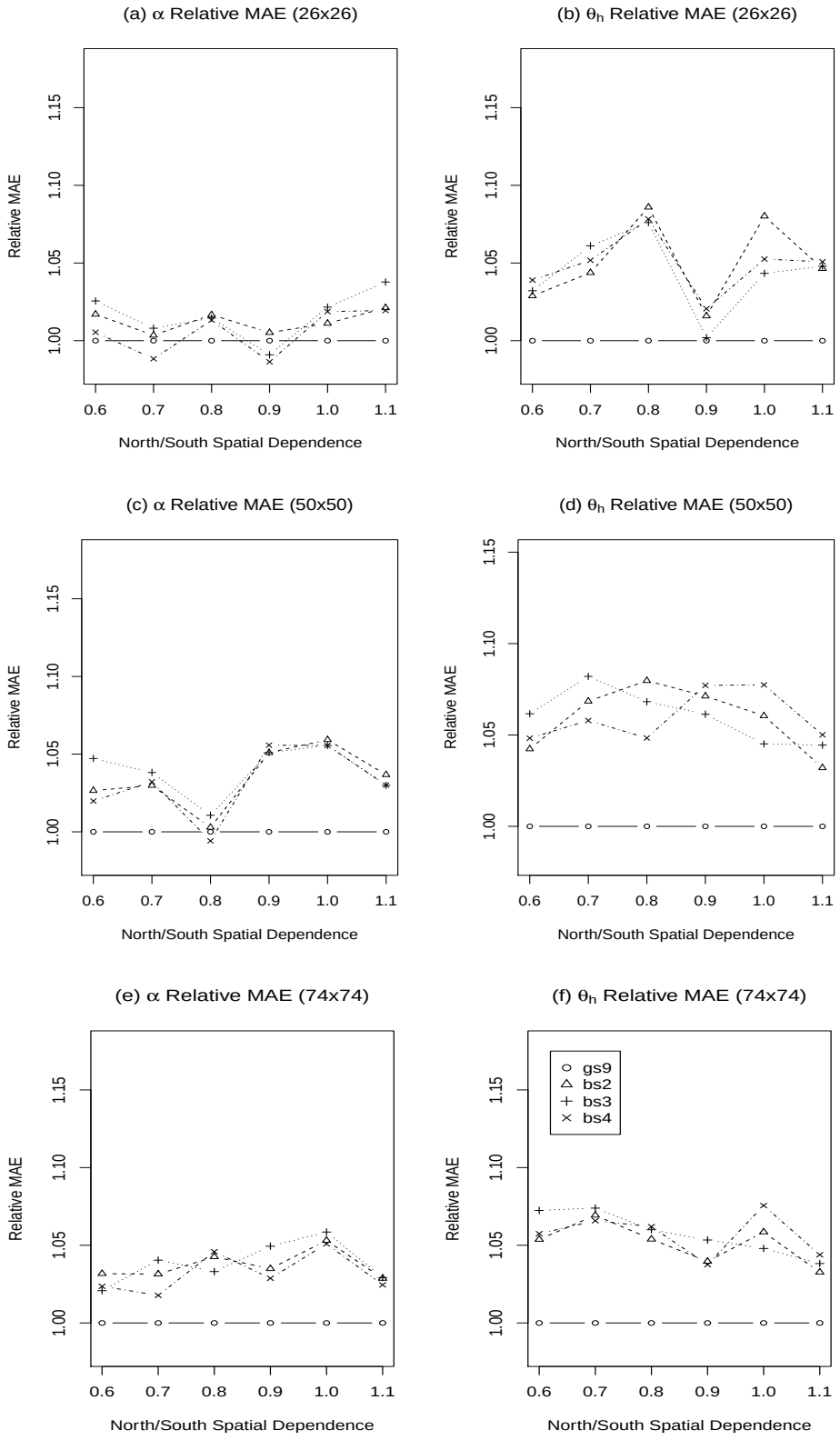


Figure A.1 (continued): Relative MAEs of α, θ_h, and θ_v when the Model is FAhv3p and the Methods of Estimation are GPL for a Group Size of 9 (gs9) and BGPL for Block Sizes of 2x2 (bs2), 3x3 (bs3), and 4x4 (bs4)

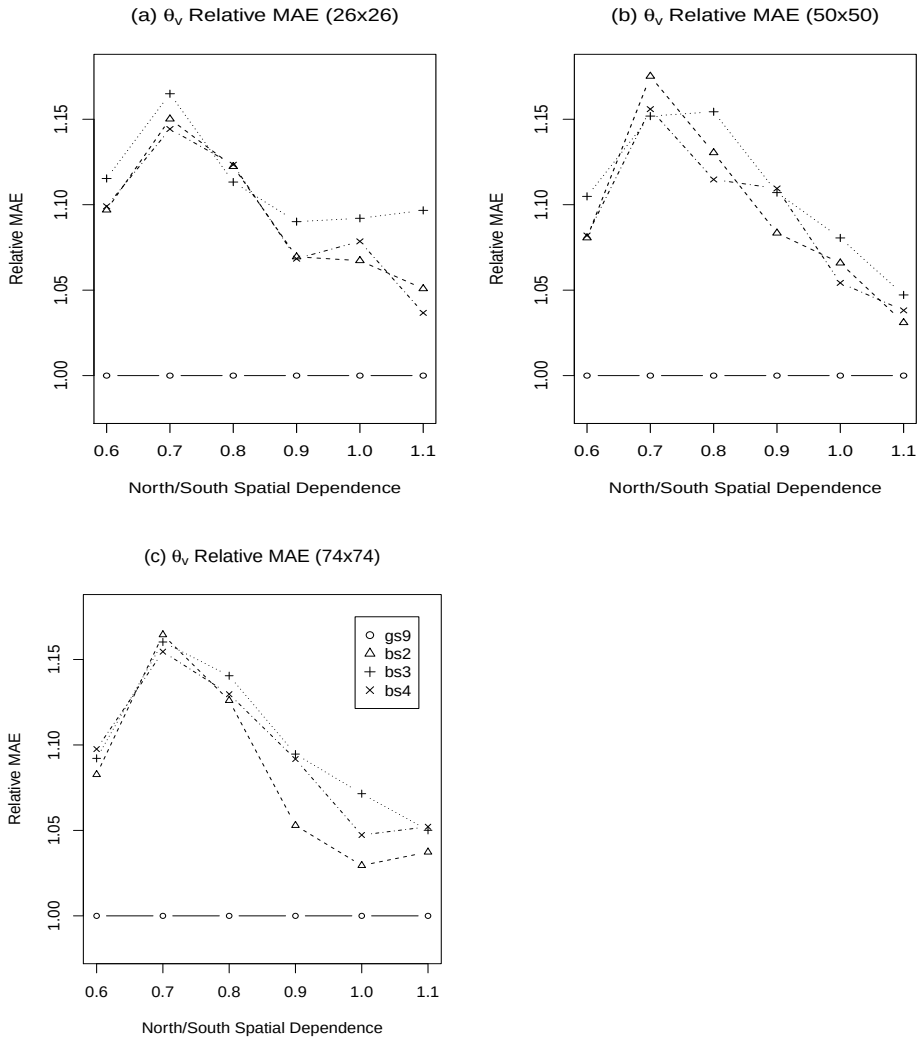


Table A.4: Proportion of Trials Converging when the Model is an Ising Model with $(\alpha, \theta) = (-3.2, 1.6)$ and the Method of Estimation is MCMCML Under the Following Reference Points: MPLE (MCMCML_{PL}), MGPLE for a Group Size of 9 (MCMCML_{gs9}), and MBGPLE for a Block Size of 4x4 (MCMCML_{bs4})

			Failed To Converge For		
Lattice Size		Method	MCMCML_{PL}	MCMCML_{gs9}	MCMCML_{bs4}
26×26	But Converged For	MCMCML_{PL}	0	0.2636	0.3333
		MCMCML_{gs9}	0.6836	0	0.6613
		MCMCML_{bs4}	0.5156	0.4273	0
50×50	But Converged For	MCMCML_{PL}	0	0.3333	0.3636
		MCMCML_{gs9}	0.7606	0	0.7847
		MCMCML_{bs4}	0.5317	0.5588	0
74×74	But Converged For	MCMCML_{PL}	0	0.2830	0.4395
		MCMCML_{gs9}	0.7226	0	0.7713
		MCMCML_{bs4}	0.5438	0.5189	0

Table A.5: Proportion of MCMCML Trials Converging for At Least One of the Three Reference Point Sets (MPLE, MGPLE for a Group Size of 9, or MBGPLE for a Block Size of 4x4) when the Model is an Ising Model Near Phase Transition

Model	26×26	50×50	74×74
Ising (-0.4,0.2)	1.000	1.000	1.000
Ising (-0.8,0.4)	1.000	1.000	1.000
Ising (-1.2,0.6)	1.000	1.000	1.000
Ising (-1.6,0.8)	1.000	1.000	1.000
Ising (-2.0,1.0)	1.000	1.000	1.000
Ising (-2.4,1.2)	1.000	1.000	1.000
Ising (-2.8,1.4)	0.992	1.000	1.000
Ising (-3.2,1.6)	0.898	0.932	0.924

Figure A.2: Relative MAEs of α and θ when the Model is an Ising Model Near Phase Transition and the Method of Estimation is MCMCML Under the Following Reference Points: MPLE (MCMCML(PL)), MGPLE for a Group Size of 9 (MCMCML(gs9)), and MBGPLE for a Block Size of 4x4 (MCMCML(bs4))

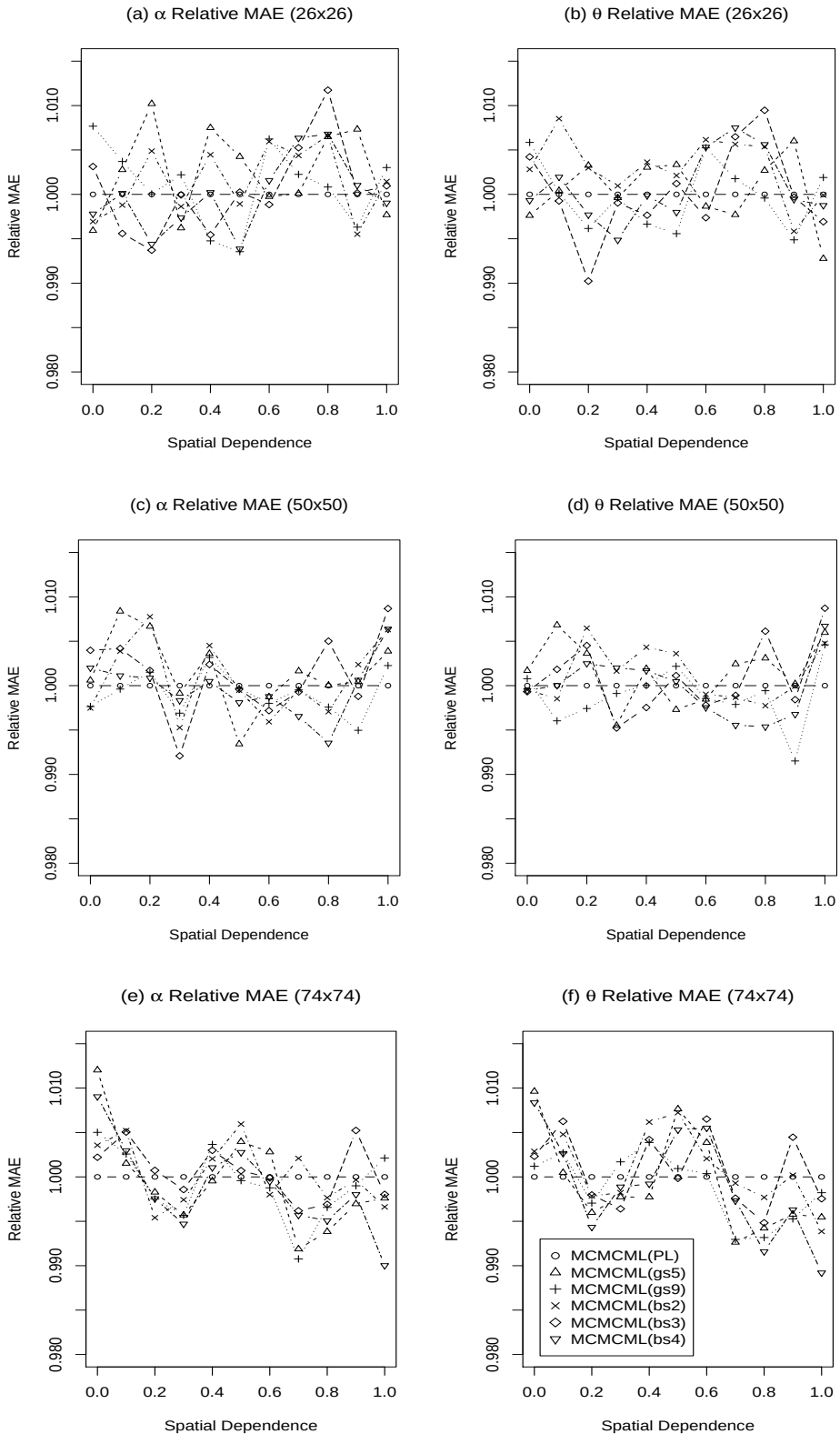


Table A.6: Sample Means and Standard Errors of α, θ, and β (Fire Covariate) when the Model is FI3p and the Methods of Parameter Estimation are PL, GPL for a Group Size of 9 (gs9), BGPL for a Block Size of 4x4 (bs4), and MCMCML

		26x26			
Model	Estimate	PL	gs9	bs4	MCMCML
	$\bar{\hat{\alpha}}$	-8.6995	-8.7342	-8.6957	-8.7033
	$\hat{SD}(\hat{\alpha})$	2.4917	2.6683	2.4570	2.5670
FI3p	$\bar{\hat{\theta}}$	0.6423	0.5951	0.6435	0.6038
(-8.6,0.8,0.065)	$\hat{SD}(\hat{\theta})$	0.5745	0.6032	0.5338	0.5323
	$\bar{\hat{\beta}}$	0.0660	0.0666	0.0660	0.0663
	$\hat{SD}(\hat{\beta})$	0.0305	0.0325	0.0299	0.0315
		50x50			
Model	Estimate	PL	gs9	bs4	MCMCML
	$\bar{\hat{\alpha}}$	-8.5868	-8.6346	-8.6378	-8.5706
	$\hat{SD}(\hat{\alpha})$	1.1641	1.1922	1.1639	1.1639
FI3p	$\bar{\hat{\theta}}$	0.7689	0.7771	0.7846	0.7591
(-8.6,0.8,0.065)	$\hat{SD}(\hat{\theta})$	0.2401	0.2351	0.2334	0.2288
	$\bar{\hat{\beta}}$	0.0648	0.0654	0.0654	0.0647
	$\hat{SD}(\hat{\beta})$	0.0142	0.0146	0.0142	0.0142
		74x74			
Model	Estimate	PL	gs9	bs4	MCMCML
	$\bar{\hat{\alpha}}$	-8.6815	-8.7427	-8.7551	-8.6806
	$\hat{SD}(\hat{\alpha})$	0.7212	0.7245	0.7142	0.7190
FI3p	$\bar{\hat{\theta}}$	0.7872	0.8041	0.8103	0.7807
(-8.6,0.8,0.065)	$\hat{SD}(\hat{\theta})$	0.1355	0.1288	0.1327	0.1271
	$\bar{\hat{\beta}}$	0.0660	0.0666	0.0667	0.0660
	$\hat{SD}(\hat{\beta})$	0.0086	0.0086	0.0085	0.0085

Table A.7: Sample Means and Standard Errors of α and θ when the Model is an Ising Model Near Phase Transition and the Methods of Estimation are PL, GPL for Group Sizes of 5 (gs5) and 9 (gs9), BGPL for Block Sizes of 2x2 (bs2), 3x3 (bs3), and 4x4 (bs4), and MCMCML

		26x26		50x50		74x74	
Model	Estimate	PL	MCMCML	PL	MCMCML	PL	MCMCML
	$\hat{\alpha}$	-0.4079	-0.4003	-0.4026	-0.4013	-0.4028	-0.4015
Ising	$\hat{SD}(\hat{\alpha})$	0.2310	0.2367	0.1168	0.1165	0.0839	0.0843
$(-0.4, 0.2)$	$\bar{\hat{\theta}}$	0.2026	0.1991	0.2001	0.1995	0.2006	0.2000
	$\hat{SD}(\hat{\theta})$	0.1093	0.1099	0.0561	0.0552	0.0398	0.0398
	$\hat{\alpha}$	-0.7942	-0.7818	-0.7998	-0.7973	-0.8038	-0.8034
Ising	$\hat{SD}(\hat{\alpha})$	0.2278	0.2270	0.1105	0.1103	0.0768	0.0757
$(-0.8, 0.4)$	$\bar{\hat{\theta}}$	0.3984	0.3927	0.4006	0.3994	0.4021	0.4020
	$\hat{SD}(\hat{\theta})$	0.1109	0.1080	0.0532	0.0524	0.0373	0.0364
	$\hat{\alpha}$	-1.1874	-1.1667	-1.2001	-1.1934	-1.2014	-1.1977
Ising	$\hat{SD}(\hat{\alpha})$	0.2425	0.2279	0.1163	0.1130	0.0802	0.0790
$(-1.2, 0.6)$	$\bar{\hat{\theta}}$	0.5932	0.5834	0.6001	0.5969	0.6014	0.5997
	$\hat{SD}(\hat{\theta})$	0.1185	0.1091	0.0560	0.0538	0.0385	0.0374
	$\hat{\alpha}$	-1.5880	-1.5600	-1.6012	-1.5915	-1.6063	-1.6009
Ising	$\hat{SD}(\hat{\alpha})$	0.2304	0.2214	0.1173	0.1086	0.0810	0.0716
$(-1.6, 0.8)$	$\bar{\hat{\theta}}$	0.7950	0.7814	0.8010	0.7964	0.8032	0.8007
	$\hat{SD}(\hat{\theta})$	0.1102	0.1039	0.0564	0.0510	0.0392	0.0341
	$\hat{\alpha}$	-2.0040	-1.9680	-2.0063	-1.9956	-1.9987	-1.9958
Ising	$\hat{SD}(\hat{\alpha})$	0.2284	0.2002	0.1099	0.0955	0.0780	0.0658
$(-2.0, 1.0)$	$\bar{\hat{\theta}}$	1.0019	0.9848	1.0034	0.9984	0.9995	0.9982
	$\hat{SD}(\hat{\theta})$	0.1138	0.0969	0.0546	0.0463	0.0385	0.0317
	$\hat{\alpha}$	-2.4004	-2.3513	-2.4019	-2.3919	-2.4044	-2.4005
Ising	$\hat{SD}(\hat{\alpha})$	0.2368	0.1796	0.1163	0.0947	0.0737	0.0582
$(-2.4, 1.2)$	$\bar{\hat{\theta}}$	1.2020	1.1776	1.2012	1.1961	1.2021	1.2002
	$\hat{SD}(\hat{\theta})$	0.1158	0.0848	0.0571	0.0449	0.0364	0.0281
	$\hat{\alpha}$	-2.8266	-2.7641	-2.8078	-2.7931	-2.8067	-2.7968
Ising	$\hat{SD}(\hat{\alpha})$	0.2412	0.1695	0.1164	0.0838	0.0799	0.0612
$(-2.8, 1.4)$	$\bar{\hat{\theta}}$	1.4128	1.3813	1.4034	1.3961	1.4033	1.3985
	$\hat{SD}(\hat{\theta})$	0.1210	0.0820	0.0581	0.0413	0.0397	0.0299
	$\hat{\alpha}$	-3.1700	-3.1158	-3.1952	-3.1819	-3.1967	-3.1870
Ising	$\hat{SD}(\hat{\alpha})$	0.2156	0.1348	0.1102	0.0862	0.0704	0.0506
$(-3.2, 1.6)$	$\bar{\hat{\theta}}$	1.5815	1.5512	1.5945	1.5845	1.5964	1.5880
	$\hat{SD}(\hat{\theta})$	0.1138	0.0706	0.0590	0.0460	0.0375	0.0283

Table A.7 (continued): Sample Means and Standard Errors of α and θ when the Model is an Ising Model Near Phase Transition and the Methods of Estimation are PL, GPL for Group Sizes of 5 (gs5) and 9 (gs9), BGPL for Block Sizes of 2x2 (bs2), 3x3 (bs3), and 4x4 (bs4), and MCMCML

		26x26		50x50		74x74	
Model	Estimate	gs5	gs9	gs5	gs9	gs5	gs9
	$\hat{\alpha}$	-0.3999	-0.3984	-0.4024	-0.4017	-0.4021	-0.4019
Ising	$\hat{SD}(\hat{\alpha})$	0.2447	0.2424	0.1246	0.1228	0.0842	0.0837
$(-0.4, 0.2)$	$\hat{\bar{\theta}}$	0.1987	0.1980	0.1999	0.1996	0.2003	0.2002
	$\hat{SD}(\hat{\theta})$	0.1160	0.1147	0.0592	0.0583	0.0400	0.0397
	$\hat{\alpha}$	-0.7861	-0.7846	-0.7988	-0.7988	-0.8031	-0.8033
Ising	$\hat{SD}(\hat{\alpha})$	0.2381	0.2345	0.1136	0.1130	0.0776	0.0771
$(-0.8, 0.4)$	$\hat{\bar{\theta}}$	0.3945	0.3940	0.3999	0.3999	0.4019	0.4020
	$\hat{SD}(\hat{\theta})$	0.1151	0.1129	0.0545	0.0542	0.0375	0.0373
	$\hat{\alpha}$	-1.1757	-1.1711	-1.1983	-1.1966	-1.1989	-1.1982
Ising	$\hat{SD}(\hat{\alpha})$	0.2484	0.2391	0.1146	0.1128	0.0803	0.0792
$(-1.2, 0.6)$	$\hat{\bar{\theta}}$	0.5876	0.5853	0.5990	0.5981	0.6001	0.5998
	$\hat{SD}(\hat{\theta})$	0.1208	0.1159	0.0553	0.0545	0.0386	0.0380
	$\hat{\alpha}$	-1.5726	-1.5657	-1.5974	-1.5952	-1.6049	-1.6039
Ising	$\hat{SD}(\hat{\alpha})$	0.2342	0.2280	0.1111	0.1074	0.0759	0.0726
$(-1.6, 0.8)$	$\hat{\bar{\theta}}$	0.7867	0.7832	0.7992	0.7981	0.8025	0.8020
	$\hat{SD}(\hat{\theta})$	0.1116	0.1083	0.0535	0.0516	0.0367	0.0351
	$\hat{\alpha}$	-1.9893	-1.9809	-2.0048	-2.0026	-2.0004	-2.0002
Ising	$\hat{SD}(\hat{\alpha})$	0.2250	0.2156	0.1020	0.0972	0.0702	0.0663
$(-2.0, 1.0)$	$\hat{\bar{\theta}}$	0.9950	0.9906	1.0027	1.0016	1.0002	1.0002
	$\hat{SD}(\hat{\theta})$	0.1107	0.1064	0.0507	0.0481	0.0345	0.0326
	$\hat{\alpha}$	-2.3874	-2.3777	-2.3992	-2.3975	-2.4053	-2.4045
Ising	$\hat{SD}(\hat{\alpha})$	0.2088	0.1909	0.1044	0.0953	0.0650	0.0592
$(-2.4, 1.2)$	$\hat{\bar{\theta}}$	1.1950	1.1899	1.1997	1.1988	1.2025	1.2021
	$\hat{SD}(\hat{\theta})$	0.1011	0.0915	0.0509	0.0464	0.0320	0.0291
	$\hat{\alpha}$	-2.8197	-2.8086	-2.8077	-2.8061	-2.8045	-2.8029
Ising	$\hat{SD}(\hat{\alpha})$	0.2191	0.1973	0.1031	0.0930	0.0693	0.0629
$(-2.8, 1.4)$	$\hat{\bar{\theta}}$	1.4087	1.4032	1.4034	1.4026	1.4022	1.4013
	$\hat{SD}(\hat{\theta})$	0.1088	0.0977	0.0516	0.0464	0.0344	0.0312
	$\hat{\alpha}$	-3.1619	-3.1602	-3.1950	-3.1954	-3.1995	-3.1999
Ising	$\hat{SD}(\hat{\alpha})$	0.1884	0.1661	0.0925	0.0819	0.0597	0.0521
$(-3.2, 1.6)$	$\hat{\bar{\theta}}$	1.5763	1.5754	1.5944	1.5946	1.5973	1.5973
	$\hat{SD}(\hat{\theta})$	0.1000	0.0879	0.0491	0.0432	0.0314	0.0272

Table A.7 (continued): Sample Means and Standard Errors of α and θ when the Model is an Ising Model Near Phase Transition and the Methods of Estimation are PL, GPL for Group Sizes of 5 (gs5) and 9 (gs9), BGPL for Block Sizes of 2x2 (bs2), 3x3 (bs3), and 4x4 (bs4), and MCMCML

		26x26			50x50			74x74		
Model	Estimate	bs2	bs3	bs4	bs2	bs3	bs4	bs2	bs3	bs4
	$\hat{\alpha}$	-0.4004	-0.4065	-0.3981	-0.4004	-0.4031	-0.4007	-0.4002	-0.4035	-0.4013
Ising	$\hat{SD}(\hat{\alpha})$	0.2434	0.2461	0.2372	0.1261	0.1238	0.1234	0.0883	0.0881	0.0874
$(-0.4, 0.2)$	$\bar{\hat{\theta}}$	0.1989	0.2019	0.1977	0.1990	0.2003	0.1992	0.1993	0.2009	0.1998
	$\hat{SD}(\hat{\theta})$	0.1153	0.1173	0.1121	0.0606	0.0598	0.0594	0.0414	0.0418	0.0410
	$\hat{\alpha}$	-0.7904	-0.7824	-0.7863	-0.8029	-0.8004	-0.8006	-0.8055	-0.8040	-0.8040
Ising	$\hat{SD}(\hat{\alpha})$	0.2380	0.2333	0.2312	0.1143	0.1169	0.1142	0.0805	0.0812	0.0806
$(-0.8, 0.4)$	$\bar{\hat{\theta}}$	0.3965	0.3924	0.3944	0.4022	0.4008	0.4011	0.4030	0.4022	0.4023
	$\hat{SD}(\hat{\theta})$	0.1158	0.1139	0.1124	0.0549	0.0568	0.0548	0.0390	0.0397	0.0389
	$\hat{\alpha}$	-1.1781	-1.1779	-1.1746	-1.1969	-1.1988	-1.1925	-1.2003	-1.2011	-1.1978
Ising	$\hat{SD}(\hat{\alpha})$	0.2377	0.2418	0.2294	0.1176	0.1181	0.1170	0.0832	0.0828	0.0826
$(-1.2, 0.6)$	$\bar{\hat{\theta}}$	0.5885	0.5886	0.5868	0.5985	0.5994	0.5965	0.6009	0.6013	0.5997
	$\hat{SD}(\hat{\theta})$	0.1160	0.1173	0.1127	0.0566	0.0574	0.0567	0.0401	0.0399	0.0399
	$\hat{\alpha}$	-1.5835	-1.5774	-1.5758	-1.5996	-1.5991	-1.5957	-1.6053	-1.6053	-1.6044
Ising	$\hat{SD}(\hat{\alpha})$	0.2314	0.2328	0.2306	0.1154	0.1143	0.1122	0.0779	0.0777	0.0753
$(-1.6, 0.8)$	$\bar{\hat{\theta}}$	0.7927	0.7894	0.7889	0.8003	0.7998	0.7982	0.8027	0.8027	0.8023
	$\hat{SD}(\hat{\theta})$	0.1107	0.1120	0.1093	0.0552	0.0549	0.0534	0.0377	0.0377	0.0363
	$\hat{\alpha}$	-1.9911	-1.9847	-1.9832	-2.0028	-2.0003	-2.0015	-1.9981	-1.9977	-1.9985
Ising	$\hat{SD}(\hat{\alpha})$	0.2179	0.2175	0.2159	0.1055	0.1024	0.1019	0.0724	0.0697	0.0705
$(-2.0, 1.0)$	$\bar{\hat{\theta}}$	0.9955	0.9924	0.9918	1.0016	1.0004	1.0010	0.9992	0.9990	0.9993
	$\hat{SD}(\hat{\theta})$	0.1077	0.1066	0.1063	0.0523	0.0503	0.0503	0.0358	0.0344	0.0346
	$\hat{\alpha}$	-2.3885	-2.3807	-2.3784	-2.4017	-2.3995	-2.4001	-2.4060	-2.4054	-2.4055
Ising	$\hat{SD}(\hat{\alpha})$	0.2151	0.1947	0.1963	0.1063	0.1014	0.0982	0.0671	0.0635	0.0608
$(-2.4, 1.2)$	$\bar{\hat{\theta}}$	1.1959	1.1924	1.1904	1.2010	1.2002	1.2002	1.2029	1.2026	1.2026
	$\hat{SD}(\hat{\theta})$	0.1043	0.0931	0.0938	0.0520	0.0490	0.0473	0.0332	0.0309	0.0297
	$\hat{\alpha}$	-2.8257	-2.8071	-2.8106	-2.8115	-2.8049	-2.8059	-2.8060	-2.8016	-2.8034
Ising	$\hat{SD}(\hat{\alpha})$	0.2222	0.2001	0.1901	0.1037	0.0948	0.0919	0.0689	0.0656	0.0633
$(-2.8, 1.4)$	$\bar{\hat{\theta}}$	1.4123	1.4022	1.4054	1.4051	1.4016	1.4024	1.4029	1.4006	1.4015
	$\hat{SD}(\hat{\theta})$	0.1104	0.0987	0.0946	0.0517	0.0471	0.0456	0.0341	0.0325	0.0309
	$\hat{\alpha}$	-3.1638	-3.1685	-3.1628	-3.1935	-3.1974	-3.1976	-3.1970	-3.1981	-3.1984
Ising	$\hat{SD}(\hat{\alpha})$	0.1853	0.1799	0.1594	0.0935	0.0892	0.0821	0.0593	0.0567	0.0526
$(-3.2, 1.6)$	$\bar{\hat{\theta}}$	1.5775	1.5802	1.5769	1.5934	1.5957	1.5954	1.5962	1.5965	1.5963
	$\hat{SD}(\hat{\theta})$	0.0980	0.0951	0.0834	0.0493	0.0467	0.0426	0.0314	0.0297	0.0274