Construction and behavior of Multinomial Markov random field models

Kim Mueller

Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/etd

Part of the Statistics and Probability Commons

Recommended Citation
Mueller, Kim, "Construction and behavior of Multinomial Markov random field models" (2010). Graduate Theses and Dissertations. 11419.
https://lib.dr.iastate.edu/etd/11419

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
Construction and behavior of Multinomial Markov random field models

by

Kim Marie Mueller

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Statistics

Program of Study Committee:
Mark S. Kaiser, Major Professor
Shauna Hallmark
Daniel Nordman
Stephen Vardeman
Huaiqing Wu

Iowa State University
Ames, Iowa
2010

Copyright © Kim Marie Mueller, 2010. All rights reserved.
# TABLE OF CONTENTS

**CHAPTER LIST OF TABLES** ........................................ iii

**CHAPTER LIST OF FIGURES** ...................................... vi

**CHAPTER 1  GENERAL INTRODUCTION** .......................... 1

**CHAPTER 2  LITERATURE REVIEW** .............................. 3

  2.1 Markov Random Field (MRF) Models .......................... 3
  2.2 Construction of Conditional Distribution Form .................. 4
  2.3 Centered Parameterizations of the Natural Parameter Function .... 5
  2.4 Estimation of Markov Random Field (MRF) Model Parameters ....... 6

**CHAPTER 3  CONSTRUCTION OF MULTINOMIAL MARKOV RANDOM FIELD MODEL** .......................... 7

  3.1 Problem Setting .............................................. 7
  3.2 Conditional Distribution Form ................................ 8
  3.3 Construction of the Multinomial MRF Model ...................... 11
  3.4 Simulation Method and Estimation ................................ 13
  3.5 Comparison of Traditional and Centered Models .................. 16
  3.6 Bounds for the Spatial Dependence Parameter .................. 20

**CHAPTER 4  MODEL BEHAVIOR** .................................. 26

  4.1 Asymmetry of Multinomial MRF Model .......................... 26
  4.2 Variances and Covariances of Conditional Expectations ............ 30
  4.3 Marginal Variances and Covariances ............................ 41
  4.4 Representation of Dependence .................................. 48
4.5 Dependence of Parameter Estimation and PMSE on Category Indices ... 52
4.6 Assignment of Category Indices ........................................... 61

CHAPTER 5 APPLICATION .................................................... 65

5.1 Introduction ................................................................. 65
5.2 Data Description ............................................................ 65
5.3 Model Formulation ......................................................... 67
  5.3.1 Response Variable ...................................................... 67
  5.3.2 Neighborhoods ......................................................... 67
  5.3.3 Conditional Probability Mass Function ......................... 67
  5.3.4 Estimation with the Pseudo-Likelihood Function .......... 68
5.4 Issues in Estimation ....................................................... 68
5.5 Comparison of Models .................................................... 70
  5.5.1 Multinomial MRF Model (Model 1) ....................... 70
  5.5.2 Multinomial MRF Model with Covariates (Model 2) ... 75
  5.5.3 Multinomial MRF Model with Median Polish (Model 3) 80

CHAPTER 6 GENERAL CONCLUSION ......................................... 87

BIBLIOGRAPHY ................................................................. 89

ACKNOWLEDGEMENTS ....................................................... 90
LIST OF TABLES

Table 4.1  MC Approximations of the Expected Values of the Parameter Estimates for \( k = (0.20, 0.30, 0.50)^T \) .................................................. 55
Table 4.2  MC Approximations of the Mean Squared Error for \( k = (0.20, 0.30, 0.50)^T \) 55
Table 4.3  MC Approximations of the Expected Values of the Prediction Mean Squared Errors for \( k = (0.20, 0.30, 0.50)^T \) .................................................. 55
Table 4.4  MC Approximations of the Expected Values of the Parameter Estimates for \( k = (0.20, 0.50, 0.30)^T \) .................................................. 56
Table 4.5  MC Approximations of the Mean Squared Error for \( k = (0.20, 0.50, 0.30)^T \) 56
Table 4.6  MC Approximations of the Expected Values of the Prediction Mean Squared Errors for \( k = (0.20, 0.50, 0.30)^T \) .................................................. 56
Table 4.7  MC Approximations of the Expected Values of the Parameter Estimates for \( k = (0.30, 0.50, 0.20)^T \) .................................................. 57
Table 4.8  MC Approximations of the Mean Squared Error for \( k = (0.30, 0.50, 0.20)^T \) 57
Table 4.9  MC Approximations of the Expected Values of the Prediction Mean Squared Errors for \( k = (0.30, 0.50, 0.20)^T \) .................................................. 57
Table 4.10 MC Approximations of the Expected Values of the Parameter Estimates for \( k = (0.30, 0.30, 0.40)^T \) .................................................. 58
Table 4.11 MC Approximations of the Mean Squared Error for \( k = (0.30, 0.30, 0.40)^T \) 58
Table 4.12 MC Approximations of the Expected Values of the Prediction Mean Squared Errors for \( k = (0.30, 0.30, 0.40)^T \) .................................................. 58
Table 4.13 MC Approximations of the Expected Values of the Parameter Estimates for \( k = (0.30, 0.40, 0.30)^T \) .................................................. 59
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.14</td>
<td>MC Approximations of the Mean Squared Error for $k = (0.30, 0.40, 0.30)^T$</td>
<td>59</td>
</tr>
<tr>
<td>4.15</td>
<td>MC Approximations of the Expected Values of the Prediction Mean Squared Errors for $k = (0.30, 0.40, 0.30)^T$</td>
<td>59</td>
</tr>
<tr>
<td>5.1</td>
<td>Marginal Means</td>
<td>71</td>
</tr>
<tr>
<td>5.2</td>
<td>Parameter Estimates for Model 1</td>
<td>75</td>
</tr>
<tr>
<td>5.3</td>
<td>Parameter Estimates for Model 2</td>
<td>76</td>
</tr>
<tr>
<td>5.4</td>
<td>Marginal Probabilities for 1980</td>
<td>80</td>
</tr>
<tr>
<td>5.5</td>
<td>Marginal Probabilities for 1985</td>
<td>80</td>
</tr>
<tr>
<td>5.6</td>
<td>Marginal Probabilities for 1990</td>
<td>80</td>
</tr>
<tr>
<td>5.7</td>
<td>Marginal Means of the Residuals</td>
<td>82</td>
</tr>
<tr>
<td>5.8</td>
<td>Parameter Estimates for Model 3</td>
<td>84</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>3.1</td>
<td>Comparison of Monte Carlo approximations of marginal expectations for traditional model and centered model to the marginal expectations for a model of independence for $\kappa_1 = 0.20$ and $\kappa_2 = 0.50$</td>
<td>19</td>
</tr>
<tr>
<td>3.2</td>
<td>Comparison of Monte Carlo approximations of marginal expectations for traditional model and centered model to the marginal expectations for a model of independence for $\kappa_1 = 0.30$ and $\kappa_2 = 0.30$</td>
<td>19</td>
</tr>
<tr>
<td>3.3</td>
<td>Monte Carlo approximations of marginal expectations for $\kappa_1 = 0.10$, $\kappa_2 \in {0.10, 0.20, \ldots, 0.80}$ and $\gamma \in {0, 0.5, 1.0, \ldots, 8}$</td>
<td>22</td>
</tr>
<tr>
<td>3.4</td>
<td>Monte Carlo approximations of marginal expectations for $\kappa_1 = 0.20$, $\kappa_2 \in {0.10, 0.20, \ldots, 0.80}$ and $\gamma \in {0, 0.5, 1.0, \ldots, 8}$</td>
<td>22</td>
</tr>
<tr>
<td>3.5</td>
<td>Monte Carlo approximations of marginal expectations for $\kappa_1 = 0.30$, $\kappa_2 \in {0.10, 0.20, \ldots, 0.80}$ and $\gamma \in {0, 0.5, 1.0, \ldots, 8}$</td>
<td>22</td>
</tr>
<tr>
<td>4.1</td>
<td>Monte Carlo approximations of variance of conditional expectations when $\kappa_1 = 0.20$ and $\kappa_2 = 0.30$</td>
<td>33</td>
</tr>
<tr>
<td>4.2</td>
<td>Monte Carlo approximations of variance of conditional expectations when $\kappa_1 = 0.20$ and $\kappa_2 = 0.50$</td>
<td>33</td>
</tr>
<tr>
<td>4.3</td>
<td>Monte Carlo approximations of variance of conditional expectations when $\kappa_1 = 0.30$ and $\kappa_2 = 0.50$</td>
<td>33</td>
</tr>
<tr>
<td>4.4</td>
<td>Monte Carlo approximations of variance of conditional expectations when $\kappa_1 = 0.30$ and $\kappa_2 = 0.30$</td>
<td>34</td>
</tr>
<tr>
<td>4.5</td>
<td>Monte Carlo approximations of variance of conditional expectations when $\kappa_1 = 0.30$ and $\kappa_2 = 0.40$</td>
<td>34</td>
</tr>
</tbody>
</table>
Figure 4.6 Monte Carlo approximations of covariance of conditional expectations when $\kappa_1 = 0.20$ and $\kappa_2 = 0.30$ .............................................. 35

Figure 4.7 Monte Carlo approximations of covariance of conditional expectations when $\kappa_1 = 0.20$ and $\kappa_2 = 0.50$ .............................................. 35

Figure 4.8 Monte Carlo approximations of covariance of conditional expectations when $\kappa_1 = 0.30$ and $\kappa_2 = 0.50$ .............................................. 35

Figure 4.9 Monte Carlo approximations of covariance of conditional expectations when $\kappa_1 = 0.30$ and $\kappa_2 = 0.30$ .............................................. 36

Figure 4.10 Monte Carlo approximations of covariance of conditional expectations when $\kappa_1 = 0.30$ and $\kappa_2 = 0.40$ .............................................. 36

Figure 4.11 Difference in absolute value of partial derivatives, $|f_{D_1,D_2}(D_{i,1},D_{i,2})| − |f_{D_1,D_2}(D_{i,1},D_{i,2})|$ .............................................. 38

Figure 4.12 Comparison of changes in conditional expectation for category 1 and category 2 to changes in conditional expectation for category 3 when $\kappa_1 = 0.10$ and $\kappa_2 = 0.30$ .............................................. 40

Figure 4.13 Monte Carlo approximations of variance of marginal expectations when $\kappa_1 = 0.20$ and $\kappa_2 = 0.30$ .............................................. 44

Figure 4.14 Monte Carlo approximations of variance of marginal expectations when $\kappa_1 = 0.20$ and $\kappa_2 = 0.50$ .............................................. 44

Figure 4.15 Monte Carlo approximations of variance of marginal expectations when $\kappa_1 = 0.30$ and $\kappa_2 = 0.50$ .............................................. 44

Figure 4.16 Monte Carlo approximations of variance of marginal expectations when $\kappa_1 = 0.30$ and $\kappa_2 = 0.30$ .............................................. 45

Figure 4.17 Monte Carlo approximations of variance of marginal expectations when $\kappa_1 = 0.30$ and $\kappa_2 = 0.40$ .............................................. 45

Figure 4.18 Monte Carlo approximations of covariance of marginal expectations when $\kappa_1 = 0.20$ and $\kappa_2 = 0.30$ .............................................. 46
Figure 4.19 Monte Carlo approximations of covariance of marginal expectations when $\kappa_1 = 0.20$ and $\kappa_2 = 0.50$ ........................................... 46
Figure 4.20 Monte Carlo approximations of covariance of marginal expectations when $\kappa_1 = 0.30$ and $\kappa_2 = 0.50$ ........................................... 46
Figure 4.21 Monte Carlo approximations of covariance of marginal expectations when $\kappa_1 = 0.30$ and $\kappa_2 = 0.30$ ........................................... 47
Figure 4.22 Monte Carlo approximations of covariance of marginal expectations when $\kappa_1 = 0.30$ and $\kappa_2 = 0.40$ ........................................... 47
Figure 4.23 Standard bounds for the Binomial MRF model dependence parameter ......................................................... 51
Figure 4.24 Monte Carlo approximations of $\gamma^*_k$, the standardized dependence parameter with $\gamma = 0.5$ ........................................... 53
Figure 4.25 Monte Carlo approximations of $\gamma^*_k$, the standardized dependence parameter with $\gamma = 2.0$ ........................................... 53
Figure 4.26 Probability of labeling the category originally indexed as the third category as the third category after fitting three Binomial MRF models . 64

Figure 5.1 Sampled locations from the North American Regional Reanalysis (NARR) data set ................................................................. 66
Figure 5.2 Image plots of $y_k(s_i)$ for $k = 1, 2, 3$ for year 1980 ................. 72
Figure 5.3 Image plots of $y_k(s_i)$ for $k = 1, 2, 3$ for year 1985 ................. 72
Figure 5.4 Image plots of $y_k(s_i)$ for $k = 1, 2, 3$ for year 1990 ................. 72
Figure 5.5 Profile of $-\log (P(\theta))$ for years 1980, 1985 and 1990 for Model 1 .... 74
Figure 5.6 Profile of $-\log (P(\theta))$ for year 1980 for Model 2 .................... 77
Figure 5.7 Profile of $-\log (P(\theta))$ for year 1985 for Model 2 .................... 78
Figure 5.8 Profile of $-\log (P(\theta))$ for year 1990 for Model 2 .................... 79
Figure 5.9 Image plots of $y_k(s_i)$ for $k = 1, 2, 3$ for year 1980 ................. 83
Figure 5.10 Image plots of $y_k(s_i)$ for $k = 1, 2, 3$ for year 1985 ................. 83
Figure 5.11 Image plots of $y_k(s_i)$ for $k = 1, 2, 3$ for year 1990 ................. 83
Figure 5.12 Profile of $-\log (P(\theta))$ for years 1980, 1985 and 1990 for Model 3 .... 85
CHAPTER 1 GENERAL INTRODUCTION

Models that are constructed from conditionally specified distributions are often applied to
data sets that possess a spatial structure, even data sets with complex dependence structures.
These conditionally specified distributions specify the distribution of a value at a location
given the values at all other locations. If the value is dependent only on values at a subset
of locations, called a neighborhood, then the resulting joint probability measure is referred
to as a Markov random field (MRF) model. When the conditionally specified distributions
are exponential family distributions, several results are available; hence, there has been much
interest in Markov random field models that have been constructed with Gaussian, Poisson,
and binomial distributions specified as the conditionally specified distributions. For the Gaussian
MRF model, the joint distribution can be written, and thus, many nice properties and
results are available for this model. For the Poisson and Binomial MRF models, the joint
distribution can only be identified up to an unknown constant; however, these models have
been studied and used to model spatial data as well. One exponential family distribution that
has not been subject to much interest in the area of MRF models as of yet is the multinomial
distribution, even though the multinomial distribution is an extension of the binomial distrib-
ution. Consequently, in this paper, we construct a MRF model with multinomial conditional
distributions and then study the behavior of this model regarding, for example, symmetry of
the model, variances and covariances of the conditional expectations and marginal variances
and covariances. Finally, this model is applied to a data set with spatial structure.

The remainder of the dissertation is organized as follows: In Chapter 2, general construction
and estimation of MRF models is reviewed. In Chapter 3, construction and estimation of
Multinomial MRF models is presented. Then the behavior of the Multinomial MRF model
is studied in Chapter 4 while Chapter 5 discusses the issues in applying a Multinomial MRF model to analyze wind speeds across Iowa and surrounding states. Finally, Chapter 6 closes with some general concluding remarks.
CHAPTER 2 LITERATURE REVIEW

2.1 Markov Random Field (MRF) Models

Markov random field models apply to spatial processes that can occur on a regular or an irregular system of sites that consist of points or regions. We will restrict the discussion to a regular system of \( n \) points that is often referred to as a regular lattice. The points (or locations denoted as \( s_i \) for \( i = 1, ..., n \)) on the lattice may be associated with observations that are continuous or discrete. A probability density or mass function is chosen to model the observable process such as, for example, Gaussian, Poisson or binomial. A neighborhood, denoted as \( N_i \), for location \( s_i; i = 1, ..., n \), is specified such that \( N_i \equiv \{ s_j : s_j \text{ is a neighbor of } s_i \} \).

On a regular grid with integer indices \( u_i \) in the horizontal coordinate and \( v_i \) in the vertical coordinate, a common neighborhood structure is a four-nearest neighbor specification, namely, \( N_i = \{ s_j : (u_j = u_i \pm 1, v_j = v_i), (u_j = u_i, v_j = v_i \pm 1) \} \). Another useful definition for neighborhoods, particularly if \( u_i \) and \( v_i \) denote physical distances from some origin, is to consider a location \( s_j \) to be a neighbor of location \( s_i \) if the distance between them is less than some specified value \( D \), that is, \( N_i = \{ s_j : d(s_i, s_j) \leq D \} \), where \( d(s_i, s_j) = \{(u_i - u_j)^2 + (v_i - v_j)^2\}^{1/2} \).

Let \( y(N_i) = \{ y(s_j) : s_j \in N_i \} \) denote the values of \( Y(s_j) \) at the neighbors of \( s_i \) for \( i = 1, ..., n \). Then, with \( [X] \) denoting the distribution of an arbitrary random variable \( X \), a Markov assumption is that, for each \( s_i; i = 1, ..., n \), the distribution of \( Y(s_i) \) given values at all other locations depends only on values at its neighbors. Specifically,

\[
[Y(s_i)] \{ y(s_j) : j \neq i \} = [Y(s_i)|y(N_i)].
\]

A Markov random field (MRF) model results from specification of the neighborhoods \( N_i \) and a conditional distribution for each variable \( Y(s_i) \) for \( i = 1, ..., n \).
2.2 Construction of Conditional Distribution Form

When a probability density or mass function can be written in exponential family form, we can apply several results. One way to write a probability density or mass function in exponential family form is

\[ f(x|\phi) = \exp \left[ \sum_{k=1}^{s} \phi_k T_k(x) - B(\phi) + C(x) \right], \]

(2.1)

where \( \phi = (\phi_1, \ldots, \phi_s)^T \) is called the natural parameter and \( \{T_k(x) : k = 1, \ldots, s\} \) is the set of minimal sufficient statistics. Once a neighborhood structure has been specified, we can write the conditional probability mass function for \( y(s_i) = (y_1(s_i), \ldots, y_h(s_i))^T \) given \( y(N_i) = (y_1(N_i), \ldots, y_h(N_i))^T \) and \( \theta \) in a form similar to (2.1), namely,

\[ f_i(y(s_i)|y(N_i); \theta) = \exp \left[ \sum_{k=1}^{s} y_k(s_i)A_{i,k} \{y(N_i); \theta\} - B_i \{y(N_i); \theta\} + C_i \{y(s_i)\} \right], \]

(2.2)

where \( A_{i,k}(\cdot) \) is known as the natural parameter function, which depends on the neighboring values \( y(N_i) \) and \( \theta \).

For one-parameter exponential families, (i.e., \( s = 1 \)), Besag (1974) showed that the natural parameter function must be of the form

\[ A_i \{y(N_i); \theta\} = \alpha_i + \sum_{s_j \in N_i} \eta_{i,j} y(s_j) \]

(2.3)

with \( \eta_{i,j} = \eta_{j,i} \). For multi-parameter exponential families, Kaiser et al. (2002) gives three different forms. One of the forms is

\[ A_{i,k} \{y(N_i); \theta\} = \alpha_{i,k} + \sum_{s_j \in N_i} \eta_{i,j,k} T_k(y(s_j)) \]

(2.4)

such that \( \eta_{i,j,k} = \eta_{j,i,k} \) for all \( i \neq j \) and \( k = 1, \ldots, s \). Often the number of parameters is reduced by assuming, for example, a single dependence parameter \( \eta \) such that \( \eta = \eta_{i,j,k} \) for all \( i, j \) and \( k \) and \( \alpha_k = \alpha_{i,k} \) for all \( i \).

There are conditions needed for a joint distribution to exist according to Kaiser and Cressie (2000). Even if the joint distribution exists, often that distribution can only be identified up to an unknown constant depending on the parameter \( \theta \). Consequently, the Markov random field model is often specified through conditional distributions. For further discussion on the joint distribution of a MRF model, see Section 3.3.
2.3 Centered Parameterizations of the Natural Parameter Function

When the observed values can only take on positive values or 0, for either 1-parameter or multi-parameter exponential families, neighboring values can only increase the natural parameter functions of (2.3) or (2.4) or leave the natural parameter function unchanged if all neighboring values are 0. The form of expressions (2.3) and (2.4) do not make clear what parameters will only affect marginal expectations and what models will only affect statistical dependence. Furthermore, for some 1-parameter exponential family MRF models, the conditional expectation at location $s_i$ for all $i$ can only be monotone increasing in the natural parameter function, $A_i \{y(N_i); \theta\}$. Thus, we would not expect $\alpha_i$ in (2.3) or $\alpha_{i,k}$ in (2.4) to represent marginal expectations. To allow $\alpha_i$ in (2.3) or $\alpha_{i,k}$ in (2.4) to model or approximately model the marginal expectations with some restrictions, we can reparameterize (2.3) (and similarly (2.4)) as

$$A_i \{y(N_i); \theta\} = \tau^{-1}(\kappa_i) + \sum_{j \in N_i} \eta_{i,j} \{y(s_i) - \kappa_j\}, \quad (2.5)$$

where $\tau^{-1}(\kappa_i)$ maps expected values into exponential family natural parameters. This parameterization is referred to as centered parameterization. For Gaussian models, (2.5) can be written as

$$A_i \{y(N_i); \theta\} = \kappa_i + \sum_{j \in N_i} \eta_{i,j} \{y(s_i) - \kappa_j\}, \quad (2.6)$$

where $\kappa_i$ is known to be the marginal expectation of location $s_i$ (Cressie 1993). For Binary MRF models, which are MRF models that specify the binomial distribution to model one observation per location, (2.5) can be written as

$$A_i \{y(N_i); \theta\} = \log \left( \frac{\kappa_i}{1 - \kappa_i} \right) + \sum_{j \in N_i} \eta_{i,j} \{y(s_i) - \kappa_j\}. \quad (2.7)$$

Kaiser et al. (2010) show that for Binary MRF models located on a transect such that $\kappa_i = \kappa$ for all locations $s_i$, $\kappa$ is nearly the marginal expectation of all locations when the dependence parameter is within specified bounds (or standard bounds). As the value of the dependence parameter increases, the marginal expectation decays to either 0 or 1. When the dependence
parameter is within these standard bounds, the centered parameterization allows the model to have components to capture marginal expectations or large-scale structure, namely, \( \tau^{-1}(\kappa_i) \), and components to represent the remaining structure or small-scale structure, namely, the dependence parameters, \( \eta_{i,j} \). If the marginal expectations across locations are not constant, covariates can be incorporated into the model such that \( \tau^{-1}(\kappa_i) = x(s_i)^T \beta \). If \( \tau^{-1}(\kappa_i) \) is nearly the marginal expectation at location \( s_i \), then \( x(s_i)^T \beta \) is nearly the marginal expectation at location \( s_i \), which then allows for a nice interpretation of \( \beta \).

### 2.4 Estimation of Markov Random Field (MRF) Model Parameters

Estimating parameters by maximizing the likelihood function is difficult because the joint probability density or mass function is not known in explicit form for many Markov random field models. However, we can find estimates based on the conditional density or mass functions. Besag (1975) suggested maximizing a pseudo-likelihood function, defined as the product of the conditional mass functions, to obtain parameter estimates. This pseudo-likelihood function may be written as

\[
P(\theta) = \prod_{i=1}^{n} f_i(y(s_i)|y(N_i); \theta),
\]

where \( f_i(y(s_i)|y(N_i); \theta) \) is given by (2.2).
CHAPTER 3 CONSTRUCTION OF MULTINOMIAL MARKOV RANDOM FIELD MODEL

3.1 Problem Setting

Suppose that \( n \) cells are created by overlaying a geographic area with a regular grid, and arbitrarily indexed by \( i = 1, \ldots, n \). Consider an observable process such that within each cell, a fixed number of events, \( m_i \) for \( i = 1, \ldots, n \), occur and each event belongs to one of \( h \) distinct categories. Let \( s_i \) denote the spatial location of cell \( i \), such as \( s_i = (u_i, v_i) \) where \( u_i \) denotes horizontal position and \( v_i \) denotes vertical position according to some convenient coordinate system in Euclidean space. For example, \( u_i \in \{1, 2, \ldots, U\} \) and \( v_i \in \{1, 2, \ldots, V\} \) might be integer indices relative to some specified origin, or \( u_i \) and \( v_i \) might be eastings and northings, respectively, from a universal trans-mercator projection. Then associate with the observable process the random variables \( Y_k(s_i) \) for \( k = 1, \ldots, h \), and \( i = 1, \ldots, n \), to represent the number of events belonging to the \( k^{th} \) category at location \( s_i \). Furthermore, let \( p_{i,k} \) represent the probability of an event belonging to category \( k \) at location \( s_i \). Then, at a specified location \( s_i \), we assume that the vector \( Y(s_i) \equiv (Y_1(s_i), Y_2(s_i), \ldots, Y_h(s_i))^T \) has a multinomial probability mass function with parameters \( p_i \equiv (p_{i,1}, p_{i,2}, \ldots, p_{i,h})^T \) such that \( p_{i,k} > 0 \) and \( \sum_{k=1}^{h} p_{i,k} = 1 \), namely,

\[
    f(y(s_i)|p_i) = \frac{m_i!}{y_{1,k}! \cdots y_{h,k}!} \left( \prod_{k=1}^{h-1} p_{i,k}^{y_{k+1}(s_i)} \right) \left( 1 - \sum_{k=1}^{h-1} p_{i,k} \right)^{m_i - \sum_{k=1}^{h-1} y_{k}(s_i)}. \quad (3.1)
\]

To formulate a Markov random field version of this multinomial model we require specification of a neighborhood \( N_i \) for each location \( s_i \) for \( i = 1, \ldots, n \), such that \( N_i \equiv \{s_j : s_j \text{ is a neighbor of } s_i\} \). On a regular grid with integer indices \( u_i \) and \( v_i \), a common neighborhood structure is a four-nearest neighbor specification, namely, \( N_i = \{s_j : (u_j = u_i \pm 1, v_j = v_i), (u_j = u_i, v_j \pm 1)\} \). An-
other useful definition for neighborhoods, particularly if $u_i$ and $v_i$ denote physical distances from some origin, is to consider a location $s_j$ to be a neighbor of location $s_i$ if the distance between them is less than some specified value $D$, that is, $N_i = \{s_j : d(s_i, s_j) \leq D\}$, where $d(s_i, s_j) = \{(u_i - u_j)^2 + (v_i - v_j)^2\}^{1/2}$. Let $y(N_i) = \{y(s_j) : s_j \in N_i\}$ denote the values of $Y(s_j)$ at the neighbors of $s_i$ for $i = 1, ..., n$. Then, with $[X]$ denoting the distribution of an arbitrary random variable $X$, a Markov assumption is that, for each $s_i$, the distribution of $Y(s_i)$ given values at all other locations depends only on values at its neighbors. Specifically,

$$[Y(s_i) | \{y(s_j) : j \neq i\}] = [Y(s_i) | y(N_i)].$$

(3.2)

A Markov random field (MRF) model results from specification of the neighborhoods $N_i$ and a conditional distribution for each variable $Y(s_i)$ for $i = 1, ..., n$.

### 3.2 Conditional Distribution Form

To formulate conditional distributions based on the form of the multinomial probability mass function, we will use the fact that the standard multinomial mass function given in (3.1) can be written in exponential family form as

$$f(x | \phi) = \exp \left[ \sum_{k=1}^{s} \phi_k T_k(x) - B(\phi) + C(x) \right],$$

(3.3)

where $\phi = (\phi_1, \ldots, \phi_s)^T$ is called the natural parameter and $\{T_k(x) : k = 1, \ldots, s\}$ is the set of minimal sufficient statistics. In the case of the multinomial probability mass function, the natural parameter is $\phi = (\log \left( \frac{p_{i,1}}{p_{i,h}} \right), \ldots, \log \left( \frac{p_{i,h-1}}{p_{i,h}} \right))^T$ such that $p_{i,k}$ represents the probability of an event belonging to category $k$ at location $s_i$. The minimal sufficient statistics are $T_k(y(s_i)) = y_k(s_i); k = 1, \ldots, h - 1$, the number of events in category $k$ at location $s_i$. Furthermore, for $y(s_i) = (y_1(s_i), \ldots, y_h(s_i))^T$, we have

$$B(\phi) = m_i \log \left( 1 - \sum_{k=1}^{h-1} p_{i,k} \right) \quad \text{and} \quad C(y(s_i)) = \log \left( \frac{m_i!}{y_1(s_i)! \cdots y_{h-1}(s_i)! (m_i - \sum_{k=1}^{h-1} y_k(s_i))!} \right),$$

where $m_i$ is the total number of events at location $s_i$.

When the density or mass function chosen to model the observable process can be written in exponential family form, the conditional probability mass function for $y(s_i) = (y_1(s_i), \ldots, y_h(s_i))^T$
conditioned on \( y(N_i) = (y_1(N_i), ..., y_h(N_i))^T \) and \( \theta \) can be written in a similar form to (3.3), namely,

\[
f_i(y(s_i)|y(N_i); \theta) = \exp \left[ \sum_{k=1}^{h-1} y_k(s_i) A_{i,k} \{y(N_i); \theta\} - B_i \{y(N_i); \theta\} + C_i \{y(s_i)\} \right], \tag{3.4}
\]

where \( A_{i,k}(\cdot) \) is known as the natural parameter function, which depends on \( \theta \). In the case of the Multinomial MRF model, the conditional probability mass function for \( y(s_i) = (y_1(s_i), ..., y_h(s_i))^T \) conditioned on \( y(N_i) = (y_1(N_i), ..., y_h(N_i))^T \) and \( \theta \) is

\[
f_i(y(s_i)|y(N_i); \theta) = \exp \left[ \sum_{k=1}^{h-1} y_k(s_i) A_{i,k} \{y(N_i); \theta\} - m_i \log \left( \frac{m_i!}{y_1(s_i)! \cdots y_{h-1}(s_i)! \cdot (m_i - \sum_{k=1}^{h-1} y_k(s_i))!} \right) \right], \tag{3.5}
\]

where \( m_i \) is the total number of events for location \( s_i \).

We will now give the form of the natural parameter function. Once the form of the natural parameter function is given, the form of \( \theta \) will follow because \( A_{i,k}(\cdot) \) is a function of \( \theta \). For one-parameter exponential families, i.e., \( s = 1 \), Besag (1974) showed that the natural parameter function must be of the form

\[
A_i \{y(N_i); \theta\} = \alpha_i + \sum_{s_j \in N_i} \eta_{i,j} y(s_j) \tag{3.6}
\]

with \( \eta_{i,j} = \eta_{j,i} \). Besag applied the above form to a series of models, including the Binomial MRF model, which is a Multinomial MRF model with only two categories. For multi-parameter exponential families, Kaiser et al. (2002) gives three different forms. One of the forms is

\[
A_{i,k} \{y(N_i); \theta\} = \alpha_{i,k} + \sum_{s_j \in N_i} \eta_{i,j,k} y_k(s_j) \tag{3.7}
\]

such that \( \eta_{i,j,k} = \eta_{j,i,k} \) for all \( i \neq j \) and \( k = 1, \ldots, h - 1 \).

Since the Multinomial MRF model is a multivariate version of the Binomial MRF model, one could justify using a direct extension of the form given by Besag for the form of the natural parameter function. This extension is

\[
A_{i,k} \{y(N_i); \theta\} = \alpha_{i,k} + \sum_{s_j \in N_i} \eta_{i,j,k} y_k(s_j), \text{ for } k = 1, ..., h - 1, \tag{3.8}
\]
which is equivalent to (3.7) since \( T_k(y(s_j)) = y_k(s_j) \) for a Multinomial MRF model. However, the natural parameter function often contains too many parameters for estimation. To reduce the number of parameters, let \( \eta_{i,j,k} = \eta \) and \( \alpha_{i,k} = \alpha_k \), which is frequently assumed in applications. These assumptions reduce the form of the natural parameter function to

\[
A_{i,k} \{ y(N_i); \theta \} = \alpha_k + \eta \sum_{s_j \in N_i} y_k(s_j), \quad \text{for } k = 1, \ldots, h - 1. \tag{3.9}
\]

From expression (3.9), we know one of the Multinomial MRF model parameters is \( \eta \). The other model parameters are determined by the form of \( \alpha_k \) for \( k = 1, \ldots, h - 1 \). Let \( \alpha_k \) be defined as \( \log \left( \frac{\kappa_k}{\kappa_h} \right) \). Then we have

\[
A_{i,k} \{ y(N_i); \theta \} = \log \left( \frac{\kappa_k}{\kappa_h} \right) + \eta \sum_{s_j \in N_i} y_k(s_j), \quad \text{for } k = 1, \ldots, h - 1, \tag{3.10}
\]

and \( \theta = \left( \log \left( \frac{\kappa_1}{\kappa_h} \right), \ldots, \log \left( \frac{\kappa_{h-1}}{\kappa_h} \right), \eta \right)^T \) such that \( \kappa_k > 0 \) and \( \sum_{k=1}^{h} \kappa_k = 1 \).

Because expression (3.5) with the natural parameter defined by (3.10) corresponds to expression (3.3), we also have

\[
A_{i,k} \{ y(N_i); \theta \} = \log \left( \frac{p_{i,k}}{p_{i,h}} \right) \tag{3.11}
\]

for \( \eta \in \mathbb{R} \). Under an independence model, (i.e., \( \eta = 0 \)), substituting expression (3.10) into (3.11) and simplifying gives

\[
A_{i,k} \{ y(N_i); \theta \} = \alpha_k = \log \left( \frac{p_{i,k}}{p_{i,h}} \right) \tag{3.12}
\]

Furthermore, when \( \eta = 0 \) and \( \alpha_{i,k} = \alpha_k \), the probability of an event belonging to category \( k \) is the same for all locations, i.e., \( p_{i,k} = p_k \) for all \( i = 1, \ldots, n \), and \( k = 1, \ldots, h \), where \( p_k \) represents the marginal probability of an event belonging to category \( k \). As a result, \( \alpha_k \) is not only defined as \( \log \left( \frac{\kappa_k}{\kappa_h} \right) \) but also equals \( \log \left( \frac{p_k}{p_h} \right) \) under independence. This implies, under independence, \( \kappa_k \) is equal to the marginal probability, \( p_k \). When the dependence parameter is not equal to zero, \( \kappa_k \) may no longer be the marginal probability, which will be further discussed in Section 3.5.

To map the natural parameter functions, \( A_{i,k} \{ y(N_i); \theta \}; k = 1, \ldots, h - 1, \) to the conditional probabilities, \( p_{i,k}; k = 1, \ldots, h, \) for location \( s_i \), recall \( A_{i,k} \{ y(N_i); \theta \} = \log \left( \frac{p_{i,k}}{p_{i,h}} \right) \) and
\[ \sum_{k=1}^{h} p_{i,k} = 1, \] which is a system of \( h \) equations with \( h \) variables representing the conditional probabilities. Solving the system of equations for the conditional probabilities results in the following forms for \( p_{i,k} \) in terms of the natural parameter functions:

\[ p_{i,k} = \frac{\exp \left\{ A_{i,k} \{ y(N_i); \theta \} \right\}}{1 + \sum_{k=1}^{h-1} \exp \left\{ A_{i,k} \{ y(N_i); \theta \} \right\}} \quad \text{for } k = 1, \ldots, h - 1, \text{ and } \]

\[ p_{i,k} = \frac{1}{1 + \sum_{k=1}^{h-1} \exp \left\{ A_{i,k} \{ y(N_i); \theta \} \right\}} \quad \text{for } k = h. \]

### 3.3 Construction of the Multinomial MRF Model

To construct the joint distribution for the Multinomial MRF model up to an unknown normalizing constant, we will follow the negpotential approach as outlined by Kaiser and Cressie (2000). The negpotential function is defined as

\[ Q(y) \equiv \log \left\{ \frac{g(y)}{g(y^*)} \right\}; \quad y \in \Omega, \tag{3.15} \]

where \( g(y) \) is the joint density or mass function and \( y^* \in \Omega \) is an arbitrary fixed value in the support of \( g \). The joint density function, \( g(\cdot) \), can then be obtained as

\[ g(y) = \frac{\exp \{ Q(y) \}}{\int_{\Omega} \exp \{ Q(t) \} \, d\nu(t)}, \tag{3.16} \]

where \( \nu(\cdot) \) is Lebesque or counting measure.

Using the specific value \( y^* = 0 \), Besag (1974) showed that the negpotential function may be written as the expansion

\[ Q(y) = \sum_{1 \leq i \leq n} H_i(y(s_i)) + \sum_{1 \leq i < j \leq n} H_{i,j}(y(s_i), y(s_j)) \\
+ \sum_{1 \leq i < j < k \leq n} H_{i,j,k}(y(s_i), y(s_j), y(s_k)) \\
+ \ldots \\
+ H_{1,2,\ldots,n}(y(s_1), y(s_2), \ldots, y(s_n)). \tag{3.17} \]

Kaiser and Cressie (2000) show result (3.17) holds for any \( y^* \in \Omega \) that satisfies a condition they called the Markov random field support condition. The MRF support condition states that for \( y^* \in \Omega \),

\[ \{ g^*_i(s_i) \} \times \Phi_i \subseteq \Omega, \tag{3.18} \]
where $\Phi_i$ is the support of $g_i(\cdot)$, the marginal probability mass function of $Y(s_1), \ldots, Y(s_{i-1}), Y(s_{i+1}), \ldots, Y(s_n)$. Besag (1974) proved his results assuming the positivity condition, which is

$$\Omega = \Omega_1 \times \Omega_2 \times \ldots \times \Omega_n,$$

(3.19)

where $\Omega_i$ is the set of possible values of $Y(s_i)$ for $i = 1, \ldots, n$. Although the positivity condition is stronger than the MRF support condition, the positivity condition holds for a large number of applications including applications of the Multinomial MRF model.

To simplify the expansion of the negpotential function in (3.17), the Hammersley-Clifford Theorem is often invoked. This theorem involves sets called cliques, which are singletons or sets of locations such that each location in the set is a neighbor of every other location in the set. The Hammersley-Clifford Theorem states that any function $H_{i,j,\ldots,h}$ in (3.17) is equal to zero unless the set of locations $\{s_i, s_j, \ldots, s_h\}$ form a clique. Besag (1974) proved this result for $y^* = 0$ under the positivity condition while Kaiser and Cressie (2000) proved this result for $y^* \in \Omega$ under the MRF support condition. If the four nearest neighbors constitute the neighborhood for each location $s_i$ on a regular lattice, then each single location and each pair of locations that are neighbors form cliques. Thus, under the four-nearest neighbor structure, all $H$-functions in (3.17) are zero except for the first-order $H_i$- and second-order $H_{i,j}$-functions.

For neighborhood structures that result in cliques of three or more members, it is common to assume pairwise-only dependence, which is the assumption that "the probability structure of the system is dependent only upon contributions from cliques containing no more than two sites" (Besag 1974, p.200). Therefore, under the assumption of pairwise-only dependence, only the first-order and second-order $H$-functions are used to construct the negpotential function in (3.17). Furthermore, the second-order $H$-functions are zero unless locations $s_i$ and $s_j$ are neighbors according to the Hammersley-Clifford Theorem.

To begin construction of the joint distribution for the Multinomial MRF model according to the negpotential approach, let $y^*(s_i) = (y^*_1(s_i), \ldots, y^*_h(s_i))^T = (0, \ldots, 0, m_i)^T$ for $i = 1, \ldots, n$, and assume pairwise-only dependence. According to the general forms of the first-order and second-order $H$-functions given by Kaiser and Cressie (2000), we have for the Multinomial
MRF model, given by expressions (3.5) and (3.10),

\[ H_i(y(s_i)) = \sum_{k=1}^{h-1} y_k(s_i)\alpha_k + \log \left( \frac{m_i!}{\prod_{k=1}^{h} y_{i,k}!} \right), \quad \text{and} \]

\[ H_{i,j}(y(s_i), y(s_j)) = \eta_{i,j} \left[ \sum_{k=1}^{h-1} y_k(s_i)y_k(s_j) \right]. \quad (3.21) \]

Substituting (3.20) and (3.21) into the expansion of the negpotential function in (3.17) yields

\[ Q(y) = \sum_{1 \leq i \leq n} \left[ \sum_{k=1}^{h-1} y_k(s_i)\alpha_k + \log \left( \frac{m_i!}{\prod_{k=1}^{h} y_{i,k}!} \right) \right] \]

\[ + \eta_{i,j} \sum_{1 \leq i < j \leq n} \left[ \sum_{k=1}^{h-1} y_k(s_i)y_k(s_j) \right] \quad (3.22) \]

with \( \eta_{i,j} = 0 \) unless locations \( s_i \) and \( s_j \) are neighbors.

In addition to verifying the Markov support condition or the stronger positivity condition and assuming pairwise dependence, there are two conditions that need to be satisfied for a joint distribution to exist and be identified according to Kaiser and Cressie (2000). The first condition states that \( H_{i,j} = H_{j,i} \), which holds in this case since (3.21) is symmetric in \( y(s_i) \) and \( y(s_j) \). The second condition is that \( \sum_\Omega \exp \{ Q(t) \} < \infty \). This condition is true for \( Q(y) \) as defined in (3.22) for any value of \( \eta \) since \( \Omega \), the support of \( Y \), is finite. Hence, the joint distribution exists and can be identified for any value of \( \eta \), but only up to an unknown normalizing constant because the computation of \( \sum_\Omega \exp \{ Q(t) \} \) is prohibitive. Since the behavior of the Multinomial MRF models cannot be investigated through the joint distribution, we will investigate the behavior through simulation. Furthermore, for the remainder of the paper, we will only consider data sets that have the same number of events occurring at each location, so that \( m_i = m \) for all \( i = 1, \ldots, n \).

### 3.4 Simulation Method and Estimation

Because, as shown in Section 3.3, the joint distribution can be identified only up to an unknown normalizing constant, we cannot simulate data sets based on the joint distribution. However, we do know the form of the full conditional distribution of \( y(s_i) \) for each location,
14

$s_i$, as these have been specified in the model. Consequently, a Gibbs sampling algorithm is a
natural choice for simulating data from the Multinomial MRF model (or any MRF model).
The steps for the Gibbs sampling algorithm are as follows.

1. Given the specified values for $\kappa_k$ for $k = 1,..,h$, generate starting values $y^{(0)}(s_i); i = 1,\ldots,n,$ using the multinomial conditional probability mass function defined by (3.5) and (3.10) with $\eta = 0$. The notation $y^{(0)}(s_i)$ denotes $(y_1(s_i),\ldots,y_h(s_i))^T$ at iteration 0.

2. For iterations $t = 1,\ldots,T$, order the locations by using a random permutation operator or the identity function applied to locations $s_i; i = 1,\ldots,n$. The random permutation operator and the identity function lead to what are known as random scan and systematic scan Gibbs sampling algorithms, respectively.

3. For each location, according to the order determined by step (2), generate $y^{(t)}(s_i)$ from the multinomial conditional probability mass function defined by (3.5) and (3.10) with $\eta$ equal to the specified value and replace $y^{(t-1)}(s_i)$ with $y^{(t)}(s_i)$.

4. Repeat steps 2 and 3 until the specified convergence criteria is met.

The Gibbs algorithm will converge to the desired joint distribution because, as shown in Section 3.3, the conditional distributions given by expressions (3.5) and (3.10) correspond to the joint distribution defined by (3.16) with $Q(y)$ as in (3.22). Given this, the positivity condition is sufficient to ensure irreducibility and aperiodicity (Liu et al., 1995). The random scan algorithm is known to be reversible (Liu et al., 1995) while the systematic scan Gibbs algorithm meets the general conditions given in Roberts and Smith (1993). Thus, for this application of simulating realizations from the joint distribution that corresponds to a conditionally specified Multinomial MRF model, the Gibbs sampling algorithm possesses the necessary properties to ensure convergence.

For all simulation studies in this paper, Multinomial Markov random fields were simulated for a spatial region $D = [0, 30] \times [0, 30]$ on a torus such that each cell is 1 unit by 1 unit. We specified $m = 100$ events per cell with each event belonging to one of three categories (i.e.,
Furthermore, the neighborhood for each cell location, \( s_i \), consists of the four nearest neighbors.

To obtain a Monte Carlo (MC) approximation of a parameter, \( \theta \), based on a Gibbs sampling algorithm, let \( \hat{\theta}_t \) be the estimate of \( \theta \) for the \( t^{th} \) simulated data set. For a total of \( T \) fields, the Monte Carlo approximation of \( \theta \) is

\[
E_T(\theta) \equiv \frac{1}{T} \sum_{t=1}^{T} \hat{\theta}_t .
\]  

(3.23)

To determine the number of fields needed for a simulation study, consider the common sample size problem in which, given the standard deviation of the sampling distribution of \( \hat{\theta} \), the sample size is chosen so that a future confidence interval has a width less than the specified maximum allowable width. We propose a similar method to determine the number of data sets needed before calculating a Monte Carlo approximation of \( \theta \). Suppose we want a 95% confidence interval for a given parameter, \( \theta \). Then the width of the confidence interval is approximately twice the margin of error or \( 4\sigma/\sqrt{T} \) where \( \sigma \) is the standard deviation of the parameter estimates, \( \hat{\theta}_t \), for \( t \in \{1, 2, \ldots\} \). Furthermore, we propose the width of the confidence interval to be less than 5% of the parameter value. Then the total number of data sets needed is

\[
T = \left( \frac{4\sigma}{0.05\theta} \right)^2 .
\]  

(3.24)

We will need to substitute estimates of \( \theta \) and \( \sigma \) into (3.24) since we do not know the true values of \( \theta \) and \( \sigma \). Although an estimate of \( \theta \) is needed to determine the number of simulated fields while the purpose of the simulation study is to estimate \( \theta \), we propose generating a specified number of fields, denoted as \( T_1 \), to obtain Monte Carlo approximations of \( \theta \) and \( \sigma \) to substitute in (3.24) for \( \theta \) and \( \sigma \), respectively. One-thousand was chosen to be a sufficient value for \( T_1 \) for the following reason. We considered the value for \( T_1 \) sufficient if the standard deviation of the Monte Carlo approximation, \( s_t/\sqrt{t} \), for \( t \in \{T_1 - c, \ldots\} \) and some constant \( c \), is monotone decreasing. In other words, we considered the value for \( T_1 \) sufficient if the increase in the number of data sets, \( t \), has a greater effect on the standard deviation of the Monte Carlo approximation, \( s_t/\sqrt{t} \), than changes in the standard deviation of the parameter.
estimates, $s_t$. Based on plots of the standard deviation of Monte Carlo approximations of different parameters, 1000 was considered a sufficient value for $T_1$.

Because the starting values are generated from a Multinomial MRF model with $\eta = 0$ in step 1 and the Markov random fields produced in step 3 are usually generated by a Multinomial MRF model with $\eta \neq 0$, a burn-in period is required to allow the dependence parameter to fully affect the data patterns before collecting data sets for study. For all simulation studies in this paper, the first 500 data sets generated in step 3 were discarded. Once the first 500 data sets were discarded, every $10^{th}$ data set was collected because data patterns in one data set may influence data patterns in the next simulated data set. Finally, steps 2 and 3 of the Gibbs sampling algorithm were repeated until a total of $T$ data sets were collected.

Estimating parameters by maximizing the likelihood function is difficult because the joint probability density or mass function is not known in explicit form for many Markov random field models. However, we can find estimates based on the conditional density or mass functions. Besag (1975) suggested maximizing a pseudo-likelihood function, defined as the product of the conditional mass functions, to obtain parameter estimates. This pseudo-likelihood function may be written as

$$P(\theta) = \prod_{i=1}^{n} f_i(y(s_i)|y(N_i); \theta),$$

where $f_i(y(s_i)|y(N_i); \theta)$ is given by (3.4). The pseudo-likelihood function was maximized by iterative method. However, values of the psuedo-likelihood are often too large for a computer to compute. Instead, the negative log of the psuedo-likelihood, $-\log(P(\theta))$, was minimized.

### 3.5 Comparison of Traditional and Centered Models

As discussed in Section 3.2, when the dependence parameter does not equal zero, $\kappa_k$ may not equal the marginal probability of an event belonging to category $k$, which means $m\kappa_k$ may not equal the marginal expectation of category $k$. However, if $m\kappa_k$ is approximately equal to the marginal expectation of category $k$, we would then be able to provide an approximate interpretation for the estimate of $m\kappa_k$ and thus, $\kappa_k$. To explore the agreement between $m\kappa_k$
and the marginal expectation of category $k$, we will obtain the Monte Carlo approximation of the marginal expectation given different sets of parameters values and compare the Monte Carlo approximation for category $k$ to $m\kappa_k$, which is the marginal expectation under independence.

First we will consider the Multinomial MRF model defined by (3.5) with the natural parameter defined by (3.10), which will be referred to as the traditional model. However, we could reparameterize (3.10) in the following manner,

$$A_{i,k} \{ \text{y}(N_i); \theta \} = \log \left( \frac{\kappa_k}{\kappa_h} \right) + \eta \sum_{s_j \in N_i} \{ y_k(s_j) - m\kappa_k \} \text{ for } k = 1, \ldots, h - 1. \tag{3.25}$$

Notice that when $\eta = 0$ and the above form for the natural parameter function is used in (3.5), expression (3.5) is equivalent to the independence model. When we consider the Multinomial MRF model defined by (3.5) with the natural parameter defined by (3.25), we have what will be referred to as the centered model.

Caragea and Kaiser (2006) compare the traditional model versus the centered model for a Binary MRF while incorporating covariates. The Binary MRF model can be considered a specific case of the Multinomial MRF model since a Binary MRF is a Multinomial MRF with only two categories and where the total number of events for any location is one, i.e., $m = 1$, for all locations. Caragea and Kaiser show that marginal expectations under the traditional model do not equal marginal expectations under the independence model. With the centered model, however, the marginal expectations are approximately equal to the marginal expectations under independence if $\eta$ is within certain bounds. This feature of the centered model allows the model to account for large-scale structure through the use of covariates that influence marginal expectations.

To compare the Monte Carlo approximations of the marginal expectations for both the traditional and centered models to the marginal expectations under independence, $m\kappa_k$, Multinomial Markov random fields were simulated under both the traditional model and the centered model according to the steps outlined in Section 3.4. The estimate of the marginal expectation for category $k$ and a given simulated data set, indexed by $t$, is defined as

$$E_t \{ Y_k(s_i) \} = \frac{1}{n} \sum_{i=1}^{n} y_{k,t}(s_i), \tag{3.26}$$
where $y_{k,t}(s_i)$ is the number of events belonging to category $k$ at location $s_i$ for field $t$. The Monte Carlo (MC) estimate of the marginal expectation based on $T$ simulated fields is then

$$E_T \{Y_k(s_i)\} = \frac{1}{T} \sum_{t=1}^{T} E_t \{Y_k(s_i)\}. \quad (3.27)$$

For the independence model, the marginal expectation is simply $m\kappa_k$ for category $k$.

For both the traditional model and centered model, data sets were generated for each of the values of $\eta$ in the set $\{-0.006, -0.005, \ldots, 0.006\}$. For the first case, let $\kappa_1$ and $\kappa_2$ equal 0.20 and 0.50, respectively. Then the marginal expectations under the independence model are 20, 50 and 30, for category 1, 2, and 3, respectively. For the second case, let both $\kappa_1$ and $\kappa_2$ equal 0.30, which means the marginal expectations under the independence model are 30, 30 and 40, for category 1, 2, and 3, respectively. After $T = 1000$ data sets were generated, the mean and standard deviation of $E_t \{Y_k(s_i)\}$ for $t = 1, \ldots, 1000$, and a given category, were substituted in (3.24) for $\theta$ and $\sigma$, respectively, to determine the number of additional data sets needed to satisfy the convergence criteria explained in Section 3.4. Since four times the standard deviation of the Monte Carlo approximation, $4\sigma_T / \sqrt{T}$, is less than 5% of its respective approximation of the marginal expectation, $E_T \{Y_k(s_i)\}$, for all Monte Carlo approximations under consideration, the convergence criteria as outlined in Section 3.4 is satisfied and no additional fields were needed. Consequently, all Monte Carlo approximations in this section are based on $T = 1000$ data sets.

Figures 3.1 and 3.2 show the discrepancy between the Monte Carlo approximations of the marginal expectations for the traditional and centered models. Rarely are the approximations of the marginal expectations under the traditional model nearly the respective marginal expectations under independence. For the centered model, however, approximations of the marginal expectations are nearly equal to the respective marginal expectations under independence regardless of the strength of spatial dependence, within the range examined. Therefore, the centered model appears to possess the property that we desire, while the traditional model does not. Consequently, any mention of the Multinomial MRF model during the remainder of the paper refers to the centered model as defined by (3.5) and (3.25).
Figure 3.1 Comparison of Monte Carlo approximations of marginal expectations for traditional model defined by (3.5) and (3.10) and centered model defined by (3.5) and (3.25) along with marginal expectations for a model of independence displayed as solid lines for \( \kappa_1 = 0.20, \kappa_2 = 0.50 \) and \( \eta \in \{-0.006, -0.005, \ldots, 0.006\} \)

Figure 3.2 Comparison of Monte Carlo approximations of marginal expectations for traditional model defined by (3.5) and (3.10) and centered model defined by (3.5) and (3.25) along with marginal expectations for a model of independence displayed as solid lines for \( \kappa_1 = 0.30, \kappa_2 = 0.30 \) and \( \eta \in \{-0.006, -0.005, \ldots, 0.006\} \)
3.6 Bounds for the Spatial Dependence Parameter

As mentioned in the previous section, the marginal expectations under the centered model will be approximately equal to the respective marginal expectations under the independence model only if the dependence parameter is within certain bounds, which was true for the illustrations of Figures 3.1 and 3.2. Since these bounds on the spatial dependence parameter, \( \eta \), will depend on the number of neighbors and the total number of events at each location, the natural parameter function given in (3.25) will be reparameterized. Let \( \gamma \) be the new dependence parameter such that

\[
\gamma = \frac{m |N_i|}{\eta}
\]

where \( |N_i| \) is the number of neighbors for location \( s_i \) and is assumed to be equal for all \( i = 1, \ldots, n \). We then have,

\[
A_{i,k} \{ y(N_i); \theta \} = \log \left( \frac{\kappa_k}{\kappa_h} \right) + \gamma \frac{1}{|N_i|} \sum_{s_j \in N_i} \left\{ \frac{y_k(s_j)}{m} - \kappa_k \right\} \quad \text{for } k = 1, \ldots, h - 1. \tag{3.28}
\]

For the remainder of the paper, any discussion of the dependence parameter will be in terms of \( \gamma \) instead of \( \eta \) when the number of neighbors for location \( s_i \) is equal for all \( i = 1, \ldots, n \). Furthermore, we will refer to the quantity \( \frac{1}{|N_i|} \sum_{s_j \in N_i} \left\{ \frac{1}{m} y_k(s_j) - \kappa_k \right\} \) as the average neighborhood deviation.

For the case of one-parameter exponential families, Kaiser (2007) developed methodology to calculate the bounds for the spatial dependence parameter. Because the conditional expectations are a function of the natural parameter functions, the conditional expectations are a function of the value of \( \kappa \) and the average neighborhood deviation. In order for the marginal expectations to be nearly the respective marginal expectations under independence, the conditional expectations should be within a reasonable range centered at the respective marginal expectations under independence, which is a function of \( \kappa \). The value of \( \kappa \) is required to have a greater impact on the value of the natural parameter function, and hence, the conditional expectations, than the average neighborhood deviation in order to restrict the range of the conditional expectations. This constraint leads to the standard bounds for \( \gamma \) defined by Kaiser as

\[
|\gamma| \leq \left( \tau(A_i \{ y(N_i); \theta \}) \in \Theta \left[ \frac{\tau(A_i \{ y(N_i); \theta \}) - \kappa_i}{A_i \{ y(N_i); \theta \} - \tau^{-1}(\kappa_i)} \right] \right)^{-1}, \tag{3.29}
\]

where \( \tau(A_i \{ y(N_i); \theta \}) \) is equal to \( E[Y(s_i)|y(N_i); \theta] \).
Using analytical means to define standard bounds for the dependence parameter for multi-parameter exponential families appears to be untractable and may even be impossible. Simulation, however, can be used to approximate the standard bounds for $\gamma$ numerically. Multinomial Markov random fields were generated for different combinations of values for $\kappa_1$, $\kappa_2$ and $\gamma$ according to the steps outlined in Section 3.4. Monte Carlo approximations of the marginal expectations were calculated according to expressions (3.26) and (3.27). Since four times the standard deviation of the Monte Carlo approximation, $4s_T/\sqrt{T}$, for $T = 1000$ is less than 5% of its respective approximation of the marginal expectation, $E_T\{Y_k(s_i)\}$, for all Monte Carlo approximations under consideration, the convergence criteria as outlined in Section 3.4 is satisfied and no additional fields were needed. Consequently, all Monte Carlo approximations in this section are based on $T = 1000$ data sets. The resulting Monte Carlo approximations of the marginal expectations are plotted in Figures 3.3 - 3.5.

For small values of $\gamma$, the MC approximations of the marginal expectations are nearly equal to $m\kappa_k$, the expected values under independence. Thus, the parameters, $\kappa_k$ for $k = 1, 2, 3$, in a model with dependence are nearly equal to their respective marginal probabilities, $p_k$. What one considers a small value of $\gamma$ depends on the values for $\kappa_1$ and $\kappa_2$ as suggested by Figures 3.3 - 3.5. As the values for $\kappa_1$ and $\kappa_2$ increase, the range of $\gamma$ for which the marginal expectations are approximately equal to the respective marginal expectations under independence decreases. For values of $\gamma$ that are outside of the range suggested by the Figures 3.3 - 3.5, the MC approximations of the marginal expectations corresponding to category 1 and category 2 are often near the endpoints of the range for the marginal expectations, which are 0 and 100 in this case, whereas the MC approximations of the marginal expectations corresponding to category 3 are usually nearly 0. Thus, for "large" values of the dependence parameter, the parameters, $\kappa_k$, are not approximately equal to their respective marginal probabilities, $p_k$. When the dependence parameter is too large, the dependence parameter allows the average neighborhood deviation to affect the natural parameter functions, $A_{i,k}\{N_i, \theta\}$ given by (3.25), to a larger degree than $\kappa_k$. If $\kappa_k$ no longer dominate the values of the natural parameter functions, then $\kappa_k$ no longer dominate the marginal expectations. Furthermore, if the average neighborhood
Figure 3.3 Monte Carlo approximations of marginal expectations for \( \kappa_1 = 0.10, \kappa_2 \in \{0.10, 0.20, \ldots, 0.80\} \) (as represented by lines from bottom to top on left side of center plot and lines from top to bottom on left side of right plot) and \( \gamma \in \{0, 0.5, 1.0, \ldots, 8\} \)

Figure 3.4 Monte Carlo approximations of marginal expectations for \( \kappa_1 = 0.20, \kappa_2 \in \{0.10, 0.20, \ldots, 0.70\} \) (as represented by lines from bottom to top on left side of center plot and lines from top to bottom on left side of right plot) and \( \gamma \in \{0, 0.5, 1.0, \ldots, 8\} \)

Figure 3.5 Monte Carlo approximations of marginal expectations for \( \kappa_1 = 0.30, \kappa_2 \in \{0.10, 0.20, \ldots, 0.60\} \) (as represented by lines from bottom to top on left side of center plot and lines from top to bottom on left side of right plot) and \( \gamma \in \{0, 0.5, 1.0, \ldots, 8\} \)
deviations influence the marginal means more than $\kappa_k$, then the marginal means for the data sets generated by the Gibbs sampling algorithm depend more on the values generated for the neighborhood locations in step 3 in Section 3.4 than $\kappa_k$. The marginal means, therefore, fluctuate between 0 and 100 for "large" values of the dependence parameter according to the values generated for the neighborhood locations.

For a Binary MRF model, according to Kaiser (2007), the marginal expectation when $\kappa < 0.50$ monotonically increases to 1 as $\gamma$ increases while values greater than 0.50 for $\kappa$ produce a marginal expectation that monotonically decrease to 0 as $\gamma$ increases. For $\kappa = 0.50$, the marginal expectation will be 0.50 for all values of $\gamma$. For the Multinomial MRF model, we expect to see similar patterns in the MC approximations of the marginal means. However, Figures 3.3 - 3.5 suggest that there are some combinations of values for $\kappa_1$ and $\kappa_2$ (usually when the value of $\kappa_1$ is close to the value of $\kappa_2$) such that the MC approximations of the marginal expectations for all categories do not monotonically increase or decrease as $\gamma$ increases. As the figures show, the MC approximation of the marginal expectation for either category 1 or category 2 is near 0 for some values of $\gamma$, near 100 for other values of $\gamma$ and somewhere between 20 and 80 for yet other values of $\gamma$ for certain values for $\kappa_1$ and $\kappa_2$. A question then is whether or not the MC approximations of the marginal expectations are approximately equal to the true marginal expectations for "large" values of the dependence parameter.

If the MC values are not actually approximating the corresponding true marginal expectations, then this might suggest that either the joint distribution does not exist, the joint distribution does exist but the moments do not exist, or the limiting distribution is not equal to the desired joint distribution under the Gibbs sampling algorithm outlined in Section 3.4. First, as shown in Section 3.3, the joint distribution does exist for all values of $\eta$, and thus, all values of $\gamma$. Second, since the support of a Multinomial MRF model is finite for all possible parameter values given the total number of events at each location, the moments are finite. Third, as discussed in Section 3.4, the limiting distribution is the same as the desired joint distribution for the Gibbs sampling algorithm.

Given what we know about the existence of the joint distribution, the existence of the mo-
ments and the Gibbs sampling algorithm, we can expect the Markov chains producing the data sets through the Gibbs sampling algorithm to converge. If the Markov chains converge, then the MC approximations of the marginal expectations should converge as well. Then, given we have simulated a sufficient number of data sets, the MC approximations of the marginal expectations should be nearly equal to the corresponding true marginal expectations. If the number of data sets needed to approximate the marginal expectations with considerable precision is quite large, then the Markov chains simulating the data sets may be slow to converge, which implies that a considerable amount of time would be needed to simulate enough data sets before the MC approximations could be expected to be nearly the true marginal expectations.

To explore the possibility that the Markov chains are slow to converge, different sets of starting values were generated for given values of \( \kappa_1, \kappa_2 \) and \( \gamma \). Then 1,000 data sets were generated from each set of starting values and the resulting MC approximations were compared. The MC approximations of the marginal expectations corresponding to either category 1 or category 2 rarely were similar. This exercise suggests that more than 1,000 data sets are needed before we can be confident that the MC approximations of the marginal expectations are nearly the true marginal expectations. The next step involved generating one set of starting values and collecting 1 million data sets to determine if the marginal means of the individual data sets for all categories vary from data set to data set. If the marginal means do not vary over the course of 1 million data sets, the rate of convergence for the Markov chains may make the method of using the Gibbs sampling algorithm to obtain MC approximations of the marginal expectations an unattractive method for large values of the dependence parameter. For all chains consisting of 1 million data sets that were simulated, we observed that the marginal means did not vary from data set to data set. In some cases, for example, if the marginal mean for category 1 was close to 0 at the beginning of the chain, the marginal mean for category 1 stayed near 0. Although this outcome is not desirable, this outcome may not be unexpected considering how the the data sets are generated according to Sec 3.4. If the marginal mean for category 1, for example, is nearly 0 (100), then almost all of the conditional expectations, \( mp_{i,1}; i = 1, ..., n \), will be nearly 0 (100). To slowly increase the marginal mean from nearly
0 to nearly 100, the conditional expectations, $m_{p_{i,1}}$, need to slowly increase from nearly 0 to nearly 100 for all locations $s_i$. To increase the conditional expectations, values generated from the multinomial probability mass function for $y_1(s_i); i = 1,..,n$, need to be consistently larger than the respective conditional expectations, a highly unlikely event. This means the probability for the marginal mean to change from nearly 0 to nearly 100 within a reasonable number of data sets is very small. Therefore, the data patterns observed in Figures 3.3 - 3.5 for "large" values of the dependence parameter most likely occur because the Markov chains are slow to converge as a result of the inability of the Gibbs sampling algorithm to quickly generate a data set with a large marginal mean for category $k$ after generating a data set associated with a small marginal mean for category $k$.

Although slow converging chains are a concern in many applications, slow converging chains are not a concern here because the goal of this section is to determine the values of $\gamma$ that produce data sets with marginal means nearly equal to the marginal expectations under independence through simulation. This can be accomplished by referring to Figures 3.3 - 3.5.
CHAPTER 4 MODEL BEHAVIOR

4.1 Asymmetry of Multinomial MRF Model

In a standard Multinomial MRF model that does not include spatial structure (i.e., the Multinomial MRF model under independence), the labels given to categories as $1, 2, ..., h$, are irrelevant. These indices may be assigned in an arbitrary manner without affecting the model structure or the properties of the model as long as the same indices are used for the parameter values. In particular, the expected values of components of the multinomial vector are the same regardless of which index is assigned to a category. We will call this a "symmetry" property of the Multinomial MRF model under independence. As will be demonstrated in this section, this symmetry property no longer holds for a Multinomial MRF model that incorporates a dependence parameter not equal to zero. In particular, the marginal moments of the category labeled $h$ do not remain unchanged if that category is re-labeled as 1, or any other value.

The Binomial MRF model is a special case of the Multinomial MRF model where there are only $h = 2$ categories. It will be shown that for the Binomial MRF model, the aforementioned symmetry property does hold. Let the two categories be arbitrarily labeled as category 1 and category 2. Also, suppose there are a total of $m$ events at each location $s_i$; $i = 1, \ldots, n$. Let $y_1(s_i)$ denote the number of events in category 1 at location $s_i$ and $y_2(s_i)$ denote the number of events in category 2 at location $s_i$ for $i = 1, \ldots, n$. Suppose category 1 is labeled as the first category. Then the natural parameter function for a centered model under the Binomial Markov random field structure of expression (3.5) is

$$A_{i,1} \{ y(N_i); \theta \} = \log \left( \frac{\kappa_1}{1 - \kappa_1} \right) + \frac{1}{|N_i|} \sum_{s_j \in N_i} \left\{ \frac{y_1(s_j)}{m} - \kappa_1 \right\}$$ (4.1)

Now suppose category 2 is labeled as the first category. The natural parameter function is
then
\[ A_{i,2} \{ y(N_i); \theta \} = \log \left( \frac{\kappa_2}{1 - \kappa_2} \right) + \gamma \frac{1}{|N_i|} \sum_{s_j \in N_i} \left\{ \frac{y_2(s_j)}{m} - \kappa_2 \right\} \]
\[ = -\log \left( \frac{1 - \kappa_2}{\kappa_2} \right) - \gamma \frac{1}{|N_i|} \sum_{s_j \in N_i} \left\{ \kappa_2 - \frac{y_2(s_j)}{m} \right\} \]
\[ = -\log \left( \frac{\kappa_1}{1 - \kappa_1} \right) - \gamma \frac{1}{|N_i|} \sum_{s_j \in N_i} \left\{ \frac{y_1(s_j)}{m} - \kappa_1 \right\} \]
\[ = -A_{i,1} \{ y(N_i); \theta \}. \tag{4.2} \]

The natural parameter functions of the two possible forms for this model are the negatives of each other.

For a given model form, conditional expectations are equal to the conditional probabilities given in (3.13) and (3.14) multiplied by the total number of events at a given location, assumed here to be \( m \) for all locations. Specifically, if category 1 is labeled as the first category, then the conditional expectations for category 1 and category 2 are

\[ E \{ Y_1(s_i) | y(N_i); \theta \} = mp_{i,1} \]
\[ = \frac{m \exp [A_{i,1} \{ y(N_i); \theta \}]}{1 + \exp [A_{i,1} \{ y(N_i); \theta \}]} \]
\[ = \frac{m}{1 + \exp [-A_{i,2} \{ y(N_i); \theta \}]} \]
\[ = \frac{m}{1 + \exp [A_{i,2} \{ y(N_i); \theta \}]} \tag{4.3} \]

and,

\[ E \{ Y_2(s_i) | y(N_i); \theta \} = mp_{i,2} \]
\[ = \frac{1}{1 + \exp [A_{i,1} \{ y(N_i); \theta \}]} \]
\[ = \frac{1}{1 + \exp [-A_{i,2} \{ y(N_i); \theta \}]} \]
\[ = \frac{\exp [A_{i,2} \{ y(N_i); \theta \}]}{1 + \exp [A_{i,2} \{ y(N_i); \theta \}]} \tag{4.4} \]

If category 2 is labeled as the first category, then

\[ E \{ Y_2(s_i) | y(N_i); \theta \} = mp_{i,2} = m \frac{\exp [A_{i,2} \{ y(N_i); \theta \}]}{1 + \exp [A_{i,2} \{ y(N_i); \theta \}]} \tag{4.5} \]
and,

$$E\{Y_1(s_i)|y(N_i); \theta\} = mp_{i,1} = m \frac{1}{1 + \exp[A_{i,2}\{y(N_i); \theta\}]}.$$  (4.6)

Since expression (4.3) is equal to expression (4.6) and expression (4.4) is equal to expression (4.5), the conditional expectations of $Y_1(s_i)$ and $Y_2(s_i)$ are the same, regardless of which is labeled as the “first” category. Consequently, estimates of $\kappa_1$, $\kappa_2$ and $\gamma$ obtained by maximizing the pseudo-likelihood are not dependent on which category is labeled as the first category and the Binomial MRF model is symmetric with respect to the labeling of the categories.

Consider now the Multinomial MRF model for three categories, for the sake of concreteness, arbitrarily labeled as category 1, category 2 and category 3. In this situation, there are two natural parameter functions which are in the form given in (3.25) for $k = 1, 2$. Conditional expectations are again equal to the conditional probabilities given in (3.13) and (3.14) multiplied by the number of events at location $s_i$, namely,

$$E\{Y_k(s_i)|y(N_i); \theta\} = mp_{i,k}$$

$$E\{Y_k(s_i)|y(N_i); \theta\} = m \frac{\exp[A_{i,k}\{y(N_i); \theta\}]}{1 + \sum_{k=1}^{h-1} \exp[A_{i,k}\{y(N_i); \theta\}]} \quad \text{for } k = 1, 2, \text{ and}$$  (4.7)

$$E\{Y_k(s_i)|y(N_i); \theta\} = mp_{i,k}$$

$$E\{Y_k(s_i)|y(N_i); \theta\} = m \frac{1}{1 + \sum_{k=1}^{h-1} \exp[A_{i,k}\{y(N_i); \theta\}]} \quad \text{for } k = 3.$$  (4.8)

As for the case of the model with 2 categories, if one switches the labels for category 1 and category 2, the conditional expectations for each of the respective categories will not change. However, the conditional expectations will change if one switches the label for category 3 with either category 1 or category 2. To show this formally, suppose the category originally labeled as category 2 is re-labeled as category 3 and vice-versa.

In what follows, let the indices on $Y_k(s_i)$, $\kappa_k$ and $A_{i,k}\{y(N_i); \theta\}$ remain unchanged so that these quantities are identical to those in (4.7) and (4.8). Denote the natural parameter functions of the re-labeled model as $B_{i,1}(\cdot)$ and $B_{i,2}(\cdot)$, which now play the roles of $A_{i,1}(\cdot)$ and $A_{i,2}(\cdot)$ in the original model, respectively. Then, in terms of the original $y_k(s_i)$, $\kappa_k$ and
The conditional expectations for $Y_i$ are re-labeled as the first category and category 1 as the third category. The implication is that respective conditional expectations given in (4.11) - (4.13).

A similar result occurs if category 3 is re-labeled as the first category and category 1 as the third category. The implication is that

$$A_{i,k} \{y(N_i); \theta\},$$

we have

$$B_{i,1} \{y(N_i); \theta\} = \log \left( \frac{\kappa_1}{\kappa_2} \right) + \gamma \frac{1}{|N_i|} \sum_{s_j \in N_i} \left\{ \frac{y_1(s_j)}{m} - \kappa_1 \right\}$$

$$= A_{i,1} \{y(N_i); \theta\} + \log \left( \frac{\kappa_3}{\kappa_2} \right), \quad \text{and} \quad (4.9)$$

$$B_{i,2} \{y(N_i); \theta\} = \log \left( \frac{\kappa_3}{\kappa_2} \right) + \gamma \frac{1}{|N_i|} \sum_{s_j \in N_i} \left\{ \frac{y_3(s_j)}{m} - \kappa_3 \right\}$$

$$= -A_{i,1} \{y(N_i); \theta\} - A_{i,2} \{y(N_i); \theta\} + \log \left( \frac{\kappa_1}{\kappa_3} \right). \quad (4.10)$$

The conditional expectations for $Y_1(s_i)$, $Y_2(s_i)$ and $Y_3(s_i)$ then become, under the re-labeled model,

$$E \{Y_1(s_i)|y(N_i); \theta\} = mp_{i,1}$$

$$= \frac{\exp [B_{i,1} \{y(N_i); \theta\}]}{1 + \exp [B_{i,1} \{y(N_i); \theta\}]} + \exp [B_{i,2} \{y(N_i); \theta\}]$$

$$= \frac{\frac{\kappa_3}{\kappa_2} \exp [A_{i,1} \{y(N_i); \theta\}] - \frac{\kappa_1}{\kappa_2} \exp [A_{i,2} \{y(N_i); \theta\]} + \frac{\kappa_1}{\kappa_3} \exp [-A_{i,1} \{y(N_i); \theta\} - A_{i,2} \{y(N_i); \theta\}]}{1 + \frac{\kappa_3}{\kappa_2} \exp [A_{i,1} \{y(N_i); \theta\}] + \frac{\kappa_1}{\kappa_3} \exp [-A_{i,1} \{y(N_i); \theta\} - A_{i,2} \{y(N_i); \theta\}]}, \quad (4.11)$$

$$E \{Y_2(s_i)|y(N_i); \theta\} = mp_{i,2}$$

$$= \frac{1}{1 + \exp [B_{i,1} \{y(N_i); \theta\}]} + \exp [B_{i,2} \{y(N_i); \theta\}]$$

$$= \frac{1}{1 + \frac{\kappa_3}{\kappa_2} \exp [A_{i,1} \{y(N_i); \theta\}] + \frac{\kappa_1}{\kappa_3} \exp [-A_{i,1} \{y(N_i); \theta\} - A_{i,2} \{y(N_i); \theta\}]} \cdot \quad (4.12)$$

and

$$E \{Y_3(s_i)|y(N_i); \theta\} = mp_{i,3}$$

$$= \frac{\exp [B_{i,2} \{y(N_i); \theta\}]}{1 + \exp [B_{i,1} \{y(N_i); \theta\}]} + \exp [B_{i,2} \{y(N_i); \theta\}]$$

$$= \frac{\frac{\kappa_3}{\kappa_2} \exp [-A_{i,1} \{y(N_i); \theta\} - A_{i,2} \{y(N_i); \theta\}] + \frac{\kappa_1}{\kappa_3} \exp [-A_{i,1} \{y(N_i); \theta\} - A_{i,2} \{y(N_i); \theta\}]}{1 + \frac{\kappa_3}{\kappa_2} \exp [A_{i,1} \{y(N_i); \theta\}] + \frac{\kappa_1}{\kappa_3} \exp [-A_{i,1} \{y(N_i); \theta\} - A_{i,2} \{y(N_i); \theta\}]} \cdot \quad (4.13)$$

Notice that the conditional expectations given in (4.7) and (4.8) are not the same as the respective conditional expectations given in (4.11) - (4.13). A similar result occurs if category 3 is re-labeled as the first category and category 1 as the third category. The implication is that
estimates of $\kappa_1$, $\kappa_2$, $\kappa_3$ and $\gamma$ obtained by maximizing the psuedo-likelihood depend on which category is labeled as category 3, the last category in the model. Hence, the Multinominal MRF model is not symmetric with respect to the labeling of the categories and parameter estimates depend on which category is labeled as the last or $h^{th}$ category.

4.2 Variances and Covariances of Conditional Expectations

A good deal of insight into the behavior of Multinomial MRF models can be gained by examining the variances and covariances of conditional expectations. To approximate these variances and covariances using Monte Carlo methods, Multinomial Markov random fields with three categories were simulated according to the steps outlined in Section 3.4 for different sets of values for the parameters, $\kappa_1$, $\kappa_2$ and $\gamma$.

The first set of values chosen for $\kappa = (\kappa_1, \kappa_2, \kappa_3)^T$ is $(0.20, 0.30, 0.50)^T$. In Section 4.1, it was demonstrated that the model behavior depends on which category is chosen as the third (last) category. Because of the asymmetry of the Multinomial MRF model, Markov random fields were simulated for each of the three permutations of the above values such that the values chosen for $\kappa_3$ for each of the three permutations are distinct. The second set of values chosen for $\kappa$ is $(0.30, 0.30, 0.40)^T$. For this set of values, Markov random fields were generated for the two permutations of the chosen values such that the values chosen for $\kappa_3$ for each of the two permutations are distinct. The dependence parameter, $\gamma$ in (3.28), was varied over the set {$\ldots, -0.50, -0.25, 0, 0.25, 0.50, \ldots$} as long as $\gamma$ is within the standard bounds as suggested by Figures 3.3 - 3.5 for given values of $\kappa_1$ and $\kappa_2$.

For each combination of parameter values, 1000 Markov random fields were simulated. Then, for each Markov random field, the conditional expectations at each location were calculated according to expressions (4.7) and (4.8). The variance of the conditional expectations for the $t^{th}$ field and $k^{th}$ category was then computed as

$$\text{Var}_t [E \{Y_k(s_i)|\mathbf{y}(N_i); \theta\}] = \frac{1}{900-1} \sum_{i=1}^{900} (m_{i,k,t} - \bar{m}_{k,t})^2, \quad (4.14)$$

where $p_{i,k,t}$ is the conditional probability for category $k$ at location $s_i$ for Markov random field $t$, and $\bar{p}_{k,t} = \frac{1}{900} \sum_{i=1}^{900} p_{i,k,t}$. The Monte Carlo approximation of the variance of the conditional
expectations for category \( k \) was computed as the average of (4.14) across the \( T \) simulated fields,

\[
\text{Var}_T [E \{ Y_k(s_i) | y(N_i); \theta \}] = \frac{1}{T} \sum_{t=1}^{T} \text{Var}_t [E \{ Y_k(s_i) | y(N_i); \theta \}].
\] (4.15)

Similarly, the covariance of conditional expectations for category \( k \) and conditional expectations for category \( l \) for a given field \( t \) was computed as

\[
\text{Cov}_t [E \{ Y_k(s_i) | y(N_i); \theta \}, E \{ Y_l(s_i) | y(N_i); \theta \}] = \frac{1}{900} \sum_{i=1}^{900} (m_{p_{1,k,t}} - m_{\bar{p}_{k,t}})(m_{p_{1,l,t}} - m_{\bar{p}_{l,t}}).
\] (4.16)

The Monte Carlo approximation of the covariance of the conditional expectations for category \( k \) and conditional expectations for category \( l \) is then

\[
\text{Cov}_T [E \{ Y_k(s_i) | y(N_i); \theta \}, E \{ Y_l(s_i) | y(N_i); \theta \}] = \frac{1}{T} \sum_{t=1}^{T} \text{Cov}_t [E \{ Y_k(s_i) | y(N_i); \theta \}, E \{ Y_l(s_i) | y(N_i); \theta \}].
\] (4.17)

Since four times the standard deviation of the Monte Carlo approximation of the variance of the conditional expectations is less than 5% of its respective approximation of the variance of conditional expectations, for all Monte Carlo approximations under consideration, the convergence criteria as outlined in Section 3.4 is satisfied and no additional fields were needed. Consequently, all Monte Carlo approximations of the variances of the conditional expectations in this section are based on \( T = 1000 \) data sets. For the same reason, all Monte Carlo approximations of the covariances of the conditional expectations in this section are also based on \( T = 1000 \) data sets.

Monte Carlo approximations of the variances of conditional expectations from (4.15) are plotted against values of the dependence parameter \( \gamma \) in Figures 4.1 - 4.5 for various sets of values for \( \kappa_1, \kappa_2 \) and \( \kappa_3 \). These figures suggest that the variance of the conditional expectations for a given category and a given value for the dependence parameter depends on which category is labeled as the 3\( ^{rd} \) (last) category. In particular, the variance of the conditional expectations is smallest for a given category when that category is labeled as the last category. Since \( p_{i,3} = 1 - p_{i,1} - p_{i,2} \), the conditional probability for category 3 for a given location is a function of the conditional probabilities for the other two categories. An increase in \( p_{i,1} \) will generally be
offset by a similar decrease in $p_{i,2}$ as suggested by the forms of $p_{i,1}$ and $p_{i,2}$ given by expressions (3.13) and (3.14). As a result, $p_{i,3}$ does not vary as much as one might initially anticipate.

Monte Carlo approximations of the covariances of the conditional expectations from (4.17) are plotted against the values of the dependence parameter $\gamma$ in Figures 4.6 - 4.8. These figures suggest that when $\kappa_1 \neq \kappa_2$ and $\gamma \neq 0$, the covariance of the conditional expectations for category 3 and the conditional expectations for the category labeled $k$ such that $\kappa_k = \min \{\kappa_1, \kappa_2\}$ is positive, while the covariance of conditional expectations is negative for all other pairs of categories. When $\kappa_1 = \kappa_2$, the covariance of conditional expectations is negative for all pairs of categories, as suggested by Figure 4.9.

The most surprising aspect of the plots in Figures 4.6 - 4.10 is that when $\kappa_1 \neq \kappa_2$, the covariance between the conditional expectations of the category corresponding to the smaller of these values and the conditional expectations of category 3 is positive. Because it is not possible to derive the covariance of the conditional expectations for any two categories in closed form except for the independence case ($\gamma = 0$), we must take an indirect approach to explain why the covariance of conditional expectations for a pair of category is sometimes positive.

We will examine the forms of the conditional expectations as functions of the average neighborhood deviation. Let $D_{i,k} \equiv \frac{1}{|N_i|} \sum_{s_j \in N_i} \left\{ \frac{y_k(s_j)}{m} - \kappa_k \right\}$ be the average neighborhood deviation for category $k; k = 1, 2, 3$. The forms of the conditional expectation in terms of $D_{i,k}$ are then

$$m p_{i,k} = \frac{\kappa_k \exp^{\gamma D_{i,k}}}{\kappa_3 + \kappa_1 \exp^{\gamma D_{i,1}} + \kappa_2 \exp^{\gamma D_{i,2}}}, \text{ for } k = 1, 2, \text{ and }$$

$$m p_{i,k} = \frac{1}{\kappa_3 + \kappa_1 \exp^{\gamma D_{i,1}} + \kappa_2 \exp^{\gamma D_{i,2}}}, \text{ for } k = 3. \quad (4.18)$$

For a fixed $D_{i,2}$, when $D_{i,1}$ increases (decreases), the conditional expectation for category 1 increases (decreases) while the conditional expectations for the other two categories decrease (increase) for a given location $s_i$. Therefore, given $D_{i,2}$, the mapping of the average neighborhood deviations into the conditional expectations induces positive dependence between category 2 and category 3 when $D_{i,1}$ changes. Similarly, for a fixed $D_{i,1}$, when $D_{i,2}$ increases (decreases), the conditional expectation for category 2 increases (decreases) while the conditional expectations for the other two categories decrease (increase). Therefore, given $D_{i,1}$,
Figure 4.1 Monte Carlo approximations of variance of conditional expectations when $\kappa_1 = 0.20$ and $\kappa_2 = 0.30$

Figure 4.2 Monte Carlo approximations of variance of conditional expectations when $\kappa_1 = 0.20$ and $\kappa_2 = 0.50$

Figure 4.3 Monte Carlo approximations of variance of conditional expectations when $\kappa_1 = 0.30$ and $\kappa_2 = 0.50$
Figure 4.4 Monte Carlo approximations of variance of conditional expectations when $\kappa_1 = 0.30$ and $\kappa_2 = 0.30$

Figure 4.5 Monte Carlo approximations of variance of conditional expectations when $\kappa_1 = 0.30$ and $\kappa_2 = 0.40$
Figure 4.6  Monte Carlo approximations of covariance of conditional expectations when $\kappa_1 = 0.20$ and $\kappa_2 = 0.30$

Figure 4.7  Monte Carlo approximations of covariance of conditional expectations when $\kappa_1 = 0.20$ and $\kappa_2 = 0.50$

Figure 4.8  Monte Carlo approximations of covariance of conditional expectations when $\kappa_1 = 0.30$ and $\kappa_2 = 0.50$
Figure 4.9  Monte Carlo approximations of covariance of conditional expectations when $\kappa_1 = 0.30$ and $\kappa_2 = 0.30$

Figure 4.10  Monte Carlo approximations of covariance of conditional expectations when $\kappa_1 = 0.30$ and $\kappa_2 = 0.40$
the mapping of the average neighborhood deviations into the conditional expectations induces positive dependence between category 1 and category 3 when $D_{i,2}$ changes.

To further explore the effects of the average neighborhood deviations on conditional expectations, consider the partial derivatives of the conditional expectations for category 3 with respect to $D_{i,1}$ and $D_{i,2}$, which are

$$f_{D_{i,1}}(D_{i,1}, D_{i,2}) \equiv \frac{\partial m_{i,3}}{\partial D_{i,1}} = \frac{-m\kappa_1 \gamma \exp^{\gamma D_{i,1}}}{(\kappa_3 + \kappa_1 \exp^{\gamma D_{i,1}} + \kappa_2 \exp^{\gamma D_{i,2}})^2} \quad \text{and}$$

$$f_{D_{i,2}}(D_{i,1}, D_{i,2}) \equiv \frac{\partial m_{i,3}}{\partial D_{i,2}} = \frac{-m\kappa_2 \gamma \exp^{\gamma D_{i,2}}}{(\kappa_3 + \kappa_1 \exp^{\gamma D_{i,1}} + \kappa_2 \exp^{\gamma D_{i,2}})^2}. \quad (4.20)$$

Figure 4.11 contains image plots of $|f_{D_{i,1}}(D_{i,1}, D_{i,2})| - |f_{D_{i,2}}(D_{i,1}, D_{i,2})|$ given $D_{i,1}$ and $D_{i,2}$ for different values of $\kappa_1$ and $\kappa_2$ under a moderate dependence structure ($\gamma = 1.6$).

When $\kappa_1 = \kappa_2$, changes in $D_{i,1}$ when $D_{i,2}$ is equal to some constant $d$ has the same effect on the conditional expectations of category 3 as changes in $D_{i,2}$ when $D_{i,1} = d$ as shown in Figure 4.11. Furthermore,

$$|f_{D_{i,1}}(D_{i,1}, D_{i,2})| - |f_{D_{i,2}}(D_{i,1}, D_{i,2})| = 0 \quad \text{when} \quad D_{i,1} = D_{i,2} \quad \text{and}$$

$$|f_{D_{i,1}}(D_{i,1}, D_{i,2})| - |f_{D_{i,2}}(D_{i,1}, D_{i,2})| > (\cdot)0 \quad \text{when} \quad D_{i,1} > (\cdot)D_{i,2}.$$

The above expressions suggest that when $D_{i,1}$ and $D_{i,2}$ change in value from location to location, changes in the conditional expectations are equally influenced by changes in $D_{i,1}$ and $D_{i,2}$. Therefore, neither the number of events in category 1 nor the number of events in category 2 dictate the covariance structure. Thus, when $\kappa_1 = \kappa_2$, the covariance structure is similar to the covariance structure of $Y(s_i)$ under independence in that the covariance of conditional expectations for any two categories is negative.

Now suppose that $\kappa_1 > \kappa_2$. Then

$$|f_{D_{i,1}}(D_{i,1}, D_{i,2})| - |f_{D_{i,2}}(D_{i,1}, D_{i,2})| = 0 \quad \text{when} \quad D_{i,1} = D_{i,2} - \frac{1}{\gamma} \log \left( \frac{\kappa_1}{\kappa_2} \right) \quad \text{and}$$

$$|f_{D_{i,1}}(D_{i,1}, D_{i,2})| - |f_{D_{i,2}}(D_{i,1}, D_{i,2})| > (\cdot)0 \quad \text{when} \quad D_{i,1} > (\cdot)D_{i,2} - \frac{1}{\gamma} \log \left( \frac{\kappa_1}{\kappa_2} \right).$$

The above expressions suggest that changes in the average neighborhood deviations for category 1 will have a greater influence on changes in the conditional expectations for all three categories.
Figure 4.11 Difference in absolute value of partial derivatives, 
\[|f_{D_{i,1}}(D_{i,1}, D_{i,2})| - |f_{D_{i,2}}(D_{i,1}, D_{i,2})|\], where \[f_{D_{i,1}}(D_{i,1}, D_{i,2})\] and \[f_{D_{i,2}}(D_{i,1}, D_{i,2})\] are defined by 4.20 and 4.21, respectively.
than changes in the average neighborhood deviations for category 2. And if the values of \( D_{i,1} \) have a greater influence on the conditional expectations than \( D_{i,2} \), then we would expect that as the conditional expectation for category 1 increases (decreases) from one location to another, then the conditional expectations for category 2 and category 3 will most likely decrease (increase) according to the expressions of the conditional expectations given in (4.18) and (4.19). These patterns in the conditional expectations will lead to positive covariance between the conditional expectations of category 2 and category 3. Similar conclusions follow when \( \kappa_1 < \kappa_2 \).

Finally, we will consider for category \( k; k = 1, 2 \), the difference between the conditional expectation given values of \( D_{i,1} \) and \( D_{i,2} \) to the conditional expectation given \( D_{i,1} = 0 \) and \( D_{i,2} = 0 \). Let \( f_1(D_{i,1}, D_{i,2}) \) denote expression (4.18) for \( k = 1 \) and \( f_2(D_{i,1}, D_{i,2}) \) denote expression (4.18) for \( k = 2 \). Then the difference between the conditional expectation given values of \( D_{i,1} \) and \( D_{i,2} \) and the conditional expectation given \( D_{i,1} = 0 \) and \( D_{i,2} = 0 \) for category \( k \) is \( f_k(D_{i,1}, D_{i,2}) - f_k(0, 0); k = 1, 2 \). Let \( g_k(D_{i,1}, D_{i,2}) = f_k(D_{i,1}, D_{i,2}) - f_k(0, 0); k = 1, 2 \). Given \( D_{i,1} \) and \( D_{i,2} \), the change in the conditional expectation for category 3 is equal to \(- (g_1(D_{i,1}, D_{i,2}) + g_2(D_{i,1}, D_{i,2})) \) since \( p_{i,3} = 1 - p_{i,1} - p_{i,2} \). The functions, \( g_k(D_{i,1}, D_{i,2}); k = 1, 2 \), were calculated for \( D_{i,k} \in \{-0.10, -0.009, -0.998, ..., 0.10\} \); \( k = 1, 2 \). Furthermore, we let \( \kappa_1 \) vary over the set \( \{0.10, 0.30, 0.50\} \) while holding \( \kappa_2 \) and \( \gamma \) constant such that \( \kappa_2 = 0.30 \) and \( \gamma = 1.0 \). The results are plotted in Figure 4.12. For the image plots on the left side of Figure 4.12, the light gray areas correspond to the values of \( D_{i,1} \) and \( D_{i,2} \) such that the change in conditional expectations for category 1, (i.e., \( g_1(D_{i,1}, D_{i,2}) \)), and the change in conditional expectations for category 3 (i.e., \(- g_1(D_{i,1}, D_{i,2}) + g_2(D_{i,1}, D_{i,2}) \)), are either both positive or both negative. The dark gray areas correspond to the values of \( D_{i,1} \) and \( D_{i,2} \) that lead to the other two cases. For the image plots on the right side of Figure 4.12, the light gray areas correspond to the values of \( D_{i,1} \) and \( D_{i,2} \) such that the change in conditional expectations for category 2, (i.e., \( g_2(D_{i,1}, D_{i,2}) \)), and the change in conditional expectations for category 3 (i.e., \(- g_1(D_{i,1}, D_{i,2}) + g_2(D_{i,1}, D_{i,2}) \)), are either both positive or both negative. The dark gray areas correspond to the values of \( D_{i,1} \) and \( D_{i,2} \) that lead to the other two cases.
Figure 4.12 Comparison of changes in conditional expectation for category $k$ for $k = 1, 2$ defined by $g_k(D_{i,1}, D_{i,2})$ to changes in conditional expectation for category 3 defined by $-(g_1(D_{i,1}, D_{i,2}) + g_2(D_{i,1}, D_{i,2}))$ when $\kappa_1 = 0.10$ and $\kappa = 0.30$ (top), $\kappa_1 = 0.30$ and $\kappa = 0.30$ (middle) and $\kappa_1 = 0.50$ and $\kappa = 0.30$ (bottom). (Light gray area represents when changes in conditional expectations for category $k$; $k = 1, 2$, and changes in conditional expectations for category 3 are both positive or both negative. Dark gray area represents all other cases.)
Notice in Figure 4.12 that as \( \kappa_1 \) increases in value while the value of \( \kappa_2 \) is held constant, the percentage of the area that is light gray decreases for the plots on the left while the percentage of the area that is light gray increases for the plots on the right. This suggests that as \( \kappa_1 \) increases in value while the value of \( \kappa_2 \) is held constant, the likelihood that there will be positive covariance of the conditional expectations for category 1 and category 3 decreases while the likelihood that there will be positive covariance of the conditional expectations for category 2 and category 3 increases. The patterns seen in Figure 4.12 are not unexpected according to the forms of the partial derivatives of the conditional expectation for category 3 with respect to \( D_{i,1} \) and \( D_{i,2} \) given by (4.20) and (4.20), respectively. Although only a finite number of values were specified for the parameters, \( \kappa_1, \kappa_2 \) and \( \gamma \), similar patterns were seen when the calculations and simulations as describe above were repeated with different parameter values.

These patterns observed in the covariance structure do affect the variances of the conditional expectations and vice versa since the variance of the conditional expectations for a given category can be written in terms of the variances and covariances of the conditional expectations of the other categories as shown below.

\[
\text{Var}(m_{p_{k,i}}) = \text{Var}\left\{ m \left( 1 - \sum_{h \neq k} p_{h,i} \right) \right\} \\
= \text{Var}\left( m \sum_{h \neq k} p_{h,i} \right) \\
= \sum_{h \neq k} \text{Var}(m_{p_{h,i}}) + 2 \sum_{h \neq k} \sum_{l \neq k, h} \text{Cov}(m_{p_{h,i}}, m_{p_{l,i}}). \tag{4.22}
\]

Specifically, in the case of three categories, expression (4.22) shows that positive (negative) covariance of conditional expectations between a pair of categories increases (decreases) the variance of the conditional expectations for the remaining category.

### 4.3 Marginal Variances and Covariances

To examine the marginal variances and covariances through simulation, Multinomial Markov random fields were simulated according to the steps outlined in Section 3.4. The values speci-
fied for \( \kappa = (\kappa_1, \kappa_2, \kappa_3)^T \) and \( \gamma \) in Section 4.2 will also be specified for \( \kappa \) and \( \gamma \) in this section.

The marginal variance for field \( t \) and category \( k \) was computed as

\[
\text{Var}_t [E \{ Y_k(s_i) \}] = \frac{1}{900 - 1} \sum_{i=1}^{900} \{ y_{k,t}(s_i) - \bar{y}_{k,t}(s_i) \}^2 ,
\]

where \( y_{k,t}(s_i) \) is the number of events in category \( k \) at location \( s_i \) for the \( t^{th} \) Markov random field and \( \bar{y}_{k,t}(s_i) = \frac{1}{900} \sum_{i=1}^{900} y_{k,t}(s_i) \). The Monte Carlo approximation of the marginal variance for category \( k \) is then

\[
\text{Var}_T [E \{ Y_k(s_i) \}] = \frac{1}{T} \sum_{t=1}^{T} \text{Var}_t [E \{ Y_k(s_i) \}] .
\]

The marginal covariance of category \( k \) and category \( l \) for a given field \( t \) was computed as

\[
\text{Cov}_t [E \{ Y_k(s_i) \}, E \{ Y_l(s_i) \}] = \frac{1}{900 - 1} \sum_{i=1}^{900} \{ y_{k,t}(s_i) - \bar{y}_{k,t}(s_i) \} \{ y_{l,t}(s_i) - \bar{y}_{l,t}(s_i) \} .
\]

The Monte Carlo approximation of the marginal covariance of category \( k \) and category \( l \) is

\[
\text{Cov}_T [E \{ Y_k(s_i) \}, E \{ Y_l(s_i) \}] = \frac{1}{T} \sum_{t=1}^{T} \text{Cov}_t [E \{ Y_k(s_i) \}, E \{ Y_l(s_i) \}] .
\]

Since four times the standard deviation of the Monte Carlo approximation of the marginal variance is less than 5% of its respective approximation of the marginal variance, for all Monte Carlo approximations under consideration, the convergence criteria as outlined in Section 3.4 is satisfied and no additional fields were needed. Consequently, all Monte Carlo approximations of the marginal variances in this section are based on \( T = 1000 \) data sets. For the same reason, all Monte Carlo approximations of the marginal covariances in this section are also based on \( T = 1000 \) data sets.

Monte Carlo approximations of the marginal variances from (4.24) are plotted against values of the dependence parameter \( \gamma \) in Figures 4.13 - 4.17 while Monte Carlo approximations of the marginal covariances from (4.26) are plotted against values of the dependence parameter \( \gamma \) in Figures 4.18 - 4.22.

As shown in Section 3.6, when \( \gamma \) is within the appropriate bounds for the centered model, then \( E(mp_{i,k}) \approx mp_k \) and \( E \{ mp_{i,k}(1 - p_{i,k}) \} \approx mp_k(1 - p_k) \), where \( p_k \) is the marginal probability for category \( k \) under the corresponding independence model such that \( p_k = \kappa_k \) for
Then the marginal variance for any given category $k$ is

$$\text{Var}\{Y_k(s_i)\} = E[\text{Var}\{Y_k(s_i) | y(N_i); \theta\}] + \text{Var}[E\{Y_k(s_i) | y(N_i); \theta\}]$$

$$= E\{mp_{i,k}(1 - p_{i,k})\} + \text{Var}(mp_{i,k})$$

$$\approx mp_k(1 - p_k) + \text{Var}(mp_{i,k}). \quad (4.27)$$

Expression (4.27) demonstrates that the marginal variance of category $k$ is approximately the sum of the marginal variance under the independence model and the variance of the conditional expectations corresponding to the given category. Furthermore, when $\gamma$ is within appropriate bounds for the centered model, then $E(mp_{i,k}p_{i,l}) \approx mp_kp_l$. Thus, the marginal covariance of any two categories $k$ and $l$ is

$$\text{Cov}(Y_{i,k}, Y_{i,l}) = E[E\{Y_k(s_i) Y_l(s_i) | y(N_i); \theta\}]$$

$$- E[E\{Y_k(s_i) | y(N_i); \theta\}] E[E\{Y_l(s_i) | y(N_i); \theta\}]$$

$$= E(m^2p_{i,k}p_{i,l} - mp_{i,k}p_{i,l}) - E(mp_{i,k})E(mp_{i,l})$$

$$= E(m^2p_{i,k}p_{i,l} - E(mp_{i,k})E(mp_{i,l}) - E(mp_{i,k}p_{i,l})$$

$$= \text{Cov}(mp_{i,k}, mp_{i,l}) - E(mp_{i,k}p_{i,l})$$

$$\approx \text{Cov}(mp_{i,k}, mp_{i,l}) - mp_kp_l, \quad (4.28)$$

which shows that the marginal covariance of category $k$ and category $l$ is approximately the sum of the covariance of $Y_k(s_i)$ and $Y_l(s_i)$ for a given location, $s_i$, under the independence model and the covariance of the conditional expectations for the given pair of categories.

The relationship between the marginal variances and the variances of conditional expectations as expressed by (4.27) is displayed by Figures 4.1 - 4.5 and 4.13 - 4.17 in terms of Monte Carlo approximations. For a given category, for example, the Monte Carlo approximations of marginal variances displayed by Figure 4.13 are approximately equal to the sum of the Monte Carlo approximations of the variances of conditional expectations displayed by Figure 4.1 and the corresponding marginal variances under independence. For this case, the marginal variances under independence are equal to 16, 21 and 25 for category 1, 2 and 3, respectively.
Figure 4.13 Monte Carlo approximations of variance of marginal expectations when \( \kappa_1 = 0.20 \) and \( \kappa_2 = 0.30 \)

Figure 4.14 Monte Carlo approximations of variance of marginal expectations when \( \kappa_1 = 0.20 \) and \( \kappa_2 = 0.50 \)

Figure 4.15 Monte Carlo approximations of variance of marginal expectations when \( \kappa_1 = 0.30 \) and \( \kappa_2 = 0.50 \)
Figure 4.16 Monte Carlo approximations of variance of marginal expectations when $\kappa_1 = 0.30$ and $\kappa_2 = 0.30$

Figure 4.17 Monte Carlo approximations of variance of marginal expectations when $\kappa_1 = 0.30$ and $\kappa_2 = 0.40$
Figure 4.18 Monte Carlo approximations of covariance of marginal expectations when $\kappa_1 = 0.20$ and $\kappa_2 = 0.30$

Figure 4.19 Monte Carlo approximations of covariance of marginal expectations when $\kappa_1 = 0.20$ and $\kappa_2 = 0.50$

Figure 4.20 Monte Carlo approximations of covariance of marginal expectations when $\kappa_1 = 0.30$ and $\kappa_2 = 0.50$
Figure 4.21  Monte Carlo approximations of covariance of marginal expectations when $\kappa_1 = 0.30$ and $\kappa_2 = 0.30$

Figure 4.22  Monte Carlo approximations of covariance of marginal expectations when $\kappa_1 = 0.30$ and $\kappa_2 = 0.40$
Similarly, the relationship between the marginal covariances and the covariances of conditional expectations as expressed by (4.28) is displayed by Figures 4.6 - 4.10 and 4.18 - 4.22 in terms of Monte Carlo approximations. For example, Figures 4.13 and 4.18 indicate the Monte Carlo approximations of marginal covariances are approximately equal to the sum of the Monte Carlo approximation of the covariance of conditional expectations and the corresponding covariance of \( Y_k(s_i) \) and \( Y_l(s_i) \) for a given location, \( s_i \), under independence for a given pair of categories. The covariances of \( Y_k(s_i) \) and \( Y_l(s_i) \) for a given location, \( s_i \), under independence, in this case, are equal to -6 for category 1 and categories 2, -10 for categories 1 and 3, and -15 for categories 2 and 3.

Figures 4.18 - 4.22 also suggest that the marginal covariances will always be negative for any given pair of categories, which was found not to be the case for the covariances of the conditional expectations. As discussed in Section 4.2, the covariance of conditional expectations for a Multinomial MRF model with three categories will be positive for one pair of categories when \( \kappa_1 \neq \kappa_2 \) and \( \gamma \neq 0 \). The Monte Carlo approximations of the covariances of the conditional expectations plotted in Figures 4.6 - 4.10 suggest that when the covariance of the conditional expectations for a given pair of categories is positive, the covariance of \( Y_k(s_i) \) and \( Y_l(s_i) \) for a given location, \( s_i \), under independence, which is always negative, will be greater in absolute value than the respective covariance of conditional expectations as long as \( \gamma \) is within the standard bounds discussed in Section 3.6. The implication is, according to expression (4.28), the marginal covariance is negative regardless of the value of the covariance of the conditional expectations for a specified range of values for \( \gamma \).

4.4 Representation of Dependence

As discussed in Section 4.2, the variance of the conditional expectations is smaller for the \( h^{th} \) category than it is for the other categories. Expression (4.27) of Section 4.3 indicates that this then may also be true for the marginal variance of \( Y_h(s_i) \). Stronger statistical dependence in MRF models is generally related to greater variance in conditional expectations than is weaker dependence. Thus, the dependence structure within the \( h^{th} \) category of a Multinomial
MRF may be weaker than for the dependence structure within the other \( h - 1 \) categories.

To examine the dependence structure within each category, Multinomial Markov random fields were simulated according to the steps outlined in Section 3.4. Markov random fields were simulated for values of \( \kappa_1, \kappa_2, \) and \( \gamma \) such that \( \kappa_1 \in \{0.10, 0.20, 0.30\}, \kappa_2 \in \{0.10, 0.20, ..., 0.90 - \kappa_1\} \) for a given value of \( \kappa_1 \) and \( \gamma \in \{0.5, 2.0\} \).

For each field, the events in all categories except for the first category were aggregated into a single category, i.e., the number of categories was reduced to two. This reduction in the number of categories effectively changes the Multinomial MRF into a Binomial MRF. Then the dependence parameter corresponding to the Binomial MRF model was estimated by maximizing the pseudo-likelihood function as described in Section 3.4. This estimate will be denoted as \( \hat{\gamma}_1 \). Then the events in all categories except for the second category were aggregated into a single category and the dependence parameter was estimated to obtain \( \hat{\gamma}_2 \). This step was repeated one more time to estimate the dependence parameter when events in all categories except for the third category were aggregated into a single category to obtain \( \hat{\gamma}_3 \).

As shown in Section 3.6, for a Multinomial MRF model, the standard bounds for the dependence parameter, \( \gamma \), depend on the values of \( \kappa_k \) for \( k = 1, ..., h - 1 \). For the Binomial MRF model, the standard bounds for the dependence parameter depend on the value of \( \kappa \) (Kaiser 2007). This finding indicates that a given value for the Binomial MRF model dependence parameter, \( \gamma \), could signify moderate dependence for some value of \( \kappa \) if the value for \( \gamma \) is not close to the standard bound. The same value for \( \gamma \) could signify strong dependence if the value for \( \gamma \) is near the standard bound corresponding to a different value of \( \kappa \). To standardize the estimates of \( \gamma_k \), the estimates were divided by their respective standard bounds so we will be able to compare the strength of the dependence structure within each category across categories.

The general form of the standard bounds for one-parameter exponential families (e.g., binomial probability mass function) is given by expression (3.29) in Section 3.6. For the Binomial MRF model,

\[
\tau(A_{i,1} \{y(N_i); \theta\}) = \frac{\exp[\tau_i \{y(N_i); \theta\}]}{1 + \exp[\tau_i \{y(N_i); \theta\}]} \quad \text{and} \quad (4.29)
\]
\[ \tau^{-1}(\kappa) = \log \left( \frac{\kappa}{1 - \kappa} \right). \]  

(4.30)

Substituting expressions (4.29) and (4.30) into (3.29) gives the standard bounds for \( \gamma \), denoted as \( \gamma_{\kappa} \), namely,

\[ \gamma_{\kappa} \equiv \sup_{\exp[A_{i,1}\{y(N_i);\theta\}] \in (0, \infty)} \left( \frac{\exp[A_{i,1}\{y(N_i);\theta\}] - \kappa}{A_{i,1}\{y(N_i);\theta\} - \log \left( \frac{\kappa}{1 - \kappa} \right)} \right)^{-1}. \]  

(4.31)

The standard bounds, \( \gamma_{\kappa} \), are plotted against \( \kappa \) for \( 0 < \kappa < 1 \) in Figure 4.23.

As in the previous sections, a simulation study was conducted to obtain Monte Carlo approximations of the dependence parameters, \( \gamma_k; k = 1, 2, 3 \). For each Multinomial MRF field simulated by the Gibbs sampling algorithm, let \( \hat{\gamma}_{k,t} \) be the estimate of the dependence parameter, \( \gamma \), in (4.1) for field \( t \), with \( k \) corresponding to the index of the category such that the events in category \( k \) were not aggregated with the events in the other two categories. The standardized estimate of \( \gamma_k \) for field \( t \), denoted as \( \gamma^*_k,t \), was computed as

\[ \hat{\gamma}^*_k,t = \frac{\hat{\gamma}_{k,t}}{\gamma_{\kappa_k}}, \]  

(4.32)

where \( \gamma_{\kappa_k} \) is the standard bound for the dependence parameter given the value of \( \kappa_k \). Then the Monte Carlo approximation of the expected standardized dependence parameter for \( T \) Markov random fields was computed as

\[ E_T(\hat{\gamma}^*_k) = \frac{1}{T} \sum_{t=1}^{T} \hat{\gamma}^*_k,t. \]  

(4.33)

After simulating 1,000 Markov random fields and substituting the mean and standard deviation of the estimates of the standardized dependence parameter into (3.24) for \( \theta \) and \( \sigma \), respectively, upwards of 175,000 Markov random fields would be needed to satisfy the convergence criteria as outlined in Section 3.4, especially when \( \gamma = 0.5 \). Due to the large amount of computational time needed to simulate 175,000 Markov random fields, only 10,000 Markov random fields were generated for each set of parameter values.

The resulting Monte Carlo approximations of the standardized dependence parameters, \( \gamma^*_k; k = 1, 2, 3 \) are plotted in Figures 4.24 - 4.25. As these figures show, the Monte Carlo approximations of the expected standardized dependence parameters corresponding to category 1 and category 2 are similar while the Monte Carlo approximations of the expected
Figure 4.23  Standard bounds, $\gamma_\kappa$, defined by (4.31) for the Binomial MRF model dependence parameter, $\gamma$ in (4.1), plotted against $\kappa$ for $0 < \kappa < 1$
standardized dependence parameter corresponding to category 3 is smaller than the Monte Carlo approximations of the expected standardized dependence parameters associated with the other two categories. This result suggests that the dependence structure within the first two categories is similar in strength, while the dependence structure within the third or last category is discernably weaker than the dependence structure associated with category 1 or category 2. As previously discussed, if the size of the variance of the conditional expectations is any indication of the strength of the dependence structure, then the weakest dependence structure will be found within category 3. For this reason, the patterns observed in the Monte Carlo approximations of the standardized dependence parameters were not unexpected.

4.5 Dependence of Parameter Estimation and PMSE on Category Indices

In practice, a model is fitted to a particular data set to assist researchers in answering questions that motivated the collection of the data set. In particular, a researcher is often interested in quantities such as parameter estimates and predictions to answer these questions. As shown in the previous sections, variances and covariances of the conditional expectations and marginal variances and covariances depend on which category is indexed as the $h^{th}$ category. Consequently, one might expect that parameter estimation, mean squared error (MSE) of parameter estimators, and prediction mean squared error (PMSE) would also be dependent on which category is indexed as the $h^{th}$ category.

To examine the dependence of parameter estimation, mean squared error of the estimators and prediction mean squared error on the indexing of the three categories, Multinomial Markov random fields were simulated according to the steps outlined in Section 3.4. First, Multinomial Markov random fields were simulated with $\kappa = (0.20, 0.30, 0.50)^T$ and $\gamma \in \{0.50, 2.0\}$. For each field that was simulated, three Multinomial MRF models were fitted to the data set. One Multinomial MRF model labeled the category originally indexed as the first category as the third category, another Multinomial MRF model labeled the category originally indexed as the second category as the third category and the last Multinomial MRF model (correctly) labeled the category originally indexed as the third category as the third category.
Figure 4.24 Monte Carlo approximations of $\gamma_k^*$, the standardized dependence parameter, given by (4.32) and (4.33) with $\gamma = 0.5$

Figure 4.25 Monte Carlo approximations of $\gamma_k^*$, the standardized dependence parameter, given by (4.32) and (4.33) with $\gamma = 2.0$
For each fitted model, estimates of $\kappa$ and $\gamma$ were recorded. Let $\hat{\theta}$ denote the vector of parameter estimates, $(\hat{\kappa}_1, \hat{\kappa}_2, \hat{\kappa}_3, \hat{\gamma})^T$. The parameter estimates were then used to calculate the predicted value for the $k^{th}$ category at location $s_i$, denoted as $\hat{y}_k(s_i)$, by substituting $\hat{\theta}$ for $\theta$ in (4.7) for $k = 1, 2$ and in (4.8) for $k = 3$. The prediction mean squared error for the $k^{th}$ category and the $t^{th}$ data set was calculated as

$$\text{PMSE}_{k,t} = \sqrt{\frac{\sum_{i=1}^{900} \{y_k(s_i) - \hat{y}_k(s_i)\}^2}{900}}.$$  (4.34)

Then the MC approximations of the expected values of the parameter estimates and PMSE$_k$, the prediction mean squared error for the $k^{th}$ category, were calculated according to (3.23). Finally, the MC approximation of the mean squared error was calculated for each parameter estimator. The mean squared error of an estimator, $\hat{\theta}$, of a parameter, $\theta$, is

$$E_\theta(\hat{\theta} - \theta)^2 = \text{Var}_\theta \hat{\theta} + (E_\theta \hat{\theta} - \theta)^2,$$  (4.35)

where $(E_\theta \hat{\theta} - \theta)^2$ is referred to as the bias of an estimator. To calculate the MC approximation of the mean squared error of an estimator, the MC approximation of the expected value of the parameter estimate, $E_T(\hat{\theta})$, was substituted for $E_\theta \hat{\theta}$ in (4.35). Then the variance of the parameter estimates obtained from the specified number of Markov random fields, $T$, was calculated and substituted in for $\text{Var}_\theta \hat{\theta}$.

The above steps were repeated for $\kappa = (0.20, 0.30, 0.50)^T$, $\kappa = (0.20, 0.50, 0.30)^T$, $\kappa = (0.30, 0.50, 0.20)^T$, $\kappa = (0.30, 0.30, 0.40)^T$ and $\kappa = (0.30, 0.40, 0.30)^T$ with $\gamma \in \{0.5, 2.0\}$. After the mean and variance of the parameter estimates from 1,000 Markov random fields was calculated and those values were substituted into (3.24) for $\theta$ and $\sigma$, respectively, $T = 7,000$ Markov random fields was determined to be the number necessary to satisfy the convergence criteria as outlined in Section 3.4 for all sets of parameter values. The results based on 7,000 Markov random fields are given in Tables 4.1 - 4.15.

As can be seen in Tables 4.1 - 4.15, for a specified value of $\gamma$ and a particular category, $k$, the MC approximations of $E(\hat{\kappa}_k)$ are very similar, especially when a weak dependent structure was specified for the model ($\gamma = 0.5$). When $\gamma = 0.5$, the MC approximations of $E(\hat{\kappa}_k)$ under the correct labeling of categories were not the closest to the true parameter values for every
Table 4.1  MC Approximations of the Expected Values of the Parameter Estimates for \( k = (0.20, 0.30, 0.50)^T \)

<table>
<thead>
<tr>
<th>Category Labeled As Third Category</th>
<th>( \gamma )</th>
<th>( E_T(\hat{\kappa}_1) )</th>
<th>( E_T(\hat{\kappa}_2) )</th>
<th>( E_T(\hat{\kappa}_3) )</th>
<th>( E_T(\hat{\gamma}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>first</td>
<td>0.5</td>
<td>0.200001</td>
<td>0.300021</td>
<td>0.499979</td>
<td>0.252905</td>
</tr>
<tr>
<td>second</td>
<td>0.5</td>
<td>0.199992</td>
<td>0.300031</td>
<td>0.499977</td>
<td>0.273697</td>
</tr>
<tr>
<td>third</td>
<td>0.5</td>
<td>0.199990</td>
<td>0.300020</td>
<td>0.499990</td>
<td>0.482506</td>
</tr>
<tr>
<td>first</td>
<td>2.0</td>
<td>0.200128</td>
<td>0.300061</td>
<td>0.499811</td>
<td>1.065665</td>
</tr>
<tr>
<td>second</td>
<td>2.0</td>
<td>0.200021</td>
<td>0.300180</td>
<td>0.499799</td>
<td>1.165560</td>
</tr>
<tr>
<td>third</td>
<td>2.0</td>
<td>0.199991</td>
<td>0.300021</td>
<td>0.499988</td>
<td>1.970461</td>
</tr>
</tbody>
</table>

Table 4.2  MC Approximations of the Mean Squared Error for \( k = (0.20, 0.30, 0.50)^T \)

<table>
<thead>
<tr>
<th>Category Labeled As Third Category</th>
<th>( \gamma )</th>
<th>( E_T(MSE_{\kappa_1}) )</th>
<th>( E_T(MSE_{\kappa_2}) )</th>
<th>( E_T(MSE_{\kappa_3}) )</th>
<th>( E_T(MSE_{\gamma}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>first</td>
<td>0.5</td>
<td>1.974x10^{-6}</td>
<td>2.600x10^{-6}</td>
<td>2.980x10^{-6}</td>
<td>1.199x10^{-1}</td>
</tr>
<tr>
<td>second</td>
<td>0.5</td>
<td>1.975x10^{-6}</td>
<td>2.594x10^{-6}</td>
<td>2.981x10^{-6}</td>
<td>1.342x10^{-1}</td>
</tr>
<tr>
<td>third</td>
<td>0.5</td>
<td>1.975x10^{-6}</td>
<td>2.594x10^{-6}</td>
<td>2.976x10^{-6}</td>
<td>1.674x10^{-1}</td>
</tr>
<tr>
<td>first</td>
<td>2.0</td>
<td>2.906x10^{-6}</td>
<td>4.314x10^{-6}</td>
<td>3.852x10^{-6}</td>
<td>9.243x10^{-1}</td>
</tr>
<tr>
<td>second</td>
<td>2.0</td>
<td>2.885x10^{-6}</td>
<td>4.348x10^{-6}</td>
<td>3.868x10^{-6}</td>
<td>7.733x10^{-1}</td>
</tr>
<tr>
<td>third</td>
<td>2.0</td>
<td>2.881x10^{-6}</td>
<td>4.345x10^{-6}</td>
<td>3.856x10^{-6}</td>
<td>7.855x10^{-1}</td>
</tr>
</tbody>
</table>

Table 4.3  MC Approximations of the Expected Values of the Prediction Mean Squared Errors for \( k = (0.20, 0.30, 0.50)^T \)

<table>
<thead>
<tr>
<th>Category Labeled As Third Category</th>
<th>( \gamma )</th>
<th>( E_T(P\hat{MSE}_1) )</th>
<th>( E_T(P\hat{MSE}_2) )</th>
<th>( E_T(P\hat{MSE}_3) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>first</td>
<td>0.5</td>
<td>16.01</td>
<td>20.94</td>
<td>24.93</td>
</tr>
<tr>
<td>second</td>
<td>0.5</td>
<td>15.98</td>
<td>21.00</td>
<td>24.93</td>
</tr>
<tr>
<td>third</td>
<td>0.5</td>
<td>15.97</td>
<td>20.92</td>
<td>24.95</td>
</tr>
<tr>
<td>first</td>
<td>2.0</td>
<td>16.47</td>
<td>21.34</td>
<td>25.05</td>
</tr>
<tr>
<td>second</td>
<td>2.0</td>
<td>16.32</td>
<td>21.75</td>
<td>25.16</td>
</tr>
<tr>
<td>third</td>
<td>2.0</td>
<td>15.97</td>
<td>20.96</td>
<td>24.99</td>
</tr>
</tbody>
</table>
Table 4.4 MC Approximations of the Expected Values of the Parameter Estimates for $k = (0.20, 0.50, 0.30)^T$

<table>
<thead>
<tr>
<th>Category Labeled As Third Category</th>
<th>$\gamma$</th>
<th>$E_T(\hat{\kappa}_1)$</th>
<th>$E_T(\hat{\kappa}_2)$</th>
<th>$E_T(\hat{\kappa}_3)$</th>
<th>$E_T(\hat{\gamma})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>first</td>
<td>0.5</td>
<td>0.200022</td>
<td>0.499984</td>
<td>0.299993</td>
<td>0.311776</td>
</tr>
<tr>
<td>second</td>
<td>0.5</td>
<td>0.200013</td>
<td>0.499991</td>
<td>0.299996</td>
<td>0.387456</td>
</tr>
<tr>
<td>third</td>
<td>0.5</td>
<td>0.200006</td>
<td>0.499986</td>
<td>0.300008</td>
<td>0.480048</td>
</tr>
<tr>
<td>first</td>
<td>2.0</td>
<td>0.200390</td>
<td>0.499882</td>
<td>0.299728</td>
<td>1.362243</td>
</tr>
<tr>
<td>second</td>
<td>2.0</td>
<td>0.200249</td>
<td>0.499929</td>
<td>0.299822</td>
<td>1.760314</td>
</tr>
<tr>
<td>third</td>
<td>2.0</td>
<td>0.199979</td>
<td>0.500046</td>
<td>0.299975</td>
<td>1.981436</td>
</tr>
</tbody>
</table>

Table 4.5 MC Approximations of the Mean Squared Error for $k = (0.20, 0.50, 0.30)^T$

<table>
<thead>
<tr>
<th>Category Labeled As Third Category</th>
<th>$\gamma$</th>
<th>$E_T(MSE_{\kappa_1})$</th>
<th>$E_T(MSE_{\kappa_2})$</th>
<th>$E_T(MSE_{\kappa_3})$</th>
<th>$E_T(MSE_{\gamma})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>first</td>
<td>0.5</td>
<td>1.956 x 10^{-6}</td>
<td>3.200 x 10^{-6}</td>
<td>2.473 x 10^{-6}</td>
<td>9.239 x 10^{-2}</td>
</tr>
<tr>
<td>second</td>
<td>0.5</td>
<td>1.959 x 10^{-6}</td>
<td>3.204 x 10^{-6}</td>
<td>2.475 x 10^{-6}</td>
<td>1.238 x 10^{-1}</td>
</tr>
<tr>
<td>third</td>
<td>0.5</td>
<td>1.958 x 10^{-6}</td>
<td>3.202 x 10^{-6}</td>
<td>2.470 x 10^{-6}</td>
<td>9.328 x 10^{-2}</td>
</tr>
<tr>
<td>first</td>
<td>2.0</td>
<td>3.855 x 10^{-6}</td>
<td>6.851 x 10^{-6}</td>
<td>3.361 x 10^{-6}</td>
<td>4.523 x 10^{-1}</td>
</tr>
<tr>
<td>second</td>
<td>2.0</td>
<td>3.749 x 10^{-6}</td>
<td>6.835 x 10^{-6}</td>
<td>3.321 x 10^{-6}</td>
<td>1.700 x 10^{-1}</td>
</tr>
<tr>
<td>third</td>
<td>2.0</td>
<td>3.759 x 10^{-6}</td>
<td>6.873 x 10^{-6}</td>
<td>3.321 x 10^{-6}</td>
<td>1.095 x 10^{-1}</td>
</tr>
</tbody>
</table>

Table 4.6 MC Approximations of the Expected Values of the Prediction Mean Squared Errors for $k = (0.20, 0.50, 0.30)^T$

<table>
<thead>
<tr>
<th>Category Labeled As Third Category</th>
<th>$\gamma$</th>
<th>$E_T(P\overline{MSE}_{1})$</th>
<th>$E_T(P\overline{MSE}_{2})$</th>
<th>$E_T(P\overline{MSE}_{3})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>first</td>
<td>0.5</td>
<td>16.02</td>
<td>24.96</td>
<td>20.94</td>
</tr>
<tr>
<td>second</td>
<td>0.5</td>
<td>16.00</td>
<td>25.03</td>
<td>20.96</td>
</tr>
<tr>
<td>third</td>
<td>0.5</td>
<td>15.97</td>
<td>24.93</td>
<td>20.96</td>
</tr>
<tr>
<td>first</td>
<td>2.0</td>
<td>16.65</td>
<td>25.56</td>
<td>21.14</td>
</tr>
<tr>
<td>second</td>
<td>2.0</td>
<td>16.58</td>
<td>26.23</td>
<td>21.41</td>
</tr>
<tr>
<td>third</td>
<td>2.0</td>
<td>15.97</td>
<td>24.92</td>
<td>20.96</td>
</tr>
</tbody>
</table>
Table 4.7 MC Approximations of the Expected Values of the Parameter Estimates for $k = (0.30, 0.50, 0.20)^T$

<table>
<thead>
<tr>
<th>Category Labeled As Third Category</th>
<th>$\gamma$</th>
<th>$E_T(\hat{\kappa}_1)$</th>
<th>$E_T(\hat{\kappa}_2)$</th>
<th>$E_T(\hat{\kappa}_3)$</th>
<th>$E_T(\hat{\gamma})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>first</td>
<td>0.5</td>
<td>0.300042</td>
<td>0.499983</td>
<td>0.199975</td>
<td>0.439618</td>
</tr>
<tr>
<td>second</td>
<td>0.5</td>
<td>0.300028</td>
<td>0.499994</td>
<td>0.199977</td>
<td>0.503650</td>
</tr>
<tr>
<td>third</td>
<td>0.5</td>
<td>0.300020</td>
<td>0.499888</td>
<td>0.199993</td>
<td>0.481193</td>
</tr>
<tr>
<td>first</td>
<td>2.0</td>
<td>0.300848</td>
<td>0.499639</td>
<td>0.199513</td>
<td>2.024996</td>
</tr>
<tr>
<td>second</td>
<td>2.0</td>
<td>0.300477</td>
<td>0.499758</td>
<td>0.199765</td>
<td>2.397088</td>
</tr>
<tr>
<td>third</td>
<td>2.0</td>
<td>0.299893</td>
<td>0.500135</td>
<td>0.199973</td>
<td>1.979421</td>
</tr>
</tbody>
</table>

Table 4.8 MC Approximations of the Mean Squared Error for $k = (0.30, 0.50, 0.20)^T$

<table>
<thead>
<tr>
<th>Category Labeled As Third Category</th>
<th>$\gamma$</th>
<th>$E_T(MSE_{\kappa_1})$</th>
<th>$E_T(MSE_{\kappa_2})$</th>
<th>$E_T(MSE_{\kappa_3})$</th>
<th>$E_T(MSE_{\gamma})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>first</td>
<td>0.5</td>
<td>2.836x10^{-6}</td>
<td>3.381x10^{-6}</td>
<td>1.852x10^{-6}</td>
<td>8.229x10^{-2}</td>
</tr>
<tr>
<td>second</td>
<td>0.5</td>
<td>2.832x10^{-6}</td>
<td>3.378x10^{-6}</td>
<td>1.853x10^{-6}</td>
<td>1.145x10^{-1}</td>
</tr>
<tr>
<td>third</td>
<td>0.5</td>
<td>2.838x10^{-6}</td>
<td>3.381x10^{-6}</td>
<td>1.851x10^{-6}</td>
<td>5.498x10^{-2}</td>
</tr>
<tr>
<td>first</td>
<td>2.0</td>
<td>8.972x10^{-6}</td>
<td>1.042x10^{-5}</td>
<td>2.479x10^{-6}</td>
<td>5.658x10^{-2}</td>
</tr>
<tr>
<td>second</td>
<td>2.0</td>
<td>8.487x10^{-6}</td>
<td>1.033x10^{-5}</td>
<td>2.274x10^{-6}</td>
<td>2.529x10^{-1}</td>
</tr>
<tr>
<td>third</td>
<td>2.0</td>
<td>8.666x10^{-6}</td>
<td>1.060x10^{-5}</td>
<td>2.276x10^{-6}</td>
<td>1.829x10^{-1}</td>
</tr>
</tbody>
</table>

Table 4.9 MC Approximations of the Expected Values of the Prediction Mean Squared Errors for $k = (0.30, 0.50, 0.20)^T$

<table>
<thead>
<tr>
<th>Category Labeled As Third Category</th>
<th>$\gamma$</th>
<th>$E_T(P\bar{MSE}_{1})$</th>
<th>$E_T(P\bar{MSE}_{2})$</th>
<th>$E_T(P\bar{MSE}_{3})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>first</td>
<td>0.5</td>
<td>21.03</td>
<td>24.99</td>
<td>15.98</td>
</tr>
<tr>
<td>second</td>
<td>0.5</td>
<td>20.99</td>
<td>25.05</td>
<td>15.98</td>
</tr>
<tr>
<td>third</td>
<td>0.5</td>
<td>20.94</td>
<td>24.95</td>
<td>15.98</td>
</tr>
<tr>
<td>first</td>
<td>2.0</td>
<td>22.19</td>
<td>25.84</td>
<td>16.43</td>
</tr>
<tr>
<td>second</td>
<td>2.0</td>
<td>21.87</td>
<td>26.32</td>
<td>16.59</td>
</tr>
<tr>
<td>third</td>
<td>2.0</td>
<td>20.97</td>
<td>24.89</td>
<td>15.98</td>
</tr>
</tbody>
</table>
Table 4.10 MC Approximations of the Expected Values of the Parameter Estimates for $k = (0.30, 0.30, 0.40)^T$

<table>
<thead>
<tr>
<th>Category Labeled As Third Category</th>
<th>$\gamma$</th>
<th>$E_T(\hat{\kappa}_1)$</th>
<th>$E_T(\hat{\kappa}_2)$</th>
<th>$E_T(\hat{\kappa}_3)$</th>
<th>$E_T(\hat{\gamma})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>first</td>
<td>0.5</td>
<td>0.300054</td>
<td>0.299988</td>
<td>0.399958</td>
<td>0.292261</td>
</tr>
<tr>
<td>second</td>
<td>0.5</td>
<td>0.300043</td>
<td>0.299998</td>
<td>0.399959</td>
<td>0.293063</td>
</tr>
<tr>
<td>third</td>
<td>0.5</td>
<td>0.300043</td>
<td>0.299984</td>
<td>0.399973</td>
<td>0.477075</td>
</tr>
<tr>
<td>first</td>
<td>2.0</td>
<td>0.300238</td>
<td>0.300080</td>
<td>0.399682</td>
<td>1.313712</td>
</tr>
<tr>
<td>second</td>
<td>2.0</td>
<td>0.300066</td>
<td>0.300251</td>
<td>0.399683</td>
<td>1.312319</td>
</tr>
<tr>
<td>third</td>
<td>2.0</td>
<td>0.300013</td>
<td>0.300035</td>
<td>0.399683</td>
<td>1.980156</td>
</tr>
</tbody>
</table>

Table 4.11 MC Approximations of the Mean Squared Error for $k = (0.30, 0.30, 0.40)^T$

<table>
<thead>
<tr>
<th>Category Labeled As Third Category</th>
<th>$\gamma$</th>
<th>$E_T(MSE_{\kappa_1})$</th>
<th>$E_T(MSE_{\kappa_2})$</th>
<th>$E_T(MSE_{\kappa_3})$</th>
<th>$E_T(MSE_{\gamma})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>first</td>
<td>0.5</td>
<td>2.743x10^{-6}</td>
<td>2.679x10^{-6}</td>
<td>2.823x10^{-6}</td>
<td>1.099x10^{-1}</td>
</tr>
<tr>
<td>second</td>
<td>0.5</td>
<td>2.745x10^{-6}</td>
<td>2.677x10^{-6}</td>
<td>2.824x10^{-6}</td>
<td>1.115x10^{-1}</td>
</tr>
<tr>
<td>third</td>
<td>0.5</td>
<td>2.745x10^{-6}</td>
<td>2.682x10^{-6}</td>
<td>2.821x10^{-6}</td>
<td>1.270x10^{-1}</td>
</tr>
<tr>
<td>first</td>
<td>2.0</td>
<td>5.261x10^{-6}</td>
<td>5.053x10^{-5}</td>
<td>3.808x10^{-6}</td>
<td>5.331x10^{-1}</td>
</tr>
<tr>
<td>second</td>
<td>2.0</td>
<td>5.198x10^{-6}</td>
<td>5.113x10^{-5}</td>
<td>3.809x10^{-6}</td>
<td>5.338x10^{-1}</td>
</tr>
<tr>
<td>third</td>
<td>2.0</td>
<td>5.235x10^{-6}</td>
<td>5.140x10^{-5}</td>
<td>3.799x10^{-5}</td>
<td>5.305x10^{-1}</td>
</tr>
</tbody>
</table>

Table 4.12 MC Approximations of the Expected Values of the Prediction Mean Squared Errors for $k = (0.30, 0.30, 0.40)^T$

<table>
<thead>
<tr>
<th>Category Labeled As Third Category</th>
<th>$\gamma$</th>
<th>$E_T(P\bar{MSE}_1)$</th>
<th>$E_T(P\bar{MSE}_2)$</th>
<th>$E_T(P\bar{MSE}_3)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>first</td>
<td>0.5</td>
<td>21.03</td>
<td>20.99</td>
<td>23.96</td>
</tr>
<tr>
<td>second</td>
<td>0.5</td>
<td>20.99</td>
<td>21.03</td>
<td>23.96</td>
</tr>
<tr>
<td>third</td>
<td>0.5</td>
<td>20.95</td>
<td>20.95</td>
<td>23.98</td>
</tr>
<tr>
<td>first</td>
<td>2.0</td>
<td>22.04</td>
<td>21.63</td>
<td>24.20</td>
</tr>
<tr>
<td>second</td>
<td>2.0</td>
<td>21.62</td>
<td>22.06</td>
<td>24.20</td>
</tr>
<tr>
<td>third</td>
<td>2.0</td>
<td>20.95</td>
<td>20.96</td>
<td>23.95</td>
</tr>
</tbody>
</table>
Table 4.13 MC Approximations of the Expected Values of the Parameter Estimates for $k = (0.30, 0.40, 0.30)^T$

<table>
<thead>
<tr>
<th>Category Labeled As Third Category</th>
<th>$\gamma$</th>
<th>$E_T(\hat{\kappa}_1)$</th>
<th>$E_T(\hat{\kappa}_2)$</th>
<th>$E_T(\hat{\kappa}_3)$</th>
<th>$E_T(\hat{\gamma})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>first</td>
<td>0.5</td>
<td>0.300009</td>
<td>0.399988</td>
<td>0.300003</td>
<td>0.348704</td>
</tr>
<tr>
<td>second</td>
<td>0.5</td>
<td>0.299996</td>
<td>0.399988</td>
<td>0.300006</td>
<td>0.372458</td>
</tr>
<tr>
<td>third</td>
<td>0.5</td>
<td>0.299990</td>
<td>0.399988</td>
<td>0.300022</td>
<td>0.481861</td>
</tr>
<tr>
<td>first</td>
<td>2.0</td>
<td>0.300448</td>
<td>0.399924</td>
<td>0.299629</td>
<td>1.582405</td>
</tr>
<tr>
<td>second</td>
<td>2.0</td>
<td>0.300197</td>
<td>0.400122</td>
<td>0.299681</td>
<td>1.735851</td>
</tr>
<tr>
<td>third</td>
<td>2.0</td>
<td>0.299984</td>
<td>0.400017</td>
<td>0.299999</td>
<td>1.984963</td>
</tr>
</tbody>
</table>

Table 4.14 MC Approximations of the Mean Squared Error for $k = (0.30, 0.40, 0.30)^T$

<table>
<thead>
<tr>
<th>Category Labeled As Third Category</th>
<th>$\gamma$</th>
<th>$E_T(MSE_{\kappa_1})$</th>
<th>$E_T(MSE_{\kappa_2})$</th>
<th>$E_T(MSE_{\kappa_3})$</th>
<th>$E_T(MSE_{\gamma})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>first</td>
<td>0.5</td>
<td>2.777x10^{-6}</td>
<td>3.161x10^{-6}</td>
<td>2.461x10^{-6}</td>
<td>9.078x10^{-2}</td>
</tr>
<tr>
<td>second</td>
<td>0.5</td>
<td>2.777x10^{-6}</td>
<td>3.153x10^{-6}</td>
<td>2.456x10^{-6}</td>
<td>1.008x10^{-1}</td>
</tr>
<tr>
<td>third</td>
<td>0.5</td>
<td>2.783x10^{-6}</td>
<td>3.157x10^{-6}</td>
<td>2.457x10^{-6}</td>
<td>8.316x10^{-2}</td>
</tr>
<tr>
<td>first</td>
<td>2.0</td>
<td>6.421x10^{-6}</td>
<td>7.691x10^{-5}</td>
<td>3.213x10^{-6}</td>
<td>2.296x10^{-1}</td>
</tr>
<tr>
<td>second</td>
<td>2.0</td>
<td>6.258x10^{-6}</td>
<td>7.675x10^{-5}</td>
<td>3.170x10^{-6}</td>
<td>1.465x10^{-1}</td>
</tr>
<tr>
<td>third</td>
<td>2.0</td>
<td>6.297x10^{-6}</td>
<td>7.750x10^{-5}</td>
<td>3.167x10^{-6}</td>
<td>1.069x10^{-1}</td>
</tr>
</tbody>
</table>

Table 4.15 MC Approximations of the Expected Values of the Prediction Mean Squared Errors for $k = (0.30, 0.40, 0.30)^T$

<table>
<thead>
<tr>
<th>Category Labeled As Third Category</th>
<th>$\gamma$</th>
<th>$E_T(P\tilde{MSE}_1)$</th>
<th>$E_T(P\tilde{MSE}_2)$</th>
<th>$E_T(P\tilde{MSE}_3)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>first</td>
<td>0.5</td>
<td>21.03</td>
<td>23.99</td>
<td>20.97</td>
</tr>
<tr>
<td>second</td>
<td>0.5</td>
<td>21.00</td>
<td>24.05</td>
<td>20.97</td>
</tr>
<tr>
<td>third</td>
<td>0.5</td>
<td>20.95</td>
<td>23.94</td>
<td>20.98</td>
</tr>
<tr>
<td>first</td>
<td>2.0</td>
<td>22.27</td>
<td>24.86</td>
<td>21.33</td>
</tr>
<tr>
<td>second</td>
<td>2.0</td>
<td>21.89</td>
<td>25.50</td>
<td>21.45</td>
</tr>
<tr>
<td>third</td>
<td>2.0</td>
<td>20.96</td>
<td>23.92</td>
<td>20.95</td>
</tr>
</tbody>
</table>
set of parameter values. When \( \gamma = 2.0 \), the MC approximations of \( E(\hat{\kappa}_k) \) under the correct labeling of categories were the closest to the true parameter values for every set of parameter values; however, the differences in MC approximations of \( E(\hat{\kappa}_k) \) for a given \( k \) were less than 0.001.

The MC approximations of \( E(\hat{\gamma}) \) for a given set of parameter values are not similar. One reason for the large differences between these MC approximations is that dependence parameter estimates were not standardized before calculating the MC approximations of the dependence parameters. Calculating the standard bounds for multi-parameter exponential families appears to be untractable or even impossible. We can only approximate the standard bounds through simulation as shown in Section 3.6. Furthermore, even if we could calculate the standard bounds, comparing standardized dependence parameter estimates between models that correctly labeled the categories and models that incorrectly labeled the categories may not be appropriate. As discussed in Section 2.4, the dependence structure is weakest in category 3, which indicates that a model that incorrectly labels the categories may not accurately characterize the dependence structure within each category. Consequently, a standardized dependence parameter estimate for a model that correctly labels the categories may not have the same meaning as a standardized dependence parameter that incorrectly labels the categories. For these reasons, comparisons of the MC approximations of the expected values of the dependence parameter is not meaningful.

The MC approximations of the MSE of \( \hat{\kappa}_k; k = 1, 2, 3 \), are not always the smallest when the categories are indexed correctly even if a strong dependence structure is present, i.e., when \( \gamma = 2.0 \). As noted in (4.35), the MSE is the sum of the variance of the estimator and the square of the bias. The MC approximation of the variance of \( \hat{\kappa}_k \) is of the order of \( 10^{-6} \), whereas the square of the bias is of the order of \( 10^{-7} \) or smaller for all sets of parameter values under consideration. This means the MC approximation of the MSE is dictated more by the MC approximation of the variance of \( \hat{\kappa}_k \) than the MC approximation of the bias of \( \hat{\kappa}_k \). Since the MC approximations of the variance of \( \hat{\kappa}_k \) are similar for a given set of parameter values, the MC approximations of the MSE of \( \hat{\kappa}_k \) are also similar. We cannot directly compare the MC
approximations of the MSE of $\hat{\gamma}$ for the same aforementioned reasons.

Finally, as with the MC approximations of $E(\hat{\kappa}_k); k = 1, 2, 3$, the MC approximations of $E(\hat{PMSE}_k); k = 1, 2, 3$, for a given set of parameter values are similar when dependence is weak ($\gamma = 0.5$). When the dependence is stronger (i.e., $\gamma = 2.0$), the MC approximations of $E(\hat{PMSE}_k)$ when the categories are correctly indexed is smaller than the respective MC approximations when the categories are incorrectly indexed for all sets of parameter values. In some cases, the value of $E_T(\hat{PMSE}_k)$ when the categories are incorrectly indexed is approximately 5%-7% larger than the value of $E_T(\hat{PMSE}_k)$ when the categories are correctly indexed. Perhaps, the Multinomial MRF model can correctly account for the dependence structure within each category when the categories are correctly indexed, which then allows the model to more accurately predict observations, on average, than a model that incorrectly indexed the categories.

4.6 Assignment of Category Indices

As shown in Sections 4.2 - 4.5, many aspects of the behavior of the model are influenced by the assignment of category indices. In particular, as shown in Section 4.5, the mean squared error of a parameter estimator and prediction mean squared error is affected by the assignment of the category indices, especially when there is a strong dependence structure present. Currently one question remains: How should one index the categories when one wishes to fit a Multinomial Markov random field model to a data set? The approach recommended in this section follows from the results in Section 4.4.

In Section 4.4, three Binomial MRF models were fitted to each simulated field. Since a Binomial MRF model is a Multinomial MRF model with two categories, the three categories need to be reduced to two categories. For the first Binomial MRF model, the events in all categories were aggregated except for the events in the category originally indexed as the first category. For the second (and third) Binomial MRF model, the events in all categories were aggregated except for the events in the category originally indexed as the second (third) category. Then the estimate of the dependence parameter, $\gamma$, was obtained and standardized.
Figures 4.24 and 4.25 show that when comparing the MC approximations of the standardized dependence parameters, the MC approximation of the expected standardized dependence parameter corresponding to the third Binomial model is the smallest. This finding indicates one could fit three Binomial MRF models and label the category which corresponds to the smallest standardized dependence parameter estimate as the $h^{th}$ or last category.

We note that a category could always be randomly selected to be indexed as the last category. The probability that one randomly chooses the correct category (out of $h$ categories) to be indexed as the last category is $1/h$. If another method is suggested for use in practice, then the probability that this method correctly identifies the category as the last category should be greater than $1/h$ in order for this method to be used in practice instead of the method of randomly indexing the categories.

To determine if the method of fitting three Binomial MRF models and indexing a category as the third category based the dependence parameter estimates is an improvement over the method of randomly indexing the categories, Markov random fields were simulated according to the steps outlined in Section 3.4. For each field, the three Binomial MRF models were fitted and the estimate for the dependence parameter was obtained. Then each dependence parameter estimate was standardized by dividing the estimate by the standard bounds given by (4.31). One-thousand Markov random fields were simulated for different sets of parameter values such that $\kappa_1 \in \{0.10, 0.20, 0.30\}$, $\kappa_2 \in \{0.10, ..., 0.90 - \kappa_1\}$ for a given value of $\kappa_1$ and $\gamma \in \{0.50, 2.0\}$. The category that is associated with the smallest standardized dependence parameter estimate was indexed as the third (or last) category. For each set of parameter values, the estimated probability of this method correctly identifying the last category is the number of times the category originally indexed as the third category was chosen to be the third category divided by 1000. Figure 4.26 depicts the probability of correctly identifying the third category for several sets of parameter values.

In general, given $\kappa_1$, the probability of correctly identifying the third category generally increases as the value for $\kappa_2$ increases. Similarly, given $\kappa_2$, the probability of correctly identifying the third category generally increases as the value for $\kappa_1$ increases. For each set of
parameter values, the probability of correctly identifying the third category is approximately equal to or greater than 0.33, which is the probability of correctly identifying the third category by randomly indexing a category as the third category. Furthermore, when the value of the dependence parameter is relatively large ($\gamma = 2.0$), the probability is larger than 0.60 and is quite often close to 1.0, a notable improvement over 0.33. These results indicate that the method proposed in this section is an improvement over the method of randomly indexing the categories when determining which category should be labeled as the last or $h^{th}$ category.
Figure 4.26  Probability of labeling the category originally indexed as the third category as the third category after fitting three Binomial MRF models.
CHAPTER 5  APPLICATION

5.1  Introduction

The state of Iowa is the second largest producer of wind energy in the United States, due to the state’s combination of topography and electric transmission lines. The topography affects wind speeds, which is one of the factors that determines whether or not a wind turbine is economically practical. Specifically, a wind turbine needs be exposed to wind speeds averaging at least 12 mph annually (Wind Energy Manual from Iowa Energy Center) to be economically practical. For day to day operations, a minimum wind speed of generally 7 mph to 10 mph is needed for the turbine to generate usable power (Wind Energy Manual from Iowa Energy Center). Because there is great interest in wind energy in the state of Iowa, a Multinomial MRF model will be fit to subsets of the North American Regional Reanalysis (NARR) data set to study wind speeds in Iowa and the surrounding states.

5.2  Data Description

The data used in this study are a subset of the North American Regional Reanalysis (NARR) data set provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at http://www.esrl.noaa.gov/psd/. Values are assimilated climate observations using the same model for the entire reanalysis period, which is 1979 to present. The subset of the NARR data set sampled for this study contains wind speeds at 10 m above the earth’s surface at three-hour intervals during the months of June, July and August for locations on an approximately 32 km by 32 km grid across Iowa and the surrounding states for years 1980, 1985, 1990, 1995 and 2000. Figure 5.1 shows the sampled locations represented by circles overlaying a map of Iowa and the surrounding states.
Figure 5.1 Sampled locations from the North American Regional Reanalysis (NARR) data set as represented by circles
5.3 Model Formulation

5.3.1 Response Variable

The response variable is wind speed, measured in meters per second and converted to miles per hour. There are 8 observations per day for the 92 days during June, July and August, for a total of 736 observations per location. To fit a Multinomial MRF field model, response categories will need to be defined because the observed variable is continuous. Once \( h \) categories have been defined, let \( W(s_i) = (W_1(s_i), W_2(s_i), ..., W_h(s_i))^T \) be a vector representing the wind speeds sampled at location \( s_i \) where \( m \) represents the total number of observations for each location, \( (m = 736) \).

5.3.2 Neighborhoods

The locations, represented by circles in Figure 5.1, nearly correspond to a spatial lattice, \( D = [0, 23] \times [0, 23] \). The neighborhood structure chosen for this application is the four-nearest neighbor specification as defined in Section 3.1, so that the neighborhood, \( N_i \), of location \( s_i \) consists of the four nearest neighbors except for those locations that are located in the outer-most rows and columns. For most of these locations, the neighborhoods contain three neighbors while for corner locations the neighborhoods contain two neighbors. Then the Markov assumption is that for each location \( s_i; i = 1, 2, ..., n \), the conditional distribution of \( Y(s_i) \) given the observed values at all other locations \( \{y(s_j) : j \neq i\} \) is dependent only on the observed values at the neighborhood locations, \( N_i \), as defined by (3.2).

5.3.3 Conditional Probability Mass Function

We specify for the Multinomial MRF model, the conditional probability mass function for \( y(s_i) = (y_1(s_i), y_2(s_i), y_3(s_i))^T \) given the values at the neighborhood locations, \( y(N_i) = (y_1(N_i), y_2(N_i), y_3(N_i))^T \), and the vector of parameters, \( \theta \), given by (3.5) with \( m_i = m \) for \( i = 1, 2, ..., n \). The natural parameter function, \( A_{i,k} \{y(N_i); \theta\}; k = 1, 2 \), is given by (3.25). For this application, discussion of the dependence parameter is in terms of \( \eta \) because the number of neighbors for location \( s_i \) is not equal for all \( i = 1, ..., n \).
5.3.4 Estimation with the Pseudo-Likelihood Function

To estimate the vector of parameters, $\theta$, we would like to maximize the pseudo-likelihood function, defined as the product of the conditional mass functions, by using an iterative procedure to find the maximum value. This pseudo-likelihood function according to Besag (1974) may be written as

$$P(\theta) = \prod_{i=1}^{n} f_i (y(s_i)|y(N_i); \theta),$$

where the conditional probability mass, $f_i (y(s_i)|y(N_i); \theta)$, is given by (3.5). Because the values of the pseudo-likelihood are too large for the iterative procedure to handle via a computer, we minimized the negative of the log pseudo-likelihood, $-\log(P(\theta))$, instead. The iterative method used is an implementation of the Nelder-Mead method.

5.4 Issues in Estimation

While fitting the Multinomial MRF model model to subsets of the NARR data set, several issues arose. The first issue arose when the categories were arbitrarily defined. When a data set is generated from a Multinomial MRF model, the covariance between category counts for any pair of categories will always be negative. Thus, when we fit a Multinomial MRF model to a data set, the marginal covariances of the data set should all be negative. Otherwise, the characteristics of the Multinomial MRF model does not accurately reflect the characteristics present in the data set and fitting a Multinomial MRF model to the data set is not desirable. For the data sets under consideration, some of the category definitions led to positive covariance of the category counts for one pair of categories. Although we currently do not know how the existence of positive marginal covariance for one or more pairs of categories affects statistical issues such as estimation, we do not recommend fitting a Multinomial MRF model to such data sets. Consequently, there are limits to how the categories can be defined in this application to ensure that the covariance of the category counts is negative for all pairs of categories.

Once the categories were defined and indexed, both large-scale structure and small-scale structure were detected in the data set. Large-scale structure describes the general structure
across all locations, whereas, small-scale structure describes the structure between each location and its neighbors apart from the large-scale structure. Unfortunately, there is no standard that distinctly separates large-scale structure from small-scale structure. We can only describe large-scale and small-scale structure in general terms. The question regarding this issue is whether we should fit a Multinomial MRF model and allow the dependence parameter to model both the large-scale and small-scale structure or account for the large-scale structure through, for example, covariates, and allow the dependence parameter to model the remaining structure.

After the model is specified, we need to confirm the iterative procedure used to find the global maximum of the psuedo-likelihood did converge at the global maximum and thus, the values returned at convergence are the parameter estimates. If the iterative procedure did not converge at the global maximum, then we have what is referred to as false convergence. False convergence can occur, for example, when the psuedo-likelihood is very flat or the psuedo-likelihood contains local maximum values. To check for false convergence, the iterative procedure should be run several times using a different set of starting values each time. The hope is that the convergence of the iterative process does not depend on the starting values. The profile of the psuedo-likelihood can also be plotted, which can give an indication of whether or not the psuedo-likelihood has a global maximum in various dimensions of the parameter vector.

If we do not have false convergence and estimates can be obtained, then we can address the final issue regarding the size of the dependence parameter estimate. According to Section 3.6, the value of $\eta$ needs to be within certain standard bounds for the marginal means of a data set to be approximately equal to the values of the parameters, $\kappa_k; k = 1, 2, 3$. As shown in Figures 3.3 - 3.5, simulation can give approximate standard bounds for the dependence parameter, $\eta$, as these bounds cannot be derived analytically. Although, once an estimate for the dependence parameter is obtained, there is a more precise method of determining if the dependence parameter estimate is within its standard bounds instead of using simulation to approximate the standard bounds. The marginal means of the original data set to which the model was fitted can be compared to the respective marginal means of data sets simulated
according to the steps below using the parameter estimates from the fitted model. If the dependence parameter estimate is within its standard bounds, then the marginal means of the simulated data sets will be approximately equal to the respective marginal means of the original data set. As the value of the dependence parameter increases, the marginal means (in terms of percent of observations in category $k$) of the simulated data sets will slowly decay to 0 or 1 as shown in Figures 3.3 - 3.5.

1. Given the estimates for $\kappa_k$ for $k = 1, \ldots, h$, obtained by maximizing the pseudo-likelihood, generate starting values $y^{(0)}(s_i); i = 1, \ldots, n$, using the multinomial conditional probability mass function defined by (3.5) and (3.25) with $\eta = 0$. The notation $y^{(0)}(s_i)$ denotes $(y_1(s_i), \ldots, y_h(s_i))^T$ at iteration 0.

2. For iterations $t = 1, \ldots, T$, order the locations by using the identity function applied to locations $s_i; i = 1, \ldots, n$.

3. For each location, according to the order determined by step (2), generate $y^{(t)}(s_i)$ from the multinomial conditional probability mass function defined by (3.5) and (3.25) with $\eta$ equal to $\hat{\eta}$. Replace $y^{(t-1)}(s_i)$ with $y^{(t)}(s_i)$.

4. Repeat steps 2 and 3 until the specified number of iterations has been completed. For each simulation in this section, 500 iterations was specified.

5. Once the number of iterations has been completed, compare the marginal mean of the simulated data set to the marginal mean of the original data set of category $k; k = 1, 2, 3$.

### 5.5 Comparison of Models

#### 5.5.1 Multinomial MRF Model (Model 1)

For this model, the conditional probability mass function for $y(s_i) = (y_1(s_i), y_2(s_i), y_3(s_i))^T$ given the values at the neighborhood locations, $y(N_i) = (y_1(N_i), y_2(N_i), y_3(N_i))^T$, and the vector of parameters, $\theta$, is given by (3.5) with $m_i = m$ for $i = 1, 2, \ldots, n$. The natural parameter function is given by (3.25). This parameterization gives $\theta = (\kappa_1, \kappa_2, \eta)^T$. 
The first step in fitting the Multinomial MRF model is to define the categories. Defining categories for this application is arbitrary since the response variable, wind speed, is continuous. Recall that a wind speed of at least 7 mph to 10 mph is needed for the turbine to generate usable power. Categories will be defined to reflect the wind speeds that are needed to make generating wind power economically feasible. Hence, wind speeds will be assigned to the first, second and third category if the wind speed is less than or equal to 7 mph, greater than 7 mph but less than or equal to 10 mph, and greater than 10 mph, respectively. When the categories were defined in this manner, the covariance of the category counts $y_k(s_i); k = 1, 2, 3,$ for any two categories is negative for years 1980, 1985 and 1990 only. We could define the categories differently for each year to ensure that the covariance of the category counts for any pair of categories is negative; however, if the goal is to compare results from year to year, the categories should be defined in the same manner for each year. Consequently, we will restrict the analysis to years 1980, 1985 and 1990. Figures 5.2 - 5.4 contains image plots which graphically depict the distribution of the number of observations in each category $y_k(s_i)$ for $k = 1, 2, 3$ across locations, $s_i; i = 1, ..., n$, for years 1980, 1985 and 1990. Table 5.1 contains the marginal mean of category $k$, i.e., the mean of $Y_k(s_i)$ for all locations and denoted as $\bar{Y}_k(s_i)$, for each year.

<table>
<thead>
<tr>
<th>Year</th>
<th>$\bar{Y}_1(s_i)$</th>
<th>$\bar{Y}_2(s_i)$</th>
<th>$\bar{Y}_3(s_i)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980</td>
<td>0.313</td>
<td>0.337</td>
<td>0.350</td>
</tr>
<tr>
<td>1985</td>
<td>0.345</td>
<td>0.327</td>
<td>0.327</td>
</tr>
<tr>
<td>1990</td>
<td>0.354</td>
<td>0.323</td>
<td>0.323</td>
</tr>
</tbody>
</table>

To index the categories for the Multinomial MRF model according to the discussion in Section 4.6, we fitted three Binomial MRF models to each data set. To standardize each of the three estimates of the dependence parameter, we divided the estimates by the standard bound of the dependence parameter corresponding to the estimate of $\kappa$ since the actual value of $\kappa$ is unknown. For 1980, the category consisting of winds greater than 10 mph should be indexed as the third category while winds less than or equal to 7 mph should be indexed as the third category for 1985 and 1990. For the remainder of this section, the categories will be
Figure 5.2 Image plots of $y_k(s_i)$ for $k = 1, 2, 3$ (from left to right) for year 1980

Figure 5.3 Image plots of $y_k(s_i)$ for $k = 1, 2, 3$ (from left to right) for year 1985

Figure 5.4 Image plots of $y_k(s_i)$ for $k = 1, 2, 3$ (from left to right) for year 1990
referred to as originally defined: category 1 (k=1) for wind speeds less than or equal to 7 mph, category 2 (k=2) for wind speeds greater than 7 mph and less than or equal to 10 mph and category 3 (k=3) for wind speeds greater than 10 mph.

Table 5.2 gives the parameter estimates of $\kappa_k; k = 1, 2, 3$ and $\eta$ for 1980, 1985 and 1990. Different starting values led to the same values at convergence, which suggests that we do not have false convergence. In addition, the profile of the pseudo-likelihood was also plotted to detect the existence of a global maximum. Plotting the profile of the pseudo-likelihood begins with calculating the value of the pseudo-likelihood over a range of values for the parameters of interest. In this case, we chose a range of values for $\kappa_1$ and $\kappa_2$ for 1980 and $\kappa_2$ and $\kappa_3$ for 1985 and 1990 and held the value of $\eta$ constant. Specifically, values of the pseudo-likelihood were calculated for $\kappa_k \in \{0.01, 0.02, 0.03, \ldots, 0.99\}; k = 1, 2$, for 1980 and $\kappa_k \in \{0.01, 0.02, 0.03, \ldots, 0.99\}; k = 2, 3$, for 1985 and 1990, and a value of 0.0013 for $\eta$. Then for each value of $\kappa_1$, the maximum value of the pseudo-likelihood across all values of $\kappa_2$ was obtained and plotted versus the value of $\kappa_1$ for 1980. Similarly, for each value of $\kappa_2$, the maximum value of the pseudo-likelihood across all values of $\kappa_3$ was obtained and plotted versus the value of $\kappa_2$ for 1985 and 1990. An ideal plot of the profile of pseudo-likelihood would be similar in shape to an upside-down U. This would indicate that there is one local maximum which is also the global maximum. Since values of the pseudo-likelihood are too large for a computer to handle, the negative log of the psuedo-likelihood, $-\log (P(\theta))$, was computed and plotted. The plots in Figure 5.5 suggest there may be more than one local maximum for each of the three years; however, the plots also suggest there is a global maximum for each of the three years. As long as reasonable starting values are used, i.e., the starting values for $\kappa_k; k = 1, 2$, for 1980 and $\kappa_k; k = 2, 3$, for 1985 and 1990 are approximately equal to the respective marginal means of the original data sets and a starting value for the dependence parameter is a small positive number or 0, the iterative process should converge near the global maximum of the psuedo-likelihood.

Next we assessed whether or not the dependence parameter estimate was within its respective standard bounds. We compared the marginal means of data sets simulated as outlined in Section 5.4 to the respective marginal means of the original data set for each year. The
Figure 5.5  Profile of $-\log (P(\theta))$ for years 1980, 1985 and 1990 (top to bottom) for Model 1
Table 5.2 Parameter Estimates for Model 1

<table>
<thead>
<tr>
<th>Year</th>
<th>$\hat{\kappa}_1$</th>
<th>$\hat{\kappa}_2$</th>
<th>$\hat{\kappa}_3$</th>
<th>$\hat{\eta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980</td>
<td>0.315</td>
<td>0.332</td>
<td>0.353</td>
<td>0.00132</td>
</tr>
<tr>
<td>1985</td>
<td>0.348</td>
<td>0.317</td>
<td>0.335</td>
<td>0.00138</td>
</tr>
<tr>
<td>1990</td>
<td>0.357</td>
<td>0.314</td>
<td>0.329</td>
<td>0.00134</td>
</tr>
</tbody>
</table>

Marginal means of the simulated data sets were not approximately equal to the respective marginal means of the original data set for each year. Based on these results, we conclude that each estimate of the dependence parameter is not within its respective standard bounds.

In a Markov random field model that incorporates a natural parameter function with a centered parameterization as shown in (3.25), we desire the dependence parameter to model the small-scale structure, whereas we desire the other parameters to model the large-scale structure. If the value of the dependence parameter, which models the small-scale structure, is too large, then the sum of the neighborhood deviations dominates the leading constant in the natural parameter function, which models the large-scale structure. This leads to the small-scale structure dictating the marginal means of the categories instead of the large-scale structure. If the large-scale structure is modeled through the parameters $\kappa_k$; $k = 1, 2$, for 1980 and $\kappa_k$; $k = 2, 3$, for 1985 and 1990, then this suggests that the large-scale structure, i.e., the marginal mean, should be relatively constant across the Markov random field for each of the three categories. As Figures 5.2 - 5.4 suggest, the marginal mean of category $k$; $k = 1, 2, 3$, is not nearly constant for each of the three years presented. The large-scale structure appears to be approximately linear from north to south. Thus, Model 1 should be reparameterized to reflect the large-scale structure present in the data set. This reparameterization leads to what will be referred to as Model 2.

### 5.5.2 Multinomial MRF Model with Covariates (Model 2)

For this model, the conditional probability mass function for $y(s_i) = (y_1(s_i), y_2(s_i), y_3(s_i))^T$ given the values at the neighborhood locations, $y(N_i) = (y_1(N_i), y_2(N_i), y_3(N_i))^T$, and the vector of parameters, $\theta$, is given by (3.5) with $m_i = m$ for $i = 1, 2, ..., n$, as with Model 1. Since
the large-scale structure present in each of the data sets is approximately linear from north to south, a north-south component will be incorporated into the natural parameter function, $A_{i,k} \{y(N_i); \theta\}$, by substituting

$$\log \left( \frac{\kappa_k}{\kappa_3} \right) = \beta_{0,k} + \beta_{1,k} r_i; \quad k = 1, 2,$$

in (3.25) such that $r_i$ is the row number for location $s_i$. Then the above system of two equations is solved for $\kappa_1$ and $\kappa_2$ such that

$$\kappa_1 = \exp^{(\beta_{0,1}+\beta_{1,1} r_i)} \left/ \left( 1 + \exp^{(\beta_{0,1}+\beta_{1,1} r_i)} + \exp^{(\beta_{0,2}+\beta_{1,2} r_i)} \right) \right. \right.$$ \text{and}

$$\kappa_2 = \exp^{(\beta_{0,2}+\beta_{1,2} r_i)} \left/ \left( 1 + \exp^{(\beta_{0,1}+\beta_{1,1} r_i)} + \exp^{(\beta_{0,2}+\beta_{1,2} r_i)} \right) \right. \right..$$

The above forms are substituted into (3.25) so $m \kappa_k; \quad k = 1, 2$, is in terms of $\beta_{0,k}$ and $\beta_{1,k}; \quad k = 1, 2$, as well. We now have $\theta = (\beta_{0,1}, \beta_{1,1}, \beta_{0,2}, \beta_{1,2}, \eta)^T$ for the vector of parameters. The parameter estimates are given in Table 5.3. In addition, the parameter estimates were converted to marginal probabilities for the first row and the last (23rd) row and are given in Tables 5.4 - 5.6.

**Table 5.3 Parameter Estimates for Model 2**

<table>
<thead>
<tr>
<th>Year</th>
<th>$\hat{\beta}_{0,1}$</th>
<th>$\hat{\beta}_{1,1}$</th>
<th>$\hat{\beta}_{0,2}$</th>
<th>$\hat{\beta}_{1,2}$</th>
<th>$\hat{\eta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980</td>
<td>-0.5369</td>
<td>-0.0356</td>
<td>-0.0308</td>
<td>0.0265</td>
<td>0.00110</td>
</tr>
<tr>
<td>1985</td>
<td>-0.1337</td>
<td>-0.0183</td>
<td>0.4471</td>
<td>-0.0440</td>
<td>0.00131</td>
</tr>
<tr>
<td>1990</td>
<td>-0.0031</td>
<td>0.0096</td>
<td>0.3433</td>
<td>-0.0390</td>
<td>0.00135</td>
</tr>
</tbody>
</table>

To check for false convergence, we plotted the profile of the pseudo-likelihood. The pseudo-likelihood was calculated for values of $\beta_{0,1}$ and $\beta_{0,2}$ in the set $\{-1.0, -0.95, ..., 1.0\}$ and values of $\beta_{1,1}$ and $\beta_{1,2}$ in the set $\{-0.10, -0.095, ..., 0.10\}$. The value of the dependence parameter, $\eta$, was set equal to 0.0012 for all three years. Then for each value of $\beta_{0,1}$, the maximum value of the pseudo-likelihood across all values of $\beta_{1,1}$, $\beta_{0,2}$ and $\beta_{1,2}$ was obtained and plotted. The process was repeated for each of the other three parameters. The four plots corresponding to the four parameters, $\beta_{0,1}$, $\beta_{1,1}$, $\beta_{0,2}$ and $\beta_{1,2}$, are given in Figure 5.6 for 1980, Figure 5.7 for 1985 and Figure 5.8 for 1990. The plots in Figures 5.6 - 5.8 indicate that the pseudo-likelihood contains one local maximum, which is also the global maximum for each of the three years.
Figure 5.6  Profile of $-\log (P(\theta))$ for year 1980 for Model 2
Figure 5.7  Profile of $-\log (P(\theta))$ for year 1985 for Model 2
Figure 5.8 Profile of $-\log(P(\theta))$ for year 1990 for Model 2
Table 5.4 Marginal Probabilities for 1980

<table>
<thead>
<tr>
<th>Row Number</th>
<th>Cat 1</th>
<th>Cat 2</th>
<th>Cat 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.257</td>
<td>0.320</td>
<td>0.424</td>
</tr>
<tr>
<td>23</td>
<td>0.361</td>
<td>0.368</td>
<td>0.272</td>
</tr>
</tbody>
</table>

Table 5.5 Marginal Probabilities for 1985

<table>
<thead>
<tr>
<th>Row Number</th>
<th>Cat 1</th>
<th>Cat 2</th>
<th>Cat 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.276</td>
<td>0.310</td>
<td>0.414</td>
</tr>
<tr>
<td>23</td>
<td>0.431</td>
<td>0.324</td>
<td>0.245</td>
</tr>
</tbody>
</table>

Table 5.6 Marginal Probabilities for 1990

<table>
<thead>
<tr>
<th>Row Number</th>
<th>Cat 1</th>
<th>Cat 2</th>
<th>Cat 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.299</td>
<td>0.295</td>
<td>0.406</td>
</tr>
<tr>
<td>23</td>
<td>0.421</td>
<td>0.337</td>
<td>0.242</td>
</tr>
</tbody>
</table>

The next issue that can now be addressed is the issue regarding the size of the dependence parameter. Data sets were simulated using the parameter estimates. If the estimate of the dependence parameter is within its respective bounds, then the category counts at each location, \( s_i \), will be similar to the respective category counts of the original data set. The category counts of the simulated data sets were not approximately equal to the respective category counts of the original data set for all years. There may still be some large-scale structure that was not modeled by the covariates. If the covariates do not account for the large-scale structure correctly, the dependence may still too strong for the estimate of the dependence parameter to be within its standard bounds. In the next section, we will discuss another method that can be used to account for the large-scale structure before a Multinomial MRF model is fitted, which will hopefully lead to an estimate of the dependence parameter that is within its standard bounds for each year.

### 5.5.3 Multinomial MRF Model with Median Polish (Model 3)

Instead of accounting for the large-scale structure through the use of covariates, we can apply median polish to the data before we fit the Multinomial MRF model described in Section
5.5.1. Median polish de-trends the data without having to specify the trend in terms of parameters. Any remaining structure in the data set will be modeled by only one parameter, the dependence parameter, \( \eta \). To apply median polish to a data set, the data need to be located on a lattice (or grid). Let \( \mathbf{X}_{p,q}(s_i) = (X_1(s_i), X_2(s_i), \ldots, X_m(s_i))^T \) represent the vector of \( m = 736 \) observations at location \( s_i \) in row \( p \) for \( p = 1, \ldots, n_r \), and column \( q \) for \( q = 1, \ldots, n_c \). Let \( M_{p,q}(s_i) \) denote the median of \( \mathbf{X}_{p,q}(s_i) \). Then consider the following model,

\[
\mu_{p,q}(s_i) = a + r_p + c_q \quad \text{for} \quad p = 1, \ldots, n_r \quad \text{and} \quad q = 1, \ldots, n_c.
\]

The constants \( \{r_p : p = 1, \ldots, n_r\} \) are the row effects and the constants \( \{c_q : q = 1, \ldots, n_c\} \) are the column effects. Now define values \( M_{p,q}^{(0)}(s_i) ; p = 1, \ldots, n_r; q = 1, \ldots, n_c, r_p^{(0)} = 0; p = 1, \ldots, n_r, c_q^{(0)} = 0; q = 1, \ldots, n_c, \) and \( a^{(0)} = 0 \). A median polish algorithm proceeds as follows:

At iteration \( t = 1, 2, \ldots \)

1. \( \Delta r_p^{(t)} = med \left\{ M_{p,q}^{(t-1)}(s_i) : q = 1, \ldots, n_c \right\} ; \quad p = 1, \ldots, n_r \)
2. \( \Delta a_c^{(t)} = med \left\{ c_q^{(t-1)} : q = 1, \ldots, n_c \right\} \)
3. \( M_{p,q}^{*(t)}(s_i) = M_{p,q}^{(t-1)}(s_i) - \Delta r_p^{(t)} ; \quad p = 1, \ldots, n_r; \quad q = 1, \ldots, n_c \)
4. \( \Delta c_q^{(t)} = med \left\{ M_{p,q}^{(t-1)}(s_i) : p = 1, \ldots, n_r \right\} ; \quad q = 1, \ldots, n_c \)
5. \( \Delta a_r^{(t)} = med \left\{ r_p^{(t-1)} + \Delta r_p^{(t)} : p = 1, \ldots, n_r \right\} \)
6. \( M_{p,q}^{(t)}(s_i) = M_{p,q}^{*(t)}(s_i) - \Delta c_q^{(t)} ; \quad p = 1, \ldots, n_r; \quad q = 1, \ldots, n_c \)
7. \( a^{(t)} = a^{(t-1)} + \Delta a_r^{(t)} + \Delta a_c^{(t)} \)
8. \( r_p^{(t)} = r_p^{(t-1)} + \Delta r_p^{(t)} - \Delta a_r^{(t)} ; \quad p = 1, \ldots, n_r \)
9. \( c_q^{(t)} = c_q^{(t-1)} + \Delta c_q^{(t)} - \Delta a_c^{(t)} ; \quad q = 1, \ldots, n_c \)

The algorithm is continued (over iterations \( t \)) until the new row medians, \( r_p; p = 1, \ldots, n_r \), and column medians, \( c_q; q = 1, \ldots, n_c \), equal zero. Usually the algorithm can be run for two
or three iterations and then stopped. For this application, the new row medians and column medians are equal to zero in ten or less iterations for each of the years so we carried out the algorithm for 10 iterations. When the algorithm is stopped, for observation $x$ at location $s_i$ located in row $q$ and column $p$, the residual is calculated as

$$ R_x(s_i) = X_x(s_i) - \hat{a} - \hat{r}_p - \hat{c}_q. \quad (5.1) $$

Then these residuals are placed into categories and the Multinomial MRF model described in Section 5.5.1 is fitted. The categories are defined as follows: category 1 contains residuals that are less than or equal to -1.5, category 2 contains residuals that are greater than -1.5 and less than or equal to 1, and category 3 contains residuals that are greater than 1. This definition creates a data set such that the covariance of counts between each pair of categories for each of the three years is negative. Figures 5.9 - 5.11 graphically depict the number of the residuals $Y_k(s_i)$ in category $k$ at location $s_i$; $i = 1, ..., n$, for years 1980, 1985 and 1990. Table 5.7 contains the marginal mean of category $k$, i.e., the mean of $Y_k(s_i)$ for all locations and denoted as $\bar{Y}_k(s_i)$, for each year.

<table>
<thead>
<tr>
<th>Year</th>
<th>$\bar{Y}_1(s_i)$</th>
<th>$\bar{Y}_2(s_i)$</th>
<th>$\bar{Y}_3(s_i)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980</td>
<td>0.324</td>
<td>0.294</td>
<td>0.382</td>
</tr>
<tr>
<td>1985</td>
<td>0.328</td>
<td>0.291</td>
<td>0.381</td>
</tr>
<tr>
<td>1990</td>
<td>0.327</td>
<td>0.294</td>
<td>0.380</td>
</tr>
</tbody>
</table>

To index the categories for the Multinomial MRF model according to the discussion in Section 4.6, we fitted three Binomial MRF models to each data set. The Multinomial MRF model was fitted with the category containing residuals greater than 1 indexed as the third category for year 1980 and the category containing residuals less than or equal to -1.5 indexed as the third category for years 1985 and 1990. For the remainder of this section, the categories will be referenced as originally defined: category 1 ($k=1$) for residuals less than or equal to -1.5 mph, category 2 ($k=2$) for residuals greater than -1.5 mph and less than or equal to 1 mph and category 3 ($k=3$) for residuals greater than 1 mph. The parameter estimates are located in Table 5.8.
Figure 5.9  Image plots of $y_k(s_i)$ for $k = 1, 2, 3$ (from left to right) for year 1980

Figure 5.10  Image plots of $y_k(s_i)$ for $k = 1, 2, 3$ (from left to right) for year 1985

Figure 5.11  Image plots of $y_k(s_i)$ for $k = 1, 2, 3$ (from left to right) for year 1990
The profile of the psuedo-likelihood for each year was plotted and is shown in Figure 5.12 to detect the existence of a global maximum. Plotting the profile of the pseudo-likelihood begins with calculating the value of the pseudo-likelihood over a range of values for the parameters of interest. As with the Model 1, we chose a range of values for $\kappa_1$ and $\kappa_2$ for 1980 and $\kappa_2$ and $\kappa_3$ for 1985 and 1990 and held the value of $\eta$ constant. Specifically, values of the pseudo-likelihood were calculated for $\kappa_k \in \{0.01, 0.02, 0.03, ... , 0.99\}$; $k = 1, 2$, for 1980 and $\kappa_k \in \{0.01, 0.02, 0.03, ... , 0.99\}$; $k = 2, 3$, for 1985 and 1990, with a value of 0.0011 chosen for $\eta$. Then for each value of $\kappa_1$, the maximum value of the pseudo-likelihood across all values of $\kappa_2$ was obtained and plotted versus the value of $\kappa_1$ for 1980. Similarly, for each value of $\kappa_2$, the maximum value across all values of $\kappa_3$ was obtained and plotted versus the value of $\kappa_2$ for 1985 and 1990. According to Figure 5.12, the profiles indicate the pseudo-likelihood has a global maximum for each year.

Next we assessed whether or not the dependence parameter estimate is within its respective standard bounds. We compared the marginal means of data sets simulated as outlined in Section 5.4 to the respective marginal means of the original data set for each year. The marginal means of the simulated data sets were approximately equal to the respective marginal means of the original data set for year 1980 only. This suggests that the estimate of the dependence parameter is within its respective standard bounds for year 1980 only.

Since the data sets fitted with Model 3 were detrended first through median polish, both the large-scale and small-scale structure across locations is weaker than the large-scale and small-scale structure in the data sets fitted with Model 1. This should lead to the standardized dependence parameter estimate under Model 3 to be less than the standardized dependence parameter estimate under Model 1. The dependence parameter estimates given in Tables 5.2
Figure 5.12 Profile of $-\log(P(\theta))$ for years 1980, 1985 and 1990 (top to bottom) for Model 3
and 5.8 are not standardized; i.e., divided by its respective standard bounds. Standard bounds for the Multinomial MRF model cannot be derived analytically, and hence, we cannot compare the estimates of $\eta$ directly. However, since the standard bounds are functions of the values of $\kappa_k; k = 1, 2, 3$, we can compare the estimates to some degree since the estimates of $\kappa_k; k = 1, 2, 3$, under Model 1 are similar to the estimates of $\kappa_k; k = 1, 2, 3$, under Model 3. For a given year, we do find that the estimate of $\eta$ under Model 3 is less than the estimate of $\eta$ under Model 1. This indicates that median polish has successfully removed some of the large-scale structure but not enough in order for each dependence parameter estimate to be within its respective standard bounds for each of the three years.

Median polish has one disadvantage when using median polish to de-trend the data before fitting the data with a Multinomial MRF model. The placement of the observations into defined categories is affected by the center and shape of the distribution of wind speeds at location $s_i$. Median polish affects the center (in this case, median) of the distribution of wind speeds at each location $s_i$ but does affect the shape of the distribution. Consequently, some large-scale structure may still be present in a data set as a result of changes in the distribution of the wind speeds across locations.
CHAPTER 6 GENERAL CONCLUSION

Models that are constructed from conditionally specified distributions are being applied to data sets that possess a spatial structure, even data sets with complex dependence structures. Much attention have been given to Gaussian, Poisson and Binomial MRF models. Although the multinomial distribution is an extension of the binomial distribution, little attention has been given to the Multinomial MRF model. Hence, the primary goal of this dissertation is to construct and study the Multinomial MRF model.

In Chapter 3, we constructed the joint distribution of the Multinomial MRF model up to an unknown constant. Although the joint distribution cannot be explicitly written, Multinomial Markov random fields can be simulated with the Gibbs sampling algorithm given the conditionally specified distributions. Within the conditionally specified distributions, the natural parameter function can be parameterized in a manner such that the marginal expectations are approximately equal to the marginal expectations under independence as long as the dependence parameter is within certain bounds. Approximate bounds were determined through simulation.

In Chapter 4, we discussed the asymmetry property of the Multinomial MRF model. The Multinomial MRF model is not symmetric, unlike the Binomial MRF model, and consequently, the behavior of the model depends on which category is labeled as the last or $h^{th}$ category. Specifically, when there are three categories and the dependence parameter is within certain bounds, the variance of the conditional expectations for the third category was noted as being smaller than the variance of the conditional expectations for the other two categories. Furthermore, the covariance between the conditional expectations of the third category and the conditional expectations of one of the other two categories was often found to be positive. How-
ever, the marginal covariances were found to be always negative. Because stronger statistical
dependence is MRF models is generally related to greater variance in conditional expectations
than is weaker dependence and we noted that the variance of the conditional expectations for
the third category was the smallest of the variances of the three categories, we studied the
dependence structure within each category under a Multinomial MRF model. We found that
the dependence structure is, on average, weakest in the third category. This result provided
a method to determine which category should be labeled as the last category when fitting a
Multinomial MRF model to data set. For data sets simulated from Multinomial MRF models,
our method performed better on average than the method of randomly selecting a category to
be indexed as the last category. Although we recommend carefully selecting a category to be
indexed as the last category, choosing the wrong category as the last category appears to not
increase the mean squared error of estimates, but will slightly increase the prediction mean
squared error, especially when the dependence structure is strong.

Issues in applying the model to analyze wind speeds across Iowa was discussed in Chapter
5. When the response variable is continuous, categories need to be carefully defined so the
resulting data set possesses characteristics that are similar to characteristics of data sets simu-
lated from the Multinomial MRF model. For example, the covariances of the category counts
should be negative for all pairs of categories. If large-scale structure is present in the data,
then fitting a Multinomial MRF model will most likely lead to an estimate of the dependence
parameter that is not within its standard bounds. Either covariates could be added to the cen-
tered natural parameter function or the large-scale structure could be removed with a method
such as median polish before the Multinomial MRF model is fitted.
BIBLIOGRAPHY


ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Mark Kaiser, for his guidance and support. I would also like to thank my husband, Mark, who is a constant source of encouragement. This work was supported in part by NSF grant DMS #0502347 EMSW21-RTG awarded to the Department of Statistics, Iowa State University.