Exploring dependence in binary Markov random field models

Kenneth William Wakeland
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/etd
Part of the Statistics and Probability Commons

Recommended Citation
https://lib.dr.iastate.edu/etd/16234

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.
Exploring dependence in binary Markov random field models

by

Kenneth William Wakeland

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
Daniel Nordman
Petruta Caragea
Jarad Neimi
Vivekananda Roy

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this dissertation. The Graduate College will ensure this dissertation is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2017

Copyright ©Kenneth William Wakeland, 2017. All rights reserved.
# TABLE OF CONTENTS

## LIST OF TABLES

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>iv</td>
</tr>
</tbody>
</table>

## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>v</td>
</tr>
</tbody>
</table>

## ABSTRACT

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>v</td>
</tr>
</tbody>
</table>

## 1 GENERAL INTRODUCTION

1.1 Markov Random Fields ........................................ 1  
1.2 Dissertation Organization ................................. 2  
1.3 Bibliography ................................................. 3

## 2 MODELING WATER VAPOR FLUX DIVERGENCE WITH MARKOV RANDOM FIELDS OVER TIME

2.1 Introduction ................................................... 4  
2.2 Model Development ............................................ 5  
  2.2.1 Static Binary random field ............................. 7  
  2.2.2 Markov random fields over time ....................... 9  
2.3 Estimation and Inference ..................................... 11  
2.4 Water Vapor Flux Divergence Data .......................... 15  
2.5 Conclusions .................................................. 18  
  2.5.1 Future Work ............................................. 20  
2.6 Bibliography ................................................. 22

## 3 NEGATIVE DEPENDENCE IN BINARY MARKOV RANDOM FIELDS

3.1 Introduction .................................................. 24  
3.2 Binary MRF Formulation ....................................... 26  
3.3 Data Patterns in MRF Models ................................. 28  
  3.3.1 Models with Positive Spatial Dependence .......... 28  
  3.3.2 Degeneracy in Binary MRF Models .................... 29  
  3.3.3 Models with Negative Spatial Dependence .......... 30  
3.4 Quantifying Data Patterns ................................... 32  
  3.4.1 Conflicts in Realizations of MRF Models .......... 32  
  3.4.2 Distribution of Run Lengths ......................... 34  
3.5 Directional Dependence ....................................... 36  
  3.5.1 Conflicts in Directional MRF Models ............... 37  
  3.5.2 Distribution of Run Lengths in Directional MRF models 39
4 CONDITIONAL EXPECTATIONS IN BINARY MARKOV RANDOM FIELDS

4.1 Introduction ............................................. 63
4.2 Model and Neighborhood Formulation ............... 65
  4.2.1 Model Formulation ................................ 65
  4.2.2 Neighborhood Formulation .................... 67
4.3 Conditional Expectation Fields ....................... 68
  4.3.1 Patterns in Conditional Expectation Fields .... 68
  4.3.2 Frequency of Conditioning Sets ............... 71
  4.3.3 Empirical Conditional Expectation .......... 72
4.4 Measures of Heterogeneity and Scale ............... 74
  4.4.1 Scale of the Conditional Expectation Field .... 75
  4.4.2 Measure of Heterogeneity .................. 76
4.5 Water Vapor Flux Reanalysis Data .................. 78
4.6 Conclusions ............................................ 79
4.7 Bibliography ........................................... 80
Appendix: Tables and Figures ............................ 82

5 GENERAL CONCLUSIONS

5.1 Summary ............................................... 87
5.2 Future Work ......................................... 89

APPENDIX. SUPPLEMENTAL INFORMATION 90
LIST OF TABLES

Table 2.1 Model Parameter Estimates obtained from the double MH algorithm applied to the water vapor flux divergence data .............................................. 20

Table 3.1 Probabilities of configurations of 4 points on a transect with positive dependence .......................................................... 45

Table 3.2 Probabilities of configurations of 4 points on a transect with negative dependence .......................................................... 46

Table 3.3 Table showing the relative frequency of the number of conflicts in a clique of size 4 .......................................................... 46

Table 3.4 Quantile of Run Lengths for rows and columns of two realizations of two MRF models .......................................................... 47

Table 3.5 Expected number of conflicts in each direction for 2 directional MRF models .......................................................... 48

Table 3.6 Mean and lag 1 autocorrelations for pseudo-likelihood estimates for Water Vapor Flux Divergence data .............................. 48

Table 4.1 The APLE for data and conditional expectation fields originating from various MRF models ........................................... 82

Table 4.2 Summary Statistics for Conditional Expectation Fields ........................................... 82

Table 4.3 Frequency distributions of the conditioning sets for 3 MRF models ........................................... 83

Table 4.4 Simple Linear Regression of $\hat{E}_i$ against $E_i$ by model ........................................... 84

Table 4.5 Summary Statistics for Scale ($b_t^*$) and heterogeneity ($H(b_t^*)$) ........................................... 84

Table A1 Monte Carlo estimates of the distribution of 4 interior points of a MRF with 4 nearest neighbors on a 10x10 lattice along with Monte Carlo standard errors ........................................... 93

Table A2 Monte Carlo estimates of the proportion of each possible conflict within a field for increasing $\eta$. The Monte Carlo errors are in parenthesis ........................................... 94
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Simple neighborhood structures on a regular lattice.</td>
<td>49</td>
</tr>
<tr>
<td>3.2</td>
<td>4 nearest neighbor models with $\kappa = 0.5$, and top-left to bottom right $\eta = 0.5, 1, 2, 4$</td>
<td>50</td>
</tr>
<tr>
<td>3.3</td>
<td>4 nearest neighbor models with $\kappa = 0.5$, and top-left to bottom right $\eta = -0.5, -1, -2, -4$</td>
<td>51</td>
</tr>
<tr>
<td>3.4</td>
<td>Empirical marginal means of fields with various $\kappa$ values and $\eta$ ranging from 0 to -20</td>
<td>52</td>
</tr>
<tr>
<td>3.5</td>
<td>Left: 4 nearest neighbors. Right: 8 nearest neighbors</td>
<td>52</td>
</tr>
<tr>
<td>3.6</td>
<td>Example patterns of cliques of size 4</td>
<td>53</td>
</tr>
<tr>
<td>3.7</td>
<td>Number of conflicts in a 3x3 field where each clique of size 4 has 2 conflicts</td>
<td>54</td>
</tr>
<tr>
<td>3.8</td>
<td>8 nearest neighbor models with $\kappa = 0.5$ and $\eta = -100$</td>
<td>54</td>
</tr>
<tr>
<td>3.9</td>
<td>Mean of Monte Carlo distributions for runs in rows of lengths 1 to 4. Top 2 plots are for four nearest neighbors, bottom 2 are for 8 nearest neighbors</td>
<td>55</td>
</tr>
<tr>
<td>3.10</td>
<td>Mean of Monte Carlo distributions for runs in columns of lengths 1 to 4. Top 2 plots are for four nearest neighbors, bottom 2 are for 8 nearest neighbors</td>
<td>56</td>
</tr>
<tr>
<td>3.11</td>
<td>Box Plot of $p_r$ values by Run Length and Dependence Strength</td>
<td>57</td>
</tr>
<tr>
<td>3.12</td>
<td>Realizations from a directional MRF model with fixed dependence in the rows and varying dependence in the columns</td>
<td>58</td>
</tr>
<tr>
<td>3.13</td>
<td>Mean of Monte Carlo distributions of run lengths for four 4 nearest neighbor MRF models with directional dependence</td>
<td>59</td>
</tr>
<tr>
<td>3.14</td>
<td>Realizations from a directional MRF model with fixed dependence in the rows and varying dependence in the columns</td>
<td>60</td>
</tr>
<tr>
<td>3.15</td>
<td>Locations used in the calculation of Water Vapor Flux Divergence data</td>
<td>61</td>
</tr>
<tr>
<td>3.16</td>
<td>Plot of the $p_r$ values from the thresholded water vapor flux divergence data</td>
<td>62</td>
</tr>
<tr>
<td>4.1</td>
<td>The left column displays contour maps for three realizations of MRF models, the right column is a corresponding slice of the surface.</td>
<td>85</td>
</tr>
<tr>
<td>4.2</td>
<td>Plot of variance between block means for block sizes $b = 3, \ldots, 100$</td>
<td>86</td>
</tr>
<tr>
<td>A1</td>
<td>Box plots for Gibbs and Perfect Samplers</td>
<td>95</td>
</tr>
</tbody>
</table>
Many problems in ecology and the environmental sciences, such as monitoring the presence/absence of a species, involve the observation of spatial binary random fields. Markov random field (MRF) models are commonly used to analyze data of this kind. It is becoming more common for these studies to include a time component as well. Markov random field models can be modified to incorporate temporal dependence whether the dependence is on a local level or through a global effect. However, it is important when working with MRF models to ensure the spatial dependence is properly specified. Over the course of three papers, this dissertation explores binary MRF models with an emphasis on arriving at an appropriate model for binary fields observed over time. In the first paper, we explore options to incorporate temporal dependence in MRF models. Our attempts to apply these to data resulted in unexpected results, namely negative spatial dependence estimates. In the second paper, we examine what negative spatial dependence means in the context of binary MRF models. We develop the run length distribution as a tool that can be used to diagnose the strength and direction of the dependence by examining data patterns in realized fields. In the third paper, we shift our focus to exploring the effect of negative dependence on the conditional expectations of MRF models. We develop the empirical conditional expectation as a method of examining the conditional expectation of binary fields in which we do not have knowledge of the true parameters used to generate the field (i.e. real data). The scale and heterogeneity of the empirical conditional expectation fields are statistics we introduce to assess the effect negative dependence has on the conditional expectation field.
CHAPTER 1. GENERAL INTRODUCTION

The motivation behind the work for this dissertation came from the models Christopher Wikle and Noel Cressie presented in their book “Statistics for Spatio-Temporal Data.” Many of the models used in this book leverage knowledge of the underlying process that was used to generate the data. Several models use partial differential equations that are known to describe or dictate how the species of animal or physical process behave. My goal with this work was to attempt to model data in which such knowledge was not available. That is, if there was no previous knowledge of how the data were generated, and we only suspected the possibility of spatial and temporal dependence, which model would we fit and how would we judge the appropriateness of this model fit? I chose from the start to focus on binary data, specifically binary data in which Markov random field models can be applied.

However, when attempting to apply these spatio-temporal Markov random field models to a real data example, we discovered the spatial structure at each time point was more complicated than what was assumed in the model. Specifically, we obtained negative spatial dependence parameter estimates. While it is generally known negative dependence can exist for binary Markov random fields (Besag, 1974), an in depth consideration of the effects this has on realizations from such models has not been performed. It was for this reason we decided to explore the effects of negative dependence in order to, ultimately, arrive at an appropriate model for the data.
1.1 Markov Random Fields

The Markov random field (MRF) models we consider in this dissertation have their root in the often cited paper by Julian Besag (Besag, 1974). In this paper, Besag provides a general formulation for MRF models from the exponential family class. This included both the joint and conditional specification for autoregressive models that he refers to as auto-normal, auto-logistic, auto-binomial and auto-Poisson. Kaiser and Cressie in (Kaiser and Cressie, 2000) provide a more general proof for the construction of these joint distributions as well as conditions in which the distributions exist. An updated parameterization for these autoregressive models is provided by (Kaiser, Caragea, and Furukawa, 2012). This parameterization uses centering to provide an interpretation of the spatial dependence which has less effect on the large scale structure parameter, provided the spatial dependence is not too large.

Estimation of these models can be done using a pseudo-likelihood method in which the joint distribution is approximated using a product of the conditional distributions for each point in the field (Besag, 1974). Such methods are necessary because of an intractable normalizing constant present in the joint distribution. Monte Carlo methods have also been developed to either estimate this normalizing constant (Geyer and Thompson, 1995) in order to use maximum likelihood methods or canceling it out (Møller et al., 2006) (Liang, 2010) for the use in Bayesian methods. The Bayesian methods require the use of an auxiliary field in order to cancel the normal constant. In (Møller et al., 2006), the auxiliary field had to be a "perfect sample", which can be obtained using a coupling in the past algorithm (Propp and Wilson, 1996). In (Liang, 2010), the auxiliary field is obtained using a one step Gibbs sample with the data as the starting point, which requires much less computation and therefore has can be used for fields contain many more points than the methods that require a perfect sample (Hughes, Haran, and Caragea, 2011).
The Bayesian methods for estimating parameters in the MRF models allow for the use of hierarchical techniques to incorporate temporal dependence. That is, incorporating a temporal process on the parameters of a MRF model for a series of fields occurring over time. The Bayesian techniques also allow for comparisons of the hierarchical modeling technique to a point-wise method of incorporating temporal dependence in MRF models (Zhu, Huang, and Wu, 2005) (Zhu et al., 2008) (Zheng and Zhu, 2012). For these reasons, we primarily use Bayesian methods when fitting MRF models to data. We also develop statistics that can be used with Bayesian methods as posterior predictive quantities.

1.2 Dissertation Organization

The next three chapters of this thesis represent work to be submitted for publication in peer-reviewed journals. In Chapter 2, I discuss both the specific spatio-temporal models I developed and worked with and the data I application I used for those models. There were complications in applying this models to this data however. These complications involved modeling the spatial dependence in the individual fields. These complications required us to more closely examine the effects of the spatial dependence parameters in binary Markov random field models.

Chapter 3 describes the first of these considerations concerning the dependence parameters of binary Markov random field models. This chapter focuses primarily on the effects of negative dependence on the patterns in the data realizations. Two concepts are developed here to assess data patterns in realizations from Markov random fields, the first is conflicts and the second is the run length distribution.

In Chapter 4, we consider the patterns negative dependence parameters induce on the conditional expectations at the locations of the binary Markov random field models. Such patterns provide a different, though important, view on the effect of negative dependence on MRF models. We develop methods to quantify the patterns in the conditional expectation
surface, as well as a method of finding the conditional expectation for a location without explicit knowledge of the MRF model.

In Chapter 5, we present overall conclusions as well as avenues of future research. This future work will explore extending spatial neighborhoods in order to capture the behavior observed from the real data as seen in Chapters 3 and 4.

1.3 Bibliography


CHAPTER 2. MODELING WATER VAPOR FLUX DIVERGENCE WITH MARKOV RANDOM FIELDS OVER TIME

A paper in preparation
Kenneth Wakeland and Mark S. Kaiser

Abstract

Any number of problems in ecology and the environmental sciences, such as monitoring the presence/absence of a species, involve the observation of spatial binary random fields at a sequence of points in time. There is often insufficient information about the scientific processes involved to incorporate a deterministic component for time evolution into a model. We consider Markov random field models with binary conditional distributions that include a stochastic evolution over time based on autoregressive structure for the large-scale model component. These models retain the flexibility of static Markov random field models for representation of spatial dependence in the small-scale model component. Bayesian estimation is accomplished through the use of what has been called the 'double Metropolis algorithm' which requires generation of auxiliary random fields, but does not require the use of perfect sampling.
2.1 Introduction

Markov random field (MRF) models formulated on the basis of conditional binary distributions have been used in any number of spatial problems that involve modeling the presence or absence of events. Also called autologistic models, these models are typically specified for sets of finite spatial locations on regular or irregular lattices. Some problems involve observation of events over time on a common spatial domain, such as occurs in environmental monitoring or disease detection programs (Zhu, Huang, and Wu, 2005) (Zheng and Zhu, 2012) (Zhu et al., 2008). Another example of observations of events over time on a common spatial domain comes from temporal sequences of meteorological variables observed over a fixed spatial domain. In this setting, we may have interest in the exceedance of a given threshold such as positive versus negative water vapor flux divergence in the Midwestern United States (Hobbs, 2014). When a spatial lattice is observed at multiple points in time, we may wish to incorporate temporal structure as well as spatial structure in a model. In this article, we examine several model structures that might be used to achieve this goal.

The autologistic model for spatial binary data (Besag, 1974) imposes a joint Markov random field for the binary data and has proved to be quite useful in analyzing these kind of data. This kind of model incorporates a relatively simple parameter for the spatial dependence. Modifications on this original method have been introduced by (Kaiser, Caragea, and Furukawa, 2012) which make use of a centered parameterization. This parameterization allows for easier interpretation of the model parameters. Both Besag’s original parameterization and the centered parameterization can be fit using the maximum pseudo-likelihood method detailed in (Besag, 1974). Other inference methods have been introduced for binary MRF including (but not limited to) Monte Carlo maximum likelihood (Geyer and Thompson, 1995), MCMC with perfect samples (Møller et al., 2006) and MCMC using a double Metropolis-Hasting algorithm (Liang, 2010). These methods have been used with great success in fitting the binary MRF model on a lattice. In this chapter, we will focus on a Bayesian
approach using the double Metropolis-Hastings algorithm of (Liang, 2010). This approach had the best results in terms of computation time and model comparisons we found when fitting the different temporal structures.

In some situations, it is natural to model structure through time using temporal statistical dependence such as autocorrelation or dynamic processes. There are several techniques by which temporal dependence can be incorporated into binary MRF models for sequences of spatial random fields over time. We consider three devices to incorporate dependence through time in sequences of binary MRF. One of these is based on an extension of traditional MRF model neighborhoods to include both spatial and temporal dimensions. This maintains the overall MRF model structure. Another approach for incorporating temporal structure is to use time-ordered dependencies for each spatial location, typically in the form of a first-order Markov property in time. This results in autoregressive structure in a point-wise manner for individual spatial locations. Both of these approaches, extending the neighborhoods and modeling point-wise dependencies over time represent temporal dependence at the level of individual spatial locations. An alternative to this is to use dynamic structures for the large-scale components of sequences of spatially structured Markov random fields.

The goal was to compare the three approaches of the previous paragraph when applied to the water vapor flux divergence data described in (Hobbs, 2014) using various posterior predictive quantities. However, the parameter estimates obtained from the models when applied to this data forced us to change direction and perform a more in depth consideration of the properties of the spatial dependence in a MRF model.

The remainder of the chapter is structure as follows. In section 2, we discuss the specific formulation of each of the models presented above as well as notation that we use throughout the rest of the paper. Section 3 presents the estimation and inference method used to fit the models to the data. In section 4, we present the water vapor flux divergence data along with the complications in our efforts to fit the models to these data. The final section we present the future research inspired by these complications.
2.2 Model Development

In this section we introduce three models for incorporating temporal structure into sequences of binary random fields. We first provide some background on Markov random fields.

Let $Y_i$ denote a random variable and $y(N_i)$ the values of "neighboring" random variables according to some pre-determined neighborhood structure (e.g., four-nearest neighbors on a regular lattice). Assume that the distribution of $Y_i$ given all other random variables in the domain of interest is the same as the distribution of $Y_i$ given only the values in $y(N_i)$. Let $\{N_i : i = 1, \ldots, n\}$ denote spatial neighborhoods and $y(N_i) = \{y(s_j) : s_j \in N_i\}$ denote values of the random variable at those locations. Further assume that the conditional distributions can be described as the Markov random field version of a one parameter exponential family (Besag, 1974)

$$f(y(s_i)|N_i) = \exp [A_i\{y(N_i)\} - B_i\{y(N_i)\} + C_i\{y(s_i)\}].$$  \hspace{1cm} (2.1)

For (2.1) to describe a binary probability mass function

$$B\{y(N_i)\} = \log(1 + \exp [A_i\{y(N_i)\}]),$$  \hspace{1cm} (2.2)

$$C_i\{y(s_i)\} = 0.$$

We will formulate a standard binary Markov random field for a single point in time before introducing structures that incorporate a temporal component.
2.2.1 Static Binary random field

In this section, let \( s_i = (u_i, v_i), \ i = 1 \ldots n \) denote spatial locations consisting of a horizontal coordinate \( u_i \) and a vertical coordinate \( v_i \). Define binary random variables \( \{Y(s_i) : i = 1, \ldots, n\} \) at locations \( s_i \).

In equation (2.1), \( A_i \) is the natural parameter function. It is the natural parameter for a conditional binary exponential family. If the maximal clique size is two, or if an assumption of pairwise only dependence has been made (Besag, 1974), a general formulation for \( A_i \) is

\[
A_i(y(N_i)) = \logit(\kappa_i) + \sum_{s_j \in N_i} \eta_{i,j}(y(s_j) - \kappa_j).
\]

(2.3)

Expression (2.3) uses the centered parameterization of (Kaiser, Caragea, and Furukawa, 2012). The parameters \( \kappa_i \) are approximately the marginal expected value for each point when \( \eta_{i,j} \) is not too large. The spatial dependence parameters \( \{\eta_{i,j} : i, j = 1, \ldots, n\} \) capture all the statistical dependence incorporated into the model.

Letting \( y_{-i} = \{y(s_j) : j \neq i\} \) expression (2.3) can be written as

\[
A_i(y_{-i}) = \logit(\kappa_i) + \sum_{j=1}^{n} \eta_{i,j}\{y(s_j) - \kappa_j\},
\]

(2.4)

where \( \eta_{i,i} = 0, \eta_{i,j} = \eta_{j,i}, \) and \( \eta_{i,j} = 0 \) unless \( s_j \in N_i \). This form lends itself to the expression of the negpotential function as

\[
Q(y) = \sum_{i=1}^{n} y(s_i, t) \left[ \log \left( \frac{\kappa_i}{1 - \kappa_i} \right) - \sum_{j \neq i} \eta_{i,j}\kappa_i \right] + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \eta_{i,j}y(s_i)y(s_j),
\]

(2.5)

and the joint probability mass function is

\[
f(y) = \frac{\exp[Q(y)]}{\sum_{\omega \in \Omega} \exp[Q(\omega)]}.
\]

(2.6)
With the formulation of a binary Markov random field at a single time point finished, we move onto models that incorporate temporal structures.

### 2.2.2 Markov random fields over time

In this section we discuss how we incorporate time into the Markov random field model introduced previously. We do this in three ways, the first is with extended neighborhoods, the second is a model with dynamic parameters, and the third is a model with point-wise dependence on the past.

#### Extended Neighborhoods

In the first model, time is incorporated in a manner that retains the structure of a single Markov random field for the entire collection of random variables involved in a problem. To do this, we will need to modify our notation. Let \( s_i = (u_i, v_i, t_i) \) where \( u_i \) and \( v_i \) are as above, and \( t_i \) is the time component. Let \( M \) denote the number of spatial coordinates at a specific time. Let \( T \) denote the total number of time points, and \( n = MT \) denote the total number of random variables in the model. We can then define \( N_{i,s} \) as the spatial neighborhood, exactly the same as before for a fixed time point, and let \( N_{i,t} = (s_j : u_j = u_i, v_j = v_i, t_j = t_i \pm N) \).

As mentioned previously, \( y(N_{i,s}) \) and \( y(N_{i,t}) \) are the values of the random variables at the neighboring locations. The neighborhood of the random variable \( y(s_i) \) can be written as \( N_i = N_{i,s} \cup N_{i,t} \).

In this case, the model for the Markov random field becomes

\[
A_{i,t}\{y(N_i)\} = \logit(\kappa_{i,t}) + \sum_{s_j \in N_{i,s}} \eta_{i,j}(y(s_j) - \kappa_{j,t}) + \sum_{s_j \in N_{i,t}} \xi_{i,j}(y(s_i) - \kappa_i). \tag{2.7}
\]

The \( \xi_{i,j} \) parameter captures the statistical dependence in the temporal dimension. As for static binary MRF models, we specify that \( \eta_{i,i} = 0, \eta_{i,j} = \eta_{j,i} \) and \( \eta_{i,j} = 0 \) if \( s_j \not\in N_{i,s} \), but
we also specify $\xi_{i,i} = 0$, $\xi_{i,j} = \xi_{j,i}$ and $\xi_{i,j} = 0$ if $s_j \not\in N_{i,t}$. Then the neg-potential function for this model may be given as

$$Q(y) = \sum_{i=1}^{n} y(s_i) \left[ \log \left( \frac{\kappa_i}{1 - \kappa_i} \right) - \sum_{j \neq i} \eta_{i,j} \kappa_i - \sum_{j \neq i} \xi_{i,j} \kappa_i \right]$$

(2.8)

$$+ \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \eta_{i,j} y(s_i) y(s_j) + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \xi_{i,j} y(s_i) y(s_j).$$

Models with Dynamic Large Scale Components

With these models, we will need to switch notations slightly in order for ease in specifying models. Namely, let $s_i = (u_i, v_i)$ where $u_i$ and $v_i$ are as defined previously. Now we define random variables $y(s_i, t)$ for $i = 1, \ldots, n$, $t = 1, \ldots, T$. $N_i$ specifies the spatial location of an individual field at a fixed time point. Then $y(N_i, t)$ are the values of the neighboring locations at time $t$.

We first consider the model for the dynamic MRF. Since we are now considering a sequence of binary random fields, the form of equation (2.3) changes slightly, but the distinctions are important

$$A_{i,t}(y(N_i, t)) = \text{logit}(\kappa_{i,t}) + \sum_{s_j \in N_i} \eta_{i,j}(y(s_j, t) - \kappa_{j,t}).$$

(2.9)

The above equation (2.9) implies we are assuming the spatial dependence is "constant" across the fields, while the $\kappa_{i,t}$ is dynamic. The specific dynamic structure we will use in our data examples is AR(1),

$$\text{logit}(\kappa_{i,t}) = \mu_i + \rho(\text{logit}(\kappa_{i,t-1}) - \mu_i) + \epsilon_{i,t}. \quad (2.10)$$

It follows then the new negpotential function for each field is given by

$$Q(y_i) = \sum_{i=1}^{n} y(s_i, t) \left[ \log \left( \frac{\kappa_{i,t}}{1 - \kappa_{i,t}} \right) - \sum_{j \neq i} \eta_{i,j} \kappa_{i,t} \right] + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \eta_{i,j} y(s_i, t) y(s_j, t). \quad (2.11)$$
This model is useful if we want to model a "global" dependence over time. That is, we know the Markov random fields are not independent, but we have little reason to believe each location depends directly on the time point before it. It is important to note the joint likelihood for all the fields does exist. This is due to the fact the conditional likelihood for each field is well defined.

The second model of this kind is what we refer to as the Point-wise dependence model. This model requires a sequence of conditionally specified Markov random fields, similar to the dynamic model. However, in this model the temporal dependence acts on each spatial location, similar to the model given by (2.7). The natural parameter function, \( A_{i,t} \), for this model is given by

\[
A_{i,t}\{y(s_i, t)\} = \logit(\kappa_{i,t}) + \sum_{s_j \in N_i} \eta_{i,j}(y(s_j, t) - \kappa_{j,t}) + \eta_{\rho}(y(s_i, t - 1) - \kappa_{i,t-1}). \tag{2.12}
\]

The negpotential for this model is given by

\[
Q(y_t) = \sum_{i=1}^{n} y_t(s_i) \left[ \log \left( \frac{\kappa_{i,t}}{1 - \kappa_{i,t}} \right) + \eta_{\rho}(y_{t-1}(s_i) - \kappa_{i,t-1}) - \sum_{j \neq i} \eta_{i,j} \kappa_{i,t} \right] + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \eta_{i,j} y_t(s_i) y_t(s_j). \tag{2.13}
\]

### 2.3 Estimation and Inference

Estimation and inference of binary MRF can be difficult. In equation (2.6) we gave the form of the joint probability mass function for a binary MRF. The normalizing constant of this PMF \( \sum_{\omega \in \Omega} \exp[Q(\omega)] \) has \( 2^n \) terms, where \( n \) is the number of locations in the MRF. This makes the normalizing constant intractable for even very modest sizes of MRF. This means the estimation techniques for these models revolve around trying to remove this normalizing constant, or estimate it without directly calculating it. The first method of dealing
with the intractable normalizing constant is maximum pseudo-likelihood. This has been the most used method of fitting MRF for both Besag’s original parameterization as well as the centered parameterization of Kaiser and Caragea. Maximum pseudo-likelihood uses a product of conditional distributions to approximate the joint distribution of the MRF. Confidence intervals and standard errors are found using parametric bootstrap. The statistical efficiency of the estimates is not as good when compared to Monte Carlo maximum likelihood or Bayesian Methods (Hughes, Haran, and Caragea, 2011). Monte Carlo maximum likelihood (MCML) uses Monte Carlo samples to estimate the normalizing constant (Geyer and Thompson, 1995). Unlike maximum pseudo-likelihood, exact errors for the parameters can be calculated (Hughes, Haran, and Caragea, 2011). The Bayesian methods for estimating MRF are the newest (Møller et al., 2006) (Liang, 2010). These techniques use auxiliary fields to cancel the normalizing constants, allowing the use of the un-normalized likelihood. The technique we will focus on for this article is Liang’s Double Metropolis Hastings algorithm (Liang, 2010).

The double Metropolis-Hasting algorithm involves using two Metropolis-Hastings updates. The first MH sample uses the prior distribution as the target. The other update uses a MH update to obtain an auxiliary field. This can be achieved by using a one step Gibbs update with the observed data as the starting point. A basic formulation of the algorithm is as follows:

1. Simulate a new sample $\theta^*$ from prior distribution $\pi(\theta)$ using the MH algorithm starting at $\theta_t$.

2. Generate an auxiliary field, $\mathbf{x}$, using $\theta^*$ using a MH update or a one step Gibbs sample using the data as the starting point.

3. Accept $\theta^*$ with probability $\min\{1, r(\theta_t, \theta^*, \mathbf{x} | \mathbf{y})\}$ where

$$r(\theta, \theta^*, \mathbf{x} | \mathbf{y}) = \frac{\pi(\theta^*)q(\theta | \theta^*, \mathbf{y})g(\mathbf{x} | \theta)g(\mathbf{y} | \theta^*)}{\pi(\theta)q(\theta^* | \theta, \mathbf{y})g(\mathbf{y} | \theta^*)g(\mathbf{x} | \theta^*)}$$

(2.14)
4. Set $\theta_{t+1} = \theta^*$ if the double MH update is accepted in step (b), and set $\theta_{t+1} = \theta_t$ otherwise.

In the above equation (2.14), $g(\cdot | \cdot)$ is the un-normalized likelihood of the MRF. More simply, it is the numerator from (2.6). The MH transition distribution is represented by $q(\cdot | \cdot, \cdot)$.

The reason we use the double MH algorithm to fit these models is it is much simpler (and more general) to implement than a perfect sampler. The double MH algorithm only needs to perform a one step Gibbs update in order to obtain an auxiliary field, rather than general a perfect sample of the field. The perfect sampling method by (Møller et al., 2006) also requires an estimate of $\theta$, most commonly taken to be a pseudo-likelihood estimate, in order to simulate the auxiliary fields. The double MH algorithm does not require that step. A Bayesian procedure was preferred for ease in model comparison, such as posterior predictive methods.

## 2.4 Water Vapor Flux Divergence Data

Water vapor flux divergence is an element of the water vapor equation and has been used to predict rainfall associated with synoptic-scale systems, as a key component in the Kuo convective parameterization scheme, and in the prediction of severe local storm prediction (Banacos and Schultz, 2005). There are different methods for calculating water vapor flux divergence (convergence), in this application, the water vapor flux divergence (convergence) is computed from reanalysis products. Specifically, the reanalysis data from 2 meteorological models: the Modern-Era Retrospective Analysis for Research and Applications (MERRA) and the North American Regional Reanalysis (NARR) are used to calculate the water vapor flux divergence. Both meteorological models were computed at 3 hour intervals for a month at a time. The final data set is calculated using both of these meteorological models over a $33 \times 33$ lattice covering a region of the central United States. Further details about the exact calculation of these data can be found in (Hobbs, 2014).
The water vapor flux divergence data at each of these locations are continuous and consisting of both positive and negative values. Positive values indicate a convergence of water vapor and negative values indicate a divergence. It is useful to know which phenomenon (convergence or divergence) is happening at a location, which leads to the choice to threshold these data at 0. Specifically, let \( Z(s_i, t) \) be the water vapor flux divergence value at location \( s_i \) and time \( t \), then we define \( Y(s_i, t) = I[Z(s_i, t) \geq 0] \), which results in a series of binary fields on a 33 x 33 lattice. The specific data we use is from May 2006. This means there are 248 binary fields that we can analyze using our methods. However, for this example, we are restricting our data to the first 50 fields. Our goal with this data is to fit a spatio-temporal model using all 248 fields using the double Metropolis-Hastings method described above.

For each of the models described in Section 2.2, we initially will consider a simple spatial dependence structure given by

\[
\eta_{i,j} = \begin{cases} 
\eta & \text{if } s_j \in N_i \\
0 & \text{otherwise}
\end{cases}.
\] (2.15)

At the start, we will consider a simple four nearest neighbors structure on a lattice given by

\[
N_i = \{ s_j : u_j = u_i \pm 1, v_j = v_i \} \cup \{ s_j : u_j = u_i, v_j = v_i \pm 1 \},
\] (2.16)

which we will use for the spatial dependence in each of the models. Based on results from (Hobbs, 2014), we incorporated large scale structure by dividing the field into four parts and estimated a mean for each quadrant. That is
\[ \kappa_{i,t} = \begin{cases} 
\alpha_t + \kappa_1 & \text{if } u_{i,t} < 17, v_{i,t} < 33 \\
\alpha_t + \kappa_2 & \text{if } u_{i,t} \geq 17, v_{i,t} < 33 \\
\alpha_t + \kappa_3 & \text{if } u_{i,t} < 17, v_{i,t} \geq 33 \\
\alpha_t & \text{if } u_{i,t} \geq 17, v_{i,t} \geq 33 
\end{cases}, \quad (2.17) \]

where \( \alpha_t \) has an AR(1) structure (2.10) for the Dynamic MRF model, and is a constant \( \alpha_t = \alpha \) for the Point-wise Dependence and Extended neighborhoods model. This leads to the following forms of the natural parameter functions:

**Extended Neighborhoods model:**

\[
A_{i,t}\{y(N_i)\} = \logit(\kappa_{i,t}) + \sum_{s_j \in N_{i,s}} \eta(y(s_j) - \kappa) + \sum_{s_j \in N_{i,t}} \xi_{i,j}(y(s_i) - \kappa), \quad (2.18)
\]

**Point-wise Dependence:**

\[
A_{i,t}\{y(s_i, t)\} = \logit(\kappa_{i,t}) + \sum_{s_j \in N_i} \eta(y(s_j, t) - \kappa) + \eta(\rho(y(s_i, t-1) - \kappa), \quad (2.19)
\]

**Dynamic MRF Model:**

\[
A_{i,t}\{y(N_i, t)\} = \logit(\kappa_{i,t}) + \sum_{s_j \in N_i} \eta(y(s_j, t) - \kappa_t), \quad (2.20)
\]

where \( \kappa_{i,t} \) is defined in (2.17).

In the case of the Extended Neighborhoods model, we assume \( \xi_{i,j} \) follows a similar convention to the spatial dependence (2.15). That is

\[
\xi_{i,j} = \begin{cases} 
\xi & \text{if } s_j \in N_{i,t} \\
0 & \text{otherwise} 
\end{cases}, \quad (2.21)
\]
where $N_{i,t}$ is defined as in section 2.2.2. The prior distributions for the first 2 cases (Extended Neighborhoods and Point-wise Dependence models) are: $\kappa \sim Unif(0, 1); \eta \sim N(0, 4); \eta_\rho \sim N(0, 4); \text{ and } \xi \sim N(0, 4)$. The prior distributions for the Dynamic MRF model are: $\mu \sim N(0, 4); \rho \sim U(-1, 1); \sigma \sim U(0, 4); \eta \sim N(0, 4)$.

In order to test our implementation of the algorithm described in the previous section, we simulated a set of binary fields with properties similar to the water vapor flux divergence data. That is, we simulate 50, 33x33 fields with an overall field mean estimated from the data and in the case of the Dynamic MRF model, estimates of $\rho$ and $\sigma^2$ as well. We varied the values of the spatial dependence for each model, however, we also varied the values of the temporal dependence for the Point-wise dependence and Extended Neighborhoods model as there is not a simple way of estimating this quantities. For each of these cases, the prior distributions were the same as described in the preceding paragraph, and we were able to recover the correct model parameters using the appropriate model. With this in hand, we used each of these models to fit the water vapor flux divergence data, the resulting estimates are presented in Table 2.1. These results illustrate a possible source of concern. That is, in each model, the spatial dependence is somewhat large, above the threshold mentioned in (Caragea and Kaiser, 2009). As concerning as this is, the spatial dependence parameters for the Point-wise dependence model and the Extended Neighborhoods model are also large, and even more troubling, the means do not match up between the models. In the Dynamic MRF model, the mean of the AR(1) process is estimated to be just below 0.5, with a slight effect in the northeast quadrant. In the Point-wise dependence model, each quadrant has a mean that is nearly exactly 0.5. In the Extended Neighborhoods model however, the mean for each quadrant is estimated to be approximately 0.94, which is considerably different from the marginal mean of each field. This disparity between the marginal mean and the $\kappa$ parameter can be a symptom of MRF models approaching degeneracy, and a common method to address this issues is to incorporate more information into the large scale structure (Kaiser and Caragea, 2009) (Kaiser, Caragea, and Furukawa, 2012).
In our example, we do not have any information on possible covariates for the spatial locations. This means if we wish to incorporate large scale structure beyond what we have done, we would need to use spatial basis functions. The specific basis function we use is the bisquare basis function (Cressie and Johannesson, 2008) given by

\[ x_k(s_i) = \begin{cases} 
1 - \left( \frac{\|s_i - r_k\|}{d} \right)^2, & \|s_i - r_k\| \leq d \\
0, & \text{otherwise} 
\end{cases} \]

where \( r_k, k = 1, \ldots, K \) are a specified number of knots, \( \beta_k \) are unknown and \( \logit(\kappa_i) = \alpha(s_i) \).

We fit each field individually with number of knots \( K = 16, 25, 36 \). This was done to attempt to explore what, if any, temporal dependence existed among the estimates of \( \beta_k \) without having to fit increasingly complex Bayesian models. However, the estimates of \( \beta_k, k = 1, \ldots, K, K = 16, 25, 36 \) were all estimated to be zero for each field. This seems to suggest there is no evidence of spatially varying large scale structure. However, another possible reason for the behavior seen in Table 2.1 could be incorrect specification of the dependence structure (Kaiser and Caragea, 2009). While we have some suspicion the temporal dependence is also too high in at least two of the models (the Extended Neighborhoods and Point-wise dependence models), we will focus our efforts on the spatial dependence for now.

To address the possible incorrect model specification of the spatial dependence, we used pseudo-likelihood methods (Besag, 1974) (Kaiser and Cressie, 2000) to fit directional models to each field individually. More information on the specific directional models can be found in Chapter 3.5, we suffice to say now that when these models were fit to these data, several fields had dependence parameters that were estimated to be negative. This led us to question what the effect of negative dependence in MRF model was, and how does it manifest in data generated from these models.
Table 2.1: Model Parameter Estimates obtained from the double MH algorithm applied to the water vapor flux divergence data

<table>
<thead>
<tr>
<th>Dynamic MRF Model</th>
<th>Parameter</th>
<th>Estimate</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \kappa_1 )</td>
<td>-0.080</td>
<td>0.239</td>
<td></td>
</tr>
<tr>
<td>( \kappa_2 )</td>
<td>-0.051</td>
<td>0.192</td>
<td></td>
</tr>
<tr>
<td>( \kappa_3 )</td>
<td>0.172</td>
<td>0.182</td>
<td></td>
</tr>
<tr>
<td>( \mu )</td>
<td>-0.166</td>
<td>0.144</td>
<td></td>
</tr>
<tr>
<td>( \rho )</td>
<td>0.532</td>
<td>0.322</td>
<td></td>
</tr>
<tr>
<td>( \sigma^2 )</td>
<td>0.026</td>
<td>0.020</td>
<td></td>
</tr>
<tr>
<td>( \eta )</td>
<td>1.184</td>
<td>0.018</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Point-wise Dependence Model</th>
<th>Parameter</th>
<th>Estimate</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td>-0.006</td>
<td>0.058</td>
<td></td>
</tr>
<tr>
<td>( \kappa_1 )</td>
<td>-0.018</td>
<td>0.082</td>
<td></td>
</tr>
<tr>
<td>( \kappa_2 )</td>
<td>0.022</td>
<td>0.077</td>
<td></td>
</tr>
<tr>
<td>( \kappa_3 )</td>
<td>0.021</td>
<td>0.081</td>
<td></td>
</tr>
<tr>
<td>( \eta )</td>
<td>1.095</td>
<td>0.018</td>
<td></td>
</tr>
<tr>
<td>( \eta_{\rho} )</td>
<td>1.473</td>
<td>0.066</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Extended Neighborhoods Model</th>
<th>Parameter</th>
<th>Estimate</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td>2.747</td>
<td>0.133</td>
<td></td>
</tr>
<tr>
<td>( \kappa_1 )</td>
<td>-0.007</td>
<td>0.081</td>
<td></td>
</tr>
<tr>
<td>( \kappa_2 )</td>
<td>-0.052</td>
<td>0.083</td>
<td></td>
</tr>
<tr>
<td>( \kappa_3 )</td>
<td>0.026</td>
<td>0.084</td>
<td></td>
</tr>
<tr>
<td>( \eta )</td>
<td>0.948</td>
<td>0.031</td>
<td></td>
</tr>
<tr>
<td>( \xi )</td>
<td>1.317</td>
<td>0.079</td>
<td></td>
</tr>
</tbody>
</table>
2.5 Conclusions

We started this chapter with the idea of constructing spatio-temporal models that do incorporate scientific knowledge of an underlying process, possibly because we do not have knowledge of such a process. With this idea in mind, we introduced three such models: the Extended Neighborhoods model which treated time as another dimension in a standard MRF model; the Point-wise dependence model which looked similar to the Extended Neighborhoods model in terms of the natural parameter function $A_{i,t}$, but requires a conditional specification to justify the joint distribution existing; and the Dynamic MRF model, which modeled temporal dependence on large scale structure.

Our attempts to apply these models to the water vapor flux divergence data produced unexpected results. That is, the temporal dependence in the Extended Neighborhood and Point-wise dependence models was large and the spatial dependence for each model was also large. The large scale dependence parameter was also much larger than the observed mean for the Extended Neighborhoods model. These facts led us to consider trying to reduce the spatial dependence by incorporating large scale structure by way of spatial basis functions. However, these spatial basis functions also produced unexpected results in that all of the coefficients of the basis function were estimated to be zero for each set of basis functions. This led us to the conclusion if there was large scale structure in this data, it was not spatially varying.

The lack of spatially varying large scale structure led us to try and incorporate additional small scale structure in order to address the discrepancies observed in the large scale structures in the three models. It was here we observed several fields with negative spatial parameter estimates. Which raised the question: What effect does a negative dependence parameter have on a binary Markov random field model?
2.5.1 Future Work

The question raised in the previous paragraph is the starting point for the next two chapters. In Chapter 3, we will address the effect a negative dependence parameter has on the data patterns in realizations from a binary MRF model. We will develop tools to identify and quantify the direction and strength of dependence that can be used in real data. We will then apply these tools to the water vapor flux divergence data in order to explore the spatial dependence in these data.

In Chapter 4 we explore the effect of negative dependence in the conditional expectation of the binary MRF model. The distribution of the conditional expectations for a binary Markov random field are examined as well as the topology of the surface created by the conditional expectations. The notion of an empirical conditional expectation is introduced in order to examine the conditional expectations of real data, for which model parameters are unknown.

The methods introduced in these two chapters help to inform us of why the specific models from this chapter failed to produce an appropriate fit to our data, and also begin to led us to models that would be more appropriate for these data.

2.6 Bibliography


Hobbs, Jonathan Michael (2014). “Characterizing diurnal and interannual variability in the atmosphere through physical and stochastic models”. In:


CHAPTER 3. NEGATIVE DEPENDENCE IN BINARY MARKOV RANDOM FIELDS

A paper in preparation
Kenneth Wakeland, Mark S. Kaiser and Daniel Nordman

Abstract

Many problems in ecology or environmental sciences, such as monitoring presence/absence of species or threshold events, involve the observation of spatial binary random fields. These types of data are commonly analyzed using a Markov Random Field (MRF) model. However, difficulties can arise when the neighborhood structures present in the data are not correctly identified or become complex. To further complicate the issue, negative dependence in these models is theoretically possible, but the effects such a value for the dependence is not well known beyond the basic case. We explore the use and issues that arise when negative dependence is utilized in MRF models, both in the simple cases and when there are directional neighborhoods present.

3.1 Introduction

The ability to obtain data over large areas has led to an increase in the use of models that incorporate spatial dependence. Spatial dependence can be modeled using either discrete or continuous spatial fields. When a discrete spatial field is appropriate, Markov random field
Markov random fields have been used in image analysis (Besag, 1986), modeling beetle outbreaks in forests (Zhu, Huang, and Wu, 2005) (Zhu et al., 2008), network analysis (Schweinberger and Snijders, 2003) and in estimating harbor seals in Disenchantment Bay (Ver Hoef and Jansen, 2007) to name a few. The discrete nature of the MRF allows for a wider range of possible dependence values, that is, negative dependence. In geostatistical models, or any continuous state space model, dependence is most naturally thought of as positive. This is exemplified by the use of the variogram/semi-variogram as one of the first tools used to measure the spatial dependence in these models. But this idea does not have an analog in MRF models, and allows us to define a repulsive/negative dependence between points.

The MRF models we will consider here were first considered by Julian Besag in his famous 1974 article. These models have been used often, but an update (Caragea and Kaiser, 2009) provided a more consistent way of looking at the large scale structure (mean/trend) and small scale structure (spatial dependence) interaction. We will use the formulation of Kaiser and Caragea for this paper. In both (Besag, 1974) and (Caragea and Kaiser, 2009) models were developed for the exponential family distributions, including the binary/binomial distributions. We will focus our efforts on the properties of the binary distribution for this paper. We will give the specific model we use in this paper in section 2.

There is a well known phenomenon that occurs with MRF models which is often called degeneracy (Kaiser, Caragea, and Furukawa, 2012) (Handcock et al., 2003) (Schweinberger and Snijders, 2003). This issues comes up when the spatial parameter in the MRF gets to be "too large" and results in fields that have a degenerate joint distribution (as in Gaussian MRF) or all one value (in the Binary MRF). There are some methods to deal with this issue (Kaiser and Caragea, 2009), the one we will concentrate our efforts on for this paper are the directional neighborhoods. We combine these directional models with the idea of negative dependence discussed above to allow the binary MRF model to describe more varied structures in the data that aren’t normally associated data coming from a MRF.
Section 2 will layout the notation and model form we will use to discuss binary Markov random fields. Section 3 will delve deeper into the effects of degeneracy and how we will refer to that phenomenon in this article. Section 4 will explore the effects of negative dependence in binary Markov random fields. Section 5 will introduce the notion of directional neighborhood structures and what effects they have on binary MRF. Section 6 will layout a simulation experiment that aims to explore and begin to explain some of the complex interactions between the topics discussed in this paper.

### 3.2 Binary MRF Formulation

We will formulate a binary Markov random field model for spatial locations \( s_i = (u_i, v_i) \) where \( u_i \) and \( v_i \) denote the horizontal and vertical coordinates for location \( s_i \), respectively. Let binary random variables \( Y(s_i) \) at locations \( s_i \) be defined according to the occurrence of some event. Let \( \{N_i : i = 1, \ldots, n\} \) denote neighborhoods for the locations, and \( y(N_i) = \{y(s_j) : s_j \in N_i\} \) be the set of values of neighbors in the set \( N_i \).

Markov random field models are formulated by specifying the full conditional distributions for the random variables at each location. The Markov assumption is that these conditional distributions depend functionally only on certain neighboring values \( y(N_i) \). That is,

\[
\Pr \left[ Y(s_i) = y(s_i) \mid \{y(s_j : j \neq i)\} \right] = f \left( y(s_i) \mid y(N_i) \right), \quad y(s_i) = 0, 1. \tag{3.1}
\]

We express binary probability mass functions in exponential family form as

\[
f \left[ y(s_i) \mid y(N_i) \right] = \exp \left[ y(s_i) A_i \{y(N_i)\} - B_i \{y(N_i)\} \right]; \quad y(s_i) = 0, 1. \tag{3.2}
\]

In (3.2), \( A_i \) is the natural parameter function and \( B_i \) controls moments. Using parameters \( \kappa_i \) and \( \eta_{ij}, i, j = 1, \ldots, n \), we will model these functions as
Above, the \( \eta_{ij} \) represent dependence parameters and \( \kappa_i \) denote “large scale” parameters controlling the probability of occurrence. Using the preceding notation, the conditional expectations of the random variables are

\[
E[Y(s_i)|\{y(s_j): j \neq i\}] = E[Y(s_i)|y(N_i)] = \frac{\exp(A_i\{y(N_i)\})}{1 + \exp(A_i\{y(N_i)\})}.
\]  

(3.5)

If all dependence parameters \( \eta_{ij} = 0 \) in (3.3), the random variables are all mutually independent, and the expected value (3.5) is \( \kappa_i \) (i.e. \( \kappa_i \) is the probability that \( Y(s_i) = 1 \) or the marginal mean \( E[Y(s_i)] = \kappa_i \)). If not all \( \eta_{ij} = 0 \), conditional and marginal expectations differ, with the \( \kappa_i \) being nearly equal to marginal expected values unless the \( \eta_{ij} \) become too large in value (Kaiser, Caragea, and Furukawa, 2012), in which case the model begins to exhibit degenerate behavior (see also (Handcock et al., 2003)).

Under certain conditions (Kaiser and Cressie, 2000), which here includes \( \eta_{ij} = \eta_{ji} \) in (3.3), the negpotential function of (Besag, 1974) can be used to identify a joint distribution corresponding to the specified conditionals (3.2). Along with an assumption of pairwise only dependence (Besag, 1974), the negpotential function for our binary conditionals model becomes, up to an additive constant,

\[
Q(y) = \sum_{i=1}^{n} y(s_i) \left[ \log \left( \frac{\kappa_i}{1 - \kappa_i} \right) - \sum_{j \neq i} \eta_{ij} \kappa_i \right] + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \eta_{ij} y(s_i)y(s_j).
\]  

(3.6)

The negpotential function (3.6) then allows us to recover the joint distribution for the random field model as

\[
\Pr(Y = y) = f(y) = \frac{\exp[Q(y)]}{\sum_{t \in \Omega} \exp[Q(t)]},
\]  

(3.7)
where \( \Omega = \{0,1\}^n \) is the joint support. The denominator of (3.7) is computationally intractable for many problems, and estimation typically relies on a pseudo-likelihood approach for frequentist analysis or the simulation of auxiliary fields for Bayesian analysis (e.g., (Hughes, Haran, and Caragea, 2011) (Møller et al., 2006)).

### 3.3 Data Patterns in MRF Models

The model of the previous section contains too many free parameters to be useful. An extreme simplification that results in only two parameters is obtained by specifying \( \kappa_i = \kappa \) for \( i = 1, \ldots, n \), and, for \( i, j = 1, \ldots, n \), setting

\[
\eta_{i,j} = \begin{cases} 
\eta & \text{if } s_j \in N_i, \\
0 & \text{otherwise}
\end{cases}
\]  

(3.8)

In this section, we make use of this reduced parameterization along with neighborhood structures specified by 4 nearest and 8 nearest neighbors on a regular lattice, see Figure 3.1.

### 3.3.1 Models with Positive Spatial Dependence

In many problems, it is natural to assume spatial dependence will be positive. Realizations from a model using four nearest neighbors are displayed in Figure 3.2 to illustrate the effect of the dependence parameter \( \eta \) increasing in magnitude. The top two panels correspond to mild (\( \eta = 0.50 \)) and moderate (\( \eta = 1.00 \)) values of \( \eta \), and exhibit the classical pattern of clustering often associated with positive spatial dependence. Each of these realizations contain about half 0’s and half 1’s, corresponding to \( \kappa = 0.50 \) in the model. These two panels illustrate the types of behavior that are typically associated with positive spatial dependence in binary Markov random fields. The idea that increasing the spatial dependence parameter, \( \eta \), results in a field in which the size of the clumps increases leads us to consider what occurs when we increase the dependence to the limit \( \eta \to \infty \).
3.3.2 Degeneracy in Binary MRF Models

Consider the lower panels of Figure 3.2. These panels illustrate realizations from models that are moving towards what is typically called “degeneracy” due to positive dependence in Markov random field models (Schweinberger and Snijders, 2003) (Kaiser, Caragea, and Furukawa, 2012) (Handcock et al., 2003). The lower left panel ($\eta = 2.0$) shows a field dominated by one value (in this case 0) and the lower right panel ($\eta = 4.0$) shows a nearly constant field, despite the model having $\kappa = 0.50$. As the magnitude of the dependence parameter(s) increases, the empirical mean in the realized fields deviates from the expected value of the independence model ($\kappa$). Full degeneracy results in a model that contains, at most, two realizations with non-zero probability, constant fields of all 0 values or all 1 values. As suggested by Figure 4 in (Kaiser, Caragea, and Furukawa, 2012), the probabilities associated with these constant fields depends on the value of kappa in the model. For $\kappa = 0.50$, the relative frequencies are half for each field (either all 1s or all 0s), while for $\kappa < 0.50$ the probability of a constant field of 1s goes to 1, and for $\kappa > 0.50$ the probability of a constant field of 0s goes to 1 (see Figure 4 and Table 2 in (Kaiser, Caragea, and Furukawa, 2012), and Table 3.1 in this article). It is important to note here, fully degenerate realizations will only occur when $\eta \to \infty$. Even for large values of $\eta$ for which the realized field is constant, there is a small probability that at least one point will be different from the others.

Models with eight nearest neighbors exhibit the same patterns of behavior as the dependence parameter $\eta$ grows larger, although the absolute magnitudes of $\eta$ needed to produce degenerate-like behavior are roughly one half of those that produce such behavior in models with four neighbors. Models with negative dependence will produce their own patterns when $\eta \to -\infty$ as well. We explore such cases in the next section.
3.3.3 Models with Negative Spatial Dependence

There has been little investigation of the behavior of Markov random field models having negative dependence, perhaps because negative spatial dependence is a less intuitive concept than positive spatial dependence. Nevertheless, understanding the effect of negative dependence on Markov random field models provides insight into the types of data patterns that are represented by such models.

We begin by examining the effect of negative dependence in a model with four nearest neighbors and a single value of $\kappa = 0.5$. Four realizations from this model with various values of negative $\eta$ are presented in Figure 3.3. As for the top row of Figure 3.2, the top row of Figure 3.3 shows realizations from mild and moderate values of $\eta$ in the left and right panels, respectively. These realizations exhibit the repulsive effect of negative dependence, which produces data patterns containing alternating, short sequences, or runs, of 0 and 1 values. The lower panels of Figure 3.3 demonstrate the effect of large negative values of $\eta$, which is to enforce alternating values of 0 and 1 at individual locations. A perfectly alternating pattern represents full degeneracy for models with negative dependence, which departs rather substantially from degeneracy in models with positive dependence discussed in the previous section. Specifically, when the model is fully degenerate in this scenario, the empirical mean will be exactly 0.50.

To confirm this assertion, consider the simple case of four points on a transect where each point has two neighbors. Table 3.2 gives the exact probability of each configuration that consists of alternating values. This situation, with negative dependence, differs from the situation with positive dependence in which limiting distribution of configurations consisted of constant values of all 0s or all 1s (Table 3.1). Figure 4 in (Kaiser, Caragea, and Furukawa, 2012) shows models with large positive dependence tend to produce fields empirical means greater than 0.5 if $\kappa < 0.5$, and less than 0.5 if $\kappa > 0.5$. In contrast, models with large negative dependence tend to produce fields having empirical means equal to 0.5 regardless of
the value of kappa, as illustrated in Figure 3.4. Similarly to models with positive dependence, however, models with negative dependence become degenerate more slowly with increasing $|\eta|$ the more $\kappa$ differs from 0.5.

As discussed previously, models with four and eight nearest neighbors behave in a similar fashion under positive dependence. This is no longer the case for models having negative dependence. In fact, a model with negative dependence and eight neighbors would seem to be counter intuitive because it contains cliques of size three in which all members are negatively related to each other. This seeming paradox can be illustrated by examining marginal covariances produced by the conditional dependencies specified in a model. The left panel of Figure 3.5 shows the direction of marginal covariances resulting from simulation of 2000 realized fields from a model with negative conditional dependencies and four nearest neighbors. The model specifies negative conditional dependencies among locations separated along cardinal directions. Those dependencies produce negative marginal covariances for the same displacements. They also produce positive marginal covariances among locations offset by inter-cardinal directions. This is in contrast to the right panel of Figure 3.5 which shows the directions of marginal covariances resulting from simulation of 2000 fields from a model with negative conditional dependencies and eight nearest neighbors. This model also specifies negative conditional dependencies in the cardinal directions and produces negative marginal covariances in those same directions. But, the model also specifies negative conditional dependencies in the inter-cardinal directions and the marginal covariances, surprisingly, are also negative. Intuitively, it would seem that positive covariances in inter-cardinal directions should be induced by negative conditional dependencies in cardinal directions. For models with four nearest neighbors this intuition appears to be correct, but not for models with eight nearest neighbors. The signs associated with directional marginal covariances in Figure 3.5 hold for all models with negative dependencies, although the strength of those covariances is influenced by the numerical value of $\eta$ in the way one would expect.
3.4 Quantifying Data Patterns

In this section, we develop two quantities to examine data patterns in realizations of Markov random field models. The first quantity, which we call conflicts, assists in providing intuition for the data patterns observed in Markov random field models, especially when the dependence parameters become large. The second quantity, which we call run length distributions, provides a method of examining data patterns in real data on a regular lattice.

3.4.1 Conflicts in Realizations of MRF Models

Any binary Markov Random Field model specifies either positive or negative conditional dependence between each possible pair of locations that are not conditionally independent. For a realized field, we define a conflict to exist for a pair of locations if the data values at those locations differ under positive dependence (0 at one location and 1 at the other) or if the pair are the same under negative dependence (both 0 or both 1). In simple models with a single dependence parameter, the proportion of location pairs that exhibit a conflict tends to decrease as the strength of the dependence increases. For example, the top left realization displayed in Figure 3.2, with \( \eta = 0.50 \), contains conflicts for 78 of the 386 non-independent pairs of locations while the upper right realization in Figure 3.2, with \( \eta = 1.0 \), contains 59 conflicts.

In what follows, we will count conflicts in the largest cliques allowed by a given neighborhood structure. For a four-nearest neighbor structure this is the same since counting conflicts in neighboring pairs as the largest cliques for this configuration are of size two. In contrast, an eight-nearest neighbor structure contains cliques of size two, three and four, so we count conflicts for each clique of size four. Examples of possible conflicts are presented for cliques of size four in Figure 3.6, the four panels of which show all possible configurations of 0 and 1 values, up to their symmetric counterparts (interchanging 1s and 0s) and rotations. Each of these four panels could exist under both positive and negative dependence. The
upper left and right panels of Figure 3.6 result in 4 conflicts for positive dependence and 2 conflicts for negative dependence. The lower left panel results in 3 conflicts for both positive and negative dependence. The lower right panel results in 0 conflicts for positive dependence and 6 conflicts for negative dependence. Notice that the number of conflicts for positive and negative dependence always sum to 6, the number of pairs of locations in the clique. Also notice that for negative dependence, it is not possible for any configuration of points in a clique of size four to have 0 conflicts, the minimum is 2 conflicts.

We return to the notion that the number of conflicts should decrease as the magnitude of the dependence parameter increases. This implies that as the dependence parameter is increased to extreme values, the number of conflicts in all cliques of size 4 will tend to 0 for positive dependence and 2 for negative dependence. To illustrate this, 2000 MRF on a 33x33 lattice were generated for each combination of $\kappa = 0.50$ and $\eta = \pm 4.0, \pm 2.5, \pm 1.5, \pm 0.5, 0.0$. Table 3.3 presents the relative frequencies of cliques of size 4 that contained 2, 3 and 6 conflicts for negative dependence or 0, 3 and 4 conflicts for positive dependence. There are 1024 cliques of size 4 in a 33x33 lattice. In addition to verifying that the probabilities of observing realized fields with more than 2 or 0 conflicts approaches 0 as the magnitude of $\eta$ increases for negative and positive dependence, respectively, Table 3.3 suggests that models with positive dependence approach a state of degeneracy more quickly than do models with negative dependence as the absolute value of $\eta$ grows large.

A realized field from an approximately degenerate model with $\eta = -100$ and eight nearest neighbors is presented in Figure 3.8. This realization, or a rotation of it, nearly always results for data generated from this model, and contains the minimum number of 2 conflicts in all cliques of size 4. Recall that the upper two panels of Figure 3.6 show two configurations of values that both result in the minimum number of 2 conflicts in cliques of size 4 for negative dependence. The pattern of Figure 3.8 corresponds to repetition of only one of these configurations, the one shown in the upper right panel of right panel of Figure 3.6. The reason for this is illustrated in Figure 3.7, which shows configurations and associated
conflicts for sets of four concatenated cliques of size 4. The configuration in the left panel of Figure 3.7, which corresponds to repetition of the upper left panel of Figure 3.6, results in 8 conflicts. In contrast, repetition of the configuration in the upper right panel of Figure 3.6 results in only 6 conflicts, as shown in the right hand panel of Figure 3.7. Such an argument can be generalized to any $k \times k$ field using the same concatenation technique. Thus, the total number of conflicts in a realized field from a model with eight nearest neighbors and negative dependence is minimized by the striped pattern in Figure 3.8.

### 3.4.2 Distribution of Run Lengths

We have seen that extending the concept of conflicts within individual cliques to entire fields results in larger-scale data patterns (e.g. Figure 3.8). These patterns appear to be related to the number and length of sequences of the same data values, either 0s or 1s. We define a run of values to be a sequence of consecutive 0s or 1s within a transect embedded in a random field. Such transects may be considered to consist rows, columns, or diagonals. For example, a 33x33 lattice would contain a potential of 32 runs of length 1 within each row, column or the main diagonal, and a varying number of runs of length 1 (from 30 to 1) along the off diagonals. Consideration of conflicts suggested that in models with increasingly strong positive dependence, long run lengths should be expected to become more common. Similarly, in models with increasingly strong negative dependence short run lengths should become more common. To illustrate this, a Monte Carlo experiment was conducted in which models consisting of 4 nearest neighbors with $\eta = \pm 0.5, \pm 1.0$, and 8 nearest neighbors with $\eta = \pm 0.25, \pm 0.5$ were simulated. For each situation, 2000 data sets were simulated and compared to data sets simulated from a model with no dependence (i.e., the independence model). The value of $\kappa$ was held fixed at 0.50 for each model used. The number of runs of length $x$ were counted within each row and summed to a field-level total, for $x = 1, 2, \ldots, 33$. The same was done separately for columns. Monte Carlo approximations to the expected number of runs of length $x$ were then computed as the average field totals across the 2000
simulated data sets. Results for $x = 1, 2, 3, 4$ are presented in Figure 3.9 for rows and Figure 3.10 for columns. These results verify that in the case of positive dependence there is a smaller expected number of runs of short lengths (1 and 2) than in the independence model. Conversely, the expected number of runs of short lengths is higher than the independence model in the case of negative dependence.

These results suggest that examining the number of short run lengths in actual data fields can provide guidance about the type and strength of dependence expected to result from a simple Markov random field model. We develop a diagnostic for this purpose as follows. For a given observed data field, let $\hat{p}$ denote the empirical mean. Simulate $M$ fields from an independence model with parameter $\hat{p}$. That is, simulate $k \times k$ values from a binary distribution with parameter $\hat{p}$ and arbitrary location labels. For a given transect definition (e.g., rows or columns) let $q_{r,m}$ be the total number of runs of length $r$ in field $m = 1, 2, \ldots, M$. Let $a_r$ be the number of runs of length $r$ in the actual data field. A quantity to compare the actual number of runs with what is expected under an independence model is then

$$p_r = \frac{1}{M} \sum_{m=1}^{M} I(a_r \leq q_{r,m}). \quad (3.9)$$

We compute this quantity for small values of $r$ ($r = 1, 2, 3, 4$) and restricting attention to rows and columns. Extensions to other runs lengths and transect definitions are possible depending on the size and shape of an observed lattice. As an example, we generated two "actual data fields" by simulated from MRF models with $\eta = \{-0.5, 0.5\}$ and $\kappa = 0.5$ in each case. The empirical means for these fields were $\hat{p} = \{0.499, 0.502\}$. In each case, 10000 fields were simulated from independence models having parameters equal to the empirical means. Resultant values of $p_r$ for the rows are reported in Table 3.4 for $r = 1, 2, 3, 4$. The model with positive dependence had fewer runs of short length (small $p_r$ for $r = 1, 2$) while the model with negative dependence had more runs of short length (large $p_r$ for $r = 1, 2$). Repeating this exercise 2000 times resulted in the distribution of $p_r$ values, $r = 1, 2$ shown in Figure 3.11. The distribution of $p_r$ displayed in this figure are once again for rows, the $p_r$
values for columns are nearly identical and thus omitted. The box plot displayed in Figure 3.11 illustrates a clear separation between the distribution of $p_r$ values for $r = 1, 2$ for each model.

### 3.5 Directional Dependence

We turn our attention from isotropic neighborhoods for Markov Random Fields to neighborhoods that incorporate directional dependence. This can be done in four nearest neighbors by using the decomposition expressed in (3.10).

\[
N_{4,i} = N_{u,i} \cup N_{v,i},
\]

\[
N_{v,i} = \{ s_j : u_j = u_i, v_j = v_i \pm 1 \},
\]

\[
N_{u,i} = \{ s_j : u_j = u_i \pm 1, v_j = v_i \}.
\]

Using such a decomposition, the natural parameter function is represented in (3.11)

\[
A_i(y(N_i)) = \tau^{-1}(\kappa) + \eta_v \left( \sum_{s_j \in N_{v,i}} [y(s_j) - \kappa] \right) + \eta_u \left( \sum_{s_j \in N_{u,i}} [y(s_j) - \kappa] \right).
\]

The eight nearest neighbor model allows for a similar decomposition, given by (3.12)
\[ N_i = N_{1,i} \cup N_{2,i} \cup N_{3,i} \cup N_{4,i}, \]

\[ N_{1,i} = \{ s_j : u_j = u_i, v_j = v_i \pm 1 \}, \]
\[ N_{2,i} = \{ s_j : u_j = u_i \pm 1, v_j = v_i \}, \]
\[ N_{3,i} = \{ (u_{i+1}, v_{i+1}), (u_{i-1}, v_{i-1}) \}, \]
\[ N_{4,i} = \{ (u_{i-1}, v_{i+1}), (u_{i+1}, v_{i-1}) \}. \] (3.12)

And similarly to the the four nearest neighbor model, the natural parameter function for the eight nearest neighbor model is given by (3.13)

\[ A_i(y(N_i)) = \tau^{-1}(\kappa) + \eta_1 \left( \sum_{s_j \in N_{1,i}} [y(s_j) - \kappa] \right) + \eta_2 \left( \sum_{s_j \in N_{2,i}} [y(s_j) - \kappa] \right) \]
\[ + \eta_3 \left( \sum_{s_j \in N_{3,i}} [y(s_j) - \kappa] \right) + \eta_4 \left( \sum_{s_j \in N_{4,i}} [y(s_j) - \kappa] \right). \] (3.13)

Focusing on the MRF models defined above, we present the remaining discussion in two parts. In the first, we use conflicts to explore the behavior of these models particularly as the magnitude of the dependence parameters increase. In the second, we apply the run length diagnostic in various directions and examine the results.

### 3.5.1 Conflicts in Directional MRF Models

It was observed in Section 4.1 that, for \( \kappa = 0.50 \), the number of conflicts in cliques of size 2 (for four nearest neighbors) and 4 (for eight nearest neighbors) would decrease as the magnitude of the dependence parameter increased. While these results were demonstrated using models with isotropic dependence, similar phenomena occur for models with directional dependencies. In the case of a model with four nearest neighbors, partitioned according to
The number of conflicts among adjacent pairs of locations in a given direction will decrease as the dependence parameter for that direction increases in magnitude. This is true for each direction separately, conditioned on a fixed strength of dependence in the other direction. When a dependence parameter in one direction grows sufficiently large, degeneracy is produced in that direction as long as the dependence in the other direction is the same sign, fixed, or growing at a slower rate. That is, with positive dependence in the given direction, values become constant in that direction. For negative dependence in the given direction, values will converge to a perfectly alternating pattern. This is illustrated in Figure 3.12. The top two panels of Figure 3.12 show realizations of fields from models with \( \eta_u = 0.2 \) and \( \eta_v \) growing from 5 on the left to 100 on the right. In the left panel, \( \eta_v \) is large enough to produce regular behavior, but \( \eta_u \) remains sufficiently influential to produce occasional reversals. In the right panel, \( \eta_v \) is large enough that \( \eta_u \) no longer affects the fixed pattern of degeneracy (which is constant values) within the columns. The lower two panels of Figure 3.12 illustrate the same phenomenon for negative \( \eta_v \) values in which case the fixed pattern of degeneracy is perfectly alternating values. Models with eight nearest neighbors demonstrate these same features, although the bookkeeping involved in tracking conflicts in four directions becomes more tedious, especially when dependence parameters are not all the same sign. This is because what constitutes a conflict depends on whether dependence is positive or negative.

To demonstrate that realizations from directional models will approach the minimum number of conflicts possible we conduct a small Monte Carlo exercise with two models. The first is the four nearest neighbor model defined by (3.11) with \( \kappa = 0.50, \eta_v = 0, 1.0, 2.0, 3.0, 4.0 \) and \( \eta_u = -\eta_v \). The second model is defined by (3.13) with \( \kappa = 0.50, \eta_1 = \eta_2 = 0, 1.0, 2.0, 3.0, 4.0 \) and \( \eta_3 = \eta_4 = -\eta_1 \). For each model, 2000 realizations were generated using a \( 10 \times 10 \) lattice and the average number of conflicts per clique were approximated. In the case of the four nearest neighbor model, two expected number of conflicts are found, one for the north-south direction, the second for the east-west direction. The results displayed in Table 3.5 exhibit the expected behavior. That is, as the dependence parameter increases in magnitude, the
expected number of conflicts in the appropriate clique size decreases, approaching the minimum number of conflicts within a clique. In this exercise, dependencies in different directions were of opposite sign. The same results hold for models in which dependencies in different directions are the same sign, with convergences to the minimum number of conflicts occurring more rapidly than for the situations included.

### 3.5.2 Distribution of Run Lengths in Directional MRF models

In Section 4.2 we showed that the number of runs of lengths one and two could be used as a diagnostic for direction and strength of dependence in isotropic models. In this section, we demonstrate that the same use may be made of run lengths in models with directional dependence by examining such runs in individual directions. In fact, interpretation of directional run lengths appear nearly unaffected by the type and strength of dependence in other directions. We conduct two Monte Carlo exercises, analogous to those presented previously in Section 4.2, to demonstrate these points. In the first Monte Carlo, we generated 2000 fields from model (3.10) with four nearest neighbors, $\kappa = 0.50$, $\eta_u = 0.50$ and $\eta_v \in \{-0.50, 0.0, 0.25, 0.50\}$ on a $33 \times 33$ lattice. The Monte Carlo approximates to expected number of runs of lengths 1, 2, 3 and 4 are presented in Figure 3.12. In the same manner as for isotropic models, there are fewer directional runs of short length for positive directional dependence, and more for negative directional dependence. When dependence is different in different directions, the expected number of runs behaves in the same manner as seen previously for isotropic models so that directions may be assessed independently. In particular, if there is no dependence in a given direction, the expected number of short runs in that direction is nearly identical to that of an independence model (e.g., top right panel of Figure 3.12). In the second Monte Carlo exercise, we make use of the diagnostic (3.9) by generating 2000 fields from model (3.11) with $\kappa = 0.50$, $\eta_u = 0.50$ and $\eta_v \in \{-0.50, 0.0, 0.25, 0.50\}$ on a $33 \times 33$ lattice, resulting in four sets of 2000 simulated fields. The distributions of $p_r$ computed for each individual field are presented in Figure 3.14. The results for the verti-
cal direction with dependence parameter $\eta_v$ are the same as they would have been for that dependence in an isotropic model. Specifically, relative to an independence model, when $\eta_v$ is negative the number of short runs are increased (top left panel) and when $\eta_v$ is positive the number of short runs are decreased (lower right panel). The degree of difference from the independence model becomes more dramatic as the magnitude of $\eta_v$ increases (lower left versus lower right panels) when $\eta_v$ is zero, the distribution of the number of short runs is identical to independence (upper right panel).

Although we have presented results only for models with four nearest neighbors, results for models with eight neighbors follow similar patterns. In addition, run lengths may also be examined along diagonal transects, and provide information on the type and strength of dependencies in those directions.

### 3.6 Water Vapor Flux Reanalysis Data

We apply the run length distribution analysis developed above to water vapor flux data obtained from (Hobbs, 2014). Water vapor flux divergence is an element of the water vapor equation and has been used to predict rainfall associated with synoptic-scale systems, as a key component in the Kuo convective parameterization scheme, and in the prediction of severe local storm prediction (Banacos and Schultz, 2005). This quantity can be calculated in different ways depending on the availability of other data. One method used upper air observations of winds and specific humidity to calculate water vapor flux divergence. Another combined windows from reanalysis with remotely sensed water vapor to complete the water vapor flux calculation (Hobbs, 2014). In this application however, the water vapor flux divergence (convergence) is computed from reanalysis products. Specifically, the reanalysis data from 2 meteorological models: the Modern-Era Retrospective Analysis for Research and Applications (MERRA) and the North American Regional Reanalysis (NARR) are used to calculate the water vapor flux divergence. Both meteorological models were computed at
3 hour intervals for a month at a time. The final data set is calculated using both of these meteorological models over a 33 x 33 lattice covering a region of the central United States, as seen in Figure 3.15. Further details about the exact calculation of these data can be found in (Hobbs, 2014).

The water vapor flux divergence data at each of these locations are continuous and consisting of both positive and negative values. Positive values indicate a convergence of water vapor and negative values indicate a divergence. It is useful to know which phenomenon (convergence or divergence) is happening at a location, which leads to the choice to threshold these data at 0. Specifically, let \( Z(s_i, t) \) be the water vapor flux divergence value at location \( s_i \) and time \( t \), then we define \( Y(s_i, t) = I [Z(s_i, t) \geq 0] \), which results in a series of binary fields on a 33 x 33 lattice. The specific data we use is from May 2006. This means there are 248 binary fields that we can use to analyze the thresholded water vapor flux. Our goal with this data is to fit a spatio-temporal model using all 248 fields using a Markov Chain Monte Carlo based on the method proposed by (Liang, 2010). However, such an analysis is computationally expensive and as such we first use the run length distribution method discussed above to explore the strength and directions of the dependencies in each direction to get a better understanding of the appropriate MRF model to use for these data. Figure 3.16 displays the results for the east-west (rows), north-south (columns), northeast-southwest and northwest-southeast (diagonal) directions. This figure seems to indicate that the dependence in all directions except the northwest-southeast diagonal are positive. In the northwest-southeast diagonal direction, the dependence fluctuates considerably and has a clustering structure that is often associated with autocorrelation. In fact, were we to fit the MRF model described in (3.13) to each field individually using pseudo-likelihood methods, each parameter appears to have at least moderate levels of autocorrelation. Table 3.6 displays the mean and lag 1 autocorrelation for each parameter. The results of both the \( p_r \) values and the pseudo-likelihood estimates point to a model that has time dependence in not only the mean, but also the spatial dependence values.
3.7 Conclusions

Negative dependence in MRF models has been known since their inception to be mathematically possible, but, outside of Poisson models, is rarely mentioned and has not been deeply examined. In our experience, negative dependence is not common, but as shown in this study, does happen with real data, which makes understanding the implications of negative dependence in MRF models important. And while negative dependence in isotropic four nearest neighbor models can be incorporated into our traditional understanding of dependence in MRF models, negative dependence in isotropic eight nearest models require us to readdress how dependence manifests in realized data fields. This led us to counting the number of conflicts within a clique which allows for an intuition that is consistent across positive, negative, and a combination of dependencies.

The notion of conflicts within a clique led us to develop a simple tool to explore the strength and direction of dependence in MRF models in order to detect if positive dependence, negative dependence or a combination of the two is present in the data. While the run length distribution tool is exploratory and only gives a basic understanding of these dependencies, it less computationally intensive than fitting many models, even more so if the method use to fit those models is Bayesian. The run length distribution concept can also be adapted for use in a Bayesian analysis as a part of a posterior predictive analysis. When the run length tool was applied to the thresholded water vapor flux data, we were able to observe negative dependence in one direction, and more so we were able to see a time effect in the strength of that dependence. This provides insight on how we might structure a model to fit this data, one that is more complex than we would normally consider.

Work in this area that is on going involves an issues that arises when counting the number of conflicts in a clique when $\kappa$ is not close to 0.50. Throughout this paper, when generating realizations from MRF models, the $\kappa$ value is close to 0.50, but when $\kappa < 0.25$ or $\kappa > 0.75$, the behavior in the conflicts begins to differ from what happens when $\kappa \in [0.25, 0.75]$. While
there is some precedent for such behavior happening (Kaiser, Caragea, and Furukawa, 2012), more work needs to be done in order to fully understand this phenomenon.

Another possible extension to the work done here is to consider the data patterns from realized MRF models when cliques of size greater than 2 are directly modeled. Specifically, what changes are observed in conflicts in cliques and within the run length distributions. And can these tools provide insight when these models are consider/used.

3.8 Bibliography


Hobbs, Jonathan Michael (2014). “Characterizing diurnal and interannual variability in the atmosphere through physical and stochastic models”. In:


### Appendix: Tables and Figures

Table 3.1: Probabilities of configurations of 4 points on a transect with positive dependence

<table>
<thead>
<tr>
<th>$\kappa$</th>
<th>$y(s_1)$</th>
<th>$y(s_2)$</th>
<th>$y(s_3)$</th>
<th>$y(s_4)$</th>
<th>$\eta = 1$</th>
<th>$\eta = 2$</th>
<th>$\eta = 4$</th>
<th>$\eta = 8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.25</td>
<td>5.54</td>
<td>94.47</td>
<td>100.00</td>
</tr>
<tr>
<td>0.9</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>66.83</td>
<td>60.43</td>
<td>1.71</td>
<td>0.00</td>
</tr>
<tr>
<td>0.8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.94</td>
<td>18.04</td>
<td>94.59</td>
<td>100.00</td>
</tr>
<tr>
<td>0.8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>45.17</td>
<td>38.00</td>
<td>1.64</td>
<td>0.00</td>
</tr>
<tr>
<td>0.7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5.21</td>
<td>24.78</td>
<td>90.27</td>
<td>99.98</td>
</tr>
<tr>
<td>0.7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>31.20</td>
<td>29.94</td>
<td>4.45</td>
<td>0.01</td>
</tr>
<tr>
<td>0.6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9.58</td>
<td>27.02</td>
<td>76.23</td>
<td>99.11</td>
</tr>
<tr>
<td>0.6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>21.80</td>
<td>27.62</td>
<td>15.73</td>
<td>0.83</td>
</tr>
<tr>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>14.96</td>
<td>27.32</td>
<td>45.04</td>
<td>49.90</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>14.96</td>
<td>27.32</td>
<td>45.04</td>
<td>49.90</td>
</tr>
<tr>
<td>0.4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>21.80</td>
<td>27.62</td>
<td>15.73</td>
<td>0.83</td>
</tr>
<tr>
<td>0.4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>9.58</td>
<td>27.02</td>
<td>76.23</td>
<td>99.11</td>
</tr>
<tr>
<td>0.3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>31.20</td>
<td>29.94</td>
<td>4.45</td>
<td>0.01</td>
</tr>
<tr>
<td>0.3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5.21</td>
<td>24.78</td>
<td>90.27</td>
<td>99.98</td>
</tr>
<tr>
<td>0.2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>45.17</td>
<td>38.00</td>
<td>1.64</td>
<td>0.00</td>
</tr>
<tr>
<td>0.2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.94</td>
<td>18.04</td>
<td>94.59</td>
<td>100.00</td>
</tr>
<tr>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>66.83</td>
<td>60.43</td>
<td>1.71</td>
<td>0.00</td>
</tr>
<tr>
<td>0.1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.25</td>
<td>5.54</td>
<td>94.47</td>
<td>100.00</td>
</tr>
</tbody>
</table>
Table 3.2: Probabilities of configurations of 4 points on a transect with negative dependence

<table>
<thead>
<tr>
<th>$\omega$</th>
<th>$P(\omega)$ (%)</th>
<th>$\eta = -5$</th>
<th>$\eta = -10$</th>
<th>$\eta = -15$</th>
<th>$\eta = -20$</th>
<th>$\eta = -50$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa$</td>
<td>$y(s_1)$ $y(s_2)$ $y(s_3)$ $y(s_4)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>1 0 1 0</td>
<td>3.81</td>
<td>11.97</td>
<td>25.04</td>
<td>37.22</td>
<td>49.96</td>
</tr>
<tr>
<td>0.9</td>
<td>0 1 0 1</td>
<td>3.81</td>
<td>11.97</td>
<td>25.04</td>
<td>37.22</td>
<td>49.96</td>
</tr>
<tr>
<td>0.8</td>
<td>1 0 1 0</td>
<td>22.29</td>
<td>43.50</td>
<td>49.03</td>
<td>49.87</td>
<td>50.00</td>
</tr>
<tr>
<td>0.8</td>
<td>0 1 0 1</td>
<td>22.29</td>
<td>43.50</td>
<td>49.03</td>
<td>49.87</td>
<td>50.00</td>
</tr>
<tr>
<td>0.7</td>
<td>1 0 1 0</td>
<td>39.89</td>
<td>49.42</td>
<td>49.97</td>
<td>50.00</td>
<td>50.00</td>
</tr>
<tr>
<td>0.7</td>
<td>0 1 0 1</td>
<td>39.89</td>
<td>49.42</td>
<td>49.97</td>
<td>50.00</td>
<td>50.00</td>
</tr>
<tr>
<td>0.6</td>
<td>1 0 1 0</td>
<td>46.64</td>
<td>49.94</td>
<td>50.00</td>
<td>50.00</td>
<td>50.00</td>
</tr>
<tr>
<td>0.6</td>
<td>0 1 0 1</td>
<td>46.64</td>
<td>49.94</td>
<td>50.00</td>
<td>50.00</td>
<td>50.00</td>
</tr>
<tr>
<td>0.5</td>
<td>1 0 1 0</td>
<td>48.06</td>
<td>49.99</td>
<td>50.00</td>
<td>50.00</td>
<td>50.00</td>
</tr>
<tr>
<td>0.5</td>
<td>0 1 0 1</td>
<td>48.06</td>
<td>49.99</td>
<td>50.00</td>
<td>50.00</td>
<td>50.00</td>
</tr>
<tr>
<td>0.4</td>
<td>1 0 1 0</td>
<td>46.64</td>
<td>49.94</td>
<td>50.00</td>
<td>50.00</td>
<td>50.00</td>
</tr>
<tr>
<td>0.4</td>
<td>0 1 0 1</td>
<td>46.64</td>
<td>49.94</td>
<td>50.00</td>
<td>50.00</td>
<td>50.00</td>
</tr>
<tr>
<td>0.3</td>
<td>1 0 1 0</td>
<td>39.89</td>
<td>49.42</td>
<td>49.97</td>
<td>50.00</td>
<td>50.00</td>
</tr>
<tr>
<td>0.3</td>
<td>0 1 0 1</td>
<td>39.89</td>
<td>49.42</td>
<td>49.97</td>
<td>50.00</td>
<td>50.00</td>
</tr>
<tr>
<td>0.2</td>
<td>1 0 1 0</td>
<td>22.29</td>
<td>43.50</td>
<td>49.03</td>
<td>49.87</td>
<td>50.00</td>
</tr>
<tr>
<td>0.2</td>
<td>0 1 0 1</td>
<td>22.29</td>
<td>43.50</td>
<td>49.03</td>
<td>49.87</td>
<td>50.00</td>
</tr>
<tr>
<td>0.1</td>
<td>1 0 1 0</td>
<td>3.81</td>
<td>11.97</td>
<td>25.04</td>
<td>37.22</td>
<td>49.96</td>
</tr>
<tr>
<td>0.1</td>
<td>0 1 0 1</td>
<td>3.81</td>
<td>11.97</td>
<td>25.04</td>
<td>37.22</td>
<td>49.96</td>
</tr>
</tbody>
</table>

Table 3.3: Table showing the relative frequency of the number of conflicts in a clique of size 4

<table>
<thead>
<tr>
<th>Conflicts</th>
<th>$\eta = 0$</th>
<th>$\eta = -0.5$</th>
<th>$\eta = -1.5$</th>
<th>$\eta = -2.5$</th>
<th>$\eta = -4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.37502</td>
<td>0.45405</td>
<td>0.59696</td>
<td>0.91486</td>
<td>0.98204</td>
</tr>
<tr>
<td>3</td>
<td>0.50006</td>
<td>0.48379</td>
<td>0.38846</td>
<td>0.08466</td>
<td>0.01795</td>
</tr>
<tr>
<td>6</td>
<td>0.12492</td>
<td>0.06216</td>
<td>0.01458</td>
<td>0.00048</td>
<td>0.00002</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conflicts</th>
<th>$\eta = 0$</th>
<th>$\eta = 0.5$</th>
<th>$\eta = 1.5$</th>
<th>$\eta = 2.5$</th>
<th>$\eta = 4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.12492</td>
<td>0.29674</td>
<td>0.98292</td>
<td>0.99921</td>
<td>0.99998</td>
</tr>
<tr>
<td>3</td>
<td>0.50006</td>
<td>0.45477</td>
<td>0.01624</td>
<td>0.00077</td>
<td>0.00002</td>
</tr>
<tr>
<td>4</td>
<td>0.37502</td>
<td>0.24849</td>
<td>0.00084</td>
<td>0.00001</td>
<td>0.00000</td>
</tr>
</tbody>
</table>
Table 3.4: Quantile of Run Lengths for rows and columns of two realizations of two MRF models

<table>
<thead>
<tr>
<th>Dependence</th>
<th>Direction</th>
<th>Length 1</th>
<th>Length 2</th>
<th>Length 3</th>
<th>Length 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\eta = -0.50$</td>
<td>Row</td>
<td>1.0000</td>
<td>0.8858</td>
<td>0.5216</td>
<td>0.0513</td>
</tr>
<tr>
<td></td>
<td>Column</td>
<td>0.9979</td>
<td>0.9520</td>
<td>0.6301</td>
<td>0.0732</td>
</tr>
<tr>
<td>$\eta = 0.50$</td>
<td>Row</td>
<td>0.0000</td>
<td>0.0890</td>
<td>0.7307</td>
<td>0.9997</td>
</tr>
<tr>
<td></td>
<td>Column</td>
<td>0.0000</td>
<td>0.0019</td>
<td>0.0604</td>
<td>0.9214</td>
</tr>
</tbody>
</table>
Table 3.5: Expected number of conflicts in each direction for 2 directional MRF models

<table>
<thead>
<tr>
<th>Model</th>
<th>0.0</th>
<th>1.0</th>
<th>2.0</th>
<th>3.0</th>
<th>4.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 nbrs, N-S</td>
<td>0.500</td>
<td>0.362</td>
<td>0.098</td>
<td>0.020</td>
<td>0.012</td>
</tr>
<tr>
<td>4 nbrs, E-W</td>
<td>0.500</td>
<td>0.361</td>
<td>0.097</td>
<td>0.020</td>
<td>0.012</td>
</tr>
<tr>
<td>8 nbrs</td>
<td>2.999</td>
<td>2.574</td>
<td>2.250</td>
<td>2.046</td>
<td>2.020</td>
</tr>
</tbody>
</table>

Table 3.6: Mean and lag 1 autocorrelations for pseudo-likelihood estimates for Water Vapor Flux Divergence data

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>$\hat{\rho}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa$</td>
<td>0.473</td>
<td>0.202</td>
</tr>
<tr>
<td>$\eta_1$</td>
<td>1.658</td>
<td>0.378</td>
</tr>
<tr>
<td>$\eta_2$</td>
<td>0.778</td>
<td>0.271</td>
</tr>
<tr>
<td>$\eta_3$</td>
<td>0.594</td>
<td>0.727</td>
</tr>
<tr>
<td>$\eta_4$</td>
<td>-0.556</td>
<td>0.542</td>
</tr>
</tbody>
</table>
(a) 4 nearest neighbors structure  \hspace{2cm} (b) 8 nearest neighbors structure

Figure 3.1: Simple neighborhood structures on a regular lattice.
Figure 3.2: 4 nearest neighbor models with $\kappa = 0.5$, and top-left to bottom right $\eta = 0.5, 1, 2, 4$
Figure 3.3: 4 nearest neighbor models with \( \kappa = 0.5 \), and top-left to bottom right \( \eta = -0.5 \), –1, –2, –4
Figure 3.4: Empirical marginal means of fields with various $\kappa$ values and $\eta$ ranging from 0 to -20

Figure 3.5: Left: 4 nearest neighbors. Right: 8 nearest neighbors
Figure 3.6: Example patterns of cliques of size 4
Figure 3.7: Number of conflicts in a 3x3 field where each clique of size 4 has 2 conflicts

Figure 3.8: 8 nearest neighbor models with $\kappa = 0.5$ and $\eta = -100$
Figure 3.9: Mean of Monte Carlo distributions for runs in rows of lengths 1 to 4. Top 2 plots are for four nearest neighbors, bottom 2 are for 8 nearest neighbors.
Figure 3.10: Mean of Monte Carlo distributions for runs in columns of lengths 1 to 4. Top 2 plots are for four nearest neighbors, bottom 2 are for 8 nearest neighbors.
Figure 3.11: Box Plot of $p_r$ values by Run Length and Dependence Strength
Figure 3.12: Realizations from a directional MRF model with fixed dependence in the rows and varying dependence in the columns
Figure 3.13: Mean of Monte Carlo distributions of run lengths for four 4 nearest neighbor MRF models with directional dependence.
Figure 3.14: Realizations from a directional MRF model with fixed dependence in the rows and varying dependence in the columns
Figure 3.15: Locations used in the calculation of Water Vapor Flux Divergence data
Figure 3.16: Plot of the $p_r$ values from the thresholded water vapor flux divergence data.
CHAPTER 4. CONDITIONAL EXPECTATIONS IN BINARY MARKOV RANDOM FIELDS

A paper in preparation
Kenneth Wakeland, Mark S. Kaiser and Daniel Nordman

Abstract

In spatial statistics, positive dependence is most often considered and studied. This is natural as many problems in spatial statistics will naturally produce positive dependence. However, negative dependence is possible when the locations are discrete, as they are in Markov random field models. It is important to understand the effects of negative dependence on Markov random field models for the instances when the parameters of these models are estimated to be negative. In this paper, we build on previous work done to explore the effects of negative dependence on binary Markov random fields by considering the effects negative dependence induced on the conditional expectations of binary Markov random fields. We use simulated data to develop tools to explore the effect of negative dependence and then apply those tools to a real data problem.

4.1 Introduction

Negative dependence in Markov Random Field (MRF) models is commonly associated with Poisson responses. This is because the spatial parameter for a MRF model with Poisson
random variables can only take on negative values (Besag, 1974). There have been attempts to introduce MRF models with positive spatial dependence (Kaiser and Cressie, 1997), but the natural state of this model is negative dependence. In (previous paper reference), negative dependence for binary MRF models were examined in some detail. In this paper, two tools were developed in order to understand the effect negative dependence on MRF models. Conflicts within a clique were used to understand the patterns in data realizations as the model approaches degeneracy. Run length distributions were used to explore the strength and sign of dependence in realizations of MRF models.

These tools are useful for exploring the effect of MRF models on the data level. However, a key aspect of MRF models is the difference between the conditional and marginal expectations. That is, the spatial dependence in MRF models is expressed through the conditional expectations, which will differ from the marginal mean unless there is no spatial dependence. With this in mind, it would be useful to examine how negative dependence impacts the conditional expectation for a binary MRF model. A cursory examination of the conditional expectation provides some conflicting insight. In the case of the standard eight nearest neighbor model with positive dependence provides intuitive results. That is, the conditional expectations are positively related to each other when using directional covariance, Moran’s I (Moran, 1950), or the APLE statistic (Li, Calder, and Cressie, 2007). However, when we consider either the standard eight nearest neighbor model with negative dependence or a 2 direction eight nearest neighbor model, with the north, south, east and west directions having positive dependence, and the inter-cardinal directions having negative dependence, the results do not match intuition. The associations between conditional expectations for this model are negative for all directions. This counter intuitive results prompts us to take a deeper look into the conditional expectations produced by each of the models and try to develop tools to help us assess the patterns in the conditional expectations produced by these MRF models.
In the next section we formally introduce the MRF models used for this paper as well as notation that will be useful for describing said models. In Section 3 we will describe the conditional expectation for the MRF model and the methods we use to explore them. In Section 4, we develop the tools we use for describing the patterns in the conditional expectations. Section 5 contains a real data example in which we apply the tools developed in Section 4. We end the paper with concluding remarks and discussion.

4.2 Model and Neighborhood Formulation

In this section we lay the ground work for Markov random field models as well as define notation for these models. We will also discuss the neighbor structures that will be useful for the subsequent sections.

4.2.1 Model Formulation

We begin by denoting spatial locations in two dimensional space $s_i = (u_i, v_i)$ where $u_i$ and $v_i$ are horizontal and vertical coordinates of $s_i$, respectively. We then define binary random variables at locations $s_i$, $Y(s_i)$, where

$$Y(s_i) = \begin{cases} 
1 & \text{if an event of interest occurs} \\
0 & \text{otherwise} 
\end{cases} \text{ for } i = 1, \ldots, n.$$

We assume that neighborhoods $\{N_i : i = 1, \ldots, n\}$ have been specified for the $s_i$ and take $y(N_i) = \{Y(s_i) : s_j \in N_i\}$ be the set of values of neighbors of $s_i$.

The formulation of binary Markov random field models proceeds through the specification of full conditional distributions for the random variables at each location. The Markov
assumption states these full conditional distributions depend functionally on only the neighboring values $y(N_i)$. That is,

$$\Pr [Y(s_i) = y(s_i) | \{y(s_j) : j \neq i\}] = f(y(s_i) | y(N_i)), \quad y(s_i) = 0, 1,$$  \hspace{1cm} (4.1)

where $f[y(s_i) | y(N_i)] = \exp [y(s_i) A_i\{y(N_i)\} - B_i\{y(N_i)\}], \quad i = 1, \ldots, n,$  \hspace{1cm} (4.2)

In (4.2)

$$A_i\{y(N_i)\} = \logit(\kappa_i) + \sum_{s_j \in N_i} \eta_{i,j}(y(s_j) - \kappa_j),$$  \hspace{1cm} (4.3)

$$B_i\{y(N_i)\} = \log(1 + \exp [A_i\{y(N_i)\}]).$$  \hspace{1cm} (4.4)

The function $A_i$ gives the natural parameter and $B_i$ controls the moments of the conditional binary distributions.

The parameters $\eta_{i,j}$ control spatial dependence between locations $s_i$ and $s_j$ and are subject to the restrictions that $\eta_{i,i} = 0$ and $\eta_{i,j} = \eta_{j,i}$ for all $i, j = 1, \ldots, n$. The $\kappa_i$ are the “large scale” parameters that control the marginal probability of occurrence at location $s_i, i = 1, \ldots, n$. The conditional expectations of the random variables are

$$E [Y(s_i) | y(N_i)] = \frac{\exp [A_i\{y(N_i)\}]}{1 + \exp [A_i\{y(N_i)\}]}, \quad i, j = 1, \ldots, n.$$  \hspace{1cm} (4.5)

If all dependence parameters $\eta_{ij} = 0$ in (4.3), the random variables are all mutually independent, and the expected value in (4.5) is simply $\kappa_i$. If not all the dependence parameters are 0, then at least some marginal and conditional expectations will differ. In these cases, $\kappa_i$ will be approximately the marginal expectation, unless dependencies are “too large” (Kaiser, Caragea, and Furukawa, 2012).
4.2.2 Neighborhood Formulation

In the previous section, we assumed that a neighborhood $N_{i}$ had been specified for each location. In this section, we define a neighborhood structure on a regular grid that will be used through the remainder of this article.

Recall that $s_{i} = (u_{i}, v_{i})$ where $u_{i}$ and $v_{i}$ are the horizontal and vertical coordinates of location $s_{i}$ respectively. Denote sets of adjacent locations on a regular lattice in the cardinal and diagonal directions as

$$
N_{i}^c = \{ s_j : u_j = u_i \pm 1, v_j = v_i \} \cup \{ s_j : u_j = u_i, v_j = v_i \pm 1 \} \quad (4.6)
$$

$$
N_{i}^d = \{ s_j : u_j = u_i \pm 1, v_j = v_i \pm 1 \}.
$$

Using (4.6) with $\kappa_{i} = \kappa$ for all $i = 1, \ldots, n$, and

$$
\eta_{i,j} = \begin{cases} 
\eta_{c} & \text{if } s_j \in N_{i}^c, \\
\eta_{d} & \text{if } s_j \in N_{i}^d, \\
0 & \text{otherwise}
\end{cases}
$$

the natural parameter function defined in (4.3) becomes

$$
A_{i}\{ y(N_{i}) \} = \text{logit}(\kappa) + \eta_{c} \sum_{s_j \in N_{i}^c} (y(s_j) - \kappa) + \eta_{d} \sum_{s_j \in N_{i}^d} (y(s_j) - \kappa). \quad (4.7)
$$

The sets of neighbors (4.6) can be used to define two of the most common neighborhood structures for MRF models, those being four and eight nearest neighbor configurations. Specifically, the four nearest neighbor structure is $N_{i}^4 = N_{i}^c$, and the eight nearest neighbor structure is $N_{i}^8 = N_{i}^c \cup N_{i}^d$. Natural parameter functions for the four nearest neighbor structure and isotropic spatial dependence result from (4.7) by setting $\eta_{d} = 0$. Similarly,
natural parameter functions for the eight nearest neighbor structure and isotropic spatial dependence is given by (4.7) with \( \eta_c = \eta_d \).

### 4.3 Conditional Expectation Fields

Expression (4.5) gives the conditional expectation for a location \( s_i \) given values in the neighborhood \( N_i \). We now denote this conditional expectation at location \( s_i \) as \( E_i(s_i) \) and define the conditional expectation field as the collection \( \mathbf{E} = E(s_i), i = 1, \ldots, n \). These values constitute a random field for which spatial dependence is induced by the dependence parameters of the MRF model (4.7) used to generate the data. In this section, we explore the properties of conditional expectation fields and how they are affected by changes in the dependence parameters of an underlying MRF model.

#### 4.3.1 Patterns in Conditional Expectation Fields

For a preliminary examination of the patterns in conditional expectation fields generated by the models of Section 4.2.2, we make use of the approximate profile likelihood estimate (APLE) (Li, Calder, and Cressie, 2007). The APLE, like Moran’s I, can be used as an estimate of the strength of spatial dependence and is given by

\[
M = \frac{Z' A Z}{Z'(A'A + \lambda' \lambda I)Z}
\]  

(4.8)

where \( Z = (\mathbf{X} - \bar{X} 1_n) \) and \( \mathbf{X} = [X_1, \ldots, X_n]' \) are the variables of interest, either a data field \( \mathbf{Y} \) or a conditional expectation field \( \mathbf{E} \). The vector \( \lambda \) contains the eigenvalues of the adjacency matrix, \( \mathbf{A} \), defined as the \( n \times n \) matrix with \( i,j \)th elements

\[
(a_{ij}) = \begin{cases} 
1 & \text{if } s_j \in N_i, \\
0 & \text{otherwise}
\end{cases}
\]
The statistic (4.8) has the same general interpretation as Moran’s I, but provides a more stable estimator of the strength of spatial dependence (Cressie and Wikle, 2015). We can calculate this statistic for each neighborhood set, $N^c = \{N^c_1, \ldots, N^c_n\}$ and $N^d = \{N^d_1, \ldots, N^d_n\}$, by using the appropriate adjacency matrix.

As an example, we calculated $M$ for both the data field, $Y$, and the conditional expectation field, $E$, for a realization from model (4.7) with $\kappa = 0.5$ and (i) $\eta_c = \eta_d = 0.5$, (ii) $\eta_c = \eta_d = -0.5$, and (iii) $\eta_c = 0.5, \eta_d = -0.5$. The statistic (4.8) is computed separately for each neighborhood set $N^c$ and $N^d$, the results are presented in Table 4.1. The signs of $M$ for the data fields are exactly what we would expect. That is, $M$ is positive when the dependence parameters are positive, and negative when the dependence parameters are negative. But the signs of $M$ do not align with our expectations when we consider the conditional expectation fields. For the conditional expectation field calculated from model with parameters (i), the signs of $M$ are positive. However, for the conditional expectation field calculated from model with parameters (ii), the $M$ value associated with $N_d$ is negative and the $M$ value for $N_c$ is positive. For the conditional expectation field calculated from model with parameters (iii), both values of $M$ are negative. If we simulate each of these models 1000 times, the signs of $M$ match those in Table 4.1 in all data fields and all but 8.2% of the conditional expectation fields from the model with parameters (ii) and 8.4% for the model with parameters (iii). The cause of the change in sign for both of these cases is the $M$ value for $N^d$ changing from negative to positive.

The discrepancies between the signs of $M$ and dependence parameters seen in Table 4.1 motivate a closer examination of the behavior exhibited by fields of conditional expectations. We begin with summary statistics presented in Table 4.2. The values in this table suggest how the phenomena described before is occurring. That is, while the centers (means and medians) of the distributions of the conditional expectation fields are the same, the inter-quartile ranges and variances are different. The variance for the two MRF models with negative dependence is approximately half of the variance of the MRF with only positive
dependence. This provides insight into the topography of the surface of the conditional expectations. Fields generated by models that include negative dependence parameters are not as smooth as those generated by models with only positive dependence parameters. Illustrations of the differences in the topography of the surfaces for these models are presented in Figure 4.1. The panels display the contour plots of the conditional expectation fields and a corresponding slice of the conditional expectation field for a fixed value of $v_i$. The first row displays the plots for the model with positive dependence parameters, and in this plot we observe wide hills and long valleys. The panel on the right of the top row displays the slice of the surface for $v_i = 25$. This plot provides a cleaner view of the smooth transition between hills and valleys seen in the contour plot. The middle row displays the contour and slice plots of the isotropic eight nearest neighbor model with negative dependence. The topography illustrated in these plots is more jagged, with sharper hills and narrower, shorter valleys. Both of the plots in the middle row indicate the hills and valleys obtain the global maximum less often than the plots in the top row even though those maximum values are the same for all the fields illustrated in Figure 4.1 (see Table 4.2). The third row displays the contour and slice plot for the two direction eight nearest neighbor model. The plots for this model are similar to the plots displayed in the middle row, although the hills are even more peaked and valleys are more narrow. The frequency of the hills and valleys that obtain the global maximum is again reduced from the field of the first row. The fact we observe less extreme values for the middle and bottom rows is consistent with the marginal variance for these two fields being lower than the field represented in the top row (see Table 4.2). Because the conditional expectation at a focal location is a function of the data values at neighboring locations, variability in the conditional expectation field is a function of variability in sets of neighboring values. These neighboring values enter conditional expectations as sums, and we will call those sums in the neighboring sets $N^c_i$ and $N^d_i$ conditioning sets.
4.3.2 Frequency of Conditioning Sets

To develop the concept of conditioning sets in a formal manner, we define a conditioning set at location $s_i$ to be

$$L_i = (a, b) = \left\{ \sum_{s_j \in N_c^i} y(s_j), \sum_{s_j \in N_d^i} y(s_j) \right\}. \quad (4.9)$$

For binary random fields, expression (4.9) has 25 possible values which allows us to construct a frequency distribution for the conditioning sets. For the purpose of illustration, we once again generate realizations from three MRF models with parameters $\kappa = 0.50$ and (i) $\eta_c = \eta_d = 0.50$, (ii) $\eta_c = \eta_d = -0.50$ and (iii) $\eta_c = 0.50 \eta_d = -0.50$. Each simulated data set was generated on a lattice of size 200x200 which included a border strip of two rows and columns. A border strip includes locations that are not included as focal points, but only for conditioning values for the interior locations. The frequency distribution of the conditioning sets was calculated, the resulting frequency distributions are given in Table 4.3. The differences between the conditioning sets for the models primarily occur at the conditioning sets that correspond to the extreme values of conditional expectation. That is, $L_i = (4, 4), (0, 0)$ correspond to the highest and lowest conditional expectation, respectively, for the model with parameters (i), and the lowest and highest conditional expectation, respectively, for the models with parameters (ii). The differences in the frequency distributions for these two points is striking, with the model (i) placing approximately 20 times the weight on these values compared to model (ii). A similar situation happens for the model with parameters (iii). In this case, $L_i = (4, 0), (0, 4)$ corresponds to the highest and lowest conditional expectations respectively. However, the frequency of these configurations is the same as the model with parameters (ii). This trend continues if we consider the second highest/lowest conditional expectation values ($L_i = (4, 3), (0, 1)$ for models (i) and (ii) and $L_i = (4, 1), (0, 3)$ for model (iii)) with the model with parameters (i) placing far more weight on these configurations.
than the models with parameters \((ii)\) and \((iii)\). These observations are consistent with the marginal variance of the conditional expectations being lower for models with parameters \((ii)\) and \((iii)\) than model with parameters \((i)\). This is not the only observation we can take from these frequency distributions however. Specifically, we will leverage this information in the next section to develop an empirical estimator for the conditional expectation.

### 4.3.3 Empirical Conditional Expectation

Throughout the previous sections, we have made use of the conditional expectation for MRF models for our assessments. However, this statistic requires us to know the parameters as well as the conditioning set of the model. This information is rarely, if ever, known, although the conditioning set is often assumed. In the previous section, we introduced the frequency of the conditioning sets and using this idea, we develop the concept of an empirical conditional expectation which we will allow us to consider fields of conditional expectations without having explicit knowledge of the model parameters. Let

\[
\hat{E}_k = \frac{\sum_{i=1}^{n} Y(s_i)I(L_i = (a_k, b_k))}{\sum_{i=1}^{n} I(L_i = (a_k, b_k))}, \tag{4.10}
\]

where \(k = 1, \ldots, K\) enumerates the possible values of the conditioning set \(L_i\) defined in \((4.9)\). In this case, \(K = 25\), though this depends on the neighborhood structure. Expression \((4.10)\) allows us to define an empirical conditional expectation for a location \(s_i\) with conditioning set values \(L_i = (a_k, b_k)\) as

\[
\hat{E}(s_i) = \hat{E}_k. \tag{4.11}
\]

To assess the accuracy of the estimator \(\hat{E}(s_i)\), we perform a Monte Carlo experiment in which we generate 2000, \(150 \times 150\) realizations from models \((i)\), \((ii)\) and \((iii)\) described in Section 3.2. For each realization, we perform a simple linear regression of \(\hat{E}(s_i)\) against \(E(s_i)\) for the appropriate model. The Monte Carlo estimates of the slope, intercept and \(R^2\) values are displayed in Table 4.4. The Monte Carlo estimate of \(R^2\) for each model are at least
0.96, the estimates of the intercepts are all approximately 0 and the estimates of the slopes are all close to 1. These results illustrate the estimator $\hat{E}(s_i)$ can be used as an estimator of $E_i(s_i)$, and only requires an assumption on the structure of the conditioning sets.

We can further generalize this procedure by splitting the neighborhoods into directional components

$$N_i^1 = \{s_j : u_j = u_i, v_j = v_i \pm 1 \},$$
$$N_i^2 = \{s_j : u_j = u_i \pm 1, v_j = v_i \},$$
$$N_i^3 = \{(u_{i+1}, v_{i+1}), (u_{i-1}, v_{i-1}) \},$$
$$N_i^4 = \{(u_{i-1}, v_{i+1}), (u_{i+1}, v_{i-1}) \},$$

where $N_i^c = N_i^1 \cup N_i^2$ and $N_i^d = N_i^3 \cup N_i^4$. Using such a decomposition, $L_i$ (4.9) becomes

$$L_i = (a, b, c, d) = \left\{ \sum_{s_j \in N_i^1} y(s_j), \sum_{s_j \in N_i^2} y(s_j), \sum_{s_j \in N_i^3} y(s_j), \sum_{s_j \in N_i^4} y(s_j) \right\}. \quad (4.13)$$

Using expression (4.13), let

$$\hat{E}_k = \frac{\sum_{i=1}^n Y(s_i) I(L_i = (a_k, b_k, c_k, d_k))}{\sum_{i=1}^n I(L_i = (a_k, b_k, c_k, d_k))}. \quad (4.14)$$

Once again, $k = 1, \ldots, K$ enumerates the possible values of the conditioning set $L_i$ defined in (4.13), and $K = 81$ for this neighborhood structure. However, we can continue to use (4.11) to define the empirical conditional expectation for the directional case at location $s_i$ with $L_i = (a_k, b_k, c_k, d_k)$. 


We assess the accuracy of this estimator using the same method as previously. That is, we generate realizations from the model

\[ A_i(y(N_i)) = \text{logit}(\kappa) + \eta_1 \left( \sum_{s_j \in N^1_i} [y(s_j) - \kappa] \right) + \eta_2 \left( \sum_{s_j \in N^2_i} [y(s_j) - \kappa] \right) + \eta_3 \left( \sum_{s_j \in N^3_i} [y(s_j) - \kappa] \right) + \eta_4 \left( \sum_{s_j \in N^4_i} [y(s_j) - \kappa] \right) \]  

(4.15)
on a 150x150 lattice with \( \kappa = 0.50, \eta_1 = \eta_2 = \eta_4 = 0.50 \) and \( \eta_3 = -0.50 \). We calculate the conditional expectations using the model parameters and then we estimate the empirical conditional expectation using (4.14), (4.11). As before, we fit a simple linear regression model with the conditional expectation as the independent variable and the empirical conditional expectation as the dependent variable for each realization. This results in Monte Carlo estimates of \( R^2 = 0.967 \), intercept 0.003 and slope 0.995. These results match well with the results using (4.10). And while these results match well, we will concentrate on the empirical conditional expectation defined in (4.10), though we will utilize (4.14) when we explore the real data problem.

The definitions of empirical conditional expectation developed in this section allows us to define and apply tools for exploring the structure of the conditional expectation field to a real data set for which we do not have explicit knowledge of these parameters. In the next section, we will define two such tools.

### 4.4 Measures of Heterogeneity and Scale

The previous section illustrated the differences in the conditional expectation fields when a negative dependence parameter is used in the model. Using \( \hat{E}(s_i) \) defined in (4.11), we can explore and diagnose patterns in conditional expectation fields of data for which we do
not have explicit knowledge of the dependence parameters. We begin by defining a measure of scale for conditional expectation fields.

### 4.4.1 Scale of the Conditional Expectation Field

Recall in Figure 4.1 the topography of the conditional expectation fields become less smooth as we progress down the rows. However, this is not the only pattern that we can observe. The last row of Figure 4.1 displays the contour and slice plots for the 2 direction MRF model with both positive and negative dependence parameters. The contour plot displays a surface for which the topography is jagged on a local level. That is, there is a considerable amount of variability when we consider small, two dimensional blocks of the field. However, while the values of the conditional expectation within the blocks are heterogeneous, if we divide the field up into many of these small blocks, the distribution of values between the blocks would look similar to each other. In contrast, if we were to divide the contour plot in the top row of Figure 4.1 into these same small blocks, the conditional expectation values within the blocks would be similar, but the distribution of values between the blocks would look different. In fact, it would require larger blocks in order for the distribution of values between blocks to look similar to each other for the realization in the top row. While this is an informal observation, we will develop this into a formal exploratory and diagnostic tool to quantify the notion of scale in the conditional expectation field.

The first step in establishing the measure of scale is to divide the field into two dimensional blocks. In order to accomplish this, we use moving blocks with a side length of \( b = 3, \ldots, k - 2 \) where each block has a total number of points \( b^2 \) and \( k \) is the number of horizontal/vertical points of the lattice. The second step in this process is to calculate the mean of each block \( b = 3, \ldots, k - 2 \), then calculate the variance of the means of the blocks, denoted as \( \sigma_b^2 \). We define the scale of the process to be

\[
b^* = \min \left( \inf \left\{ b : \sigma_b^2 < \tau^* \right\}, k \right),
\]  

(4.16)
where \( t^* \) is a threshold we define. The value of \( b^* \) is \( k \) if the between block variance never falls below the threshold \( t^* \).

To illustrate this process, we generate three realizations from MRF on a \( 33 \times 33 \) lattice with parameters \( \kappa = 0.50 \) and (i) \( \eta_c = \eta_d = 0.50 \), (ii) \( \eta_c = \eta_d = -0.50 \), and (iii) \( \eta_c = 0.50 \), \( \eta_d = -0.50 \). We then calculate the empirical conditional expectation field for each of these realizations. We apply the process described in the preceding paragraph to this data and the results are given in Figure 4.2. The results are striking and consistent with the informal observations noted previously. That is, the separation between the curves when the block size is low is pronounced with the between block variability of the means being smaller for the model with parameters (iii) than the model with parameters (ii), which are in turn smaller than the model with parameters (i). If we use the threshold \( t^* = e^{-8} \), the scale for the model with parameters: (i) is 21; (ii) is 11; (iii) is 5. The threshold here is arbitrary, but from Figure 4.2 we can see the ordering here will stay the same no matter what threshold we choose. That is, any threshold value will result in the scale of the conditional expectation field increasing from (iii) to (ii) to (i). With the notion of scale defined here, we transition to the measure of heterogeneity.

4.4.2 Measure of Heterogeneity

In the previous section we defined the notion of scale as the the largest block in which the variance of the block mean conditional expectation is less than a threshold \( t^* \). However, this is only one measure we can use to quantify the phenomena illustrated in Figure 4.1. In this section, we introduce the measure of heterogeneity for the conditional expectation field as a function of of the measure of scale we introduced in the previous section. We will once again use the estimator \( \hat{E}(s_i) \) in efforts to utilize this tool for real data applications.
We refer to heterogeneity primarily as it concerns to how rough the conditional expectation field is. With this in mind, we define our measure of the heterogeneity of a block based on a common measure of the roughness of a curve (Ramsay, Hooker, and Graves, 2009)

\[ R(A_j) = \frac{1}{|A_j|} \int_{A_j} \left( \frac{d^2 E(s)}{du^2} \right)^2 + \left( \frac{d^2 E(s)}{dv^2} \right)^2 + \left( \frac{d^2 E(s)}{dudv} \right)^2 d\mu(u \times v), \quad (4.17) \]

where the area \( A_j \) is a block with size \( b \). Using (4.17), we define the heterogeneity of the conditional expectation field to be

\[ H(b^*) = \frac{1}{[k-(b^*+1)]^2} \sum_{j=1}^{[k-(b^*+1)]^2} R(A_j), \quad (4.18) \]

where \( b^* \) is the scale of the field given in (4.16) and \( k \) is the length of the side of the lattice. The second derivatives of the conditional expectation field in (4.17) are found using finite difference methods (Olver, 2014) (Chaudhry, 2007). Recall the scales, \( b^* \), for the models with parameters \( \kappa = 0.50 \) and (i) \( \eta_c = \eta_d = 0.50 \), (ii) \( \eta_c = \eta_d = -0.50 \), and (iii) \( \eta_c = 0.50 \), \( \eta_d = -0.50 \) are (i) \( b^*_{(i)} = 21 \); (ii) \( b^*_{(ii)} = 11 \); \( b^*_{(iii)} = 5 \). Using these scale values we calculate \( H(b^*) \) in (4.18) for each model to be: (i) \( H(b^*_{(i)}) = 0.0975 \); (ii) \( H(b^*_{(ii)}) = 0.2336 \); (iii) \( H(b^*_{(iii)}) = 0.4186 \). These values of heterogeneity confirm the observations made with regards to Figure 4.1. The model with isotropic positive dependence, (i), has a heterogeneity value much lower than the models that incorporate negative dependence, (ii) and (iii). The model with parameters (ii) has a lower heterogeneity value than model with parameters (iii), which we also observed previously.

The methods of measuring scale and heterogeneity that we have described here allow us to quantify the patterns we observed in the contour and slice plots given in Figure 4.1. We can now apply these methods to a real data problem to diagnose the patterns seen in the empirical conditional expectation fields.
4.5 Water Vapor Flux Reanalysis Data

We now turn our attention to water vapor flux divergence data obtained from (Hobbs, 2014) and Chapter 2. This data is computed from reanalysis products, specifically the reanalysis data from 2 meteorological models: the Modern-Era Retrospective Analysis for Research Applications (MERRA) and the North American Regional Reanalysis (NARR). Both of these models were computed at 3 hour intervals for a month at a time on a $33 \times 33$ lattice that covers a region of the center United States. The specific month we will use for our demonstration is May 2005. Further details on this data can be found in (Hobbs, 2014) and Chapter 2.

The water vapor flux divergence/convergence data are continuous, however, it is of use to consider whether the value of the water vapor flux is positive (convergence) or negative (divergence). With this in mind we threshold the data at 0. That is, let $Z(s_i, t)$ be the value of the water vapor flux divergence at location $s_i$ and time $t$. We then define $Y(s_i, t) = I[Z(s_i, t) \geq 0]$. This results in a $33 \times 33$ binary Markov random field at $t = 1, \ldots, 248$ time points, as there are 31 days each with 8 time points in a day. An curious aspect of this data is the lack of a spatially varying large scale structure on the conditional expectation. That is, during the course of attempting to fit naive models to this data, we employed a radial spatial basis function similar to those found in (Hobbs, 2014). However, the estimated parameters from these basis functions were all estimated to be zero. With our end goal being to develop a Bayesian model that incorporates the spatial and temporal dependence in which we fit using the Markov Chain Monte Carlo method described by (Liang, 2010), we need to more carefully explore the spatial dependence of this data. We build off the results in Chapter 2 along with the tools developed in the previous sections to accomplish this goal.

The tools developed in Chapter 2 led to the conclusion the model given in expression (4.15) might be more appropriate for these data than the model given in (4.7). Using (4.14) and (4.11), we find the empirical conditional expectation field for each time point
$t = 1, \ldots, 248$. We then calculate the scale, $b_t^*$, and the heterogeneity, $H(b_t^*)$, for each $t = 1, \ldots, 248$ time points. The results are summarized in Table 4.5. In comparing these results to the results from Section 4, we see scales, $b_t^*$, larger than the model with all positive spatial dependence parameters (model (i)), but the heterogeneity values, $H(b_t^*)$, are also larger, on average, than the heterogeneity values observed in the model with both positive and negative dependence parameters (model (iii)). The scale and heterogeneity values also display moderate levels of lag 1 autocorrelation, 0.360 and 0.694 for scale and heterogeneity respectively. This autocorrelation is also observed in the run length results from Chapter 2. These results and the results from Chapter 2 indicate the presence of at least moderate temporal dependence, along with a complicated spatial dependence structure. That is, the results from this paper and Chapter 2 indicate the spatial dependence parameters likely have temporal dependence. However, the large values of both $b_t^*$ and $H(b_t^*)$ could indicate a more complicated spatial structure than a directional eight nearest neighbor model, as there seems to be a lack of identifiable large scale structure on the conditional expectation.

4.6 Conclusions

In this chapter, as well as Chapter 2, we have seen that negative dependence in Markov random field models can induce contradictory, or counter-intuitive, behavior. In Chapter 2, the focus was placed on patterns in the data, while throughout this article we focused on patterns in the conditional expectation fields. When negative dependence is incorporated in a MRF model, we observed counter-intuitive behavior in the induced dependence in the of the conditional expectation field, which we measured using the APLE statistic. In the process of exploring the patterns in the conditional expectation fields we developed the empirical conditional expectation field in order to explore the patterns in the conditional expectation fields of data in which we do not have explicit knowledge of the true model. We then used
the empirical conditional expectation field to find the statistics we refer to as the scale, $b^*$, and heterogeneity, $H(b^*)$.

We then applied the methods for finding the scale and heterogeneity to the thresholded water vapor flux divergence data. When we compare the behavior observed in the water vapor flux divergence data to the simulated data of the previous section, we observe the scale, $b_t^*$, was larger, on average, than the largest scale in the simulated data and the heterogeneity, $H(b_t^*)$, was also greater, on average, than the largest heterogeneity. This is unusual as we saw the scale and heterogeneity had an inverse relationship in the simulated data. This would seem to indicate there is a more complicated structure in the thresholded water vapor flux divergence than previously suggested. A topic for future studies would explore these more complicated dependence structures and extend the techniques used here for such dependence structures.

### 4.7 Bibliography


Durbin, James (1973). *Distribution theory for tests based on the sample distribution function*. SIAM.

Hobbs, Jonathan Michael (2014). “Characterizing diurnal and interannual variability in the atmosphere through physical and stochastic models”. In:


# Appendix: Tables and Figures

Table 4.1: The APLE for data and conditional expectation fields originating from various MRF models

<table>
<thead>
<tr>
<th>Data Fields</th>
<th>( \eta_c )</th>
<th>( \eta_d )</th>
<th>( N^c )</th>
<th>( N^d )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>0.085</td>
<td>0.086</td>
<td></td>
</tr>
<tr>
<td>-0.5</td>
<td>-0.5</td>
<td>-0.039</td>
<td>-0.060</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>-0.5</td>
<td>0.047</td>
<td>-0.055</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conditional Expectation Fields</th>
<th>( \eta_c )</th>
<th>( \eta_d )</th>
<th>( N^c )</th>
<th>( N^d )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>0.203</td>
<td>0.183</td>
<td></td>
</tr>
<tr>
<td>-0.5</td>
<td>-0.5</td>
<td>0.132</td>
<td>-0.027</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>-0.5</td>
<td>-0.125</td>
<td>-0.025</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.2: Summary Statistics for Conditional Expectation Fields

<table>
<thead>
<tr>
<th>( \eta_c )</th>
<th>( \eta_d )</th>
<th>( \bar{E} )</th>
<th>Var((E))</th>
<th>Min</th>
<th>( Q_1 )</th>
<th>Median</th>
<th>( Q_3 )</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>0.511</td>
<td>0.046</td>
<td>0.119</td>
<td>0.378</td>
<td>0.500</td>
<td>0.731</td>
<td>0.881</td>
</tr>
<tr>
<td>-0.5</td>
<td>-0.5</td>
<td>0.503</td>
<td>0.0215</td>
<td>0.119</td>
<td>0.378</td>
<td>0.500</td>
<td>0.622</td>
<td>0.881</td>
</tr>
<tr>
<td>0.5</td>
<td>-0.5</td>
<td>0.500</td>
<td>0.0218</td>
<td>0.119</td>
<td>0.378</td>
<td>0.500</td>
<td>0.622</td>
<td>0.881</td>
</tr>
</tbody>
</table>
Table 4.3: Frequency distributions of the conditioning sets for 3 MRF models

<table>
<thead>
<tr>
<th>$L_i = (a_i, b_i)$</th>
<th>$\eta_c = 0.50$</th>
<th>$\eta_c = -0.50$</th>
<th>$\eta_c = 0.50$</th>
<th>$\eta_d = -0.50$</th>
<th>logit[$A{y(N_i)}$]</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0,0)</td>
<td>3.37</td>
<td>0.15</td>
<td>0.49</td>
<td>0.119</td>
<td>0.881 0.500</td>
</tr>
<tr>
<td>(0,1)</td>
<td>4.88</td>
<td>0.65</td>
<td>1.24</td>
<td>0.182</td>
<td>0.818 0.378</td>
</tr>
<tr>
<td>(0,2)</td>
<td>3.02</td>
<td>1.51</td>
<td>1.36</td>
<td>0.269</td>
<td>0.731 0.269</td>
</tr>
<tr>
<td>(0,3)</td>
<td>0.94</td>
<td>1.29</td>
<td>0.80</td>
<td>0.378</td>
<td>0.622 0.182</td>
</tr>
<tr>
<td>(0,4)</td>
<td>0.10</td>
<td>0.46</td>
<td>0.15</td>
<td>0.500</td>
<td>0.500 0.119</td>
</tr>
<tr>
<td>(1,0)</td>
<td>3.19</td>
<td>1.26</td>
<td>2.25</td>
<td>0.182</td>
<td>0.818 0.622</td>
</tr>
<tr>
<td>(1,1)</td>
<td>7.72</td>
<td>5.20</td>
<td>6.87</td>
<td>0.269</td>
<td>0.731 0.500</td>
</tr>
<tr>
<td>(1,2)</td>
<td>7.53</td>
<td>8.58</td>
<td>8.50</td>
<td>0.378</td>
<td>0.622 0.378</td>
</tr>
<tr>
<td>(1,3)</td>
<td>3.52</td>
<td>7.08</td>
<td>5.33</td>
<td>0.500</td>
<td>0.500 0.269</td>
</tr>
<tr>
<td>(1,4)</td>
<td>0.65</td>
<td>2.21</td>
<td>1.29</td>
<td>0.622</td>
<td>0.378 0.182</td>
</tr>
<tr>
<td>(2,0)</td>
<td>1.65</td>
<td>3.01</td>
<td>3.00</td>
<td>0.269</td>
<td>0.731 0.731</td>
</tr>
<tr>
<td>(2,1)</td>
<td>6.73</td>
<td>10.57</td>
<td>10.98</td>
<td>0.378</td>
<td>0.622 0.622</td>
</tr>
<tr>
<td>(2,2)</td>
<td>10.53</td>
<td>15.70</td>
<td>15.26</td>
<td>0.500</td>
<td>0.500 0.500</td>
</tr>
<tr>
<td>(2,3)</td>
<td>7.30</td>
<td>10.62</td>
<td>10.70</td>
<td>0.622</td>
<td>0.378 0.378</td>
</tr>
<tr>
<td>(2,4)</td>
<td>1.95</td>
<td>3.18</td>
<td>3.15</td>
<td>0.731</td>
<td>0.269 0.269</td>
</tr>
<tr>
<td>(3,0)</td>
<td>0.64</td>
<td>2.32</td>
<td>1.28</td>
<td>0.378</td>
<td>0.622 0.818</td>
</tr>
<tr>
<td>(3,1)</td>
<td>3.56</td>
<td>6.99</td>
<td>5.28</td>
<td>0.500</td>
<td>0.500 0.731</td>
</tr>
<tr>
<td>(3,2)</td>
<td>7.87</td>
<td>8.40</td>
<td>8.69</td>
<td>0.622</td>
<td>0.378 0.622</td>
</tr>
<tr>
<td>(3,3)</td>
<td>8.32</td>
<td>5.44</td>
<td>6.93</td>
<td>0.731</td>
<td>0.269 0.500</td>
</tr>
<tr>
<td>(3,4)</td>
<td>3.53</td>
<td>1.19</td>
<td>2.43</td>
<td>0.818</td>
<td>0.182 0.378</td>
</tr>
<tr>
<td>(4,0)</td>
<td>0.11</td>
<td>0.55</td>
<td>0.18</td>
<td>0.500</td>
<td>0.500 0.881</td>
</tr>
<tr>
<td>(4,1)</td>
<td>1.04</td>
<td>1.33</td>
<td>0.75</td>
<td>0.622</td>
<td>0.378 0.818</td>
</tr>
<tr>
<td>(4,2)</td>
<td>3.19</td>
<td>1.39</td>
<td>1.40</td>
<td>0.731</td>
<td>0.269 0.731</td>
</tr>
<tr>
<td>(4,3)</td>
<td>5.30</td>
<td>0.72</td>
<td>1.25</td>
<td>0.818</td>
<td>0.182 0.622</td>
</tr>
<tr>
<td>(4,4)</td>
<td>3.35</td>
<td>0.19</td>
<td>0.44</td>
<td>0.881</td>
<td>0.119 0.500</td>
</tr>
</tbody>
</table>
Table 4.4: Simple Linear Regression of $\hat{E}_i$ against $E_i$ by model

<table>
<thead>
<tr>
<th>$\eta_c$</th>
<th>$\eta_d$</th>
<th>$R^2$</th>
<th>$b_0$</th>
<th>$b_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>0.980</td>
<td>0.982</td>
<td>0.0088</td>
</tr>
<tr>
<td>-0.5</td>
<td>-0.5</td>
<td>0.960</td>
<td>0.966</td>
<td>0.0170</td>
</tr>
<tr>
<td>0.5</td>
<td>-0.5</td>
<td>0.986</td>
<td>0.999</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

Table 4.5: Summary Statistics for Scale ($b_t^*$) and heterogeneity ($H(b_t^*)$)

<table>
<thead>
<tr>
<th>Quantile</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>0%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b_t^*$</td>
<td>24.44</td>
<td>3.11</td>
<td>13</td>
<td>23</td>
<td>25</td>
<td>26</td>
<td>33</td>
</tr>
<tr>
<td>$H(b_t^*)$</td>
<td>0.47</td>
<td>0.13</td>
<td>0.23</td>
<td>0.39</td>
<td>0.45</td>
<td>0.52</td>
<td>1.06</td>
</tr>
</tbody>
</table>
Figure 4.1: The left column displays contour maps for three realizations of MRF models, the right column is a corresponding slice of the surface.
Figure 4.2: Plot of variance between block means for block sizes $b = 3, \ldots, 100$
CHAPTER 5. GENERAL CONCLUSIONS

5.1 Summary

This dissertation introduced several methods of examining dependence in binary random fields. In Chapter 2, efforts were made to incorporate temporal dependence into Markov random field models. This was done in three ways: the first method simply treated time as another dimension in the Markov random field model; the second method used a conditional specification to incorporate temporal dependence on each point in the field; the third method used a hierarchical structure to incorporate a temporal process on the parameters of the Markov random field model. While we were able to successfully fit simulated data with these models, we obtained unexpected results when they were applied to the water vapor flux divergence data. That is, we obtained spatial and temporal dependence that was large and in one case, the large scale parameter did not match the marginal mean. Our attempts at fitting a large scale structure with spatial basis functions resulted in all zero parameter estimates, and when we fit directional dependence models, we obtained another unexpected result, that is, negative dependence.

Chapter 3 was inspired by the negative dependence result obtained by fitting the directional models to the water vapor flux divergence data. In this chapter, we examined patterns in data realizations from Markov random field models with negative dependence parameters. This examination was done using two quantities, the number of conflicts in a clique and the distribution of run lengths. The number of conflicts within a clique was introduced to serve as a bridge between our intuition of degenerate patterns in Markov random field mod-
els constructed using a four nearest neighbor structure and negative or positive dependence and Markov random field models constructed using an eight nearest neighborhood structure with negative dependence. The run length distribution tool was inspired by the resulting patterns in the number of conflicts within a clique. The run length distribution tool allowed us to determine the strength and direction of the dependence in a specific direction. When this was applied to the water vapor flux divergence data, we observed evidence of negative dependence in one direction, the northeast-southwest direction specifically. There was also evidence of auto-correlation in this quantity, which could imply a temporal structure needs to be incorporated on the spatial dependence parameters.

In Chapter 3, we examined the patterns in the data generated by Markov random field models, in Chapter 4, we provided a different method of examining the effects negative dependence. In this chapter, we explore the effect of negative spatial dependence parameters on the conditional expectations of Markov random field models. In order to examine the conditional expectations at each location, the empirical conditional expectation field was developed in order to obtain estimates of the conditional expectation without requiring knowledge of the true parameters used to generate the data. Along with the empirical conditional expectation, notions of scale and heterogeneity of the conditional expectation field were developed to quantify the striking, and counter intuitive, differences between the conditional expectation fields resulting from models only positive dependence and those which have at least one negative dependence parameter. In general, models with only positive dependence parameters have larger values for scale and smaller values for heterogeneity than models with at least one negative dependence parameter. These tools were once again applied to the water vapor flux divergence data and, once again, unexpected behavior was noticed. That is, the scale parameter was larger, on average, than the scale parameters from a model with only positive dependence, but the heterogeneity was also larger than the models with one positive and one negative dependence parameter. This led to the conclusion there is more structure in this data than can be explained with these simple models.
5.2 Future Work

The results of Chapters 3 and 4 indicate extensions to the models used in Chapter 2. In Chapter 3, the run length distribution tool indicated there was evidence of autocorrelation in spatial dependence. Incorporating temporal dependence in the spatial dependence parameter can be done most easily using the same method as the Dynamic MRF model in Chapter 2. However, this adds significant computational complexity to the model estimation and it is not clear at this time what temporal dependence structure should be used (i.e. which ARIMA model, or another dependence structure).

Chapter 4 also indicated there is evidence of autocorrelation in the spatial dependence, however, the results from this chapter also provided evidence the spatial dependence might not be able to be explained by the eight nearest neighborhood models used in this dissertation. Possible extensions to this work include moving beyond eight nearest neighborhoods to 24 nearest neighbors, a model we have termed to contain a second ring. Another possible extension to the eight nearest neighborhood model used here is to introduce terms for cliques of size greater than two. That is, we would have terms for interactions, not just between pairs, but between triples, quadruples, etc. as well. These two extensions are not mutually exclusive. There is work to be done to inspect what effect models with 24 nearest neighbors, 3-way/4-way dependence, or a combination of both have on run length distributions, scale and heterogeneity, or if there are more meaningful measures for these models.

Once this work is done, we would return to the work already done in Chapter 2, and, with the results of Chapters 3, 4 and this future work, find a more appropriate model to fit the water vapor flux divergence data.
APPENDIX. SUPPLEMENTAL INFORMATION

There is a concern about how we generate realizations from the Markov random field models that we discuss throughout this document. Specifically, we generated realizations of binary MRF models for this document using a Gibbs sampler in which we leverage the conditional independence of non-neighboring locations in a MRF model. We refer to these non-neighboring values as concliques (Kaiser, Lahiri, and Nordman, 2012). Using the notation from (Kaiser, Lahiri, and Nordman, 2012), we perform two (for 4 nearest neighborhood structures) or four (for eight nearest neighborhood structures) updates for each iteration of the Gibbs sample of the field. When we simulate a realization using a model with a 4 nearest neighborhood structure for example, we first update locations that are members of conclique 1 given the values of members in conclique 2. The next update uses the values of the members of conclique 1 to update the values of the members of conclique 2. The result of these two updates represents one iteration of the Gibbs sampler. For a model using 8 nearest neighbors, the only change is instead of two updates, four updates per iteration are required. That is, the values of the members of conclique 1 are updated using the values of the members of the other concliques. Then the values of the members of conclique 2 are updated using the values of the other concliques and so on. As is typical with a Gibbs sampler, a burn-in was used, and for all of the simulations in this document, the burn-in was set to 1000 Gibbs iterations.

The concern raised with the use of this method is we do not know that the fields we obtain are draws from the joint distribution for the specified MRF model. To address this concern, we compare our Gibbs sampler with results we obtain using an exact sampler (Hughes and
This sampler is known to produce an exact sample from the joint of a specified binary MRF model. The purpose of this appendix is to examine in what situations these samplers give us the same output and in which situations does the output differ.

The first quantity we will examine is the joint distribution of the interior 2x2 block of a MRF on a 10x10 lattice with a four nearest neighbors structure. The interior points chosen were \( i = 45, 46, 55, 56 \). These points fall directly in the middle of the lattice and have a similar interpretation to the four points on a circle described in Section 3.3. However, we cannot directly compute the joint distribution of these points, so we estimate the joint distribution using Monte Carlo. That is, we generate 10000 realizations from a MRF model on a 10x10 lattice, a four nearest neighbors structure, and parameters \( \kappa = 0.50, \eta = 0.5 \).

Using these 10000 realizations, we calculate a joint distribution for those 4 interior points. We then repeat this procedure 1000 times to produce our Monte Carlo approximation to the joint. The results of this are displayed in Table A1. It is clear from this table the two samplers produce results well within the Monte Carlo error.

The second quantity focuses on the idea conflicts in a clique of size four. For this quantity, we attempt a parallel analysis to the analysis whose results are presented in Table 3.3, with a couple of changes. Specifically, we generate 2000 realizations of MRFs on a 10x10 lattice with eight nearest neighbors and parameters \( \kappa = 0.70 \) and \( \eta = \{\pm0.5, \pm1.0, \pm3.0, \pm6.0\} \), and calculate the proportion of cliques of size four with 0, 3, and 4 conflicts for the realizations with positive dependence and 2, 3 and 6 conflicts for realizations with negative dependence for each generated field. The change in \( \kappa \) is due to the fact the exact sampler algorithm we used could not produce realizations with \( \kappa = 0.50 \) and \( \eta > 4.0 \). Our choice of \( \eta = \{\pm0.5, \pm1.0, \pm3.0, \pm6.0\} \) is also affected by this phenomenon, as realizations on a 10x10 lattice with \( \kappa = 0.70 \) and \( \eta > 6.0 \) can be produced, but the algorithm requires 5 or more minutes to produce one realization. The results of this Monte Carlo experiment are presented in Table A2. Once again we see the the Monte Carlo estimates of the proportion of conflicts
for each sampling method are nearly identical with the differences being well within the Monte Carlo Error.

The last quantity we use to determine differences between the two methods of generating realizations for MRF is the Monte Carlo experiment described in Section 4.2. We repeat this Monte Carlo experiment with both the Gibbs sampler and the exact sampler. The results are presented in Figure A1. And again, we see the distributions produced by these samplers appear to be identical.

With the results of these three Monte Carlo experiments above, we feel it is justified to continue to use the Gibbs sampler for the results of this paper.

Bibliography


Table A1: Monte Carlo estimates of the distribution of 4 interior points of a MRF with 4 nearest neighbors on a 10x10 lattice along with Monte Carlo standard errors.

<table>
<thead>
<tr>
<th>$\omega$</th>
<th>$P(\omega)$ (%)</th>
<th>MC Error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Exact</td>
<td>Gibbs</td>
</tr>
<tr>
<td>0,0,0,0</td>
<td>10.063</td>
<td>10.049</td>
</tr>
<tr>
<td>0,0,0,1</td>
<td>6.050</td>
<td>6.055</td>
</tr>
<tr>
<td>0,0,1,0</td>
<td>6.051</td>
<td>6.058</td>
</tr>
<tr>
<td>0,0,1,1</td>
<td>6.045</td>
<td>6.052</td>
</tr>
<tr>
<td>0,1,0,0</td>
<td>6.060</td>
<td>6.053</td>
</tr>
<tr>
<td>0,1,0,1</td>
<td>3.645</td>
<td>3.645</td>
</tr>
<tr>
<td>0,1,1,0</td>
<td>6.058</td>
<td>6.045</td>
</tr>
<tr>
<td>0,1,1,1</td>
<td>6.050</td>
<td>6.043</td>
</tr>
<tr>
<td>1,0,0,0</td>
<td>6.053</td>
<td>6.052</td>
</tr>
<tr>
<td>1,0,0,1</td>
<td>6.052</td>
<td>6.056</td>
</tr>
<tr>
<td>1,0,1,0</td>
<td>3.632</td>
<td>3.638</td>
</tr>
<tr>
<td>1,0,1,1</td>
<td>6.048</td>
<td>6.055</td>
</tr>
<tr>
<td>1,1,0,0</td>
<td>6.048</td>
<td>6.035</td>
</tr>
<tr>
<td>1,1,0,1</td>
<td>6.054</td>
<td>6.057</td>
</tr>
<tr>
<td>1,1,1,0</td>
<td>6.043</td>
<td>6.052</td>
</tr>
<tr>
<td>1,1,1,1</td>
<td>10.048</td>
<td>10.056</td>
</tr>
</tbody>
</table>
Table A2: Monte Carlo estimates of the proportion of each possible conflict within a field for increasing $\eta$. The Monte Carlo errors are in parenthesis.

<table>
<thead>
<tr>
<th>Sampler</th>
<th># Conflicts</th>
<th>$\eta = -0.50$</th>
<th>$\eta = -1.00$</th>
<th>$\eta = -3.0$</th>
<th>$\eta = -6.0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gibbs</td>
<td>2</td>
<td>0.282 (0.0014)</td>
<td>0.323 (0.0014)</td>
<td>0.418 (0.0017)</td>
<td>0.588 (0.0018)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.505 (0.0013)</td>
<td>0.548 (0.0013)</td>
<td>0.557 (0.0016)</td>
<td>0.41 (0.0018)</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.213 (0.0015)</td>
<td>0.129 (0.0011)</td>
<td>0.025 (5e-04)</td>
<td>0.002 (1e-04)</td>
</tr>
<tr>
<td>Exact</td>
<td>2</td>
<td>0.279 (0.0014)</td>
<td>0.322 (0.0014)</td>
<td>0.418 (0.0016)</td>
<td>0.582 (0.0017)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.507 (0.0013)</td>
<td>0.549 (0.0013)</td>
<td>0.557 (0.0016)</td>
<td>0.416 (0.0017)</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.214 (0.0015)</td>
<td>0.129 (0.001)</td>
<td>0.025 (5e-04)</td>
<td>0.002 (1e-04)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sampler</th>
<th># Conflicts</th>
<th>$\eta = 0.50$</th>
<th>$\eta = 1.0$</th>
<th>$\eta = 3.0$</th>
<th>$\eta = 6.0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gibbs</td>
<td>0</td>
<td>0.286 (0.0019)</td>
<td>0.882 (0.0013)</td>
<td>1 (0)</td>
<td>1 (0)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.464 (0.0014)</td>
<td>0.101 (0.0011)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.25 (0.0015)</td>
<td>0.017 (4e-04)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Exact</td>
<td>2</td>
<td>0.288 (0.0019)</td>
<td>0.883 (0.0013)</td>
<td>1 (0)</td>
<td>1 (0)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.464 (0.0013)</td>
<td>0.1 (0.0011)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.247 (0.0015)</td>
<td>0.016 (4e-04)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>
Figure A1: Box plots for Gibbs and Perfect Samplers