2013

Methods for modeling and forecasting wind characteristics

Lisa Bramer
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/13605

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.
Methods for modeling and forecasting wind characteristics

by

Lisa M. Bramer

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:
Petrută C. Caragea, Co-major Professor
Mark S. Kaiser, Co-major Professor
Alicia Carriquiry
Daniel J. Nordman
Eugene S. Takle

Iowa State University
Ames, Iowa
2013
Copyright © Lisa M. Bramer, 2013. All rights reserved.
DEDICATION

_Service to others is the rent you pay for your room here on Earth_

_-Muhammad Ali_

I dedicate this work to Brian Flagstad who embodied this quotation and whose service has made a lasting difference in my life. His caring nature, support, willingness to listen, and humor were unmatched. As a teacher, he aided in the growth of my interest and capability in mathematics. More importantly, the positive impact he made on me personally is unmistakably evident to me everyday. Much like the marker smears he left on my books and papers, these influences are sure to remain throughout my life.
# TABLE OF CONTENTS

| LIST OF TABLES                      | vii |
| LIST OF FIGURES                    | ix |
| ACKNOWLEDGEMENTS                   | xv |

## CHAPTER 1. INTRODUCTION            | 1  |
1.1 Background                      | 1  |
1.2 Overview                        | 2  |
  1.2.1 Bias corrected wind speed forecasts | 2  |
  1.2.2 Bivariate modeling of wind vector components’ bias | 2  |
  1.2.3 Spatial properties of wind speed and practical modeling strategies | 3  |

## CHAPTER 2. HIERARCHICAL MODELS FOR THE BIAS OF METEOROLOGICAL FORECASTS OF WIND SPEED AND DIRECTION | 4  |
2.1 Introduction & Motivation       | 4  |
2.2 Average Meteorological Forecast (AMF) Model | 8  |
2.3 Data-Driven Models             | 10 |
  2.3.1 Time Series Model Description | 10 |
  2.3.2 Training Period for Time Series Model Fitting | 11 |
  2.3.3 Data Transformation         | 13 |
  2.3.4 Model Selection             | 14 |
  2.3.5 Time Series Model Performance | 15 |
2.4 Model Output Statistics (MOS) Method | 19 |
  2.4.1 OCF Model Description       | 19 |
  2.4.2 OCF Model Implementation    | 20 |
### Chapter 2: Incorporating Bias Adjustment of Wind Speed Forecasts

#### 2.5 Hierarchical Model for Bias Adjustment of Wind Speed Forecasts

- **2.5.1 Hierarchical Model Description** ........................................ 25
- **2.5.2 Prior Distribution Specifications** ....................................... 27
- **2.5.3 Conditional Posteriors of the Hierarchical Model** .................. 27
- **2.5.4 Monte Carlo Method Implementation** .................................... 28
- **2.5.5 Combining Model Forecasts** ............................................ 29
- **2.5.6 Hierarchical Model Performance** ....................................... 35

#### 2.6 Incorporating Wind Direction into the Hierarchical Model

- **2.6.1 Modeling Circular Variables** ............................................ 41
- **2.6.2 Model Implementation** .................................................. 42
- **2.6.3 Obtaining Bias-Corrected Wind Speed Forecasts** .................... 44
- **2.6.4 Obtaining Bias-Corrected Wind Direction Forecasts** ................. 44
- **2.6.5 Performance of the Hierarchical Model Incorporating Direction** .... 46

#### 2.7 Discussion and Future Work ............................................... 51

### Chapter 3: Incorporating Dependence in a Hierarchical Model for the Bias of Meteorological Wind Vector Forecasts

- **3.1 Background and Motivation** .............................................. 53
- **3.2 Bivariate Model for Wind Speed and Direction** ......................... 55
  - **3.2.1 General Model Description** ........................................ 55
  - **3.2.2 Correlation Structure** ............................................... 58
  - **3.2.3 Priors** .......................................................... 61
- **3.3 Conditional Posterior Distributions & Parameter Estimation** ........ 62
- **3.4 Simulation Study Setup** ................................................ 63
  - **3.4.1 Values Necessary for Simulation** .................................. 63
  - **3.4.2 General Steps for Simulation Studies** .............................. 64
- **3.5 Simulation Studies for Determining Parameter Estimation Adequacy** ... 65
  - **3.5.1 Strength of Dependence** ........................................ 66
3.5.2 Sample Size ............................................................. 72
3.5.3 Interaction of Sample Size and Correlation Strength ......... 73
3.6 Consequences of Improper Model Specification .................. 78
3.7 Forecasting Performance on Observed Data ......................... 79
3.8 Discussion/Future Work ................................................. 82

CHAPTER 4. SPATIAL MODELING OF THE WIND SPEED PROCESS

AT THE FARM-LEVEL ..................................................... 84
4.1 Introduction & Motivation .............................................. 84
4.2 Data ................................................................. 86
4.3 General Markov Random Field (MRF) Model ....................... 88
4.4 Unidirectional Gaussian MRF Model ............................... 90
4.5 Model Estimation ...................................................... 90
4.6 Neighborhood Radius Selection ..................................... 91
  4.6.1 Parameter Space of \( \eta \) ........................................ 92
  4.6.2 Simulation Study: Estimated Dependence Parameter Sensitivity .... 93
  4.6.3 Simulation Study: Neighborhood Selection Criterion ....... 95
  4.6.4 Radius Selection Based on Observed Data .................. 96
4.7 Directional Gaussian MRF Model .................................... 101
  4.7.1 Accounting for Wind Direction ................................ 101
  4.7.2 Model Specification ............................................. 102
4.8 Hybrid Gaussian MRF Model ....................................... 105
4.9 Simulation Based Model Comparisons ............................... 107
  4.9.1 Marginal Means ................................................. 109
  4.9.2 Marginal Variances ............................................ 111
  4.9.3 Conditional Expectations ..................................... 114
  4.9.4 Generalized Spatial Residuals ................................. 115
  4.9.5 Neighborhood Structure Choice Related to Weather Conditions .. 123
4.10 Tolerance Angle Selection .......................................... 126
4.11 Model Estimation on Observed Data .............................. 128
CHAPTER 5. GENERAL CONCLUSIONS ........................................... 139

5.1 General discussion ....................................................... 139
5.2 Recommendation for future research ................................. 139

APPENDIX A. CONDITIONAL POSTERIOR DERIVATIONS FOR HI-
ERARCHICAL MODEL ......................................................... 141

APPENDIX B. CONDITIONAL POSTERIOR DERIVATIONS FOR BI-
VARIATE HIERARCHICAL MODEL ....................................... 148

BIBLIOGRAPHY ............................................................... 164
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Number of times time series model was selected by AICc criterion</td>
<td>14</td>
</tr>
<tr>
<td>2.2</td>
<td>Summary statistics of the difference in mean absolute error (MAE) values for OCF, persistence, and average meteorological forecast over 102 cases</td>
<td>22</td>
</tr>
<tr>
<td>2.3</td>
<td>Summary statistics of the difference in mean absolute error (MAE) values for hierarchical, OCF, persistence, and average meteorological forecast over 102 cases</td>
<td>38</td>
</tr>
<tr>
<td>2.4</td>
<td>Summary statistics of the difference in mean absolute error (MAE) values for vector hierarchical, hierarchical, OCF, persistence, and average meteorological forecasts over 102 cases</td>
<td>48</td>
</tr>
<tr>
<td>3.1</td>
<td>Value ranges for parameters, forecasts, and observed values for simulation purposes</td>
<td>64</td>
</tr>
<tr>
<td>3.2</td>
<td>Summary statistics of $\kappa$ for 200 simulated datasets per value of $\kappa$</td>
<td>70</td>
</tr>
<tr>
<td>3.3</td>
<td>Mean square error (MSE) of parameter estimates for all 200 simulated cases, over various sample sizes</td>
<td>73</td>
</tr>
<tr>
<td>3.4</td>
<td>Mean absolute relative errors for $\alpha_{1j}$ over various sample size and dependence strength values</td>
<td>74</td>
</tr>
<tr>
<td>3.5</td>
<td>Mean absolute relative errors for $\alpha$ over various sample size and dependence strength values</td>
<td>75</td>
</tr>
<tr>
<td>3.6</td>
<td>Average MAE values of 24 hour-ahead forecasts over 200 simulations</td>
<td>78</td>
</tr>
<tr>
<td>3.7</td>
<td>Mean absolute error values for regression parameters over 200 datasets simulated from the independence model</td>
<td>79</td>
</tr>
</tbody>
</table>
3.8 Summary statistics of the difference in mean absolute error (MAE) values for vector hierarchical, bivariate vector hierarchical, hierarchical, and average meteorological forecasts over 102 cases 82

4.1 Summary of model characteristics. 108

4.2 Summary statistics of mean absolute error (MAE) values for simulated marginal means across 165 cases. 111

4.3 Summary statistics of mean absolute error (MAE) values for simulated marginal variances across 165 cases. 113

4.4 Summary statistics for conditional expectation MAE values, computed over 165 cases, for all models. 114

4.5 Models 3 and 4 Kolmogorov-Smirnov p-value comparisons by mean wind strength. 125

4.6 Summary statistics of mean absolute error (MAE) values for simulated marginal means, for varying tolerance possibilities (Model 3), across 172 cases. 127

4.7 Summary statistics for Kolmogorov-Smirnov p-values, for varying tolerance choices (Model 3), across 172 cases. 127

4.8 Number of hourly cases for which there is evidence of a lack-of-fit at varying levels of significance and all considered tolerance values. 128

4.9 Comparison of dependence parameter magnitudes for 172 estimated cases. 133
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Mean absolute error (MAE) values of the time series model forecasts, averaged over 98 cases,</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>for varying training period lengths.</td>
<td></td>
</tr>
<tr>
<td>2.2</td>
<td>Time series models’ estimates of the data transformation parameter, $\kappa$, values for 98</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>cases.</td>
<td></td>
</tr>
<tr>
<td>2.3</td>
<td>Time series model mean absolute error (MAE) values for a 24 hour ahead period, by case.</td>
<td>16</td>
</tr>
<tr>
<td>2.4</td>
<td>Mean absolute error (MAE) values for time series model by hour-ahead forecast, for 98 cases.</td>
<td>16</td>
</tr>
<tr>
<td>2.5</td>
<td>Time series model mean absolute error (MAE) values for a 5 hour ahead period, by case.</td>
<td>17</td>
</tr>
<tr>
<td>2.6</td>
<td>Time series, average meteorological forecast, and persistence model forecasts for cases 10 and</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>31.</td>
<td></td>
</tr>
<tr>
<td>2.7</td>
<td>Model output statistics model, persistence model, and average meteorological forecast mean</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>absolute error (MAE) values for a 24 hour ahead period, by case.</td>
<td></td>
</tr>
<tr>
<td>2.8</td>
<td>Model output statistics, average meteorological forecasts’, and persistence model forecasts</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>for cases 49 and 82.</td>
<td></td>
</tr>
<tr>
<td>2.9</td>
<td>Model output statistics, average meteorological forecast, and persistence model forecasts for</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>cases 55 and 50.</td>
<td></td>
</tr>
<tr>
<td>2.10</td>
<td>Prior predictive distributions of the mean bias, $b_{j,t}$, for each of the twelve meteorological models for hour 54 on November 11, 2008.</td>
<td>30</td>
</tr>
</tbody>
</table>
Prior predictive distributions of the bias-corrected wind speed forecast, $y^*_j,t$, for each of the twelve meteorological models for hour 54 on November 11, 2008.

Prior predictive distribution of the bias-corrected wind speed forecast, $y^*_j,t$, resulting from a finite mixture distribution with equal weight parameters for hour 54 on November 11, 2008.

Prior predictive distribution of the bias-corrected wind speed forecast, $y^*_j,t$, resulting from a finite mixture distribution with inverse variance weight parameters, for hour 54 on November 11, 2008.

Hierarchical model forecasts and forecast intervals for both weighting schemes for case 97: September 5, 2010.

Hierarchical average model, hierarchical weighted model, persistence model, and average meteorological forecast mean absolute error (MAE) values for a 24 hour ahead period, by case.

Hierarchical average model, hierarchical weighted model, average meteorological forecast, and persistence model forecasts for cases 47 and 42.

Hierarchical average model, hierarchical weighted model, average meteorological forecast, and persistence model forecasts for case 78: August 26, 2009.

Hierarchical model, persistence model, and average meteorological forecast mean absolute error (MAE) values for a 24 hour ahead period, by case.

Hierarchical model, average meteorological forecast, and persistence model forecasts for cases 70 and 95.


Hierarchical model and average meteorological forecast wind direction forecasts for cases 35 and 40.
\section*{2.22} Hierarchical model forecasts, forecast intervals, and average meteorological forecast wind direction forecasts for cases 35 and 40.  

\section*{2.23} Vector hierarchical model, hierarchical model, persistence model, and average meteorological forecast mean absolute error (MAE) values for a 24 hour ahead period, by case.  

\section*{2.24} Hierarchical model, vector hierarchical model, and average meteorological forecast wind speed forecasts for cases 55 and 84.  

\section*{2.25} Vector hierarchical model, hierarchical model, and average meteorological forecast wind speed forecasts with 95\% intervals for case 75: August 23, 2009.  

\section*{2.26} Vector hierarchical model, hierarchical model, and average meteorological forecast mean absolute error (MAE) values for a 24 hour ahead period, by case.  

\section*{3.1} Cartesian plane sections of interest.  

\section*{3.2} Correlation between $u$ and $v$ produced from a 10 degree window of direction values centered at $\theta$.  

\section*{3.3} Sample correlation of wind vector bias components by section for 102 cases.  

\section*{3.4} MCMC chain for $\kappa = -0.9$.  

\section*{3.5} MCMC convergence for 3 chains when $\kappa = -0.9$ is used for simulation.  

\section*{3.6} MCMC sampling chains for parameters when $\kappa = 0.9$, red lines indicate true value of parameter used to simulate data.  

\section*{3.7} MCMC chains for $\kappa$ for various dependence strengths, red lines indicate true value of parameter used to simulate data.  

\section*{3.8} Mean relative errors for parameters estimates over 200 simulated datasets for varying values of $\kappa$.  

\section*{3.9} Mean relative errors for correlation function parameters estimates over 200 simulated datasets for varying values of $\kappa$.  
3.10 Boxplots of the mean difference for regression parameters over varying magnitudes of $\kappa$ when $T = 30$. ........................................ 76
3.11 Boxplots of the mean difference for correlation function parameters over varying magnitudes of $\kappa$ when $T = 30$. .............................. 77
3.12 MAE values for 24 hour-ahead forecast periods of 200 simulated datasets, by model ............................................................... 80
3.13 Hierarchical average model, hierarchical independent vector model, bivariate hierarchical model, and average meteorological forecast mean absolute error (MAE) values for a 24 hour ahead period, by case. .... 81
3.14 Observed wind direction used in training period (black lines) compared to average meteorological forecast directions used in forecast period (red lines). .................................................. 82

4.1 Locations of wind turbines for the wind farm ......................... 86
4.2 Interpolated average hourly wind speed values (m/s) for the wind farm on January 8, 2009, 0700 (LST). ................................. 88
4.3 Estimated $\eta$ values across neighborhood radii for nine combinations of $\eta$ and $\delta$. The horizontal line represents the true $\eta$ used to generate simulated data, and the red point represents the estimated value of $\eta$ at the true value of $\delta$ used to generate simulated data. ............... 94
4.4 Estimated negative log-psedolikelihood values across neighborhood radii for nine combinations of $\eta$ and $\delta$. Red points represent the estimated value at the true value of $\delta$ used to generate simulated data. .............. 97
4.5 Average estimated negative log-psedolikelihood values across 250 simulated datasets and neighborhood radii for nine combinations of $\eta$ and $\delta$. Red points represent the estimated value at the true value of $\delta$ used to generate simulated data. ....................... 98
4.6 Estimated negative log-psedolikelihood values for two observed hours, across six neighborhood radii. .................................. 99
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.7</td>
<td>Average estimated negative log-pseudolikelihood values across datasets for six neighborhood radii.</td>
</tr>
<tr>
<td>4.8</td>
<td>Difference in estimated negative log-pseudolikelihood value and minimum value, across observed cases, for six neighborhood radius values.</td>
</tr>
<tr>
<td>4.9</td>
<td>Illustrations of unidirectional MRF neighborhoods for turbine 50.</td>
</tr>
<tr>
<td>4.10</td>
<td>Average hourly wind speed values (m/s) for the wind farm for several different wind directions. Wind direction is given by the arrow at the bottom left of each plot.</td>
</tr>
<tr>
<td>4.11</td>
<td>Illustration of directional neighborhoods for turbine 50.</td>
</tr>
<tr>
<td>4.12</td>
<td>Illustration of hybrid neighborhood for turbine 50.</td>
</tr>
<tr>
<td>4.13</td>
<td>Marginal means of 500 simulated datasets for each model, for two example cases.</td>
</tr>
<tr>
<td>4.14</td>
<td>Marginal means of 500 simulated datasets for each model, for two example cases where Model 2 performs poorly.</td>
</tr>
<tr>
<td>4.15</td>
<td>Marginal variances of 500 simulated datasets for each model corresponding to June 19, 2008, 1000 (LST).</td>
</tr>
<tr>
<td>4.16</td>
<td>Marginal variances and means of 500 simulated datasets for each model corresponding to October 18, 2009, 1700 (LST).</td>
</tr>
<tr>
<td>4.17</td>
<td>Boxplots of MAE values of conditional expectations compared to observed data values over 165 cases.</td>
</tr>
<tr>
<td>4.18</td>
<td>Empirical cdfs for all models, for two hourly cases.</td>
</tr>
<tr>
<td>4.19</td>
<td>Empirical cdfs of all models based on observed data and one simulated dataset for May 27, 2009, 0500 (CST).</td>
</tr>
<tr>
<td>4.20</td>
<td>Empirical cdfs for Model 2 based on observed data (blue line) and simulated datasets (purple lines) for two example cases.</td>
</tr>
<tr>
<td>4.21</td>
<td>Model 1 and 2 p-values corresponding to Kolmogorov-Smirnov test, $T_1$, for 165 cases.</td>
</tr>
<tr>
<td>4.22</td>
<td>Model 1 and 3 and 2 and 3 p-values corresponding to Kolmogorov-Smirnov test statistics, $T_1$, for 165 cases.</td>
</tr>
<tr>
<td>Section</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>4.23</td>
<td>Model 3 and 4 p-values corresponding to Kolmogorov-Smirnov test statistics, (T_1), for 165 cases. 123</td>
</tr>
<tr>
<td>4.24</td>
<td>Model 3 and 4 p-values, for Kolmogorov-Smirnov test, versus mean wind speed. 125</td>
</tr>
<tr>
<td>4.25</td>
<td>Estimated parameter values for (\alpha) and (\tau^2) plotted against observed mean marginal wind speeds. 129</td>
</tr>
<tr>
<td>4.26</td>
<td>Estimated parameter values for (\eta_1) and (\eta_2) plotted against observed mean marginal wind speeds. 130</td>
</tr>
<tr>
<td>4.27</td>
<td>Estimated parameter values for (\alpha) and (\tau^2) plotted against observed marginal wind speed variance. 131</td>
</tr>
<tr>
<td>4.28</td>
<td>Estimated parameter values for (\eta_1) and (\eta_2) plotted against observed marginal wind speed variance. 132</td>
</tr>
<tr>
<td>4.29</td>
<td>Estimated dependence parameter values for 172 hourly cases. 133</td>
</tr>
<tr>
<td>4.30</td>
<td>Wind vectors corresponding to hourly cases where (\hat{\eta}_1 &lt; 0). 134</td>
</tr>
<tr>
<td>4.31</td>
<td>Time series plots of (\hat{\eta}_1) for two stretches of consecutive hourly observations. 136</td>
</tr>
<tr>
<td>4.32</td>
<td>Sample autocorrelation plots for the differenced (\hat{\eta}_1) values for two stretches of consecutive hourly observations. 136</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

There are many people who have been influential throughout this journey and without whom none of this would have been possible. It is certain that I would not have come this far and arrived at the end of this process in one piece without the following people.

I would like to thank my advisors for their guidance in the writing of this dissertation. I am sincerely grateful for Dr. Petruţa Caragea who has served as an advisor and mentor throughout my graduate career at Iowa State. I would also like to thank my family and friends for their love and support. Their confidence in my abilities never waivers, no matter what endeavors I choose to tackle. Many thanks to Brian Hornbuckle for exposing me to the possibility of graduate school and to the Hornbuckle family for their kindness. I never imagined that an internship would lead to an extended stay in Iowa for graduate school and finding a second family there. I am grateful for their warmth and continued support.

There have been many struggles and bumps in the road throughout this process and times when I lost confidence. I am very much indebted to Kim Flagstad for her kindness, advice, and support that has continued throughout the years. She has acted as a tremendous example of strength and selflessness and has been a source of hope and compassion in my life. Last but certainly not least, I would like to thank Joann Seeman Smith who has seen me through the good and the bad of this process and life as it happened. She has been instrumental in much of my growth in balancing both academic and personal challenges. I am extremely grateful for her genuineness, compassion, honesty, and humor throughout my journey.
CHAPTER 1. INTRODUCTION

1.1 Background

The rising costs and negative environmental effects of traditional, nonrenewable energy sources have led to increased research regarding the viability of alternative energy sources. Wind has been the fastest growing source of electricity generation in the world since the 1990s (Schreck et al., 2008). Additionally, wind occurs naturally and is renewable with the added benefit of being more environmentally friendly than traditional energy sources. However, in comparison to many other countries, wind energy resources in the United States remain largely untapped (Wiser and Bolinger, 2012). The viability of wind in the energy market is hindered by the fact that there is currently no cost effective method for storing the energy that a wind turbine produces, and utility companies must declare the amount of energy they will produce prior to the actual production.

In a recent report, the U.S. Department of Energy (DOE) established a desired scenario in which wind energy will provide 20% of the nations electrical energy by the year 2030. In addition to this goal, the DOE has called for and funded a large number of research studies to investigate various areas related to wind. These areas include but are not limited to: wind speed forecasting, post-processing meteorological wind speed forecasts, wind power forecasts, and a general understanding of wind behavior in time and space (Department of Energy, 2008).

In this dissertation we present developments in the modeling and understanding of wind characteristics, including:

1. wind speed and wind vector forecasts through the modeling of the bias of meteorological
2. the relationship between wind speed and direction through a bivariate wind vector bias model for forecasting

3. the spatial behavior of wind at a farm-level

4. methods for addressing practical issues related to modeling wind speed spatially

In the remainder of this chapter we provide an overview of the papers comprising this dissertation. Additionally, Chapter 5 provides a summary of the dissertation and discussion of future plans.

1.2 Overview

1.2.1 Bias corrected wind speed forecasts

Many methods for the post-processing and bias correction of physical meteorological wind speed forecasts have been developed over the last 20 years. The method proposed here models bias directly based on a hierarchical model developed by Tebaldi et al. (2005). We further extend this framework to simultaneously model wind speed and direction. Comparisons of model performance are made relative to traditional types of models used in wind speed forecasting. Additionally, this approach produces forecasts at an elevation equivalent to turbine hub height and provides prediction intervals and point forecasts for each hour.

1.2.2 Bivariate modeling of wind vector components’ bias

Wind speed and direction are thought to be related throughout the meteorological community. Numerous studies have attempted to model bivariate distributions of these variables, however, few studies have forecasted speed and direction simultaneously based on meteorological forecasts. Recent developments by Sloughter et al. (2013), Schuhen et al. (2012), and Pinson et al. (2009) model the relationship between wind speed and direction with the use of meteorological model forecasts. The method proposed here models the bias of wind speed and
direction, through wind vector components, rather than the observed quantities directly. Properties of the model are assessed through simulation studies, and model forecasts are compared to independent bias correction model forecasts.

1.2.3 Spatial properties of wind speed and practical modeling strategies

The understanding of wind speed spatially at the scale of a wind farm is especially useful to utility companies and can lead to better wind power forecasting abilities. Minimal research has been conducted at this scale due to a lack of available data. This work uses data observed at the turbine-level to investigate the behavior of wind speeds and possible dependence structures. Additionally, methods for making practical modeling decisions are proposed, as well as methods for assessing model performance, on an irregular spatial domain.
CHAPTER 2. HIERARCHICAL MODELS FOR THE BIAS OF
METEOROLOGICAL FORECASTS OF WIND SPEED AND
DIRECTION

2.1 Introduction & Motivation

Concerns regarding the negative consequences arising from the burning of fossil fuels, evidence of global warming, and rising fuel prices have led to a push for the development of renewable energy sources. In an effort to promote “green energy”, the U.S. Department of Energy (DOE) established the goal of wind energy providing 20% of the nation’s electrical energy by the year 2030 (Department of Energy, 2008). The presence of wind energy in the electricity market continues to increase in the U.S., as evidenced by the 16% growth in cumulative wind power capacity in 2011 (Wiser and Bolinger, 2012). Despite this growth, wind energy provided only 2.29% of all energy in the United States in 2011 (Department of Energy, 2011).

The reliability of wind is one of the major limiting factors of wind energy as a source of energy in the electricity market. Wind power is directly related to wind speed, which is variable in both space and time. Additionally, there is currently no cost-efficient method for storing wind energy produced by a wind turbine, so it must be introduced into the electrical grid immediately. Because utility companies must declare the amount of energy they will produce at a future time, to ensure projected energy demands will be met, knowledge about future wind behavior is necessary in order for wind energy to be viable (Schreck et al., 2008). The financial implications of wind forecasting are also of great consequence. According to Schreck et al. (2008), errors in forecasted wind speeds of only 1%, for a 100-MW wind facility, can lead to a loss of $12,000,000 over the lifetime of the facility. Thus, the ability to accurately and precisely
forecast wind speeds has become increasingly important as more wind power is introduced into electricity markets.

The majority of literature regarding wind speed forecasting can be broken down into three types of models: time series models, model output statistics (MOS), and approaches combining meteorological forecasts with probabilistic models.

Statistical models consistently outperform other models when making very short-term (1 to 3 hours ahead) forecasts (Giebel et al., 2003). These methods take past values and sometimes other explanatory variables into account for making predictions. Forecast methods such as Kalman filters and autoregressive moving average (ARMA) models have been investigated for making short-term forecasts. Kalman filters have been shown to provide better forecasts than persistence models for forecast periods less than an hour ahead (see, for example, Kalman (1960), Bossanyi (1985), and Louka et al. (2008)). Several researchers have shown that ARMA models are useful for wind speed simulation (Chou and Corotis (1981) and Kamal and Jafri (1997)). A large amount of research has been done on the ability of the family of ARMA models to forecast wind speeds. For example, Brown et al. (1984), Torres et al. (2005), and Nfaouï et al. (1996) accounted for the non-normality and seasonality of wind speed by transforming and standardizing wind speed observations and then fit ARMA models to the transformed data. Bivona et al. (2011) used a seasonal autoregressive integrated moving average (SARIMA) structure to model transformed wind speed values. While, Liu et al. (2011) used a generalized autoregressive conditional heteroskedasticity (GARCH) model to account for the heteroskedasticity of wind speed caused by turbulence. Other authors have introduced time-varying coefficients into regression models to forecast wind speed (e.g. Reikard (2008)) or regime-switching models in space and time (e.g. Hering and Genton (2010)). While the specific details of these models differ from study to study, the results of studies consistently indicate that statistical models perform well when forecasting wind speeds only a few hours in advance. However, these models are severely limited in their ability to forecast wind speed for periods more than a few hours ahead (Monteiro et al., 2009).
When wind speed forecasts for forecast horizons longer than a few hours in advance are desired, it is necessary to use alternative forecasting methods to purely statistical models. Numerical weather prediction (NWP) models, based on physical principles, were developed to make meteorological predictions and forecasts. Current meteorological models are capable of making predictions for numerous meteorological variables, including wind speed, from 1 to 72 hours ahead. However, systematic biases are often present in NWP forecasts, as there is uncertainty in unknown, initial atmospheric conditions and parameterizations of physical processes required for NWP models (Giebel et al., 2003). It is common for ensembles of NWP models with varying conditions, models and physical parameterizations to be run and combined when generating wind speed forecasts. Ensembles of physical forecasts are more accurate than single members (Toth and Kalnay (1993) and Stensrud et al. (2000)). Therefore, much research has been conducted investigating methods of bias correction and combining ensemble member forecasts.

Typically, studies have developed methods of post-processing NWP forecasts using model output statistics. Woodcock and Engel (2005) presented a MOS method, operational consensus forecasts (OCF), that applied a weighted average bias adjustment, where weights for bias components were determined by a training period of data. Hibon and Evgeniou (2005) and Fritsch et al. (2000) showed that combining models by OCF methodology to obtain a forecast was more accurate, on average, than any best model for a particular day. Additionally, Engel and Ebert (2007) used the OCF method to produce bias-corrected forecasts for several meteorological variables, including wind speed, in the Australian Region. Howard and Clark (2007) used physical models of wind flow to correct wind speed forecasts. While, Hart et al. (2004) used multiple linear regression and stepwise model selection to downscale wind speed forecasts to locations of interest. Thorarinsdottir and Gneiting (2010) implemented a heteroskedastic censored regression model to improve NWP wind speed forecasts. MOS methodology has been shown to improve NWP wind speed forecasts across different NWP models and locations throughout the world. However, the magnitude of improvement, the time scale, and the choice
Depending on the given energy market, small errors in wind speed forecasts can lead to significant financial losses. When the potential for financial profit/loss is large, it may be useful for an energy company to have the ability to quantify uncertainty in bias-corrected wind speed forecasts or to have access to probabilities associated with wind speed exceeding a certain threshold. In these cases, the ability to obtain a full probability distribution function (pdf) is desirable. However, most traditional MOS methods only have the capability of producing point forecasts. Gneiting et al. (2005) introduced a method for obtaining predictive distributions based on ensemble model output statistics (EMOS). Additionally, Thorarinsdottir and Gneiting (2010) implemented this method in the context of a heteroskedastic censored regression model to obtain the predictive distribution of wind speed forecasts. Bayesian model averaging (BMA) was developed to post-process ensembles while producing predictive pdfs (Gneiting and Raftery, 2005). Sloughter et al. (2010) used BMA to improve upon NWP forecasts and obtain predictive pdfs for the maximum wind speed for locations in the Pacific Northwest. These methods led to improvements in NWP forecasts while also providing a full predictive pdf for their respective forecasts.

This work is done in partnership with an electric utility company located in the Midwest. Electricity market traders have the need for improving the accuracy of wind speed forecasts, as they must make a bid of wind energy that will be produced, on an hourly basis, for the following day. Our utility company partner obtains meteorological forecasts for 1 to 54 hour-ahead forecasts, where the first 30 hours of forecasts are used as a training period and the last 24 hours are combined to produce wind speed forecasts for the next day, for which energy bids are made. Most research on wind speed forecasting makes use of data and forecasts at a height of 10 m, the height at which official wind observations available to the public are taken. In their report, the DOE declared the need for observations, and model validation and improvement at the level of turbine hub heights. At the 10 m height, winds are greatly affected by surface friction and are usually not representative of winds at the level of interest for a wind turbine.
Here, we have meteorological observations, including wind speed, available at a height of 80 m (the hub height of the utility company’s wind turbines). We develop a hierarchical structure to model the bias of meteorological wind speed forecasts, on an hourly basis, within the context of our partnering utility company’s forecasting problem.

The method proposed here is based on a model first introduced by Tebaldi et al. (2005) in the context of combining information from a collection of atmosphere-ocean general circulation model output and observations to predict temperature changes. The model developed by Tebaldi et al. (2005) allows for ensemble member-specific variability in temperature change forecasts. Additionally, the individual precision parameters are used to determine weights of ensemble members when forecasts are combined. We have adapted this model to combine wind speed forecasts’ biases from an ensemble consisting of 12 members, relevant covariate information, and observations within a Bayesian framework. Furthermore, we have extended this framework to simultaneously model wind speed and direction. Our approach is novel in that it produces forecasts at a higher elevation than previously considered (80 m instead of 10 m), is based on a short training period (hourly data, for 30 hours), for a longer temporal window (hourly, for 24 hours ahead), and proposes a simple model to combine observed and forecasted wind speed, direction, and covariates to produce point forecasts as well as prediction intervals for each hourly forecast.

2.2 Average Meteorological Forecast (AMF) Model

Numerical weather models based on meteorological physical processes have been developed by meteorologists and can be used to generate wind speed forecasts. These models can also produce forecasts for additional weather variables, such as wind direction, temperature, and pressure. The Weather Research and Forecasting (WRF) model is a widely used numerical weather prediction model developed by the National Oceanic and Atmospheric Administration (NOAA) and the National Center for Atmospheric Research (NCAR). For this study, the WRF model was used to produce forecasts with a 10 km horizontal resolution and a vertical resolution
of 80 m (as well as 40 m and 120 m). The model was run for a total of 102 cases, between June 2008 and September 2010, where a “case” is a 54 hour period for which forecasts are produced. Note that not all cases correspond to consecutive day model runs. All model runs were done using models run at Iowa State University, and specific details regarding model specification and runs are given in Deppe (2011) and Deppe et al. (2013).

In order to run the WRF model for a desired resolution, additional inputs are required, such as a planetary boundary layer (PBL) scheme. The planetary boundary layer is the lowest part of the atmosphere and has an upper boundary between 2 km and 200 m, depending on atmospheric conditions. Turbulence and exchanges of heat and moisture occur within the PBL due to the interaction with buildings, trees, and other objects located on the Earth’s surface. These interactions must be modeled within the WRF model, especially since the height at which we are interested in making forecasts, 80 m, is located in the PBL. Unfortunately, physical equations representing what happens at the PBL lead to sets of equations fewer than the number of unknowns. As a result of this closure problem, several PBL schemes have been developed by meteorologists by making assumptions about the PBL and surface interaction. The WRF model was run 6 times for each case. Each model run utilized a different PBL schemes. The PBL schemes used in this study were: the Yonsei University scheme (YSU), the Mellor-Yamada-Janjic scheme (MYJ), the Quasi-Normal Scale Elimination PBL scheme (QNSE), the Mellor-Yamada Nakanishi and Niino level 2.5 PBL scheme (MYNN 2.5), the Mellor-Yamada Nakanishi and Niino level 3.0 PBL scheme (MYNN 3.0), and the Pleim PBL scheme.

In addition to a PBL scheme, a set of boundary and initial conditions, for the future time period in which you wish to make a forecast, must be entered into the WRF model to produce the desired forecasts. The Global Forecast System (GFS) model and North American Mesoscale (NAM) model are two global meteorological models. Forecasts from these global models are used to give initial and boundary conditions to the lower resolution WRF model of interest. The WRF model was run using both the GFS and NAM global models for each PBL scheme, leading to a total of 12 meteorological forecast models, sometimes referred to as
“ensemble members.”

One common approach used in meteorology is to use the sample average of the 12 ensemble forecasts at a given time to be the forecasted wind speed for the time of interest. We consider this as a baseline model, in addition to the persistence model in Section 2.3.

### 2.3 Data-Driven Models

Persistence forecasting is a basic technique that takes the last observed wind speed to be the forecasted wind speed for the forecasting window of interest. Persistence models are often used as “straw man” models in wind speed forecasting, especially for very short-term forecasting windows, when assessing the forecasting performance of a proposed model (Monteiro et al., 2009). Here, we use the persistence model as a baseline for model performance.

Wind speed observations are made over time, so it is natural to consider fitting a time series model to the hourly wind speed data. The time series model can then be used produce forecasts for future time periods. Several researchers have attempted to use the autocorrelation of wind speed observations to produce forecasts. These efforts have typically been done without accounting for the non-Gaussian distribution of wind speed (Chou and Corotis, 1981) and others accounting for this but only examining pure autoregressive models (Brown et al., 1984). Here we apply a method introduced by Torres et al. (2005) to account for the non-Gaussian distribution of wind speed while considering a combination of autoregressive and moving average processes in the model. This model does not make use of any meteorological model output.

#### 2.3.1 Time Series Model Description

Let $y_t$ denote the observed wind speed at hour $t$; the distribution of hourly wind speeds is represented by a Weibull distribution as given in expression (2.1). In order to use traditional time series methods, based on the assumption of normality, a transformation needs to be applied
to the observed data. For each case, we estimate the parameter $\kappa$ from the training data set using maximum likelihood.

\[ P(y_t) = \frac{\kappa}{\lambda} \left( \frac{y_t}{\lambda} \right)^{\kappa-1} e^{-\left( \frac{y_t}{\lambda} \right)} \quad y_t \geq 0 \]  

(2.1)

Dubey (1967) demonstrated that the Weibull distribution can be viewed as approximately Normal when $\kappa = 3.6$. Thus, each observation in the training data set of hourly wind speeds is raised to the power $m = \hat{\kappa}/3.6$, where $\hat{\kappa}$ is the maximum likelihood estimate of $\kappa$. The data is then standardized to eliminate any daily seasonality that may exist. The standardized values, $y_t^*$, for each training dataset are obtained by the following:

\[ y_t^* = \frac{y_t' - \mu(t)}{\sigma(t)} \]  

(2.2)

where $y_t'$ is the transformed data value at hour $t$, $\mu(t)$ is the mean of transformed speed values during hour $t$ and $\sigma(t)$ is the standard deviation of transformed speed values during hour $t$. We then fit an ARMA model to the data after it is transformed and standardized ($y_t^*$). The process for selecting a model, for a particular data set, is described in Section 2.3.4. Forecasts for $y_{t+1}^*, \ldots, y_{t+24}^*$ are produced using the selected ARMA model with the observed, transformed data. Forecasts for the wind speed, $y_{t+1}, \ldots, y_{t+24}$, are obtained by back-transforming $y_{t+1}^*, \ldots, y_{t+24}^*$.

### 2.3.2 Training Period for Time Series Model Fitting

It is important to determine an appropriate length of time for the training dataset, i.e. a period of time which is representative of seasonal wind behavior. Torres et al. (2005) use data from previous years that occur in the same month of interest as their desired forecast time. The choice of a calendar month is somewhat arbitrary, so here we consider data immediately prior to the time period for which we wish to produce forecasts. We consider training periods from two to thirty days, with 24 hourly observations made in each day. In order to implement
Figure 2.1 Mean absolute error (MAE) values of the time series model forecasts, averaged over 98 cases, for varying training period lengths.

Although we could choose to fit the ARMA model to any time period in our data, we examine the cases for which we have meteorological forecasts so its performance may be compared to baseline models. The first four cases of the 102 for which we have meteorological forecasts for are close enough to the beginning of the dataset, so we do not have the ability to go 30 days back. Thus, for the time series model we consider the last 98 of our 102 cases. For each of the 98 cases and each training period length, a mean absolute error (MAE) between the ARMA model and observed data is calculated. Figure 2.1 gives the average MAE value, over the 98 cases, for the different training periods.
Figure 2.1 shows that the average MAE decreases as the training period length increases. The average MAE begins to level off at around a 11 to 13 days. For training periods longer than 11 to 13 days, there is marginal improvement in the average MAE values. Based on these results, we choose to be a bit conservative and set the training period to a length of 15 days for all future ARMA analysis.

### 2.3.3 Data Transformation

Many studies have suggested that the Normal distribution is not the best choice to model wind speed. However, numerous studies have demonstrated that a two-parameter Weibull distribution describes wind speed well, over various locations throughout the world. Figure 2.2 displays the distribution of the estimated shape parameter, $\hat{\kappa}$, for all 98 cases. A shape parameter value very close to 3.6 would indicate that the distribution of wind speeds could be considered approximately Normal (Dubey, 1967).
Figure 2.2 shows that all of the $\hat{\kappa}$ values for our data were lower than Dubey’s $\kappa$ value of 3.6, represented by the vertical red line in the figure. The smallest estimated shape parameter value that we observed over the 98 cases was $\hat{\kappa} = 1.518$, while the median estimated value was $\hat{\kappa} = 2.570$. A majority of the estimated shape parameter values, 70 of 98 values (71.5%), had a value less than 3. These results support the findings of previous studies which have indicated that the Normal distribution is likely not the best choice for describing the marginal distribution of wind speed.

### 2.3.4 Model Selection

A corrected Akaike information criterion (AICc) is used to determine the most adequate ARMA model. We consider candidate ARMA models that make up the collection of ARMA models where the order of the autoregressive and moving average processes are at most 5 each. Once an ARMA model is selected, a Ljung-Box test is performed on the residuals of the time series to ensure that the model has adequately captured the autocorrelation structure of the series. Table 2.1 summarizes the number of times that each candidate model was selected over the 98 cases considered. The most commonly selected ARMA models were ARMA(1,1), ARMA(3,3), and ARMA(4,3). A pure autoregressive model was selected in only 3 cases while a pure moving average model was never selected. Additionally, all of the selected models passed the Ljung-Box test and were determined to adequately describe the autocorrelation of their respective time series.

<table>
<thead>
<tr>
<th>AR</th>
<th>MA</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>3</td>
<td>13</td>
<td>5</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>3</td>
<td>8</td>
<td>12</td>
<td>7</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>8</td>
<td>11</td>
<td>3</td>
<td>3</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 2.1 Number of times time series model was selected by AICc criterion.
2.3.5 Time Series Model Performance

With the length of the training period determined and an ARMA model selected for each case, we forecast 24 hour-ahead wind speeds as described in Section 2.3.4. Figure 2.3 displays the MAE, computed from the 24-hour ahead forecast period, for the ARMA model on a case by case basis. Figure 2.3 also shows the MAE over 24 hours for the persistence model, as described in Section 2.3, and the MAE over 24 hours for the AMF, as described in Section 2.2, for each case.

Overall, the methods based on observed data (ARMA and persistence models) perform worse than the AMF model forecasts. More specifically, the ARMA method had a smaller MAE than the AMF in only 31 (31.6%) of the 98 cases and had a smaller MAE than the persistence model in 48 (49.0%) of the 98 cases.

Due to the short-term forecasting nature of ARMA models, it is not meaningful to compare models over a 24 hour period. It is more appropriate to look at a shorter period of time to compare forecasting performance of the three models, as we know the ARMA model is limited in its ability to produce long-term forecasts. Figure 2.4 shows the forecasting performance of the ARMA model by hour. Here the MAE of the 98 cases is calculated for each hour-ahead forecast from 1 hour-ahead to 24 hours-ahead. The MAE values increase as the hours ahead forecasted increases for 1 to 10 hour-ahead forecasts. There is a notable increase in MAE between 5 and 8 hour-ahead time periods, thus we will consider evaluating the ARMA model performance for the time periods up to 5 hours-ahead. Figure 2.5 displays the MAE, computed from the 5 hour-ahead forecast period for: the ARMA, average meteorological forecast, and persistence models, on a case by case basis.

In general, using a shorter 5 hour-ahead forecast period to compute a model’s MAE for a given case, the ARMA model has smaller MAE values than for a 24 hour-ahead forecast period. Under this forecast criterion, the models driven by observed data perform better than the AMF
Figure 2.3  Time series model mean absolute error (MAE) values for a 24 hour ahead period, by case.

Figure 2.4  Mean absolute error (MAE) values for time series model by hour-ahead forecast, for 98 cases.
Figure 2.5 Time series model mean absolute error (MAE) values for a 5 hour ahead period, by case.

for a majority of the cases. The ARMA model now outperforms the AMF in 62 (63.3%) of the 98 cases. However, despite the improvement of the ARMA model relative to the AMF, the persistence model has a smaller MAE value than the ARMA model in 52 (53.1%) of the 98 cases. Thus, the ARMA model is unable to provide compelling evidence of improvement over baseline methods in these cases, even when the forecast time window is shortened to a length more appropriate for time series models.

The shortcomings of the ARMA model, in the short-term forecasting time period, become clearer when we look at some of the individual 54 hour time periods. Figure 2.6 illustrates the behavior observed in many of the cases where the ARMA model performed poorly relative to the other models. The plot shows the observed wind speed, AMF, ARMA model’s forecasts, and persistence model’s forecasts over a 54 hour period. The observed wind speed is decreasing in the last few hours of the training period, as a result the ARMA model’s initial forecasts continue this pattern while the observed wind speed increases. We also see that the meteorological forecasts predict this increasing wind speed. In general the ARMA model is able to describe the short-term behavior of the wind speed when there are no major changes or shifts in the wind speed’s behavior. However, when there are major changes in the wind speed’s behavior, the ARMA model is unable to capture these shifts and its performance often suffers.
The ARMA model has the ability to provide wind speed forecasts solely based on observed data and is able to account for the non-normality of wind speeds and any daily seasonality that may exist in wind speeds. When looking at a short-term forecasting period of 5 hours or less, the ARMA model forecasts outperform the AMF in many cases. Although the ARMA model provides some advantages when modeling wind speed, there are also several disadvantages to using this model. Even in the short-term forecasting period, the ARMA model has higher MAE values than the simple persistence model in a majority of the 98 cases. The ARMA model is also unable to capture shifts in wind speed behavior, because the model takes into account observed data but not future meteorological forecasts. Finally, the forecasting capability of the ARMA model deteriorates markedly after 5 hours and is not suitable for making forecasts over longer periods. For these reasons, we will not consider the ARMA model as a viable option for accomplishing our objective of forecasting 24 hours in advance.
2.4 Model Output Statistics (MOS) Method

The use of purely statistical models on observed wind speeds is limited by the inability to model future shifts in wind speed that meteorological models are able to capture. Model Output Statistics is a technique that is used to post-process forecasts from numerical weather prediction (NWP) models. In wind speed forecasting, MOS methodology attempts to correct for systematic biases in NWP models. We now consider methods that make use of meteorological forecasts and Model Output Statistics (MOS) to determine how meteorological ensemble members should be combined to produce a single wind speed forecast for a given time.

2.4.1 OCF Model Description

Here we consider the Operational Consensus Forecasts (OCF) model proposed by Engel and Ebert (2007) and test the utility of the model on our data. Observed data and meteorological model forecasts are available as hourly wind speeds. Using OCF, forecasts are calculated as weighted averages of meteorological forecasts. The wind speed forecast given by meteorological model \( j \) at hour \( t \) is denoted as \( y_{jt} \), and the corresponding observed wind speed value as \( y_{0t} \). For a training period dataset, the bias for each meteorological model at a given time is calculated as:

\[
b_{jt} = y_{jt} - y_{0t},
\]

(2.3)

where \( j = 1, \ldots, M \) and \( t = 1, \ldots, T \). Next, the overall bias of each meteorological model, \( b_j \), is calculated as:

\[
\hat{b}_j = (Q_{1j} + 2Q_{2j} + Q_{3j})/4,
\]

(2.4)

where \( Q_{1j}, Q_{2j}, \) and \( Q_{3j} \) are the first, second, and third sample quartiles of \( b_{j1}, \ldots, b_{jT} \). The bias-corrected error of model \( j \) at time \( t \) is denoted as \( \hat{e}_{jt} \) and is calculated as:
\[ \hat{e}_{jt} = y_{jt} - \hat{b}_{j} - y_{0t}. \]  

(2.5)

The MAE for model \( j \) is denoted as \( \text{MAE}_j \), and is calculated as:

\[ \text{MAE}_j = \frac{1}{T} \sum_{t=1}^{T} |e_{jt}|. \]  

(2.6)

The weighting parameter for a given meteorological model is then calculated as:

\[ \hat{w}_j = (\text{MAE}_j)^{-1} \left[ \sum_{j=1}^{M} (\text{MAE}_j)^{-1} \right]^{-1}. \]  

(2.7)

The OCF forecast at time \( t \), \( y_t^* \), is given as:

\[ y_t^* = \sum_{j=1}^{M} \hat{w}_j (y_{jt} - \hat{b}_j). \]  

(2.8)

### 2.4.2 OCF Model Implementation

The OCF model was run to produce 24 hour-ahead forecasts for all 102 cases. In order to align with the objective of this work, a training period length of 30 hours was used to implement the OCF model. Forecasts from both NAM and GFS meteorological models were used; the 6 different PBL schemes were used for both the GFS and NAM initial/boundary conditions, for a total of 12 meteorological models.

### 2.4.3 OCF Model Performance

We evaluate the performance of the OCF model by calculating the MAE for the 24 hour forecast period, for each case. Additionally, we calculate the MAE for the persistence and AMF for this period as well. Figure 2.7 displays the results of these calculations. In general, the persistence model tends to have much larger MAE values than the other two models, across
Both the OCF and AMF models produce smaller MAE values than the persistence model for a majority of the cases. More specifically, the MAE for the OCF and AMF models are smaller than the MAE for the persistence model for 70 (68.6%) and 67 (65.7%) of the 102 cases. The OCF model produces smaller MAE values than the AMF in some cases but produces larger MAE values in other cases. The OCF model outperformed the average meteorological forecast model in 55 (53.9%) of the 102 cases.

In a majority of the cases, the difference in MAE values for these two models is small. To get a clearer picture of a model’s behavior relative to others, we calculate the difference in MAE values, for each of the 102 cases, for pairs of models. Table 2.2 gives the five number summaries of differences in MAE values for each of these pairs of models. A majority of the difference in MAE values between the AMF model and OCF model have a magnitude much

Figure 2.7 Model output statistics model, persistence model, and average meteorological forecast mean absolute error (MAE) values for a 24 hour ahead period, by case.
Table 2.2  Summary statistics of the difference in mean absolute error (MAE) values for OCF, persistence, and average meteorological forecast over 102 cases.

<table>
<thead>
<tr>
<th>Difference</th>
<th>Min.</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meteorological - OCF</td>
<td>-1.73</td>
<td>-0.181</td>
<td>0.031</td>
<td>0.371</td>
<td>3.763</td>
</tr>
<tr>
<td>Persistence - OCF</td>
<td>-2.516</td>
<td>-0.074</td>
<td>0.529</td>
<td>1.203</td>
<td>4.510</td>
</tr>
<tr>
<td>Persistence - Meteorological</td>
<td>-3.337</td>
<td>-0.233</td>
<td>0.502</td>
<td>1.170</td>
<td>4.112</td>
</tr>
</tbody>
</table>

smaller in comparison to differences with each of these models and the persistence model. More specifically, the middle seventy-five percent of the differences in MAE values between the AMF and OCF MAE values fall between -0.181 and 0.371 m/s, giving an interquartile range of 0.552 m/s. On the other hand, the interquartile ranges for differences in MAE values between the persistence model and the AMF and the OCF model are more than twice as large at 1.403 m/s and 1.277 m/s, respectively.

The small difference between the two models is seen in many cases. Figure 2.8 shows the observed wind speed and three forecast models for two examples of such cases. Figure 2.8 shows examples of common types of cases where the AMF has small bias during the training period. As a result, the OCF forecasts only make minor bias corrections during the forecast period. Additionally, Figure 2.8(b) illustrates the inability of the OCF method to make significant improvements to the AMF model when the meteorological forecasts perform decently during the training period.

There are cases where the OCF model either significantly improves or worsens the performance of the AMF model, and we would like to investigate under what circumstances these situations occur. Figure 2.9 shows the observed wind speed and three forecast models for two cases, illustrating of each of these situations. Figure 2.9(a) shows an example case of when the OCF method improves the AMF model. The key characteristic of such situations is that the bias behavior of the average meteorological forecast model is consistent from the training period to the forecasting period. In other words, when the meteorological model has an overall positive bias during the training period and the forecasting period, the OCF model will improve the AMF when the bias of the meteorological model is also positive during the period of time...
Figure 2.8 Model output statistics, average meteorological forecasts’, and persistence model forecasts for cases 49 and 82.

for which we wish to produce forecasts. However, when the bias behavior of AMF model does not remain consistent over the two periods, the OCF model often adjusts the forecasts in the wrong direction. Figure 2.9(b) gives an example of one such situation. Here the meteorological model overpredicts the wind speed during the training period, however the model grossly under predicts the wind speed during the forecasting period. The OCF model makes an adjustment during the forecasting period by further decreasing the meteorological models’ forecasts due to the over prediction during the training period.

The OCF model provides an improvement in forecasts over the persistence model but only outperforms the AMF in just over half of the cases. The OCF model does make use of meteorological forecasts and is capable of capturing shifts in wind speed that are predicted by the meteorological model. Additionally, because it makes use of the meteorological forecast models, the OCF model has the capability to make longer range forecasts. We have seen that the OCF model performs well when the bias of the average meteorological forecast model behaves consistently from the training period to the forecasting period. However, the OCF most often
follows the AMF closely. As a result, the OCF model usually does not produce large improvements in the meteorological models’ forecast and can often produce forecasts with large MAE values when the meteorological model’s bias changes from over predicting to underpredicting (or vis-versa) from the training to forecasting periods. The OCF model does not require us to make any distributional assumptions about the wind speed data but this leads to forecasts with no quantification of the forecasts’ uncertainty. In the context of this work, it is desirable to have the ability to estimate uncertainty and probabilities associated with our forecasts so energy bids can be made with the ability to make conservative bids, depending upon the energy market at the time.

2.5 Hierarchical Model for Bias Adjustment of Wind Speed Forecasts

The ability to estimate uncertainties associated with wind speed forecasts is important when energy companies are making bids associated with cost and potential monetary loss. Several authors have developed methods that produce wind speed forecasts and quantify the uncer-
tainment of their forecasts. In order to quantify uncertainty in our forecasts, we must be willing to make some assumptions about the distribution of wind speed. Gneiting et al. (2005) and Hering and Genton (2010) used a regime-switching model which assumed that wind speed followed a truncated Normal distribution. Thorarinsdottir and Gneiting (2010) also assumed that wind speed could be described by a truncated Normal distribution while using censored regression to predict wind speed. Finally, Sloughter et al. (2010) modeled wind speed as a mixture of gamma models, with a gamma probability distribution function for each meteorological model, using Bayesian model averaging (BMA).

Meteorological forecasts are generated using the same overall dynamic physical model, and it is reasonable to assume that the bias of each of these models has the same general underlying process. Hence, we propose a method to model the bias of the meteorological ensemble members using a hierarchical statistical structure. Additionally, using this hierarchical structure allows us to model the variability of the bias for each meteorological model. To begin, we assume that wind speed can be adequately represented by a Normal distribution. Although a Normal distribution is likely not the best choice to model the distribution of wind speed, we begin with this assumption to ease the computation process. Possible alternatives and changes to using the Normal distribution to represent wind speed are discussed in Section 2.7.

2.5.1 Hierarchical Model Description

We let $Y_{0,t}$ and $Y_{j,t}$ represent the random variables associated with the observed and forecasted wind speed at time $t = 1, \ldots, T$, where $j$ indexes a given meteorological model for $j = 1, \ldots, M$. The forecasted wind shear and temperature difference at time $t$ are represented by $x_{1j,t}$ and $x_{2j,t}$, respectively. Here, wind shear is defined as the difference of wind speeds at 40 m and 120 m Additionally, we let $\mu_t$ represent the true wind speed at time $t$ and $b_{j,t}$ denote the mean bias of meteorological model $j$ at time $t$.

In the following we use $N(\mu, \sigma^2)$ to denote the Normal distribution with mean $\mu$ and variance $\sigma^2$ and use $G(\alpha, \beta)$ to represent the gamma distribution whose density is $\frac{\beta^\alpha}{\Gamma(\alpha)}y^{\alpha-1} \exp\{-\beta y\}$. 
We assume that wind speed at time $t$ is observed with values centered at the true wind speed $\mu_t$ with variability $\lambda_0$:

$$Y_{0,t} \sim N(\mu_t, \lambda_0^{-1}), \quad (2.9)$$

where $\lambda_0$ is known. We also assume that the true bias of model $j$ at time $t$, $Y_{j,t} - \mu_t$, is Normally distributed as:

$$(Y_{j,t} - \mu_t) \sim N(b_{j,t}, \lambda_j^{-1}). \quad (2.10)$$

These assumptions allow for unique means and variances for the bias of each meteorological model and treat the biases across meteorological models to be independent from one another. The inverse variance of the true bias for each model is assumed to follow a Gamma distribution:

$$\lambda_j \sim G(\alpha, \beta), \quad (2.11)$$

where $\alpha$ and $\beta$ are known. Here we note that although we are allowing for the variance of the bias to vary from one meteorological model to another, we assume that the underlying distribution of $\lambda_j$ is the same across meteorological models. The 6 NAM meteorological models stem from the same physical model but with different perturbations of initial conditions, so the assumption of a common underlying variance structure for the bias seems reasonable; the same is true for the GFS meteorological models. Finally, we assume that the mean bias is linearly related to the wind shear and temperature difference and allow for different linear relationships for each meteorological model:

$$b_{j,t} = \alpha_{0j} + \alpha_{1j} x_{1j,t} + \alpha_{2j} x_{2j,t} + w_{j,t}, \quad (2.12)$$

where $w_{j,t} \sim N(0, \tau^{-1})$. 

2.5.2 Prior Distribution Specifications

The parameters $\alpha_{0j}, \alpha_{1j}, \alpha_{2j}, \mu_t,$ and $\tau$ require the specification of prior distributions. The prior distribution of the regression parameters $\alpha_{0j}, \alpha_{1j},$ and $\alpha_{2j}$ are chosen to be conjugate prior distributions. Additionally, the prior distribution of each regression parameter is centered at 0 with an inverse variance, $\lambda_\alpha^{-1},$ chosen so that each prior is diffuse but proper. The prior distribution for $\mu_t$ is chosen to be conjugate and centered around 6 meters per second, the average wind speed for Iowa according to the National Oceanic and Atmospheric Administration (NOAA). The variance, $\lambda_\mu,$ was chosen so that negative wind speed values would have negligible probability. Finally, a conjugate prior for the regression error-variance, $\tau,$ is specified.

The prior distributions for $\alpha_{0j}, \alpha_{1j}, \alpha_{2j}, \mu_t,$ and $\tau$ are given as:

\[
\alpha_{0j}, \alpha_{1j}, \alpha_{2j} \sim N(0, \lambda_\alpha^{-1}) \tag{2.13}
\]
\[
\mu_t \sim N(M_\mu, \lambda_\mu^{-1}) \tag{2.14}
\]
\[
\tau \sim G(c, d) \tag{2.15}
\]

where $j = 1, \ldots, M$ and $t = 1, \ldots, T.$

2.5.3 Conditional Posteriors of the Hierarchical Model

In order to obtain a random sample from the joint probability distribution of $\tau, \alpha_{0j}, \alpha_{1j}, \alpha_{2j}, \mu_t,$ and $b_{j,t}$ for $j = 1, \ldots, M$ and $t = 1, \ldots, T,$ we implement a Monte Carlo sampling technique and update the parameters using a Gibbs sampling method. Explicit calculation of the posterior distributions can be found in Appendix A. The resulting conditional posterior distributions are as follows:
The starting value of the inverse variance of the bias was set as $\lambda = 10$. The mean wind speed was set to an initial value of 6 m/s, $\alpha = 0$. All regression parameters and mean bias parameters were set to an initial value of 0, i.e., $\lambda_\alpha = 0$, $\mu_\beta = 0$, $\lambda_\beta = 0$, $\mu_\alpha = 0$, $\lambda_\mu = 0$, $\alpha_0 = 0$, $\beta_0 = 0$, $\mu_0 = 0$, $\lambda_0 = 0$, $\lambda_\mu = 0$, $\alpha_\mu = 0$, $\beta_\mu = 0$, $\alpha_\beta = 0$, $\beta_\beta = 0$, $\alpha_\lambda = 0$, $\beta_\lambda = 0$, $\mu_\lambda = 0$, $\lambda_\mu = 0$, $\alpha_\mu = 0$, $\beta_\mu = 0$, $\alpha_\beta = 0$, $\beta_\beta = 0$, and $\alpha_\lambda = 0$, $\beta_\lambda = 0$. Additionally, we have to choose starting values for the parameters we wish to estimate. The starting value of the inverse variance of the bias was set as $\lambda_j = 1$ for all $j = 1, \ldots, M$. All regression parameters and mean bias parameters were set to an initial value of 0, i.e., $\alpha_0 = \alpha_{1j} = \alpha_{2j} = 0$ for all $j = 1, \ldots, M$ and $b_{jt} = 0$ for all $j = 1, \ldots, M$ and $t = 1, \ldots, T$. The mean wind speed was set to an initial value of 6 m/s, $\mu_t = 6$ for all $t = 1, \ldots, T$. Finally, the hyperparameter values were set to the following values: $\alpha = 1, \beta = 2, \lambda_\mu = 0.5, \lambda_\alpha = 10, \lambda_0 = 0.5, c = 1$, and $d = 1$.

The initial 30 hours for a case is used as the training data for that particular case. Then, a Markov chain Monte Carlo (MCMC) method is used to obtain conditional posterior distributions of the parameters in Section 2.5.3. Because all posterior distributions are proportional to a closed, known distribution, a Gibbs’ sampler is used to update parameter values. Hyperparameter values were set to the following values: $\alpha = 1, \beta = 2, \lambda_\mu = 0.5, \lambda_\alpha = 10, \lambda_0 = 0.5, c = 1$, and $d = 1$.
the starting value for the regression error-variance is set as $\tau = 1$.

For each case, the MCMC algorithm was run for 5,000 iterations with the first 1,000 iterations treated as a burn-in period.

2.5.5 Combining Model Forecasts

The hierarchical model produces a total of twelve posterior distributions for the regression parameters in expression (2.12) (as well as the precision of the mean bias, $\lambda_j$), one posterior for each meteorological model ensemble member. Realizations from each of these posteriors are generated from the MCMC algorithm. Given the forecasted covariates, $x_{1,j,t}$ and $x_{2,j,t}$, and simulated parameter values $\alpha^{(k)}_{0,j,t}$, $\alpha^{(k)}_{1,j,t}$, $\alpha^{(k)}_{2,j,t}$, $\lambda_j^{(k)}$, and $\tau^{(k)}$ from the kth MCMC step (after the burn-in period) are used to simulate realizations of $b_j^{(k)}$ from the posterior given in expression 2.22. This is done for each meteorological model $j$ for a given future time $t$ at each iteration of the MCMC algorithm. The set of simulated $b_j^{(1)}$, $b_j^{(2)}$, $\ldots$, $b_j^{(k)}$ form realizations from twelve prior predictive distributions of $b_j$, one for each meteorological model $j$. Given the MCMC algorithm implemented in Section 2.5.4, each simulated distribution has a total of 4,000 observations. Figure 2.10 displays the prior predictive distributions of the mean bias for each of the meteorological models at a given time for case 42.

For each future time period, each meteorological model has a fixed wind speed forecast. The prior predictive distributions of the bias-corrected wind speed forecasts, $y_{j,t}^*$ are obtained by horizontally shifting the prior predictive distribution of each $b_j,t$ by its respective wind speed forecast. Figure 2.11 shows the prior predictive distribution for $y_{j,t}^*$ for each meteorological model. These distributions correspond to the prior predictive distributions of $b_{j,t}$ presented in Figure 2.10.

It must be determined how these distributions should be combined to produce one wind speed forecast at time $t$ along with uncertainty estimates for this forecast. The prior predictive
Figure 2.10  Prior predictive distributions of the mean bias, $b_{j,t}$, for each of the twelve meteorological models for hour 54 on November 11, 2008.
Figure 2.11  Prior predictive distributions of the bias-corrected wind speed forecast, \( y_{j,t}^* \), for each of the twelve meteorological models for hour 54 on November 11, 2008.
distributions can be combined to form one distribution for the bias-corrected wind speed at
time $t$, $y_t^*$, using a finite mixture distribution. This prior predictive distribution of $y_t^*$ is given as:

$$p(y_t^* \mid \cdot) = \sum_{j=1}^M w_j \cdot p(y_{j,t}^* \mid \cdot),$$  \hspace{1cm} (2.23)

where $w_j > 0$, $\sum_j w_j = 1$, and $p(y_{j,t}^* \mid \cdot)$ denotes the prior predictive distribution of the bias-corrected wind speed forecast. expression (2.23) requires the specification of weights $w_j$ for the mixing components. As a first step, we consider the straightforward approach of giving all of the mixing distributions equal weight by setting $w_j = 1/12$ for $j = 1, \ldots, 12$. Figure 2.12 shows the prior predictive distribution of the bias-corrected wind speed forecast that results from a finite mixture distribution with equal weighting parameters. The mean of the realizations from the prior predictive distribution is taken to be the point estimate of the bias-corrected wind speed forecast. These estimates will be referred to as the “hierarchical average” forecasts for the case of equal weight parameters.

We would also like to consider more sophisticated weighting schemes for the finite mixture distribution. One such method is to determine weights based on the variance of each prior predictive distribution, where distributions with more variability are weighted less heavily. The weighting parameter for each meteorological model is determined as:

$$w_j = (\delta_{j,t})^{-1} \left[ \sum_{j=1}^M (\delta_{j,t})^{-1} \right]^{-1},$$  \hspace{1cm} (2.24)

where $\delta_j$ is the variance of the simulated prior predictive distribution of $y_{j,t}^*$ for meteorological model $j$ at time $t$. Figure 2.13 shows the prior predictive distribution of the bias-corrected wind speed forecast that results from a finite mixture distribution with weights determined using expression (2.24). A point estimate for the weighting scheme presented in expression (2.24) is taken to be the mean of the simulated prior predictive distribution. Estimates obtained from this methodology will be referred to as “hierarchical weighted” forecasts. For this case,
the hierarchical average wind speed forecast was 7.75 m/s, and the hierarchical weighted wind speed forecast was 8.32 m/s.

In addition to point estimates, the hierarchical model provides uncertainty estimates for the bias-corrected forecasts, in both the hierarchical average and weighted methods. This allows us to calculate interval estimates for the forecast or calculate probabilities of interest. For example, wind turbines are often turned off or curtailed if the wind speed exceeds a certain threshold. The probability that the wind exceeds this threshold may be of interest to utility companies and can be calculated using the simulated prior predictive distribution. Figure 2.14 shows 95% interval bounds (based on 0.025 and 0.975 percentiles of the simulated distribution) for the hierarchical average and hierarchical weighted forecasts over the forecast period for case 97. The two weighting methods produce similar forecasts and intervals for the forecasts in this case.
Figure 2.13  Prior predictive distribution of the bias-corrected wind speed forecast, $y_{j,t}^*$, resulting from a finite mixture distribution with inverse variance weight parameters, for hour 54 on November 11, 2008.

Figure 2.14  Hierarchical model forecasts and forecast intervals for both weighting schemes for case 97: September 5, 2010.
2.5.6 Hierarchical Model Performance

In the 102 cases we evaluate the performance of the hierarchical model (both average and weighted) relative to the OCF, persistence, and average meteorological forecast models. One criterion used to assess performance is to calculate the MAE for the 24 hour forecast period, separately for each case. Additionally, we calculate the MAE for the OCF, persistence, and AMF for this period as well. Figure 2.15 shows the MAE values for these models for all 102 cases, however the OCF model’s performance is left off the plot for easier reading.

The MAE values for the hierarchical average and hierarchical weighted models tend to be very close to one another in most cases. If we calculate the difference in MAE values between the weighting schemes (average - weighted) for all 102 cases, the minimum difference is -0.52 m/s, and the maximum difference is 0.23 m/s. Additionally, the first and third quantiles are -0.045 m/s and 0.027 m/s, respectively. Figure 2.16 shows the forecasts of both weighting
Figure 2.16 Hierarchical average model, hierarchical weighted model, average meteorological forecast, and persistence model forecasts for cases 47 and 42.

(a) Case 47: November 14, 2008

(b) Case 42: November 4, 2008

methods for two cases. The hierarchical weighted model provides the biggest improvement over the hierarchical average model on November 14, 2008 (case 47), shown in Figure 2.16(a). The hierarchical weighted model gives the worst performance compared to the hierarchical average model on November 4, 2008 (case 42), shown in Figure 2.16(b). The two weighting schemes produce similar forecasts in many of the cases, which leads to small differences in MAE values. Although there are some differences in forecasts for the two weighting methods, a majority of the forecasts have little difference between them; Figure 2.17 shows an example of a typical case.

The hierarchical weighted model produces similar forecasts to the hierarchical average model and does not provide any significant improvements. Thus, we only consider the hierarchical average model in the following performance analysis. We calculate the MAE for the hierarchical, OCF, persistence, and AMF models. Figure 2.18 displays the results of these calculations, omitting the OCF results for easier reading. In general, the persistence model tends to have much larger MAE values than the two other models (as well as the OCF model), across the 102 cases. Additionally, the hierarchical model has a smaller MAE value than the AMF in a
Figure 2.17 Hierarchical average model, hierarchical weighted model, average meteorological forecast, and persistence model forecasts for case 78: August 26, 2009.

In particular, the hierarchical model’s forecast has a smaller MAE value than the AMF in 71 (69.6%) of the 102 cases. This is almost 20% more than the number of cases in which the OCF model was able to outperform the AMF. The hierarchical model produced a smaller MAE value than the OCF model and persistence model in 67 (65.7%) and 82 (80.4%) of the 102 cases. While it is true that the hierarchical model gives forecasts that produce smaller MAE values than the AMF in a large majority of all cases, it is also important to look at the magnitude of the improvements that the hierarchical model provides. Table 2.3 gives the five number summary for the difference in MAE values between the hierarchical and the three other models, across all 102 models.

The first and third quantiles of the difference in MAE values between the hierarchical and AMF are 0.478 m/s and -0.027 m/s, respectively. This leads to an interquartile range for the difference in MAE values of 0.505 m/s. This is similar to the interquartile range for differences
Figure 2.18 Hierarchical model, persistence model, and average meteorological forecast mean absolute error (MAE) values for a 24 hour ahead period, by case.

<table>
<thead>
<tr>
<th>Difference</th>
<th>Min.</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meteorological - Hierarchical</td>
<td>-0.763</td>
<td>-0.027</td>
<td>0.240</td>
<td>0.478</td>
<td>1.657</td>
</tr>
<tr>
<td>Meteorological - OCF</td>
<td>-1.73</td>
<td>-0.181</td>
<td>0.031</td>
<td>0.371</td>
<td>3.763</td>
</tr>
<tr>
<td>Persistence - Hierarchical</td>
<td>-1.810</td>
<td>0.083</td>
<td>0.660</td>
<td>1.259</td>
<td>4.726</td>
</tr>
<tr>
<td>Persistence - OCF</td>
<td>-2.516</td>
<td>-0.074</td>
<td>0.529</td>
<td>1.203</td>
<td>4.510</td>
</tr>
<tr>
<td>Persistence - Meteorological</td>
<td>-3.337</td>
<td>-0.233</td>
<td>0.502</td>
<td>1.170</td>
<td>4.112</td>
</tr>
</tbody>
</table>

Table 2.3 Summary statistics of the difference in mean absolute error (MAE) values for hierarchical, OCF, persistence, and average meteorological forecast over 102 cases.
in MAE values between the OCF model and AMF, which was 0.552 m/s. While the magnitude of improvements being made to forecasts is similar to the OCF model, the hierarchical model makes improvements more frequently. We also note that in cases where the hierarchical model’s forecasts produce a larger MAE value than the AMF, the magnitude of this difference is generally smaller than it is for the OCF model.

Figure 2.19 shows two cases where the hierarchical model improves upon the meteorological forecasts. Figure 2.19(a) gives an example of the hierarchical model’s ability to improve upon the meteorological forecasts when the behavior of the bias of the meteorological forecast is consistent across the training and forecasting periods. This is a case where the OCF model’s forecasts also produced an MAE value smaller than the AMF. On the other hand, Figure 2.19(b) shows a case where the behavior of the bias for the AMF is not consistent. However, the hierarchical model has a smaller MAE than the AMF. This is one example of many similar cases where the OCF model performed worse than the meteorological forecast due to the inconsistency of the bias’ behavior, but the hierarchical model is able to produce smaller MAE values, although not overly large improvements, in the same case.

Of course not all hierarchical model forecasts produced smaller MAE values than the AMF. Figure 2.20 displays the case that produced the worst difference in MAE values between the hierarchical model and AMF. This is a case where the meteorological forecast is consistently over predicting the wind speed during the training period. The AMF does really well during the forecasting period and leaves little room for improvement, with the exception of one large shift in wind speed that the meteorological forecast does not capture. In this case, the inconsistency in the behavior of the bias leads to a diminished forecasting capability by the hierarchical model.

2.6 Incorporating Wind Direction into the Hierarchical Model

We attempt to improve upon the hierarchical model presented in Section 2.5 by introducing information about wind direction into our model. Wind direction is widely accepted as
Figure 2.19  Hierarchical model, average meteorological forecast, and persistence model forecasts for cases 70 and 95.

Figure 2.20  Hierarchical model, persistence model, and average meteorological forecast model forecasts for case 95: September 3, 2010.
being related to wind speed by the meteorological community, especially given the knowledge of prevailing and dominant winds in different geographic locations. For example, the Midwest is known to have prevailing westerly winds. However, the use of wind direction in statistical models is not straightforward, as it is an angular variable with observed values falling on a circle.

The prediction of wind speed and direction are important in many meteorological applications, especially for the purposes of wind energy prediction. Much research has been done regarding the forecasting of wind speed and wind direction independently. Despite the fact that wind speed and direction have been found to be related, little research can be found on the forecasting of these two quantities simultaneously. This model attempts to address the simultaneous forecasting problem.

2.6.1 Modeling Circular Variables

A few standard methods for modeling wind direction or a wind vector exist. One such method is to use a distribution with the same domain as the circular variable. Common choices for this approach are the Von Mises and Wrapped Normal distributions. A second approach is to use the Cartesian coordinates to represent the wind vector. Another approach is break the wind vector up into two vector components. This method gives one component of the form \( u = r \cos \theta \) and the other component is \( v = r \sin \theta \), where \( r \) represents the wind speed and \( \theta \) represents the wind direction.

Several studies have worked to develop a bivariate distribution to represent the wind speed and direction jointly. Several authors have used an Isotropic or Anisotropic Gaussian model to fit a joint probability distribution to wind speed and direction data, where the direction component is modeled using a mixture of Von Mises distributions (Erdem and Shi (2011b), Carta et al. (2008), and Weber (1991)). Fisher and Lee (1994) used a Wrapped Normal distribution to model wind direction. The ability to forecast wind direction in a univariate manner has been done by Bao et al. (2010) using Bayesian model averaging. Additionally, Modlin et al. (2012)
used a Wrapped Normal distribution in the context of circular-circular regression to produce direction forecasts.

A few studies have developed models to produce bivariate wind speed and direction forecasts (Cripps et al. (2005), Erdem and Shi (2011a), Klausner et al. (2009)). However, all of these methods used temporal dependence to generate vert short-term forecasts. Erdem and Shi (2011a) used a bivariate ARMA time series structure to model the vector components of wind, Cripps et al. (2005) used a space-time model structure on the vector components of wind, and Klausner et al. (2009) modeled the vector components and Cartesian coordinates using a times series approach for similar days. These studies have made progress in the way of forecasting wind speed and direction jointly, but due to the nature of the models, are not able to make forecasts over the time span which we desire.

2.6.2 Model Implementation

Here, we combine information about wind speed and direction (both observed and forecasted) by examining the vector components of the wind. We define two vectors as:

\[
\begin{align*}
\mathbf{u} & = r \cos \theta \\
\mathbf{v} & = r \sin \theta,
\end{align*}
\] (2.25) (2.26)

where \( r \) represents the wind speed and \( \theta \) represents the wind direction. Literature provides no clear consensus on whether it is reasonable to treat the vector components of \( \mathbf{u} \) and \( \mathbf{v} \) as independent or not. As a starting point, we assume that the vector components are independent. The validity of this assumption will be discussed in Section 2.7.

The vector components defined in expressions 2.25 and 2.26 will be referred to as the x and y component of wind. We let \( U_{0,t} \) and \( V_{0,t} \) represent the random variables associated with the observed x and y vector components at time \( t = 1, \ldots, T \). The forecasted x and y wind vectors
at time \( t = 1, \ldots, T \) for meteorological model \( j = 1, \ldots, M \) are represented by \( U_{j,t} \) and \( V_{j,t} \), respectively. The forecasted wind shear and temperature difference at time \( t \) are represented by \( x_{1j,t} \) and \( x_{2j,t} \), respectively. Additionally, we let \( \mu_u \) and \( \mu_v \) represent the true x and y wind components at time \( t \) and \( b_{uj,t} \) and \( b_{vj,t} \) denote the mean x and y component bias of meteorological model \( j \) at time \( t \).

The model for the bias of the x and y wind components is similar to the model in Section 2.5.1. More specifically, for the x wind component, \( U_{0,t} \) the model is formulated as:

\[
U_{0,t} \sim N(\mu_u, \lambda_u^{-1}), \\
(U_{j,t} - \mu_u) \sim N(b_{uj,t}, \lambda_u^{-1}), \\
\lambda_u \sim G(\alpha_u, \beta_u), \\
b_{uj,t} = \alpha_{0uj} + \alpha_{1uj}x_{1j,t} + \alpha_{2uj}x_{2j,t} + w_{uj,t}, \\
w_{uj,t} \sim N(0, \tau_u^{-1}). \tag{2.27}
\]

Similarly, the model for the y wind component, \( V_{0,t} \) is:

\[
V_{0,t} \sim N(\mu_v, \lambda_v^{-1}), \\
(V_{j,t} - \mu_v) \sim N(b_{vj,t}, \lambda_v^{-1}), \\
\lambda_v \sim G(\alpha_v, \beta_v), \\
b_{vj,t} = \alpha_{0vj} + \alpha_{1vj}x_{1j,t} + \alpha_{2vj}x_{2j,t} + w_{vj,t}, \\
w_{vj,t} \sim N(0, \tau_v^{-1}). \tag{2.28}
\]

The prior distributions for \( \alpha_{0uj}, \alpha_{1uj}, \alpha_{2uj}, \mu_u, \alpha_{0vj}, \alpha_{1vj}, \alpha_{2vj}, \mu_v \) and \( \tau \) are chosen to be conjugate priors, similar to the prior distributions described in Section 2.5.1. The estimation of the model parameters through an MCMC algorithm, for both components, is the same as described in Sections 2.5.3 and 2.5.4.
2.6.3 Obtaining Bias-Corrected Wind Speed Forecasts

The hierarchical model produces a total of twelve posterior distributions for the regression parameters related to the mean bias (as well as the precision of the mean bias, $\lambda_j$), one posterior for each meteorological model ensemble member, for each wind speed component separately. Realizations from the relevant posterior distributions, forecasted covariates, and component forecasts are used to produce prior predictive distributions for bias-corrected $x$ and $y$ components. For a given time $t$, each pair of bias-corrected components is back-transformed to obtain a wind speed forecast. The back-transformation is given as:

$$y^*_{j,t} = \sqrt{\left(u^*_{j,t}\right)^2 + \left(v^*_{j,t}\right)^2},$$

(2.29)

where $u^*_{j,t}$ and $v^*_{j,t}$ are the bias-corrected $x$ and $y$ wind speed components, respectively. This calculation is done for each iteration of the MCMC algorithm, for each meteorological model, resulting in twelve prior predictive distributions of wind speed forecasts. We implement both weighting schemes presented in Section 2.5.5 to obtain one prior predictive distribution for the bias-corrected wind speed forecasts. The similarities between forecasts from the two weighting methods were also observed in the hierarchical model for vector components, thus we will only present results from the equal weighting method.

2.6.4 Obtaining Bias-Corrected Wind Direction Forecasts

The hierarchical model for the vector components also allows us to obtain bias-corrected wind direction forecasts. Given the bias-corrected components, the bias-corrected wind direction value at time $t$ for meteorological model $j$, $\theta^*_{j,t}$, can be calculated as:

$$\theta^*_{j,t} = \begin{cases} 
\tan^{-1}\left(\frac{v^*_{j,t}}{u^*_{j,t}}\right) & \text{for } u^*_{j,t} > 0, v^*_{j,t} > 0 \\
\tan^{-1}\left(\frac{v^*_{j,t}}{u^*_{j,t}}\right) + \pi & \text{for } u^*_{j,t} < 0 \\
\tan^{-1}\left(\frac{v^*_{j,t}}{u^*_{j,t}}\right) + 2\pi & \text{for } u^*_{j,t} > 0, v^*_{j,t} < 0
\end{cases}.$$  

(2.30)
The prior predictive distribution, using equal weights, for the bias-corrected wind direction is obtained by calculating the average direction of the meteorological models’ forecasts for each realization of the MCMC algorithm. Realizations of the average direction, $\tilde{\theta}_t^*$, are computed as the circular average of all bias-corrected meteorological direction forecasts, for time $t = 1, \ldots, T$.

The circular average is defined as (Mardia and Jupp, 2000):

$$
\tilde{\theta}_t^* = \begin{cases} 
\tan^{-1}(\bar{S}/\bar{C}) & \text{for } \bar{S} > 0, \bar{C} > 0 \\
\tan^{-1}(\bar{S}/\bar{C}) + \pi & \text{for } \bar{C} < 0 \\
\tan^{-1}(\bar{S}/\bar{C}) + 2\pi & \text{for } \bar{S} < 0, \bar{C} > 0
\end{cases}, \quad (2.31)
$$

where

$$
\bar{C} = \frac{1}{M} \sum_{j=1}^{M} \cos(\theta_{j,t}^*)
$$

$$
\bar{S} = \frac{1}{M} \sum_{j=1}^{M} \sin(\theta_{j,t}^*). \quad (2.32)
$$

Here $\theta_{j,t}^*$ represents the bias-corrected wind direction forecast at time $t = 1, \ldots, T$ for meteorological model $j = 1, \ldots, M$. Finally, point estimates for the bias-corrected wind direction forecast are taken to be the mean direction of the prior predictive distribution. Additionally, the average direction of the meteorological models’ forecasts, computed using expressions (2.31) and (2.32) can be used as a baseline model to compare to the hierarchical model.

The absolute distance between two circular variables, $\theta$ and $\theta^*$, is defined as (Mardia and Jupp, 2000):

$$
\min\{|\theta - \theta^*|, 2\pi - |\theta - \theta^*|\}. \quad (2.33)
$$

The absolute errors, as defined in expression (2.33), and MAE for the 24 hour forecast period are computed for the meteorological average direction and hierarchical model. The hierarchical model has smaller MAE values than the meteorological average for 56 of the 102
cases. Figure 2.21 shows two cases where the hierarchical model performs better than the meteorological average.

The hierarchical model also allows us to calculate interval estimates, based on the simulated prior predictive distribution, for the bias-corrected direction forecasts. Figure 2.22 displays the point and interval estimates, corresponding to the cases presented in Figure 2.21, produced by the hierarchical model for wind direction.

2.6.5 Performance of the Hierarchical Model Incorporating Direction

We evaluate the performance of the hierarchical model for the vector components of wind calculating the MAE for the 24 hour forecast period, for each case. Figure 2.23 shows the MAE values for both the hierarchical model for vector components and wind speed, as well as the MAE values for the AMF and persistence model. In a majority of the cases, the hierarchical model incorporating wind direction outperforms the persistence model. Additionally, the hierarchical model for wind vector components has smaller MAE values than the AMF and OCF models for 59 (57.84%) and 54 (52.94%) of the 102 cases. However, the hierarchical model
Figure 2.22 Hierarchical model forecasts, forecast intervals, and average meteorological forecast wind direction forecasts for cases 35 and 40.

for the wind vector components outperforms the hierarchical model for wind speed in only 39 (38.24%) of the 102 cases.

Although these results seem discouraging at first glance, we see in Figure 2.23 that differences in the MAE for the two hierarchical models are often very small. We further investigate the magnitude of these differences by looking at summary statistics for the difference in MAE values between the two models. Table 2.4 gives the five number summary for the difference in MAE values for several pairs of models. The middle 75% of differences in MAE values between the hierarchical model and the hierarchical model for wind vector components are between -0.286 and 0.071 m/s. A majority of relatively small differences in MAE values indicates that although the hierarchical model incorporating direction has a higher MAE than the original hierarchical model in 85 cases, this number may be a bit misleading as the two models appear to be performing in a similar manner. Additionally, the hierarchical model for the vector components produces bias-corrected wind direction forecasts, as presented in Section 2.6.4.

Figure 2.24 shows two cases where the vector hierarchical model has smaller MAE values than the hierarchical model and the AMF model. Additionally, Figure 2.25 displays forecasts
Figure 2.23  Vector hierarchical model, hierarchical model, persistence model, and average meteorological forecast mean absolute error (MAE) values for a 24 hour ahead period, by case.

<table>
<thead>
<tr>
<th>Difference</th>
<th>Min.</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hierarchical - Vector Hierarchical</td>
<td>-1.763</td>
<td>-0.286</td>
<td>-0.113</td>
<td>0.071</td>
<td>0.479</td>
</tr>
<tr>
<td>Meteorological - Vector Hierarchical</td>
<td>-1.244</td>
<td>-0.209</td>
<td>0.006</td>
<td>0.321500</td>
<td>1.196</td>
</tr>
<tr>
<td>Meteorological - Hierarchical</td>
<td>-0.763</td>
<td>-0.027</td>
<td>0.240</td>
<td>0.478</td>
<td>1.657</td>
</tr>
<tr>
<td>Persistence - Vector Hierarchical</td>
<td>-2.477</td>
<td>-0.1284</td>
<td>0.405</td>
<td>1.3150</td>
<td>5.2030</td>
</tr>
<tr>
<td>Persistence - Hierarchical</td>
<td>-1.810</td>
<td>0.083</td>
<td>0.660</td>
<td>1.259</td>
<td>4.726</td>
</tr>
</tbody>
</table>

Table 2.4  Summary statistics of the difference in mean absolute error (MAE) values for vector hierarchical, hierarchical, OCF, persistence, and average meteorological forecasts over 102 cases.
Figure 2.24 Hierarchical model, vector hierarchical model, and average meteorological forecast wind speed forecasts for cases 55 and 84.

for both the vector hierarchical and hierarchical models along with their respective 95% forecast intervals for an example case in August 2009. In this case, the intervals are wider for the vector hierarchical model than the intervals for the hierarchical model, especially towards the end of the forecast period. This difference in widths is seen in a large number of the 102 cases and is likely due to variability present in both bias-corrected vector components leading to greater variability in the back-transformed wind speed forecasts.

Figure 2.26 displays the MAE values for two hierarchical models and AMF only, for ease of reading. Here, we also highlight some properties of the meteorological forecast cases. In the 102 cases, there are three sets of forecasts that are made up of ten or more consecutive day forecasts, and these periods of consecutive days are identified by gray boxes in Figure 2.26. The largest stretch of consecutive day models runs was from September 30, 2008 to November 8, 2008. Interestingly, a large number cases, 20 out of 35, where the hierarchical wind vector model gave improvements over the original hierarchical occurred during this time period of consecutive day forecasts.

Further investigation into wind characteristics for September 30, 2008 to November 8,
Figure 2.25  Vector hierarchical model, hierarchical model, and average meteorological forecast wind speed forecasts with 95% intervals for case 75: August 23, 2009.

Figure 2.26  Vector hierarchical model, hierarchical model, and average meteorological forecast mean absolute error (MAE) values for a 24 hour ahead period, by case.
2008 and how these characteristics may differ from other time periods is needed. Additionally, the validity of assuming independence between the $u$ and $v$ wind components should also be evaluated. The performance of the hierarchical model for $u$ and $v$ may be improved by allowing for dependence between the two wind components.

### 2.7 Discussion and Future Work

We have applied and extended the Bayesian bias-correction statistical model originally introduced by Tebaldi et al. (2005) for quantifying uncertainty in the analysis of multimodel regional climate ensembles, to forecasting wind speed and direction at a location in the Midwestern United States. We used a relatively short training period (only 30 hours) to produce hourly forecasts for a relatively long prediction window (24 hours). We have successfully combined observed data and forecasts from a 12-member meteorological ensemble (ran at Iowa State University) along with covariate information to produce 24 hour-ahead forecasts (and prediction intervals) that, in most cases, outperform the standard methods for obtaining forecasts in the energy production industry (i.e. persistence model, or the simple average of the meteorological forecasts.) The largest improvements were obtained when there were no sudden shifts in the bias between the training and the forecast period. We have considered two different weighting schemes to combine individual ensemble member forecasts and prediction intervals. We have also proposed a simple extension to model simultaneously the bias of the wind speed and direction.

While developing the current model, we have made several assumptions that may not be realistic for a larger range of scenarios. For example, the parameters $\alpha$ and $\beta$ are assumed to be known but should probably be estimated, as reasonable values for the parameters for the variance of meteorological forecasts is not intuitive. This is not completely straightforward due to the nature of the conditional posterior of $\alpha$ (it has no closed form and the function proportional to the posterior often gives a very sharp distribution). Exploratory work shows that the arbitrary choices of $\alpha$ and $\beta$ has little to no effect on the posterior of the mean bias.
However, if a weighted average of forecast models using the variability of each model’s mean bias is needed, estimating these parameters would be important.

When considering the construction of the different weights, we noted that the very small difference in the resulting forecasts is attributed to the assumption that the prior distributions for $\lambda_j$ of expression (2.10) are the same across all ensemble members. One possible extension would be to allow the parameters $\alpha$ and $\beta$ of the prior distribution $G(\alpha, \beta)$ (in expression (2.11)) to differ for the various meteorological models. Additionally, we might consider weighting schemes based on forecast attributes other than variability. One such possibility would be to include the relative skill of the forecasts during the training period, as well as the variability in forecasts when determining weights.

Another assumption made in the current work is that the bias in the wind vector components are independent of each other. The hierarchical model using the vector forecasts as inputs did not perform as well in forecasting as the hierarchical model using only wind speed forecasts. However, using the vector approach allows us to produce bias corrected wind direction forecasts, which cannot be obtained using the model only considering wind speed. A natural extension would be to allow for a correlation structure that correctly captures the possible dependence in the two components, perhaps improving the forecasting ability of the vector hierarchical model.

Finally, considering a distribution for the wind speed that captures the expected non-Normal behavior of the wind data could provide an improvement over the current forecasts. Possible distributions include, but are not limited to, truncated Normal, Weibull, and Gamma distributions. However, the assumption of normality seems reasonable in the case of the vector hierarchical model.
CHAPTER 3. INCORPORATING DEPENDENCE IN A HIERARCHICAL MODEL FOR THE BIAS OF METEOROLOGICAL WIND VECTOR FORECASTS

3.1 Background and Motivation

Previous research on wind forecasting has largely focused on wind speed. A large amount of this research has placed an emphasis on using forecasts from meteorological ensembles in this process; a rather extensive review of these types of studies was given in Chapter 2. Meteorological models require the specification of initial conditions and face issues with incomplete physical equations often leading to biased ensemble forecasts (Hamill and Colucci, 1997). Thus, recent wind speed forecasting has focused on developing methods for post-processing or combining these ensemble forecasts to correct for bias.

It is widely accepted by the meteorological community that wind direction and speed are often related. Early research dealing with wind speed and direction worked to develop a bivariate distribution for representing these variables jointly (see for example, Erdem and Shi (2011b), Carta et al. (2008), Weber (1991), Fisher and Lee (1994), and Bao et al. (2010)). Some studies have developed models to produce bivariate wind speed and direction forecasts using time series models to generate very short-term forecasts (Erdem and Shi (2011a), Cripps et al. (2005), and Klausner et al. (2009)).

Very recently, a few authors have incorporated meteorological ensemble forecasts in bivariate wind speed and direction models. Pinson (2012) represented near-surface wind through two wind components. The distribution of the vector components were modeled using a bivariate
Normal distribution where the mean of each component was defined as a weighted average of the corresponding ensemble component forecasts. Additionally, the author allowed for dependence between the components through the use of a general correlation parameter. This research was unique in that parameters were estimated and updated based on the previous time period’s estimated parameters rather than using a moving training window. Sloughter et al. (2013) made 48 hour-ahead forecasts of wind speed and direction by representing the quantities as vector components. In a manner similar to Pinson (2012), the mean wind vector components are given as a linear combination of ensemble vector component forecasts. Once the mean structure is determined, transformed errors are modeled with a bivariate Normal distribution. The variance structure of the distribution is assumed to be constant across ensemble members, and correlation between the vector components is permitted. Finally, Schuhen et al. (2012) also develop a model for wind components. The observed components are assumed to follow a bivariate Normal distribution where the mean of each component is a linear combinations of ensemble forecasts for that component. However, unlike the previous two works, Schuhen et al. (2012) use a very specific correlation structure that takes the form of a cosine function with wind direction as the input angle.

In this work, we propose a bivariate model leading to forecasts of the wind vector components. Unlike the aforementioned researchers, we do not model the wind components directly but instead model the bias of the meteorological component forecasts. Sloughter et al. (2013) implemented Bayesian model averaging (BMA) in estimating the error structure of the forecasts, while we specify a complete Bayesian model for the biases. Like the three works mentioned above, we use a bivariate Normal distribution to represent the overall structure of our data and biases. Finally, our model will specify a correlation structure similar to the one implemented by Schuhen et al. (2012).

In Chapter 2 we presented a model for forecasting wind speed and direction, through the bias of the wind components, assuming independence. This model provided the added benefit of producing wind direction forecasts. However, in a majority of the cases presented, the model
incorporating wind speed and direction did not perform better than the model that considered
wind speed only. There are several possible explanations for these results, and we investigate
possible improvements to the model by allowing for dependence between the bias components.
The primary goal of this work is to extend concepts of the model presented in Section 2.5 while
allowing for correlation between the $u$ and $v$ wind components’ biases. Before implementing
the model on observed data and forecasts, several simulation scenarios are generated to char-
acterize the behavior of the model under varying conditions and examine the conditions under
which parameter estimates are reasonable. Additionally, we use simulation to evaluate the
consequences of mis-specifying the forecasting model (i.e. what happens if we fit the bivariate
model to independent data) under varying conditions.

### 3.2 Bivariate Model for Wind Speed and Direction

We combine information about wind speed and direction (both observed and forecasted)
by examining the vector components of the wind. We define two vectors as:

$$\begin{align*}
  u &= r \cos \theta \\
  v &= r \sin \theta,
\end{align*}$$

where $r$ represents the wind speed and $\theta$ represents the wind direction. The quantities $u$
and $v$ in expression (3.1) are also referred to as the $x$ and $y$ vector components.

#### 3.2.1 General Model Description

Given the previous research by Sloughter et al. (2013), Pinson (2012), and Schuhen et al.
(2012) and preliminary investigation, it is likely reasonable to assume that the wind vector
components and bias components follow a bivariate Normal distribution. With this in mind,
we let $Y_{0t}$ and $\theta_{0t}$ represent the observed wind speed and direction at time $t = 1, \ldots, T$,
respectively. Then, we define the observed $x$ and $y$ wind components at time $t$, $U_{0t}$ and $V_{0t}$, as:
\[ U_{0t} = Y_{0t} \cos \theta_{0t}, \]
\[ V_{0t} = Y_{0t} \sin \theta_{0t}. \]  

(3.2)

We assume that the observed wind vector components follow a bivariate Normal distribution. More specifically, we assume:

\[
\begin{pmatrix}
U_{0t} \\
V_{0t}
\end{pmatrix}
\sim
\text{MVN}_2\left(\begin{pmatrix} \mu_{0t} \\ \nu_{0t} \end{pmatrix}, \Sigma_0 \right),
\]

(3.3)

where

\[
\Sigma_0 = \begin{pmatrix}
\lambda_0 & \rho_0 \lambda_0 \\
\rho_0 \lambda_0 & \lambda_0
\end{pmatrix}.
\]

(3.4)

Further, let \( Y_{j,t} \) and \( \theta_{j,t} \) represent the forecasted wind speed and direction at time \( t = 1, \ldots, T \) for meteorological model \( j = 1, \ldots, M \). We use the forecasted wind speed and direction to define the forecasted x and y wind components, \( U_{j,t} \) and \( V_{j,t} \), in a manner parallel to expression (3.2) and allow for the bias of the vector components, for a given meteorological model, to be correlated. The explicit structure of the bias of forecasted wind vector components is:

\[
\begin{pmatrix}
U_{j,t} \\
V_{j,t}
\end{pmatrix}
- \begin{pmatrix}
\mu_{0t} \\
\nu_{0t}
\end{pmatrix}
\sim
\text{MVN}_2\left(\begin{pmatrix} b_{uj,t} \\ b_{vij,t} \end{pmatrix}, \Sigma_{j,t} \right).
\]

(3.5)

The above assumptions allow for unique means and variances for the vector biases of each meteorological model and treat the bias of the meteorological vector forecasts as independent from one another. We define the covariance structure of the bias of the forecasted components for meteorological model \( j = 1, \ldots, M \) at time \( t = 1, \ldots, T \) as:

\[
\Sigma_{j,t} = \begin{pmatrix}
\lambda_j & \rho_t \lambda_j \\
\rho_t \lambda_j & \lambda_j
\end{pmatrix},
\]

(3.6)
and the variance of the true-bias vector component for each model is assumed to follow an Inverse Gamma distribution:

\[ \lambda_j \sim IG(c, d), \quad (3.7) \]

where \( c \) and \( d \) are known. Although we are allowing for the variance of the bias, of both components, to vary from one meteorological model to another, we assume that the underlying distribution of \( \lambda_j \) is the same across meteorological models. The 6 NAM meteorological models stem from the same physical model but with different perturbations of initial conditions, so the assumption of a common underlying variance structure for the bias seems reasonable; the same is true for the GFS meteorological models. Additionally, the variances of the bias of the x and y components, for a given meteorological model, are assumed to be equal. Careful consideration must be given to appropriate specification of the correlation between the wind vector bias components and is discussed further in Section 3.2.2.

We then assume that the mean bias of the vector components are linearly related to co-
variates \( x_{1j,t} \) and \( x_{2j,t} \), which we use to represent the forecasted wind shear and temperature difference for meteorological model \( j = 1, \ldots, M \) at time \( t = 1, \ldots, T \). We also allow for the linear relationship between the bias of a meteorological component to vary from one meteorological model to another and from one vector component to the other. These assumptions can be explicitly written as:

\[
\begin{align*}
    b_{uj,t} &= \alpha_{0j} + \alpha_{1j}x_{1j,t} + \alpha_{2j}x_{2j,t} + \epsilon_{j,t}, \\
    b_{vj,t} &= \eta_{0j} + \eta_{1j}x_{1j,t} + \eta_{2j}x_{2j,t} + \omega_{j,t},
\end{align*}
\]

where

\[
\begin{align*}
    \epsilon_{j,t} &\sim N(0, \tau^{-1}), \\
    \omega_{j,t} &\sim N(0, \gamma^{-1}).
\end{align*}
\]
3.2.2 Correlation Structure

A naive approach to modeling the correlation between wind vector components would be to estimate the correlation as a constant across all values of wind speed and wind direction. However, upon the results of a closer examination, we consider an alternative approach. First, consider dividing the Cartesian plane into eight sections of equal area. We divide the plane by vectors bisecting each quadrant. Figure 3.1 shows the eight sections of interest.

In order to illustrate the relationship between vector components, a set of wind speed and direction values are simulated independently. The $u$ and $v$ components are then calculated as given in expression (3.1). Finally, for a given angle, $\theta$, the correlation between all vector components produced from an angle in the range $(\theta - 5, \theta + 5)$ (in degrees) is calculated. This calculation is repeated for all integer-valued angles from 0 to 359 degrees, where the angle of interest is located at the center of the window of angles, $(\theta - 5, \theta + 5)$, considered. Figure 3.2 illustrates the results of this simulation by giving the correlation between $u$ and $v$ for a 10
The strength of the relationship between $u$ and $v$ is a periodic function of the direction $\theta$. We examine the relationship between the meteorologically forecasted wind direction and the correlation between the bias of $u$ and $v$ components for the 102 meteorological cases discussed in Section 2.2. We first divide the hourly data into 16 subsets by bisecting the eight regions displayed in Figure 3.1. The subset of an observation is determined based on the forecasted wind direction $\theta_{j,t}$ for $j = 1, \ldots, M$ and $t = 1, \ldots, T$. For our data, $M = 12$ and $T = 54$ for each of the 102 cases. Then, for each subset of data, we calculate the observed wind vector biases as:

$$
\text{bias}_{U_{j,t}} = u_{j,t} - u_{0t},
\text{bias}_{V_{j,t}} = v_{j,t} - v_{0t},
$$

(3.10)

where $u_{j,t}$ and $v_{j,t}$, and $u_{0t}$ and $v_{0t}$ are the forecasted vector values for model $j$ and time $t$.
and the observed vector values at time $t$, respectively. Finally, for each section of the Cartesian plane, we compute the correlation between the bias vector components for the corresponding subset of data. Figure 3.3 displays the results of these calculations by plotting the correlation against the circular average angle of each section. Note, that in our application, 0 degrees corresponds to a northerly wind (wind coming from the North) and 90 degrees corresponds to an Easterly wind. Therefore, wind directions between 0 and 45 degrees correspond to Area 1, wind directions between 45 and 90 degrees correspond to Area 2, etc. The general pattern of the correlation over wind direction looks similar to the plot based on simulated wind vector components in Figure 3.2 but with a smaller amplitude. Differences in amplitude and the imperfect pattern arise from the fact that we are examining the vector bias components rather than the observed vector components themselves.

Based on this finding, we propose modeling the correlation between the bias vector components as:
\[ \rho_t = \kappa \sin (2\pi \theta_0 t \alpha - \beta) \quad | \kappa | \leq 1, \quad (3.11) \]

in conjunction with the model specified in Section 3.2.1.

### 3.2.3 Priors

Several parameters require the specification of prior distributions. The prior distribution of the regression parameters \( \alpha_{0j}, \alpha_{1j}, \alpha_{2j}, \eta_{0j}, \eta_{1j}, \) and \( \eta_{2j} \) are chosen to be conjugate prior distributions. Additionally, the prior distribution of each regression parameter centered at 0 with an inverse variance, \( \lambda_r^{-1} \), chosen so that each prior is diffuse but proper. More exactly the regression parameter priors are given as:

\[
\begin{align*}
\alpha_{0j} &\sim N(0, \lambda_r^{-1}) \Rightarrow \Pi(\alpha_{0j}) = \left( \frac{\lambda_r}{2\pi} \right)^{1/2} \exp \left\{ -\frac{\lambda_r}{2} (\alpha_{0j})^2 \right\}, \\
\alpha_{1j} &\sim N(0, \lambda_r^{-1}) \Rightarrow \Pi(\alpha_{1j}) = \left( \frac{\lambda_r}{2\pi} \right)^{1/2} \exp \left\{ -\frac{\lambda_r}{2} (\alpha_{1j})^2 \right\}, \\
\alpha_{2j} &\sim N(0, \lambda_r^{-1}) \Rightarrow \Pi(\alpha_{2j}) = \left( \frac{\lambda_r}{2\pi} \right)^{1/2} \exp \left\{ -\frac{\lambda_r}{2} (\alpha_{2j})^2 \right\}, \\
\eta_{0j} &\sim N(0, \lambda_r^{-1}) \Rightarrow \Pi(\eta_{0j}) = \left( \frac{\lambda_r}{2\pi} \right)^{1/2} \exp \left\{ -\frac{\lambda_r}{2} (\eta_{0j})^2 \right\}, \\
\eta_{1j} &\sim N(0, \lambda_r^{-1}) \Rightarrow \Pi(\eta_{1j}) = \left( \frac{\lambda_r}{2\pi} \right)^{1/2} \exp \left\{ -\frac{\lambda_r}{2} (\eta_{1j})^2 \right\}, \\
\eta_{2j} &\sim N(0, \lambda_r^{-1}) \Rightarrow \Pi(\eta_{2j}) = \left( \frac{\lambda_r}{2\pi} \right)^{1/2} \exp \left\{ -\frac{\lambda_r}{2} (\eta_{2j})^2 \right\}, 
\end{align*}
\]

(3.12)

for \( j = 1, \ldots, M \). The prior distributions for \( \mu_0t \) and \( \nu_0t \) and the regression inverse error-variances, \( \tau \) and \( \gamma \) are chosen to be conjugate:
\[ \tau \sim G(a, b) \Rightarrow \Pi(\tau) = \frac{b^a}{\Gamma(a)} \tau^{a-1} \exp\{-b\tau\}, \]

\[ \gamma \sim G(a, b) \Rightarrow \Pi(\gamma) = \frac{b^a}{\Gamma(a)} \gamma^{a-1} \exp\{-b\gamma\}, \]

\[ \mu_{0t} \sim N(M_\mu, \lambda_{\mu}^{-1}) \Rightarrow \Pi(\mu_{0t}) = \left(\frac{\lambda_{\mu}}{2\pi}\right)^{1/2} \exp\left\{-\frac{\lambda_{\mu}}{2} (\mu_{0t} - M_\mu)^2\right\}, \]

\[ \nu_{0t} \sim N(M_\nu, \lambda_{\nu}^{-1}) \Rightarrow \Pi(\nu_{0t}) = \left(\frac{\lambda_{\nu}}{2\pi}\right)^{1/2} \exp\left\{-\frac{\lambda_{\nu}}{2} (\nu_{0t} - M_\nu)^2\right\}, \]  

\[ \text{(3.13)} \]

for \( t = 1, \ldots, T \). Finally, uniform priors are specified for \( \kappa, \alpha, \) and \( \beta \), as these parameters are either mathematically or practically constrained to fixed intervals of possible values:

\[ \kappa \sim U(e, f) \Rightarrow \Pi(\kappa) = \frac{1}{f - e} \mathbb{1}_{-1 \leq \kappa \leq 1}, \]

\[ \alpha \sim U(g, h) \Rightarrow \Pi(\alpha) = \frac{1}{h - g} \mathbb{1}_{-2\pi \leq \alpha \leq 2\pi}, \]

\[ \beta \sim U(p, q) \Rightarrow \Pi(\beta) = \frac{1}{q - p} \mathbb{1}_{-2\pi \leq \beta \leq 2\pi}. \]  

\[ \text{(3.14)} \]

### 3.3 Conditional Posterior Distributions & Parameter Estimation

In order to obtain a random sample from the joint probability distribution of the parameters of interest we first derive the conditional posterior distributions up to a proportionality constant using Bayes' theorem:

\[ p(x \mid y) \propto p(y \mid x) \cdot \Pi(x), \]  

\[ \text{(3.15)} \]

where \( x \) denotes the parameter of interest, \( y \) denotes other relevant parameters and data values, and \( \Pi(x) \) is the prior distribution of \( x \). The conditional posterior distribution of all parameters that we wish to estimate can be obtained in a closed distributional form with the exception of \( \kappa, \alpha, \) and \( \beta \). The normalization constants for the posterior distributions of these three parameters are not easily derived as the posteriors are not of a well known distributional form. Therefore, we implement a random walk Metropolis-Hastings algorithm for each of these three parameters within a Gibbs sampling algorithm. Samples from the conditional
posterior distributions for the other parameters of interest can be obtained with a Gibbs sampling method. The explicit form and calculation of the posterior distributions can be found in Appendix B.

3.4 Simulation Study Setup

3.4.1 Values Necessary for Simulation

In order to simulate data from the bivariate model, we must specify values for $\tau$, $\gamma$, $\kappa$, $\alpha$, and $\beta$, and choose the number of hours for which we have observations, $T$ (in application, this is our training period), and the number of ensemble members $M$. For this work, we are interested in the application of this model to our forecasted data, thus the number of ensemble members will stay fixed at $M = 12$ for any future simulations. We also need to specify values for $\alpha_0, \alpha_1, \alpha_2, \eta_0, \eta_1, \eta_2$, for $j = 1, \ldots, 12$. Finally, in order to simulate from the model we must have covariate values, $x_1,j,t$ and $x_2,j,t$, for $j = 1, \ldots, 12$ and $t = 1, \ldots, T$, as well as values for the observed wind speed, $y_0,t$, and observed wind direction $\theta_0,t$.

Rather than arbitrarily choosing values for the aforementioned parameters, covariates, and observations, we determine reasonable values that are likely to occur in our data. For the parameters discussed, we determine plausible ranges of values by looking at estimated parameter values from the hierarchical vector model presented in Section 2.6. We also examine the observed and forecasted data to determine reasonable ranges of values for $x_1,j,t$, $x_2,j,t$, and $y_0,t$. Table 3.1 gives summary statistics for these values as observed or as estimated from the hierarchical vector model over all 102 cases. We take the 2.5th and 97.5th percentile values of each to be the range of values we will consider for the respective parameter, covariate, or observation of interest. We also consider all wind directions from 0 to 360 degrees when determining values for the observed $x$ and vertical $y$ components.
Table 3.1  Value ranges for parameters, forecasts, and observed values for simulation purposes.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>2.5 Percentile</th>
<th>Mean</th>
<th>97.5 Percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed Speed (m/s)</td>
<td>2.417</td>
<td>7.534</td>
<td>20.04</td>
</tr>
<tr>
<td>$\alpha_{0j}$</td>
<td>-2.507</td>
<td>0.007</td>
<td>2.712</td>
</tr>
<tr>
<td>$\alpha_{1j}$</td>
<td>-1.291</td>
<td>-0.033</td>
<td>1.349</td>
</tr>
<tr>
<td>$\alpha_{2j}$</td>
<td>-2.142</td>
<td>-0.0271</td>
<td>1.701</td>
</tr>
<tr>
<td>$\lambda^{-1}_{j}$</td>
<td>0.012</td>
<td>1.752</td>
<td>3.274</td>
</tr>
<tr>
<td>$x_{1j,t}$</td>
<td>-4.260</td>
<td>2.688</td>
<td>9.585</td>
</tr>
<tr>
<td>$x_{2j,t}$</td>
<td>-1.336</td>
<td>0.225</td>
<td>7.707</td>
</tr>
</tbody>
</table>

3.4.2 General Steps for Simulation Studies

In the following section we outline the general steps taken when conducting any of the following simulation studies. First, we define some notation; let $p$ denote the parameter/value or set of parameters/values of interest. That is, $p$ is/are the attribute(s) whose effect we wish to investigate. Let $\Phi$ denote the set of parameters/values that are required for simulation but will be fixed when varying conditions of $p$. Steps for conducting our simulation studies are as follows.

1. Randomly choose $N$ sets of parameters/values $\Phi$ by selecting values in an appropriate manner from the set of predetermined ranges specified in Table 3.1 or specified in the model structure (e.g. $|\kappa| \leq 1$). Details on specific value selection follow.

   - **Forecasted Wind Shear $x_{1j,t}$**: Values for forecasted covariates are often similar across meteorological models for a given time $t$. In an effort to simulate this structure:

     (a) Randomly select a value for wind shear, $x_{1}^*$, within the acceptable range of values.

     (b) Randomly sample $M = 12$ wind shear values from a Normal distribution with a mean equal to $x_{1}^*$ and standard deviation equal to 1.33 m/s. Note, the value 1.33 m/s was the average standard deviation of forecasted wind shear across models, at a given time, across the 102 cases.

     (c) Set the 12 wind shear values sampled in Step (b) to be $x_{1j,t}$.

     (d) Repeat Steps (a) through (c) for until $t = T$. 

• *Forecasted Temperature Difference* $x_{2jt}$: Values for forecasted covariates are often similar across meteorological models for a given time $t$. In an effort to simulate this structure:

(a) Randomly select a value for temperature difference, $x^*_2$, within the acceptable range of values.

(b) Randomly sample $M = 12$ temperature difference values from a Normal distribution with a mean equal to $x^*_2$ and standard deviation equal to 0.51 m/s. Note, the value 0.51 m/s was the average standard deviation of forecasted temperature differences across models, at a given time, across the 102 cases.

(c) Set the 12 temperature difference values sampled in Step (b) to be $x_{2jt}$.

(d) Repeat Steps (a) through (c) for until $t = T$.

• *All Other Model Inputs*: Values of any other values are randomly sampled from the range of acceptable values.

• Note that some values were chosen as “nice” rounded numbers for the purposes of making plots easier to read when displaying results.

2. Simulate a dataset for each combination of: the unique set of values $p$ being considered and a set of random parameters/values, $\Phi$, resulting in $N$ simulated datasets for each unique set of values $p$. Note that the $N$th dataset simulated for each unique set of $p$ was simulated under the exact same conditions with the exception of the values of $p$.

3. Estimate model parameters for all simulated datasets, using the MCMC algorithm described above, letting algorithm run for 2,000 iterations each time.

4. Each parameter within a dataset is estimated as the mean of the sample from the posterior distribution, after discarding the a burn-in period of 1,000 iterations.

3.5 **Simulation Studies for Determining Parameter Estimation Adequacy**

Before attempting to implement our model on observed data, we investigate the feasibility of parameter estimation for the bivariate model. We wish to investigate and characterize the
behavior and quality of model parameter estimates for varying conditions of interest. In particular, we are interested in the effect of: the strength of the dependence between the vector components, sample size, and possible interaction between these two factors.

3.5.1 Strength of Dependence

The major difference between the model proposed for wind component bias in Chapter 2 and this model is the independence or dependence assumed by the respective model. It is reasonable to assume that the strength of correlation between the bias components could have an effect on the quality of parameter estimation for the bivariate model. To begin, we create situations in which parameter estimates should be dependable by specifying a very large sample size (i.e. a large number of samples in our training data). For the following simulations, we set the number of observed hours of data equal to $T = 500$.

The magnitude of the correlation between the bias of the wind vector components is dictated by the value of $\kappa$ in expression (3.11) and the direction of the wind. With the number of hours of observations set, we consider varying values for $\kappa$ and use $\kappa = -0.9, -0.7, -0.5, -0.3, -0.1, 0.1, 0.3, 0.5, 0.7, \text{ and } 0.9$ to evaluate the model’s ability to estimate parameters under for varying correlation strengths. We set the values of $\alpha$ and $\beta$ equal to 1 and 0, respectively. Additionally, we fix the values of regression parameters as: $\alpha_{0j} = -0.75, \alpha_{1j} = 0.5, \alpha_{2j} = -0.25, \eta_{0j} = 0.75, \eta_{1j} = 0.4, \text{ and } \eta_{2j} = -0.5$ for $j = 1, \ldots, M$. The values of the variability in the bias of wind components are set to $\lambda_j = 1$ for $j = 1, \ldots, M$. Finally, covariate values, and observed wind speed and direction were randomly sampled from intervals of plausible values as given in Table 3.1. For each value of $\kappa$, we leave all parameter values, covariate values, etc. at their fixed values. We then simulate $N = 200$ datasets of forecasted wind component values from the model with a burn-in period of 1,000 iterations. As a last step, the parameter values are estimated for each of the 200 datasets.

Figure 3.4 displays one MCMC sampling chain (for one of the $N = 200$ datasets), post
burn-in period, for $\kappa$ when data were simulated with $\kappa$ set equal to -0.9, a case where the maximum value of correlation is large in magnitude. The sampling algorithm generates values of $\kappa$ centered close to the true value with a small amount of variability in these values. More specifically, the mean value of $\kappa$ for this chain is equal to 0.9034.

It should also be noted that the starting values for parameter estimates are varied and three chains are run for a dataset to ensure convergence of parameter estimates. Figure 3.5 gives an example of the estimated value of $\kappa$ for three chains at different starting values.

We wish to evaluate the model’s ability to estimate other parameters in the model for this example. Figure 3.6 displays the sampling chains for various other parameters with the true value used for simulation listed above the subfigure and the horizontal, red lines representing the true values. For an amplitude large in magnitude ($\kappa = -0.9$ in this case) we see that parameter estimates are in line with the “true values.” Results over the additional simulated datasets for $\kappa = -0.9$ with varying parameter and covariate values show similar results.

We expect that parameter estimates would be well behaved for a model with strong cor-
Figure 3.5 MCMC convergence for 3 chains when $\kappa = -0.9$ is used for simulation.

(a) $\lambda_1 = 1$  
(b) $\alpha = 1$  
(c) $\beta = 0$

(d) $\alpha_{01} = -0.75$  
(e) $\alpha_{11} = 0.5$  
(f) $\alpha_{21} = -0.25$

Figure 3.6 MCMC sampling chains for parameters when $\kappa = 0.9$, red lines indicate true value of parameter used to simulate data.
relation. We now proceed by investigating the effect of varying the amplitude parameter on parameter estimation capabilities. Figure 3.7 shows sampling chains for $\kappa = -0.9, -0.5, -0.1, 0.1, 0.5, \text{and} 0.9$. With a sample size/training period of $T = 500$, all parameter chains look reasonable across the various levels of $\kappa$. However, we do note an increased variability in the sampling chain for smaller values of $\kappa$.

We further investigate parameter estimates of the correlation amplitude for varying levels of $\kappa$ by looking at parameter estimates over all 200 datasets for each $\kappa$ of interest. The sample mean of each sampling chain (post burn-in) is taken as the estimate of the correlation amplitude, $\hat{\kappa}$. Table 3.2 gives summary statistics for $\hat{\kappa}$ across varying levels of $\kappa$. In general, the mean of sampling chains produce reliable estimates of $\kappa$ regardless of the magnitude correlation.
Finally, we examine parameter estimates for the other parameters of interest over varying values of $\kappa$. Figure 3.8 displays the relative difference for various regression parameter estimates for the $N = 200$ cases over several values of $\kappa$. In this case, we examine the regression parameter estimates for one meteorological model ($j = 5$), but results are similar across meteorological models. We define the mean relative error to be:

$$MRD = \frac{\hat{p} - p}{p}, \quad (3.16)$$

where $p$ is the parameter value used to simulate data and $\hat{p}$ is the estimate based on the simulated data. On average, the regression parameter estimates are very close to the true value which was used in the simulation, and this is regardless of the strength of correlation.

Figure 3.9 displays the relative difference for the other two parameters of the correlation function, $\alpha$ and $\beta$ for various values of $\kappa$. On average, the estimates tend to be close to the actual value used for simulation. However, the variability in the errors of the estimates is much larger for small magnitude values of $\kappa$ than it is for estimates when $\kappa$ is large in magnitude. Additionally, we see that the behavior of estimates for the corresponding negative and positive values of $\kappa$ are similar. Estimation of parameters related to the correlation function is less precise when the amplitude of the correlation is very small.

<table>
<thead>
<tr>
<th>$\kappa$</th>
<th>Min.</th>
<th>Q1</th>
<th>Median</th>
<th>Mean</th>
<th>Q3</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-0.9$</td>
<td>-0.906</td>
<td>-0.896</td>
<td>-0.894</td>
<td>-0.894</td>
<td>-0.891</td>
<td>-0.882</td>
</tr>
<tr>
<td>$-0.5$</td>
<td>-0.509</td>
<td>-0.501</td>
<td>-0.498</td>
<td>-0.498</td>
<td>-0.495</td>
<td>-0.587</td>
</tr>
<tr>
<td>$-0.1$</td>
<td>-0.056</td>
<td>-0.070</td>
<td>-0.097</td>
<td>-0.096</td>
<td>-0.111</td>
<td>-0.128</td>
</tr>
<tr>
<td>$0.1$</td>
<td>0.053</td>
<td>0.073</td>
<td>0.095</td>
<td>0.098</td>
<td>0.109</td>
<td>0.125</td>
</tr>
<tr>
<td>$0.5$</td>
<td>0.492</td>
<td>0.499</td>
<td>0.501</td>
<td>0.502</td>
<td>0.504</td>
<td>0.513</td>
</tr>
<tr>
<td>$0.9$</td>
<td>0.919</td>
<td>0.923</td>
<td>0.924</td>
<td>0.924</td>
<td>0.926</td>
<td>0.928</td>
</tr>
</tbody>
</table>

Table 3.2 Summary statistics of $\hat{\kappa}$ for 200 simulated datasets per value of $\kappa$. amplitudes.
Figure 3.8 Mean relative errors for parameters estimates over 200 simulated datasets for varying values of $\kappa$. 
3.5.2 Sample Size

The estimates of parameters are likely to degrade with reduced sample sizes, $T$. We investigate the extent to which our parameter estimates are affected for various reductions in training period length (sample size). To begin, we fix $\kappa$ at a value of 0.9. The interaction of varying the strength of correlation and sample size will be investigated in Section 3.5.3. Additionally, the values of other inputs, with the exception of sample size $T$, that are needed to simulate from the model are randomly selected and then fixed at their respective values. For each sample size, we simulate $N = 200$ datasets and estimate all relevant parameters. Table 3.3 gives the mean square error (MSE) of parameter estimates for the $N = 200$ simulated datasets per sample size $T$, for varying sample sizes. In the table, regression parameters for the mean bias of each component ($\alpha_{0j}, \alpha_{1j}, \text{etc.}$) are presented for only one meteorological $j$, however the results seen here are similar for all regression parameters. The MSE values for sample sizes greater than $T = 100$ are quite small. In general, estimates for a majority of the parameters are reasonably accurate on a consistent basis for samples of $T = 30$ or more. When $T = 20$, some of the estimates are more accurate than others, and parameter estimates degrade markedly for nearly all parameters when $T = 10$. This is not surprising due to the fact that the number of parameters that need to be estimated is much greater than the number of time points for this case.
Table 3.3  Mean square error (MSE) of parameter estimates for all 200 simulated cases, over various sample sizes.

<table>
<thead>
<tr>
<th></th>
<th>$T = 10$</th>
<th>$T = 20$</th>
<th>$T = 30$</th>
<th>$T = 50$</th>
<th>$T = 100$</th>
<th>$T = 200$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha = 1.5$</td>
<td>0.0324</td>
<td>0.026</td>
<td>0.0014</td>
<td>0.001</td>
<td>0.001</td>
<td>0.0002</td>
</tr>
<tr>
<td>$\kappa = 0.9$</td>
<td>0.2084</td>
<td>0.1351</td>
<td>0.0247</td>
<td>0.0069</td>
<td>0.0081</td>
<td>0.0003</td>
</tr>
<tr>
<td>$\beta = 0.5$</td>
<td>0.2416</td>
<td>0.1311</td>
<td>0.0094</td>
<td>0.0025</td>
<td>0.0006</td>
<td>0.0004</td>
</tr>
<tr>
<td>$\alpha_0j = 0.6$</td>
<td>1.4697</td>
<td>1.0669</td>
<td>0.8363</td>
<td>0.1521</td>
<td>0.0624</td>
<td>0.0421</td>
</tr>
<tr>
<td>$\alpha_{1j} = -0.5$</td>
<td>5.6151</td>
<td>0.5554</td>
<td>0.2927</td>
<td>0.0490</td>
<td>0.0159</td>
<td>0.0112</td>
</tr>
<tr>
<td>$\alpha_{2j} = 0.4$</td>
<td>2.2169</td>
<td>1.6705</td>
<td>1.4528</td>
<td>0.4237</td>
<td>0.2680</td>
<td>0.1231</td>
</tr>
<tr>
<td>$\eta_0j = 0.55$</td>
<td>0.1104</td>
<td>0.1146</td>
<td>0.0806</td>
<td>0.0612</td>
<td>0.0414</td>
<td>0.0339</td>
</tr>
<tr>
<td>$\eta_{1j} = 0.25$</td>
<td>3.9699</td>
<td>0.2148</td>
<td>0.1166</td>
<td>0.0640</td>
<td>0.0490</td>
<td>0.0019</td>
</tr>
<tr>
<td>$\eta_{2j} = -0.5$</td>
<td>0.6334</td>
<td>0.4962</td>
<td>0.4478</td>
<td>0.1991</td>
<td>0.0915</td>
<td>0.0574</td>
</tr>
<tr>
<td>$\lambda_j = 0.5$</td>
<td>0.2057</td>
<td>0.0584</td>
<td>0.0223</td>
<td>0.0219</td>
<td>0.004</td>
<td>0.0025</td>
</tr>
</tbody>
</table>

3.5.3 Interaction of Sample Size and Correlation Strength

When applying the bivariate model to observed data, the value of $\kappa$ is neither known nor fixed over various cases. In Section 3.5.1, we saw that parameter estimates are adequate and stable for a large sample size $T$. In practice, we choose the length of the training period (number of samples) $T$. Therefore, it is particularly important to know how well we are able to estimate parameters over various dependence strengths for different choices of training period length, $T$. A training length of $T = 30$ is of particular interest to us as this is what was used when implementing the bivariate model presented in Chapter 2. Thus, we will examine $T = 30$ closely for the bivariate model specified in Sections 3.2.1 and 3.2.2 and assess whether implementing this model to our forecasting cases from Chapter 2 is feasible.

We implement a simulation study and consider $T$ and $\kappa$ to be the parameters of interest that we will vary. More specifically, we consider combinations of the two values for $T = 10, 20, 30, 50, 75,$ and $100$ and $\kappa = -0.9, -0.7, -0.5, -0.3, -0.1, 0.1, 0.3, 0.5, 0.7,$ and $0.9$. However, since we are interested in the magnitude of the amplitude of the correlation parameter, we will present results for $| \kappa | = 0.1, 0.3, 0.5, 0.7,$ and $0.9$ by combining results from $\kappa = 0.9$ and -$0.9$, combining $\kappa = 0.7$ and -$0.7$, and so on.
Table 3.4 Mean absolute relative errors for $\alpha_{1j}$ over various sample size and dependence strength values.

| $T$ | $|\kappa|$ | 0.9 | 0.7 | 0.5 | 0.3 | 0.1 |
|-----|-----------|-----|-----|-----|-----|-----|
| 10  |           | 0.427 | 0.446 | 0.469 | 0.459 | 0.470 |
| 20  |           | 0.087 | 0.088 | 0.093 | 0.099 | 0.101 |
| 30  |           | 0.076 | 0.095 | 0.086 | 0.084 | 0.091 |
| 50  |           | 0.074 | 0.082 | 0.079 | 0.072 | 0.085 |
| 100 |           | 0.051 | 0.059 | 0.048 | 0.051 | 0.067 |
| 200 |           | 0.021 | 0.028 | 0.039 | 0.031 | 0.043 |

Parameters are estimated for the $N = 200$ simulated datasets per each combination of $\kappa$ and $T$. Table 3.4 gives the mean absolute relative difference calculated as:

$$MARD = \frac{1}{N} \sum_{i=1}^{N} \frac{|\hat{p} - p|}{p},$$

for the regression parameter $\alpha_{1j}$. Although these results are for only one of the many regression parameters, they are generally representative of what was observed across all regression parameters. Parameter estimates are relatively stable across values of $\kappa$ for a fixed sample size, and parameter estimates degrade as the sample size $T$ decreases. Additionally, the parameter estimate accuracy drops off a considerable amount between $T = 10$ and $T = 20$.

Table 3.5 gives the mean absolute relative error for parameter estimates of $\alpha$. Similar to the pattern observed in Table 3.4, the parameter estimates degrade as the sample size decreases, but the drop off in accuracy between $T = 10$ and $20$ is much larger than it was for the regression parameter estimation. This is likely due to the difficulty in estimating parameters from the correlation function, that we saw earlier, when $T$ is small and $\kappa$ is small.

Due to the nature of our observed data, we are particularly interested in our ability to estimate parameters when $T = 30$. Figure 3.10 displays boxplots of the relative error (expression (3.16)) for the regression parameters of the model over various values of $\kappa$. The parameter estimates are close to the actual value of each parameter of interest, and the magnitude of $\kappa$
Table 3.5  Mean absolute relative errors for $\alpha$ over various sample size and dependence strength values.

<table>
<thead>
<tr>
<th>$T$</th>
<th>$\kappa$</th>
<th>0.9</th>
<th>0.7</th>
<th>0.5</th>
<th>0.3</th>
<th>0.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1.341</td>
<td>1.361</td>
<td>1.402</td>
<td>1.464</td>
<td>1.408</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.337</td>
<td>0.352</td>
<td>0.371</td>
<td>0.380</td>
<td>0.386</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.214</td>
<td>0.220</td>
<td>0.228</td>
<td>0.231</td>
<td>0.229</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.074</td>
<td>0.082</td>
<td>0.079</td>
<td>0.072</td>
<td>0.095</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.051</td>
<td>0.069</td>
<td>0.063</td>
<td>0.061</td>
<td>0.073</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>0.037</td>
<td>0.031</td>
<td>0.034</td>
<td>0.043</td>
<td>0.047</td>
<td></td>
</tr>
</tbody>
</table>

does not seem to affect the accuracy or variability of our parameter estimates.

On the other hand, Figure 3.11 shows boxplots of the relative error for the parameters $\alpha$ and $\beta$ of the correlation function. The model’s ability to estimate these parameters is good on average. However, as the magnitude of the amplitude parameter $\kappa$ decreases the variability in the parameter estimate increases noticeably.

Although the quality of parameter estimates can be compared, we are ultimately interested in making future forecasts based on the model fit with a training set of data. We evaluate the model’s forecasting efficacy over the combinations of training length and $\kappa$ specified at the beginning of this section (3.5.3). Instead of simulating datasets of length $T$ and then estimating parameters, we simulate datasets with length $T + 24$ and fit the model with the first $T$ observations of the dataset and set aside the last 24 observations simulated for evaluating forecasting performance. We take this approach simulating $N = 200$ datasets for each combination of $T$ and $\kappa$. This allows us to evaluate the model’s ability to make forecasts, over various combinations of sample size and correlation strength, when we know that the model is correctly specified.

The hierarchical model produces a total of twelve posterior distributions for the each of the regression parameters in expressions (3.8) (as well as the precision of the mean bias, $\lambda_j$), one posterior for each meteorological model ensemble member. Realizations from each of these posteriors is generated from the MCMC algorithm. The forecasted covariates, $x_{1j,t}$ and $x_{2j,t}$,
Figure 3.10  Boxplots of the mean difference for regression parameters over varying magnitudes of $\kappa$ when $T = 30$. 
Figure 3.11  Boxplots of the mean difference for correlation function parameters over varying magnitudes of $\kappa$ when $T = 30$. 

and simulated parameter values $\alpha_{(k)}$, $\alpha_{(k)}^{(1)}$, $\alpha_{(k)}^{(2)}$, $\eta_{(k)}$, $\eta_{(k)}^{(1)}$, $\eta_{(k)}^{(2)}$, $\lambda_{(k)}^{(1)}$, $\lambda_{(k)}^{(2)}$, $\gamma_{(k)}$ and $\tau_{(k)}$ from the $k$th MCMC step (after the burn-in period) are used to simulate realizations of $b_{u_j,t}^{(k)}$ and $b_{v_j,t}^{(k)}$. This is done for each meteorological model $j$ for a given future time $t$ using the circular average of the meteorological models’ wind direction forecast as the wind direction $\theta_{to}$ in the correlation function, at each iteration of the MCMC algorithm. The set of simulated $b_{u_j,t}^{(1)}, b_{u_j,t}^{(2)}, \ldots, b_{u_j,t}^{(k)}$ and $b_{v_j,t}^{(1)}, b_{v_j,t}^{(2)}, \ldots, b_{v_j,t}^{(k)}$ form realizations from twelve prior predictive distributions of $b_{u_j,t}$ and $b_{v_j,t}$, one for each meteorological model $j$. For each future time period, each meteorological model has fixed wind component forecasts. The prior predictive distributions of the bias-corrected wind component forecasts, $u_{j,t}^{*}$ and $v_{j,t}^{*}$, are obtained by horizontally shifting the prior predictive distribution of each $b_{u_j,t}$ and $b_{v_j,t}$ by its respective wind component forecast. The prior predictive distributions are then combined to form one distribution, for each component, for the bias-corrected component at time $t$, using equal weights in expression (2.23). For a given time $t$, each pair of bias-corrected horizontal and vertical components is back-transformed to obtain a wind speed forecast. The back-transformation is given as:

$$y_{j,t}^{*} = \sqrt{(u_{j,t}^{*})^2 + (v_{j,t}^{*})^2}, \quad (3.18)$$

where $u_{j,t}^{*}$ and $v_{j,t}^{*}$ are the bias-corrected horizontal and vertical wind speed components.
Table 3.6 Average MAE values of 24 hour-ahead forecasts over 200 simulations.

| $T$ | $|\kappa|$ | 0.9 | 0.7 | 0.5 | 0.3 | 0.1 |
|-----|-----|-----|-----|-----|-----|-----|
| 10  | 1.160 | 1.169 | 1.194 | 1.187 | 1.189 |
| 20  | 0.587 | 0.601 | 0.628 | 0.565 | 0.571 |
| 30  | 0.371 | 0.380 | 0.382 | 0.379 | 0.384 |
| 50  | 0.374 | 0.352 | 0.380 | 0.337 | 0.368 |
| 100 | 0.255 | 0.246 | 0.267 | 0.228 | 0.253 |
| 200 | 0.235 | 0.251 | 0.224 | 0.263 | 0.238 |

Table 3.6 gives the average mean absolute error for the 200 datasets simulated for each combination of $T$ and $|\kappa|$. A similar pattern to what was seen in parameter estimation is seen here. The MAE values remain relatively constant over values of $|\kappa|$ given a sample size $T$. Additionally, the MAE values for a sample size of 10 are almost twice as large as the MAE values for a sample size of 20.

### 3.6 Consequences of Improper Model Specification

In Section 3.5.3 we examined the skill of the bivariate model in making forecasts, for varying sample sizes and dependence levels, when the model was correctly specified. We accomplished this by simulating observations used for forecasting evaluation were simulated from the exact model used to estimate parameters. In practice, we do not know the process that generates the observed data and specifying an inappropriate model is likely to have consequences. In particular, we are interested in comparing the hierarchical wind vector bias model assuming independence to the model allowing for dependence. Therefore, we consider the consequences of fitting the dependence model on data generated independently.

We simulate 200 datasets with $\kappa = 0$ (i.e. bias components are independent) for a training period of 30 hours (plus 24 hours to be set aside for forecasting validation). Then, we fit the model assuming independence and the bivariate model to each simulated data set. Table 3.7
shows the mean absolute relative difference of parameter estimates from the fixed parameter values used for simulation, for each of the models. Both models produce similar absolute differences across the regression parameters.

Although the parameter estimates appear relatively equivalent for both models, we also examine the bivariate model’s estimate of $\kappa$. The mean and median estimates of $\kappa$ over the 200 datasets were 0.0033 and 0.0075, respectively. Additionally, the minimum and maximum estimates of $\kappa$ were -0.0925 and 0.0128, respectively. Overall, the bivariate model was able to identify that the datasets were simulated independently, as evidenced by the small $\kappa$ estimates.

Ultimately, we are concerned with the bivariate model’s ability to make accurate forecasts if data are simulated from the independence model. Forecasts are made by both models for a 24 hour-ahead period for each of the 200 simulated datasets. Then, for each dataset and each model, we calculate the MAE of our forecasts over the 24 hour period. Figure 3.12 shows boxplots of the 200 MAE values for both models. The models’ performances are comparable across the 200 cases in which data came from an independent process.

### 3.7 Forecasting Performance on Observed Data

We have seen that a training period of 30 observations is a large enough sample to estimate the model parameters adequately and produce reasonable forecasts. Additionally, we have seen that the bivariate model is able to estimate and forecast in cases where the wind vector bias components are independent. We now evaluate the model’s performance when making forecasts for observed data. We are especially interested in the bivariate model’s performance in

<table>
<thead>
<tr>
<th>Model</th>
<th>$\alpha_{0j}$</th>
<th>$\alpha_{1j}$</th>
<th>$\alpha_{1j}$</th>
<th>$\eta_{0j}$</th>
<th>$\eta_{1j}$</th>
<th>$\eta_{2j}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independence</td>
<td>0.012</td>
<td>0.024</td>
<td>0.044</td>
<td>0.038</td>
<td>0.054</td>
<td>0.030</td>
</tr>
<tr>
<td>Bivariate</td>
<td>0.017</td>
<td>0.036</td>
<td>0.039</td>
<td>0.036</td>
<td>0.067</td>
<td>0.028</td>
</tr>
</tbody>
</table>

Table 3.7  Mean absolute error values for regression parameters over 200 datasets simulated from the independence model.
Figure 3.12 MAE values for 24 hour-ahead forecast periods of 200 simulated datasets, by model comparison to the performance of the hierarchical vector bias model presented in Chapter 2.

The bivariate hierarchical model for wind component bias was run on the same data used in Chapter 2. This data consists of 102 cases, between June 2008 and September 2010, where a “case” is a 54 hour period for which meteorological forecasts are produced and observed data is also available. We fit the model using the first 30 hours of each case, and then make forecasts for the remaining 24 hours.

We evaluate the performance of the hierarchical model for the vector components modeled allowing for dependence using the MAE for the 24 hour forecast period, for each case. Figure 3.13 shows the MAE values for both the hierarchical model for vector components modeled under independence and dependence, as well as the MAE values for the AMF and hierarchical model of wind speed. The hierarchical model for the wind vector components under dependence outperforms the AMF model in 66 (64.7%) of the 102 cases. The hierarchical model for wind vector components allowing for dependence has smaller MAE values than the independence model for 64 (62.7%) of the 102 cases. Additionally, the bivariate model has a smaller MAE than the hierarchical model for wind speed in and 55 (53.9%) of the 102 cases.
Table 3.8 gives the five number summary for the difference in MAE values for several pairs of models. On average, the bivariate hierarchical vector model performs better than the hierarchical model using wind speed, the independent hierarchical vector model, and the meteorological average. The bivariate model makes improvements over the independence model often, however this model is less convincing when compared to the hierarchical speed and average meteorological models.

It is interesting to note that the we notice a pattern in the worst cases (in terms of MAE) for the bivariate hierarchical model. In these cases, when we look at the observed directions used in the training period for the model they cover only a small portion of all possible angles, and the wind directions in the forecasting period tend to fall outside of this window. Figure 3.14 shows the directions used for training and directions used in the forecasting period for
Table 3.8 Summary statistics of the difference in mean absolute error (MAE) values for vector hierarchical, bivariate vector hierarchical, hierarchical, and average meteorological forecasts over 102 cases

<table>
<thead>
<tr>
<th></th>
<th>Min.</th>
<th>Q1</th>
<th>Median</th>
<th>Mean</th>
<th>Q3</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hierarchical - Bivar. Hier.</td>
<td>-4.28</td>
<td>-0.427</td>
<td>0.058</td>
<td>0.007</td>
<td>0.432</td>
<td>2.76</td>
</tr>
<tr>
<td>Vect. Hier. - Bivar. Hier</td>
<td>-1.627</td>
<td>-0.212</td>
<td>0.302</td>
<td>0.276</td>
<td>0.524</td>
<td>3.578</td>
</tr>
<tr>
<td>Met. Avg. - Bivar. Hier</td>
<td>-3.679</td>
<td>-0.240</td>
<td>0.252</td>
<td>0.202</td>
<td>0.493</td>
<td>3.443</td>
</tr>
</tbody>
</table>

Figure 3.14 Observed wind direction used in training period (black lines) compared to average meteorological forecast directions used in forecast period (red lines).

(a) Case 97: September 5, 2010  
(b) Case 81: November 5, 2009

3.8 Discussion/Future Work

We have proposed a structure for modeling the bias of forecasted wind vector components with a correlation structure. The correlation structure of wind vector bias, for this application, was shown to have a sinusoidal pattern, leading to the use of a cosine function to model the correlation. We have shown that parameter estimation is reasonable for sample sizes of 20 to
30 or larger, when we know we have specified the model correctly. Additionally, parameter estimates for parameters involved in the mean structure of the model are stable, regardless of the magnitude of the amplitude dependence parameter $\kappa$. Some difficulties in parameter estimation occur when the training period is short and the dependence is weak. However, the ability to make accurate wind speed forecasts is fairly stable over varying sample size and dependence conditions.

Additionally, we have shown that the bivariate vector model is able to identify when data comes from an independent process, leading to small estimates of $\kappa$. The bivariate model is able to make forecasts comparable to the vector model assuming independence for simulated independent data. The bivariate model was fit to observed data and used to make forecasts on the 102 cases that were used in Chapter 2. The bivariate model did provide improvements over the vector model assuming independence. When compared to the hierarchical model and meteorological average model, the bivariate model performed better in slightly more than half of all cases.

One situation that appears to lead to poor performance by the bivariate model is specific to the wind direction. In situations where the training wind directions don’t cover a large enough range of angles or angles of relevance to the forecasting period, the bivariate model performed substantially worse than the other models. This suggests that a longer training period may need to be considered. Additionally, we may consider pursuing an alternative weighting scheme when combining bias corrected vector components from different meteorological models. Finally, additional investigation regarding potential alternative covariates may be considered.
CHAPTER 4. SPATIAL MODELING OF THE WIND SPEED PROCESS
AT THE FARM-LEVEL

4.1 Introduction & Motivation

Wind is an important component of future clean renewable technologies. A better understanding of the spatial process of wind speed spatial process would be helpful for wind utility companies in making wind power forecasts and planning future wind farm layouts. In a 2008 report, the U.S. Department of Energy (DOE) emphasized the need for research focusing on understanding wind variability at the level of a wind power facility, allowing for better wind power forecasting and increasing the viability of wind power as an alternative energy source. More specifically, they highlight the demand for understanding disparities between the wind power produced at a turbine compared to the wind power forecasted by manufacturers’ specification and possible causes for these disparities (Schreck et al., 2008).

Several researchers have focused on the spatial modeling of wind speeds for the purpose of forecasting at spatial locations. Cripps et al. (2005) developed a hierarchical spatio-temporal model to combine and post-process meteorological model wind speed forecasts. This research focused on weather stations located throughout Sydney Harbour, a spatial scale much larger than that of a typical wind farm. Cutler et al. (2007) proposed a method for scaling and transforming wind speed forecasts for the purpose of assessing the probability of large-scale shifts in wind behavior at the wind farm level. Šaltytė Benth and Šaltytė (2011) modeled spatial dependencies of wind in Lithuania using a Gaussian random field.

Other researchers have proposed models to make short-term forecasts at spatial locations
based on observed data. Boukhanovsky et al. (2003) and Malmberg et al. (2005) implemented traditional autoregressive (AR) structures to model time-series of wind speeds. As an extension, Ailliot et al. (2006) introduced time-varying coefficients into an AR model in an effort to describe behavior of wind fields. Hering and Genton (2010) used a regime-switching model to spatially and temporally model wind speeds and make 2 hour-ahead forecasts for three spatial locations of interest, located in the Columbia River Valley. This area has the unique characteristic of wind directions predominately occur along one direction due to topography.

This work is done using data from a partner utility company in the Midwest. The main issues addressed in this work are also motivated by the current forecasting procedures of our utility company partner. For a given time, the company receives a single wind speed forecast from multiple models, representing one location, for the entire wind farm. These forecasts are combined to produce one forecast for the hour of interest; the combination of these forecasts is a separate issue which we have discussed in Chapters 2 and 3. Once a single forecast is obtained, it is assumed that wind speed is constant across the farm, and forecasted energy is obtained using a speed-to-power conversion algorithm. The assumption that wind speed will be constant across the wind farm is not believed to be an accurate representation of actual wind behavior. However, this is currently the most reasonable strategy as the understanding of the spatial behavior of wind is severely limited. A better understanding of the spatial behavior of wind over the farm is needed to implement alternative approaches that better match reality.

This work aims to model wind speed spatially and gain a better understanding of the wind speed process at the spatial scale of a wind farm. In particular, the interaction of wind turbines conditioned on wind speed and direction are of particular interest. A majority of research on the spatial behavior of wind is conducted at a much larger scale than is desired by utility companies. One of the major limitations in the ability of researchers to understand wind behavior at a farm-level is that little data is available at the appropriate spatial scale. The contributions of this work are novel in that we have data observed at a turbine-level, at hub-height (80 m). In addition to investigating and increasing the understanding of wind behavior at the farm-level,
we address several practical statistical issues in the application of spatial models. We implement a Markov random field model for describing the behavior of wind speed in space. Several choices must be made in the modeling process. It is usually desirable to make these decisions based on scientific understanding of the process of interest. We propose and use alternative criteria to make these decisions as little is known about the wind speed process at the farm level.

### 4.2 Data

Data for a wind farm located in the Midwest was provided by our partnering utility company. The farm has a total of 171 wind turbines; Figure 4.1 shows the locations of the turbines. Note that turbine locations were rescaled, by subtracting the minimum Easting and Northing coordinates from each location, to improve the readability of figures.

The observed hourly average wind speed is available at a turbine-level during a two year period from January 1, 2009 to December 31, 2010, for a total of 17,520 wind speed values per turbine. Additionally, the observed hourly average wind speed, wind direction, and tempera-
ture are available for a meteorological tower located at the edge of the wind farm. Although data is available for a two-year stretch, there are many time periods when data is unavailable for several turbines for a variety of reasons: ice storms, data retrieval errors, etc. We consider time periods where at least 168 of the 171 turbines have valid data available; a total of 9,125 hourly time periods have data that satisfy this criterion. Analyzing a large number of time points of data is beyond the scope of this research. We wish to obtain sets of data that are representative of various wind speed and direction conditions. Thus, we use the observed average wind speed (over all turbines) as a criterion for determining strata and take a stratified random sample of size 100 from the 9,125 possible data sets. We also choose 75 hourly data sets from two stretches of time when data is available for many consecutive hours. In total, we consider 175 hours of data (some consecutive) where each hour has data available for at least 168 turbines.

Figure 4.2 shows an interpolated map of observed hourly average wind speeds over the wind farm for January 8, 2009, 0800 (LST). The wind direction at the meteorological tower is displayed in the top left-hand corner of the plot. The plot shows variability in the observed wind speed from turbine to turbine. From the wind speed map shown and numerous other example times, it is clear that the assumption of constant wind speed over the farm, for a given hour, is not realistic.

Understanding the behavior of wind speeds over a wind farm is of particular interest, as understanding of the spatial process could lead to better wind speed, and consequently wind power, forecasts at individual turbines throughout the wind farm. To better understand the wind speed structure, we propose fitting a model to the data to that accounts for the spatial components of wind speed. To begin, we develop a model for a static hourly dataset.
4.3 General Markov Random Field (MRF) Model

A Markov random field is a natural model to consider in the context of modeling wind speeds at a farm-level. In general, we can think of wind speeds in terms of large and small scale structure. Here the large scale structure can be thought of as the overall mean wind speed for the farm, and small scale structure is represented by the variability of wind speeds from turbine to turbine. Additionally, the wind speed at a particular turbine can be affected by wake effects and turbulence created by nearby turbines. An MRF allows us to model the wind speed at a particular turbine conditional upon other neighboring turbines, and the specification of neighboring turbines lends flexibility in choices for model specification.

It has been well documented that the marginal distribution of wind speed, for a particular location, is not adequately represented by a Normal distribution (e.g. Thorarinsdottir and Gneiting (2010), Torres et al. (2005), and Bivona et al. (2011)). Many authors have used a Weibull distribution (e.g. Torres et al. (2005)), or a truncated Normal distribution (e.g. Sloughter et al. (2010)) to represent the marginal distribution of wind speeds or used a power
transformation on the data before assuming normality. However, we are concerned with the conditional distribution of wind speeds, and the assumption of normality in this context is not unreasonable. Therefore, we proceed by proposing a Gaussian MRF for modeling our data. The general form of this model follows.

Let \( \{ s_i : i = 1, \ldots, n \} \) represent the set of locations associated with the wind turbines on the wind farm; here, \( n = 171 \) if data is available for all turbines. The wind speed at location \( s_i \), at a particular point in time, is denoted as \( y(s_i) \) for \( i = 1, \ldots, n \). Given this notation and assuming constant conditional variance across locations, the conditional distribution of \( y(s_i) \) is given as:

\[
f(y(s_i) \mid \{ y(s_j) : j \neq i \}) = \frac{1}{\sqrt{2\pi\tau^2}} \exp \left[ - \frac{1}{2\tau^2} \left\{ y(s_i) - \mu(\{ y(s_j) : j \neq i \}) \right\}^2 \right]. \tag{4.1}
\]

We assume that the large scale structure of wind speeds is given as an overall constant marginal mean wind speed and variations from the marginal mean are conditional on wind speeds at other locations. More specifically, we assume that:

\[
\mu(\{ y(s_j) : j \neq i \}) = \alpha + \sum_{j=1}^{n} c_{i,j} (y(s_j) - \alpha), \tag{4.2}
\]

where \( c_{i,i} = 0 \), and \( c_{i,j} = c_{j,i} \) are necessary to ensure the model is valid. Additionally, we assume that the Markov property is valid here. In other words, we assume that the conditional distribution of wind speed, at a given location, is only dependent on wind speeds at neighboring locations:

\[
f(y(s_i) \mid \{ y(s_j) : j \neq i \}) = f(y(s_i) \mid \{ y(s_j) : s_j \in \mathbf{N}(s_i) \}), \tag{4.3}
\]

where \( \mathbf{N}(s_i) \) denotes the collection of locations in the neighborhood of location \( s_i \).
4.4 Unidirectional Gaussian MRF Model

In order to implement the model presented in Section 4.3, a neighborhood structure must be specified. One of the common four- or eight-nearest neighborhood structures cannot be used in this application. The irregular spatial domain of the wind farm and a nearest-neighbor structure leads to violations in the assumption of dependence symmetry (i.e. \( c_{i,j} = c_{j,i} \)), which is necessary for a valid MRF model. Instead we specify the neighborhood of location \( s_i \), denoted as \( N(s_i) \), for \( i = 1, \ldots, n \) by a distance criterion:

\[
N(s_i) = \{ s_j : i \neq j; d_{i,j} \leq \delta \},
\]

(4.4)

where \( d_{i,j} \) is the Euclidean distance between locations \( s_i \) and \( s_j \), and \( \delta \) is a pre-specified distance. Any reasonable choice of \( \delta \) (i.e. values which result in most locations having at least one neighbor and no more than half of all locations being neighbors) leads to unequal numbers of neighbors from location to location. We specify one dependence parameter that can be interpreted across all pairs of locations, \( \eta \), by letting \( c_{i,j} \) to be of the form:

\[
c_{i,j} = \begin{cases} 
1_{\{i \neq j\}} \frac{\eta}{n_i+n_j} & \text{for } s_j \in N(s_i) \\
0 & \text{for } s_j \notin N(s_i)
\end{cases},
\]

(4.5)

where \( n_i \) and \( n_j \) are the number of neighbors for locations \( s_i \) and \( s_j \), respectively, and \( N(s_i) \) is the neighborhood defined in expression (4.4).

4.5 Model Estimation

Once a model is fully specified, we wish to estimate the model’s parameters. The parameter estimates which minimize the likelihood function for Gaussian conditional models, such as the one given in Section 4.4, can be derived exactly for \( \alpha \) and \( \tau^2 \), and the estimate of \( \eta \) can then be obtained by the minimization of its negative log profile likelihood function (Cressie, 1993). However, as models become more complex (e.g. specifying more than one neighborhood structure in a model), deriving the normalizing constant for the exact likelihood function becomes
increasingly difficult. Thus, several alternative likelihood-based estimation procedures have been developed. Parameter estimation based on the exact likelihood function for neighborhood structures proposed in Sections 4.7.2 and 4.8 is very difficult, thus we use estimation based the pseudolikelihood. To be consistent, we also use the pseudolikelihood estimation methodology for the unidirectional model in Section 4.4.

The pseudolikelihood function for a continuous conditional MRF is given as:

\[
f(\eta) = \prod_{i=1}^{n} f(y(s_i) \mid \{y(s_j) : j \neq i\}; \eta),
\]

where \( \eta \) is the vector comprised of all model parameters (Besag, 1975). We obtain the maximum pseudolikelihood estimator by minimizing the negative log-pseudolikelihood function, denoted as \( F(\eta) \), where:

\[
F(\eta) = -\sum_{i=1}^{n} \log f(y(s_i) \mid \{y(s_j) : j \neq i\}; \eta).
\]

This estimation method will be used for all models discussed in this work.

4.6 Neighborhood Radius Selection

The choice of a radius, \( \delta \), for the neighborhood structure given in expression (4.4) is often made based on knowledge of the application or may be determined arbitrarily. In this case, there is minimal intuition for the choice of a neighborhood radius, as little research has been done on the spatial behavior of wind at the farm-level and wind behavior can vary greatly across different landscapes, regions, etc. For this particular application, distances less than 750 m lead to very few or no neighbors for most locations. Therefore, the minimum neighborhood distance that we consider is 750 m. Distances greater than 2200 m lead to several locations having neighborhood sizes equal to approximately one third of the total number of locations. Thus, the maximum neighborhood distance that we consider is 2200 m. We wish to develop a more formal method for determining an appropriate neighborhood size within the bounds
discussed above. To begin, we conduct a simulation study to investigate the sensitivity of the
dependence parameter estimate relative to the choice of neighborhood size and evaluate the a
proposed criterion’s ability to identify the “true” neighborhood size which was used to simulate
the data of interest.

4.6.1 Parameter Space of $\eta$

It is necessary to determine the range of plausible values of $\eta$ before any simulations can
be done. The joint distribution of $y = \{y(s_1), y(s_2), \ldots, y(s_n)\}$ for a conditionally specified
Gaussian MRF model is given by:

$$f(y) = (2\pi)^{n/2} |M|^{-1/2} |I - C|^{1/2} \exp \left\{ -\frac{1}{2} (y - \alpha)' M^{-1} (I - C)(y - \alpha) \right\}, \quad (4.8)$$

where $M$ is an $n \times n$ diagonal matrix with non-zero elements equal to $\tau^2$ and $C$ is an $n \times n$
matrix whose $(i,j)$th element is $c_{i,j}$ (Cressie, 1993). Additionally, since $C$ can be rewritten as
$C = \eta H$ where $H$ is a symmetric matrix whose $(i,j)$th element is $c_{i,j}/\eta$, it can be further shown
that:

$$|I - C| = \prod_{i=1}^{n} (1 - \eta h_i), \quad (4.9)$$

where $h_i$ are the eigenvalues of $H$ (Cressie, 1993). The bounds for $\eta$ must be chosen to
ensure that the product in expression (4.9) is positive. This is ensured if $\eta$ satisfies:

$$h_1^{-1} < \eta < h_n^{-1}, \quad (4.10)$$

where $h_1$ and $h_n$ are the minimum and maximum eigenvalues, respectively (Cressie, 1993).
When evaluating the dependence parameter bounds for neighborhood radii from 750 m to
2200 m the bounds for $\eta$ differ slightly across different neighborhood sizes, however using an
approximate rule of $|\eta| < 2$ satisfies all scenarios and gives a general guideline for choosing
dependence parameter values for the simulation of data sets.
4.6.2 Simulation Study: Estimated Dependence Parameter Sensitivity

Given the range of neighborhood sizes that we wish to consider and the constraints on values of \( \eta \), we choose three neighborhood radii: \( \delta = 900, 1300, \) and \( 2000 \) m to represent a “small”, “medium”, and “large” neighborhood, respectively. We expect the dependence parameter to be positive for a unidirectional MRF in this particular application. Therefore, we choose three dependence parameter values: \( \eta = 1, 1.5, \) and \( 1.9 \) to represent cases of “weak”, “moderate”, and “strong” spatial dependence. Combining each neighborhood size and dependence parameter value leads to nine combinations from which to simulate data. For each simulated dataset, we evaluate our ability to identify the “true” neighborhood radius used to simulate the data. For each of the nine combinations a total of 250 datasets was simulated, and for each dataset we estimate the unidirectional MRF model for neighborhood sizes ranging from \( 750 \) m to \( 2200 \) m by increments of \( 50 \) m, leading to a total of \( 30 \) values of \( \delta \) for which we estimate the model for each simulated data set.

As a first step, we evaluate the stability of dependence parameter estimates in relation to neighborhood radius choice. Figure 4.3 shows the parameter estimates related to one simulated data set, across the range of neighborhood sizes, for each of the nine combinations of \( \eta \) and \( \delta \). The true values of \( \delta \) and \( \eta \) from which the data were simulated are given above each plot. Additionally, the value of \( \eta \) used to simulate the data set is represented by a horizontal line in each plot, and the value of \( \eta \) estimated at the true value of \( \delta \) is plotted in red. Figure 4.3 shows that the estimated values of \( \eta \) are close to the values used for simulation when the dependence parameter is estimated using the correct radius, across all the combinations of \( \delta \) and \( \eta \) used to simulate data. Additionally, the estimates of \( \eta \) are fairly robust to the radius used for estimation. However, choosing a neighborhood radius that is much smaller than the truth often leads to underestimation of the dependence parameter. Figure 4.3 gives results for only one set of simulated data, however similar results were seen across data simulations.
Figure 4.3  Estimated $\eta$ values across neighborhood radii for nine combinations of $\eta$ and $\delta$. The horizontal line represents the true $\eta$ used to generate simulated data, and the red point represents the estimated value of $\eta$ at the true value of $\delta$ used to generate simulated data.
4.6.3 Simulation Study: Neighborhood Selection Criterion

Simulation shows that estimates of $\eta$ are somewhat stable to the choice of neighborhood radius, unless a radius that is much too small is chosen. However, in practice, the “true” value of the dependence parameter is never known. Thus, evaluating dependence parameter values estimated over a range of radius choices will not aid in choosing a neighborhood radius for the model, and another criterion is necessary for choosing a neighborhood size.

For a fixed neighborhood size, estimated parameter values are the set of values which minimize the negative log-pseudolikelihood function given in expression (4.7). We propose using the negative log-pseudolikelihood value as a criterion for selecting the neighborhood radius, $\delta$. We first evaluate the adequacy of this criterion through a simulation study. The study is conducted in a similar manner to the study in section 4.6.2. We consider a total of nine combinations of $\delta$ and $\eta$ and simulate 250 datasets for each of these combinations. Then, for each combination and each dataset, model parameters are estimated for a set of radius values and the corresponding estimated negative log-pseudolikelihood values are recorded. The smallest radius value used in estimation is 750 m, and we estimate parameters for every radius in increments of 50 m up to the maximum radius of 2200 m. Figure 4.4 shows the estimated negative log-pseudolikelihood values for one simulated data set per combination of $\delta$ and $\eta$; the estimated negative log-pseudolikelihood that corresponds to the “true” radius value from which data was simulated is displayed as a red point. In some cases, the value corresponding to estimation with the “true” value of $\delta$ gives the minimum negative log-pseudolikelihood value across all radii for which we produce estimates. Although the true value of $\delta$ does not always produce the minimum criterion value, Figure 4.4 shows that the true radius produces one of the smallest estimated negative log-pseudolikelihood values. These results were typical across all simulated datasets. Figure 4.5 shows the average estimated negative log-pseudolikelihood value for the 250 simulations for each radius within each combination of $\eta$ and $\delta$ used for simulation; again, the values estimated at the true neighborhood radius are shown with red points. These results show that on average the radius used to generate the datasets produce the smallest neg-
ative log-pseudolikelihood values for the moderate and strong dependence parameter values. In the case of the weak dependence parameter value, the estimated negative log-pseudolikelihood value for the true radius choice is very near the smallest of these values across all radius choices.

### 4.6.4 Radius Selection Based on Observed Data

Simulation studies showed that the estimated negative log-pseudolikelihood is indicative of an appropriate neighborhood radius choice. Therefore, we proceed by evaluating this criterion on observed data to select an appropriate neighborhood size. As described in Section 4.6, we consider neighborhood radii ranging from 750 to 2200 m. We then divide the range of possible radius values uniformly to obtain six possible radii. The number of possible radius values is limited to ensure that differences in estimated negative log-pseudolikelihood are largely due to neighborhood size rather than variability in simulations, estimation, etc.

We estimate the parameters of the unidirectional Gaussian MRF model, at each of the six proposed radii for all 175 hours of data. Additionally, the estimated negative log-pseudolikelihood value is recorded for each dataset and radius value. Figure 4.6 displays these values for two observed hours. Figure 4.6 (a) and (b) shows the minimum values occurring at 1,910 and 1,040 m respectively.

While each dataset may yield a different radius that produces the minimum estimated negative log-pseudolikelihood value, we wish to understand if any of the radii consistently produce the smallest or near the smallest value of interest. Figure 4.7 shows the average estimated negative log-pseudolikelihood value across all datasets for the six proposed radius values. We see that the smallest values occur at radii of 1,330 and 1,620 m.

Although, Figure 4.7 gives us some indications as to which neighborhood radii tend to produce the lowest values, it does not account for differences in pseudo-likelihood score values from cases to case. Comparisons within observed cases can be considerably affected by outliers,
Figure 4.4 Estimated negative log-pseudolikelihood values across neighborhood radii for nine combinations of $\eta$ and $\delta$. Red points represent the estimated value at the true value of $\delta$ used to generate simulated data.
Figure 4.5 Average estimated negative log-psedolikelihood values across 250 simulated datasets and neighborhood radii for nine combinations of $\eta$ and $\delta$. Red points represent the estimated value at the true value of $\delta$ used to generate simulated data.
Figure 4.6  Estimated negative log-pseudolikelihood values for two observed hours, across six neighborhood radii.

Figure 4.7  Average estimated negative log-pseudolikelihood values across datasets for six neighborhood radii.
Figure 4.8 Difference in estimated negative log-pseudolikelihood value and minimum value, across observed cases, for six neighborhood radius values.

if any occur. In order to evaluate each radius’ value within a case, we look at the difference between the estimated value at each radius and the minimum estimated value for a particular case. Figure 4.8 displays boxplots of this difference for each radius value:

$$F_i(\hat{\eta}) - \min_i F_i(\hat{\eta}),$$

(4.11)

where $F(\hat{\eta})$ is defined in expression (4.7) and $i = 1, \ldots, 6$ denotes the proposed radius, across all observed datasets of interest. The smallest values occur for radius values between 1,330 and 1,910 m. Based on this criterion, the best radius to use is between 1,330 and 1,910 m. Based on these results, we will consider two unidirectional neighborhood structures with $\delta$ set at 1,330 and 1,910 m. Figure 4.9 (a) and (b) shows these neighborhoods for turbine 50 for $\delta = 1,330$ and 1,910 m, respectively. The solid circle shows the neighborhood boundary as defined by the choice of $\delta$. Turbine 50 is indicated by a red square, neighbors are represented by blue triangles, and non-neighbors are represented by black dots.
Figure 4.9 Illustrations of unidirectional MRF neighborhoods for turbine 50.

4.7 Directional Gaussian MRF Model

4.7.1 Accounting for Wind Direction

Many researchers have postulated and confirmed that wind direction and its interaction with topography and turbine layout can play a large role in the wind speed observed and wind power generated at an individual turbine. An increased understanding of the effect of wind direction (and other factors) on wind power production can aid in wind farm planning and forecasting corrections. The European Union (EU) has funded a number of studies which investigate wake effects and turbulence profiles which cite wind direction as one of the important factors that play a role in power production (Schreck et al., 2008). Additionally, the Danish have developed a statistical wind power forecasting system which uses wind direction as one of the inputs (Cutler et al., 2007). Pinson et al. (2009) use a speed-to-power model in which both wind speed and direction are model inputs. In a review of progress made in the way of wind power forecasting, Costa et al. (2008) also list wind direction as one of the factors that affect wind power output.

Figure 4.10 gives several examples of the interpolated wind speed map for a variety of wind
directions. The wind direction at the meteorological tower is displayed in the top left-hand corner of each plot. These plots suggest a possible wind direction effect on the spatial distribution of the wind speeds. Given previous research and preliminary investigation, we hypothesize that the direction of wind may have an effect on the wind speed process at the wind farm of interest. The direction that each wind turbine is facing, which can change to maximize the wind speed hitting the turbine blades, is recorded and acts as a proxy for the observed wind direction. Since turbulence and wake effects can affect the direction of an individual turbine, we may consider taking the circular average of all turbines’ observed directions. However, to avoid dealing with variability and uncertainty in wind direction at the turbine-level, we use the wind direction at the meteorological tower near the farm as the observed wind direction.

4.7.2 Model Specification

Incorporating direction as a covariate in a model is not straightforward, because it is a circular variable. We propose a neighborhood structure which divides the neighborhood into two directional neighborhoods: one neighborhood in the direction of the wind with a tolerance region, and the other in the orthogonal direction with a tolerance region.

Let $x_{s_i}$ and $x_{s_j}$ denote the Easting and Northing coordinates of location $s_i$. Further, denote the distance between two locations, $s_i$ and $s_j$, in the Easting and Northing directions as $d_x(s_i, s_j)$ and $d_y(s_i, s_j)$ respectively, where:

$$\delta_x(s_i, s_j) = x(s_j) - x(s_i),$$
$$\delta_y(s_i, s_j) = y(s_j) - y(s_i).$$ (4.12)

The angle of the vector between $s_i$ and $s_j$ and can be calculated as:
Figure 4.10  Average hourly wind speed values (m/s) for the wind farm for several different wind directions. Wind direction is given by the arrow at the bottom left of each plot.
\[ \theta_{i,j} = \begin{cases} 
\tan^{-1}\left( \frac{\delta_y(s_i, s_j)}{\delta_x(s_i, s_j)} \right) & \text{for } d_x(s_i, s_j) > 0, d_y(s_i, s_j) > 0 \\
\tan^{-1}\left( \frac{\delta_y(s_i, s_j)}{\delta_x(s_i, s_j)} \right) + \pi & \text{for } d_x(s_i, s_j) < 0 \\
\tan^{-1}\left( \frac{\delta_y(s_i, s_j)}{\delta_x(s_i, s_j)} \right) + 2\pi & \text{for } d_x(s_i, s_j) > 0, d_y(s_i, s_j) > 0 
\end{cases} \]  

(4.13)

as given by Mardia and Jupp (2000). Additionally, it would be rare for any turbine to be at an angle from another turbine that is exactly equal to any one value. Therefore, to be a neighbor of another turbine in the direction of the wind we define a tolerance region. If a location is within the overall distance, \( \delta \), and within the angular tolerance region of the turbine in question, then it is a neighbor of the turbine in the direction of the wind. Let \( \theta^* \) denote the observed wind direction and \( \gamma \) denote a one-sided angular tolerance. First, we define the angular tolerance region of a location, \( s_i \) as:

\[
\theta_\gamma = \{ \theta_{i,j} : (\theta^* - \gamma \leq \theta_{i,j} \leq \theta^* + \gamma) \cup ([\theta^* + \pi] - \gamma \leq \theta_{i,j} \leq [\theta^* + \pi] + \gamma) \}.
\]  

(4.14)

We then define two neighborhoods for location \( s_i \) more formally as:

\[
N_1(s_i) = \{ s_j : i \neq j; d_{i,j} \leq \delta; \theta_{i,j} \in \theta_\gamma \},
\]

\[
N_2(s_i) = \{ s_j : i \neq j; d_{i,j} \leq \delta; \theta_{i,j} \notin \theta_\gamma \},
\]  

(4.15)

where \( d_{i,j} \) is the Euclidean distance between locations \( s_i \) and \( s_j \). The directional Gaussian MRF model is given by:

\[
f(y(s_i) \mid \{ y(s_j) : j \neq i \}) = \frac{1}{\sqrt{2\pi\tau^2}} \exp \left[ -\frac{1}{2\tau^2} \left\{ y(s_i) - \mu (\{ y(s_j) : j \neq i \}) \right\}^2 \right].
\]  

(4.16)

The mean function is given as:

\[
\mu (\{ y(s_j) : j \neq i \}) = \alpha + \sum_{j=1}^{n} c_{1i,j} (y(s_j) - \alpha) + \sum_{j=1}^{n} c_{2i,j} (y(s_j) - \alpha),
\]  

(4.17)

where
where \( n_{1i}, n_{1j} \) and \( n_{2i}, n_{2j} \) are the number of neighbors for locations \( s_i \) and \( s_j \) in neighborhoods 1 and 2, respectively. A tolerance angle of \( \gamma = 30 \) degrees will be used in implementing and estimating the above model. Section 4.10 further discusses the steps taken for choosing an appropriate value of \( \gamma \) for the model. Additionally, a radius of 1,330 m is used for the directional model, and Figure 4.11 shows the neighborhoods for turbine 50. Turbine 50 is indicated with a red square, and the observed wind direction is represented by the solid line. Locations in neighborhood 1 are represented by green triangles, locations in neighborhood 2 are represented by blue asterisks, and non-neighbors are represented by black dots. The boundary generated by the choice of radius is represented by the large, black circle in the plot.

4.8 Hybrid Gaussian MRF Model

In Section 4.6.4, we narrowed the range of plausible radius values to either 1,330 or 1,910 m for the unidirectional Gaussian MRF, and we consider the directional Gaussian MRF model at a radius of 1,330 m. As preliminary analysis suggests, it seems reasonable that the turbine layout and wind may have an effect on nearby turbines. We hypothesize that one possible scenario for turbine interaction would be that turbines at slightly farther distances may have information relevant to the turbine of interest but do not interact with wind direction due to the larger lag between turbines. To evaluate this hypothesis we propose a directional model with a small radius and a unidirectional neighborhood as an outer ring surrounding the direc-
More formally, the model can be written as follows:

\[
f (y(s_i) \mid \{y(s_j) : j \neq i\}) = \frac{1}{\sqrt{2\pi\tau^2}} \exp \left[ -\frac{1}{2\tau^2} \left\{ y(s_i) - \mu (\{y(s_j) : j \neq i\}) \right\}^2 \right],
\]

(4.20)

where the mean function is given as:

\[
\mu (\{y(s_j) : j \neq i\}) = \alpha + \sum_{j=1}^{n} c_{1i,j} (y(s_j) - \alpha) + \sum_{j=1}^{n} c_{2i,j} (y(s_j) - \alpha) + \sum_{j=1}^{n} c_{3i,j} (y(s_j) - \alpha).
\]

(4.21)

The dependence parameters are:

\[
c_{1i,j} = \begin{cases} 
\eta_1/(n_{1i} + n_{1j}) & \text{for } s_j \in N_1(s_i) \\
0 & \text{for } s_j \notin N_1(s_i) \\
0 & \text{for } i = j
\end{cases}
\]

(4.22)
where $n_{1i}, n_{1j}, n_{2i}, n_{2j},$ and $n_{3i}, n_{3j}$ are the number of neighbors for locations $s_i$ and $s_j$ in neighborhoods 1, 2, and 3, respectively. Finally, we specify two radius values, $\delta_1$ and $\delta_2$, such that $\delta_1 < \delta_2$, and neighborhoods for location $s_i$ are defined as follows:

$$
N_1(s_i) = \{s_j : i \neq j; d_{i,j} \leq \delta_1; \theta_{i,j} \in \theta_\gamma \},
$$
$$
N_2(s_i) = \{s_j : i \neq j; d_{i,j} \leq \delta_2; \theta_{i,j} \in \theta_\gamma \},
$$
$$
N_3(s_i) = \{s_j : \delta_1 < d_{i,j} \leq \delta_2 \},
$$

where $d_{i,j}$ is the Euclidean distance between locations $s_i$ and $s_j$. We implement this model with $\delta_1 = 1,330$ m, $\delta_2 = 1,910$ m, and $\gamma = 30$ degrees. Figure 4.12 shows the neighborhood for turbine 50. Turbine 50 is indicated with a red square and the observed wind direction is represented by the solid line. Locations in neighborhood 1 are represented by green triangles, locations in neighborhood 2 are represented by blue asterisks, locations in neighborhood 3 are represented by purple diamonds, and non-neighbors are represented by black dots.

### 4.9 Simulation Based Model Comparisons

With possible neighborhood structures specified, we wish to compare the performance of these models in characterizing the spatial process of wind speed at the wind farm. As a first step, we fit all four models to the 175 hourly datasets as described in Section 4.2. For brevity
Figure 4.12 Illustration of hybrid neighborhood for turbine 50.

<table>
<thead>
<tr>
<th>Structure</th>
<th>$\delta_1$</th>
<th>$\delta_2$</th>
<th>$\gamma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>unidirectional</td>
<td>1,330 m</td>
<td>—</td>
</tr>
<tr>
<td>Model 2</td>
<td>unidirectional</td>
<td>1,910 m</td>
<td>—</td>
</tr>
<tr>
<td>Model 3</td>
<td>directional</td>
<td>1,330 m</td>
<td>—</td>
</tr>
<tr>
<td>Model 4</td>
<td>hybrid</td>
<td>1,330 m</td>
<td>1,910 m</td>
</tr>
</tbody>
</table>

Table 4.1 Summary of model characteristics.

(especially in figures), we will refer to the univariate models with the small and large radii as Model 1 and Model 2, respectively, the directional model as Model 3, and the hybrid model as Model 4. Table 4.1 summarizes the main attributes of each model.

The pseudolikelihood dependence parameter estimates fall slightly outside of the parameter space in 10 of the 175 cases chosen. These estimates can occur as maximization based pseudolikelihood is inexact and can be subject to numerical instability (Cressie, 1993). We will not use these ten cases for comparing models, as parameter estimates outside of the parameter space lead to an invalid MRF model and unrealistic simulated data sets. When evaluating a model’s performance, we consider how well the model captures both the large- and small-scale structure of the observed data.
4.9.1 Marginal Means

One statistic that characterizes the large-scale structure of the wind process is the marginal mean wind speed. We examine each model’s ability to characterize the overall mean wind speed in the presence of small-scale variation. The following procedure is employed for each model. For each set of hourly data, the model’s estimated parameters are used to simulate 500 data sets from the model of interest. A burn-in period of 1,000 iterations is used, and datasets are separated by 20 iterations. Then, for each simulated data set the mean wind speed across turbines is calculated. The means of simulated datasets can then be compared to observed marginal means for the corresponding observed dataset.

Figure 4.13 gives results for two hourly datasets. A boxplot constructed from the 500 simulated marginal means is displayed for each model and the observed marginal mean is represented by a red horizontal line. Date and time information of the observed dataset are given below each plot. In some cases, all four models do a good job in capturing the overall mean structure; Figure 4.13(a) shows one example of such an hour. There are also a few cases when all four models don’t capture the large scale mean structure as well as in other cases. Figure 4.13(b) shows an example of when all four models miss the mark a bit; it should be noted that these cases occurred very few times in the 165 hourly cases examined.

More generally, Model 2 does consistently poorer than other models in representing the large-scale mean structure of the wind process. Figure 4.14 shows two examples where Model 2 does not capture the marginal mean structure well. Again, box plots represent the 500 simulated marginal means for each model and the mean observed wind speed across all turbines is represented by a red horizontal line.

In order to compare model performance across all datasets, we compute the absolute difference between the marginal mean of each simulated dataset and the marginal mean of the corresponding observed dataset for each model. More explicitly we compute:
Figure 4.13 Marginal means of 500 simulated datasets for each model, for two example cases.

Figure 4.14 Marginal means of 500 simulated datasets for each model, for two example cases where Model 2 performs poorly.
Table 4.2 Summary statistics of mean absolute error (MAE) values for simulated marginal means across 165 cases.

<table>
<thead>
<tr>
<th>Model</th>
<th>Min.</th>
<th>Q1</th>
<th>Median</th>
<th>Mean</th>
<th>Q3</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>0.018</td>
<td>0.059</td>
<td>0.094</td>
<td>0.160</td>
<td>0.170</td>
<td>1.020</td>
</tr>
<tr>
<td>Model 2</td>
<td>0.015</td>
<td>0.075</td>
<td>0.155</td>
<td>0.438</td>
<td>0.634</td>
<td>5.228</td>
</tr>
<tr>
<td>Model 3</td>
<td>0.018</td>
<td>0.061</td>
<td>0.095</td>
<td>0.158</td>
<td>0.170</td>
<td>1.807</td>
</tr>
<tr>
<td>Model 4</td>
<td>0.015</td>
<td>0.069</td>
<td>0.120</td>
<td>0.276</td>
<td>0.399</td>
<td>1.626</td>
</tr>
</tbody>
</table>

Table 4.2 gives the summary statistics for the mean absolute error of the marginal means, as defined in expression (4.26), across all 165 hourly cases for each model. All four models are able to capture large-scale mean structure fairly well for a majority of cases. However, there are some cases where Model 2 does considerably worse than the other three models. In particular, the worst performance by Model 2 yields a MAE value more than two times as large as the worst case MAE for any other model.

\[
MAE(\bar{y})_{m,p} = \frac{1}{Q} \sum_{q=1}^{Q} |\bar{y}_{m,p,q} - \bar{y}_p|, \tag{4.26}
\]

where \(\bar{y}_{m,p,q}\) represents the marginal mean of simulation \(q\) for hourly case \(p\) and model \(m\). Additionally, \(\bar{y}_p\) denotes the observed marginal mean for hourly case \(p\); here, \(m = 1, 2, 3,\) or \(4\), \(p = 1, \ldots, 165,\) and \(q = 1, \ldots, 500\). Table 4.2 gives the summary statistics for the mean absolute error of the marginal means, as defined in expression (4.26), across all 165 hourly cases for each model. All four models are able to capture large-scale mean structure fairly well for a majority of cases. However, there are some cases where Model 2 does considerably worse than the other three models. In particular, the worst performance by Model 2 yields a MAE value more than two times as large as the worst case MAE for any other model.

**4.9.2 Marginal Variances**

In addition to being concerned with the mean wind speed at the large-scale level, we are also interested in a model’s ability to capture the overall variability observed in hourly cases. In a manner similar to Section 4.9.1, we use simulations generated from estimated parameters to assess this characteristic. For each set of hourly data, the model’s estimated parameters are used to simulate 500 data sets from the model of interest. A burn-in period of 1,000 iterations is used, and datasets are separated by 20 iterations. Then, for each simulated data set the variance of wind speeds across turbines is calculated. We then compare these quantities to the wind speed variance of all turbine speeds of the observed dataset of concern. This procedure is done for each model. In general, all four models perform fairly well when looking at marginal variances. Figure 4.15 gives an example of one case where all four models do well in capturing
In order to summarize model performance and compare across cases, we look at the MAE for the sample variance values. More specifically, we define:

\[
MAE(s^2)_{m,p} = \frac{1}{Q} \sum_{q=1}^{Q} \left| s^2_{m,p,q} - s^2_p \right|,
\]

where \( s^2_{m,p,q} \) represents the marginal variance of simulation \( q \) for hourly case \( p \) and model \( m \). Also, \( s^2_p \) denotes the observed marginal variance for hourly case \( p \), where, \( m = 1, 2, 3, \) or \( 4, p = 1, \ldots, 165 \) and \( q = 1, \ldots, 500 \) for each model. Table 4.3 gives the summary statistics for the MAE of the marginal variances, as defined in expression (4.27), across all 165 hourly cases for each model. In a majority of cases, all four models are able to large-scale mean structure fairly well. However, in some cases Model 2 produces simulated datasets in which the marginal variance is inflated.
Table 4.3 Summary statistics of mean absolute error (MAE) values for simulated marginal variances across 165 cases.

<table>
<thead>
<tr>
<th></th>
<th>Min.</th>
<th>Q1</th>
<th>Median</th>
<th>Mean</th>
<th>Q3</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>0.009</td>
<td>0.033</td>
<td>0.055</td>
<td>0.067</td>
<td>0.089</td>
<td>0.276</td>
</tr>
<tr>
<td>Model 2</td>
<td>0.009</td>
<td>0.037</td>
<td>0.061</td>
<td>0.088</td>
<td>0.096</td>
<td>2.010</td>
</tr>
<tr>
<td>Model 3</td>
<td>0.009</td>
<td>0.033</td>
<td>0.054</td>
<td>0.071</td>
<td>0.085</td>
<td>0.954</td>
</tr>
<tr>
<td>Model 4</td>
<td>0.009</td>
<td>0.038</td>
<td>0.061</td>
<td>0.078</td>
<td>0.096</td>
<td>0.632</td>
</tr>
</tbody>
</table>

Figure 4.16 Marginal variances and means of 500 simulated datasets for each model corresponding to October 18, 2009, 1700 (LST).

Figure 4.16 shows the case for which Model 2 has its maximum MAE value, of the 165 cases, for marginal variances. Figure 4.16 (a) shows box plots of the simulated marginal variances for the four models, where the observed marginal variance is represented by the red horizontal line. This plot shows the very large disparity between Model 2’s simulated marginal variances and the observed marginal variance as well as the simulated variances for the other three models. It’s natural to wonder if something went wrong with Model 2 in general for this case. Figure 4.16 (b) shows the simulated marginal means for the same case which shows that Model 2 has done a fairly good job of capturing the observed marginal mean even though the simulated marginal variances are too large.
Min. Q1 Median Mean Q3 Max.
---
Model 1 0.251 0.529 0.754 0.894 1.080 3.261
Model 2 0.260 0.660 1.047 1.256 1.519 4.840
Model 3 0.301 0.604 0.859 0.966 1.146 3.498
Model 4 0.267 0.644 0.985 1.125 1.422 4.302

Table 4.4 Summary statistics for conditional expectation MAE values, computed over 165 cases, for all models.

### 4.9.3 Conditional Expectations

It is necessary for a good model to capture the large-scale behavior of wind speed at the farm-level. However, the efficacy of a model in representing the small-scale variation of wind speeds is of particular interest, as this is the aspect of a MRF model that may give insight into the wind speed process and interactions between turbines over various weather conditions. The conditional expectations of wind speed at locations is one method to look at the small-scale variation of the model. Thus, if a model is adequately representing the small-scale variation of the wind speed process, the conditional expectations of the model will be close to the observed wind speeds of the corresponding locations.

For a Gaussian MRF model, the conditional expectation of location $s_i$ is $\mu \left( \{ y(s_j) : j \neq i \} \right)$. The conditional expectations for location $s_i$ are given by expressions (4.2), (4.17), and (4.21) for Models 1 and 2, Model 3, and Model 4, respectively. For each dataset, we compute the conditional expectations of all locations and compute the MAE of these conditional expectations by computing the average absolute difference between the expectation and the observed value at the corresponding location. Figure 4.17 shows the results of these computations for 165 cases for each model. The MAE values tend to be slightly higher for Models 2 and 4. Table 4.4 gives the summary statistics for each model’s MAE values more explicitly. Models 1 and 3 perform similarly, while Models 2 and 4 do not perform as well, although differences do not become apparent until looking at MAE values for the worst cases of each model. Based on a measure of central tendency such as the mean or median, all of the models perform in well.
Figure 4.17  Boxplots of MAE values of conditional expectations compared to observed data values over 165 cases.

4.9.4 Generalized Spatial Residuals

Some traditional models, such as linear regression, which rely on the assumption of independence allow for the use of residuals to assess the adequacy of the model. However, these methods involving residuals cannot be directly applied to conditionally specified models, as the independence assumption is not valid. An alternative method to assess the goodness of fit (GOF) of a conditionally specified model was developed by Kaiser et al. (2012). The authors demonstrate that sets of locations can be selected such that generalized spatial residuals calculated from these sets are independent. Thus, this allows for some of the traditional, formal tests for GOF to be used. We implement the methods proposed by Kaiser et al. (2012) to assess the GOF of our four proposed models.

To begin, Kaiser et al. (2012) defines a conclique as a “set of locations such that no location in the set is a neighbor of any other location in the set.” Additionally, a collection of concliques is a minimal conclique cover “if it contains the smallest number of concliques needed to partition the set of all locations.” (Kaiser et al., 2012). The specification of concliques and minimal conclique covers is fairly straightforward for a regular lattice spatial domain with an \( n \)-nearest neighborhood (i.e. \( n = 4, 8 \), etc.) structure. For our application, we need only to consider the
unidirectional model for $\delta = 1,330$ m and $\delta = 1,910$ m, as these sets will also be concliques for Models 3 and 4 respectively. Finding a minimal clique cover for an irregular spatial domain such as ours is a complex problem. Therefore, we develop an algorithm to find concliques with as many locations as possible but do not attempt to find a minimal clique cover for this application. The algorithm used to identify concliques is as follows.

Consider a unidirectional model with a fixed radius value. For any location $s_i$ we take the following steps:

1. Set your current location to $s_i$ and add this location to the clique set, denoted as $C_{s_i}$.

2. Define a set of candidate locations as $A_{s_i} = \{s_j : j = 1, \ldots, n; j \neq i\}$.

3. Exclude any locations that are neighbors of $s_i$ from the candidate set. More specifically:

$$A_{s_i} = A_{s_i} \cap \{s_j : s_j \in N(s_i)\},$$

where $N(s_i)$ is the neighborhood of $s_i$.

4. Compute the Euclidean distance between $s_i$ and each member of $A_{s_i}$.

5. Select the location, $s_j^*$, corresponding to the smallest distance, and add this location to $C_{s_i}$ and remove the location from $A_{s_i}$.

6. Repeat Steps 3 through 5 for $s_j^*$ until $A_{s_i}$ is an empty set.

The above steps are performed for all $s_i$ in the spatial domain as the starting location for the algorithm, leading to a set of $n$ possible concliques, where $n$ is the number of turbines. Given our method for finding concliques, we are able to identify two nonoverlapping concliques with the maximum number of members across all possible concliques. The sets of concliques are unique to the neighborhood radius choice. The structure of Model 1 gives two concliques, each with 31 locations, while the structure of Model 2 gives two concliques, each comprised of
For a conditionally specified Gaussian model, the generalized spatial residuals, denoted as $U(s)$, can be obtained by performing a probability integral transform (PIT) on a set of locations $s = \{s_i : i = 1, \ldots, n\}$. More explicitly:

$$U(s) = F(y(s) \mid \{y(t) : t \in N(s)\}), \quad (4.28)$$

where $F(\cdot \mid \cdot)$ is a Normal cumulative distribution function (cdf). Kaiser et al. (2012) showed that if the spatial process $Y(s)$ comes from the conditionally specified model of interest, then the generalized spatial residuals $\{U(s) : s \in C_j\}$, where $\{C_j ; j = 1, \ldots, q\}$ are the collection of concliques, are iid Uniform(0,1).

The empirical distribution of the generalized spatial residual, for clique $j$, is defined as:

$$G_j(u) = \frac{1}{|C_j|} \sum_{s \in C_j} I(U(s) \leq u). \quad (4.29)$$

Further, we can test the null hypothesis that our data represent a partial sample from a particular process model with unknown parameters. We compute the estimated generalized residuals $\hat{U}(s)$ as in expression (4.28) using the Normal cdf with estimated parameter values being used. Assuming that we have reasonable parameter estimates and a smooth cdf function, then $\hat{U}(s)$ will be approximately Uniform(0,1) if the process model is reasonably specified. Finally, Kaiser et al. (2012) proposed computing four test statistics, each combining information from all concliques to assess the goodness of fit for the proposed model. More specifically, the four test statistics are given as:
\[ T_1 = \max_{j=1,\ldots,q} \sup_{u \in [0,1]} | G_j(u) - u | \]
\[ T_2 = \left( \frac{1}{q} \sum_{j=1}^{q} \left( \sup_{u \in [0,1]} | G_j(u) - u | \right) \right) \]
\[ T_3 = \max_{j=1,\ldots,q} \left( \int_0^1 | G_j(u) - u |^r \, du \right)^{1/r} \]
\[ T_4 = \frac{1}{q} \sum_{j=1}^{q} \left( \int_0^1 | G_j(u) - u |^r \, du \right)^{1/r}, \quad (4.30) \]

where \( r \in [1, \infty) \), and in our application we use two concliques (i.e. \( q = 2 \)). Test statistics \( T_1 \) and \( T_2 \) are based on the Kolmogorov-Smirnov test statistic, and test statistics \( T_3 \) and \( T_4 \) are based on the Cramer von Mises test statistic. The tests based on the Komogorov-Smirnov test statistic are based on the maximum vertical distance between the empirical cdf and the reference Uniform distribution’s cdf. We compute these four test statistics for each proposed model on an observed dataset. In order to compute a p-value for each test statistic, for a given model, we perform a parametric bootstrap as follows:

1. Estimate the set of model parameters, \( \hat{\eta} \) from the observed data.

2. Generate a starting set of observed data \( y^*(s_1), \ldots, y^*(s_n) \) from the conditional model using \( \hat{\chi} \), allowing for a burn-in period.

3. Generate a set of observed data \( y^{(1)}(s_1), \ldots, y^{(1)}(s_n) \) from the conditional model using \( \hat{\chi}^{(1)} \), with a small number of iterations separating the simulated data from the starting data values.

4. Estimate the model parameters, \( \hat{\chi}^{(1)} \) from the simulated data.

5. Compute the generalized spatial residuals based on the pre-specified concliques, simulated data, and estimated model parameters, \( \hat{\eta}^{(1)} \).

6. Compute the test statistics defined in expression \( (4.30) \), \( T_1^{(1)}, T_2^{(1)}, T_3^{(1)}, \) and \( T_4^{(1)} \), based on the generalized spatial residuals from Step 5.
7. Repeat Steps 2 through 6 for \( M \) iterations to obtain \( T_1 = \left\{ T_1^{(1)}, T_1^{(2)}, \ldots, T_1^{(M)} \right\} \), and \( T_2, T_3, \) and \( T_4 \) similarly.

8. Calculate the p-value for each test as \( \frac{1}{M} \sum_{w=1}^{M} \mathbb{1}(T_1^{(w)} \geq T_1^*) \), where \( T_1^* \) is the observed test statistic calculated based on observed data(similarly for test statistics 2, 3, and 4).

This parametric bootstrap is implemented for the 165 datasets and all four models using \( M = 200 \). The four test statistics are various methods for quantifying the discrepancy between the empirical cdf and the Uniform(0,1) cdf that should dictate the behavior of the generalized spatial residuals if the model is adequate. Figure 4.18 shows the empirical cdfs of the four models for two cases. Figure 4.18(a) shows a case when all four models produce similar empirical cdfs, while Figure 4.18(b) displays a case when the empirical cdfs of Models 1 and 3 will lead to smaller observed test statistics than those for Models 2 and 4.

Figure 4.19(a) displays the empirical cdf based on observed data for one hourly case. The corresponding empirical cdf based on one dataset simulated with the estimated model parameters is shown in Figure 4.19. In practice, test statistics are calculated based on empirical cdfs generated based on each dataset (observed and simulated) along with many others from other
Figure 4.19  Empirical cdfs of all models based on observed data and one simulated dataset for May 27, 2009, 0500 (CST).

Figure 4.20 shows two cases where the empirical cdf, based on Model 2, for the observed data is plotted with the empirical cdfs from many simulated datasets. Figure 4.20(a) gives one example where the empirical cdfs based on simulated data produces many cases where the maximum vertical distance from the reference distribution is greater than the maximum distance of the empirical cdf based on observed data. In this instance, assuming the remainder of simulations and the empirical cdfs generated by using the second conclique produced similar results, the p-value for any one of the test statistics would be large. Figure 4.20(b) gives an example of the opposing situation. In this case, the empirical cdf for Model 2 has a larger maximum distance from the reference distribution than nearly all of the bootstrap empirical cdfs. In this case, assuming agreement of all simulations and the second conclique, we would expect to yield a small p-value corresponding to the test statistics of interest and conclude that we have evidence that Model 2 has a lack-of-fit in this instance.

Rather than attempting to assess the goodness-of-fit of the four models visually, we ex-
amine the p-values calculated for each of the test statistics of interest, given in expression (4.30). In the following discussion, we address only test statistic, $T_1$, as other test statistics yield very similar results. Figure 4.21 shows the p-values from, all 165 hourly cases, for Models 1 and 2. A majority of the p-values are much larger than any reasonable level of significance for both models, and neither model produces p-values consistently smaller or larger than the other. On the other hand, Figure 4.22 compares p-values corresponding to Model 3 to p-values from Models 1 and 2. Again, a majority of the p-values for Models 1, 2, and 3 are large, indicating that in most cases, we do not have evidence that any of the models are inadequately representing the data. However, Model 2 does produce noticeably more small p-values than Models 1 and 2, that would result in concluding a lack-of-fit. Additionally, the p-values for Model 3 are consistently larger than the p-values for Models 1 and 2 for a majority of the cases.

Although none of Models 1, 2, or 3 produce p-values of concern in most cases, it does look as though Model 3 may provide an adequate fit for the data more consistently than Models 1 and 2. Thus, we proceed by comparing Model 4 to only Model 3. Figure 4.23 shows the p-values for Models 3 and 4; both models produce larger p-values than the other model in about half of the cases.
Figure 4.21 Model 1 and 2 p-values corresponding to Kolmogorov-Smirnov test, $T_1$, for 165 cases.

(a) Model 3 vs Model 1  
(b) Model 2 vs Model 3

Figure 4.22 Model 1 and 3 and 2 and 3 p-values corresponding to Kolmogorov-Smirnov test statistics, $T_1$, for 165 cases.
cases, suggesting that these models perform similarly in terms of the Kolmogorov-Smirnov test.

Based on the above criteria, Model 2 is inadequate more frequently than the other models for this particular wind speed process. Additionally, Models 1, 3, and 4 all appear to do an adequate job capturing the large-scale structure of the wind speed process. However, goodness-of-fit tests, including results for the three test statistics not explicitly presented, indicate that Models 3 and 4 may capture the small-scale structure of the wind speed process better than Model 1. Therefore, we proceed with the following discussion focusing on Models 3 and 4 only.

### 4.9.5 Neighborhood Structure Choice Related to Weather Conditions

As both Models 3 and 4 appear to capture large- and small-scale structure of the wind speed process comparably well. One might wonder if one model performs better than the other for any particular weather conditions. Figure 4.24 displays the p-values corresponding to the Kolmogorov-Smirnov tests conducted for 165 hourly cases against the mean wind speed across all turbines for each case. A similar plot of the wind speed variance against p-values yields similar results.
At a first look, no patterns are apparent from visually inspecting the plot. However, one interesting feature can be found by dividing the range of observed mean wind speeds into three categories. We refer to cases with mean wind speeds less than or equal to 5.5 m/s to be “slow” wind speed cases, and cases with mean wind speeds greater than 12 m/s to be “fast” wind speed cases. Additionally, the remaining cases with mean wind speeds greater than 5.5 m/s and less than or equal to 12 m/s will be referred to as “moderate” wind speed cases. When hourly cases are divided by this criteria, we can examine the number of cases, for each category, in which a model yields a higher/lower p-value than the other model. Figure 4.5 summarizes the number of cases in which the given model produces a larger p-value than the other model divided by the three wind categories. The percentage of cases is also given in parentheses, conditional on the number of cases in the category of interest. Table 4.5 shows that for “slow” wind speed conditions, Model 3 yield the larger p-value in more than two-thirds of the slow wind speed cases. On the other hand, Model 4 produces the larger p-value in nearly 80% of cases when the overall wind conditions are fast. The two models are split fairly evenly for the moderate wind speed cases.

These patterns match what we might expect when thinking about models from a physical perspective. For stronger winds, we would expect that the extent to which turbines influence the wind speeds of surrounding turbines might extend to a larger distance. Model 4 includes the outer, unidirectional neighborhood that extends the size of the neighborhood from the Model 3 directional model, and this neighborhood more closely matches what we might expect physically. Likewise, we would expect that the dependence between turbines may not extend to larger distances when the overall wind speeds are slow, and Model 3, with the smaller radius, does produce larger p-values than Model 4 in a majority of these cases.
Figure 4.24  Model 3 and 4 p-values, for Kolmogorov-Smirnov test, versus mean wind speed.

Table 4.5  Models 3 and 4 Kolmogorov-Smirnov p-value comparisons by mean wind strength.
4.10 Tolerance Angle Selection

Diagnostics of model efficacy indicate that it is beneficial to allow for a directional neighborhood structure in our application. Directional models require one additional decision: determining a reasonable value for the tolerance angle, $\gamma$, in expressions (4.15) and (4.25). We proceed by examining only Model 3, as the directional neighborhood structure for Model 4 is equivalent, and we assume that tolerance angle choices made for Model 3 will have similar consequences for Model 4.

When the tolerance angle, $\gamma$, is set to 45 degrees, the two directional neighborhoods have equal areas. Since we are most interested in dependence along the direction of the wind, we will not consider tolerance values greater than 45 degrees. Also, preliminary investigation shows that tolerance values less than 20 degrees produce empty neighborhood sets for a large number of turbines. For these reasons, we limit our range of $\gamma$ values of practical interest to $\{\gamma : 20 \leq \gamma \leq 45\}$. Even with a narrowed range of tolerance values, we prefer not to choose $\gamma$ arbitrarily, as the possible effects of such a choice are unknown to us. Therefore, we employ some of the criteria for evaluating models that were used in Section 4.9.

We formally assess directional models for $\gamma = 20, 25, 30, 35, 40,$ and 45 degrees. We examine each model’s ability to characterize the overall wind speed in the presence of small-scale variation by simulating 1,000 datasets with parameters corresponding to each tolerance choice, for 172 datasets. As in Section 4.9.1, the means of simulated datasets can then be compared to observed marginal means for the corresponding observed dataset. Note, computational difficulties with obtaining dependence parameters within the parameter space occur in only three cases, thus, we have a total of 172 hourly cases available. Table 4.6 gives summary statistics for the MAE for simulated marginal means compared to corresponding observed marginal means, for the possible tolerance values. Model 2 captures the marginal mean structure of the wind speed process similarly across the possible tolerance value choices. All tolerance choices result in reasonable marginal mean wind speeds.
Table 4.6 Summary statistics of mean absolute error (MAE) values for simulated marginal means, for varying tolerance possibilities (Model 3), across 172 cases.

<table>
<thead>
<tr>
<th>γ</th>
<th>Min.</th>
<th>Q1</th>
<th>Median</th>
<th>Mean</th>
<th>Q3</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0.004</td>
<td>0.090</td>
<td>0.077</td>
<td>0.140</td>
<td>0.267</td>
<td>1.918</td>
</tr>
<tr>
<td>25</td>
<td>0.002</td>
<td>0.065</td>
<td>0.063</td>
<td>0.109</td>
<td>0.270</td>
<td>1.741</td>
</tr>
<tr>
<td>30</td>
<td>0.018</td>
<td>0.061</td>
<td>0.095</td>
<td>0.158</td>
<td>0.170</td>
<td>1.807</td>
</tr>
<tr>
<td>35</td>
<td>0.003</td>
<td>0.017</td>
<td>0.058</td>
<td>0.233</td>
<td>0.478</td>
<td>2.015</td>
</tr>
<tr>
<td>40</td>
<td>0.002</td>
<td>0.091</td>
<td>0.099</td>
<td>0.186</td>
<td>0.5001</td>
<td>1.925</td>
</tr>
<tr>
<td>45</td>
<td>0.017</td>
<td>0.074</td>
<td>0.057</td>
<td>0.207</td>
<td>0.523</td>
<td>2.181</td>
</tr>
</tbody>
</table>

Table 4.7 Summary statistics for Kolmogorov-Smirnov p-values, for varying tolerance choices (Model 3), across 172 cases.

<table>
<thead>
<tr>
<th>γ</th>
<th>Min.</th>
<th>Q1</th>
<th>Median</th>
<th>Mean</th>
<th>Q3</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0.01</td>
<td>0.275</td>
<td>0.450</td>
<td>0.488</td>
<td>0.656</td>
<td>1.00</td>
</tr>
<tr>
<td>25</td>
<td>0.01</td>
<td>0.325</td>
<td>0.540</td>
<td>0.525</td>
<td>0.740</td>
<td>1.00</td>
</tr>
<tr>
<td>30</td>
<td>0.03</td>
<td>0.321</td>
<td>0.478</td>
<td>0.535</td>
<td>0.780</td>
<td>1.00</td>
</tr>
<tr>
<td>35</td>
<td>0.01</td>
<td>0.300</td>
<td>0.485</td>
<td>0.511</td>
<td>0.711</td>
<td>1.00</td>
</tr>
<tr>
<td>40</td>
<td>0.01</td>
<td>0.280</td>
<td>0.448</td>
<td>0.467</td>
<td>0.650</td>
<td>1.00</td>
</tr>
<tr>
<td>45</td>
<td>0.00</td>
<td>0.2438</td>
<td>0.383</td>
<td>0.410</td>
<td>0.526</td>
<td>0.995</td>
</tr>
</tbody>
</table>

We evaluate each tolerance value’s ability to model the small-scale structure by computing $T_1$ in expression (4.30) for the observed data and comparing these values to bootstrapped test statistics. Table 4.7 gives the summary statistics for the Kolmogorov-Smirnov test statistics for each tolerance value, across the 172 cases. Small p-values indicate a lack-of-fit for the model. For a majority of the cases, the p-values yielded by the test are above any reasonable level of significance that we might set for all tolerance values considered.

More specifically, Table 4.8 gives the number of cases (out of 172) for which a lack-of-fit would be indicated by the p-value, for various levels of significance $\alpha^*$ and tolerance choices. The choices of tolerance at each end of the range we are willing to consider lead to the most cases where we have evidence of a lack-of-fit, and this becomes especially apparent for a significance level of $\alpha^* = 0.10$. Here the number of cases indicating a lack-of-fit by the model is more than double the number of instances for middle values of $\gamma$. Choosing a level of significance of $\alpha^* = 0.2$, tolerance values of 30 and 35 yield considerably fewer cases, than other tolerance
values, where the test indicates that the model is inadequate. All tolerance value choices, along with Model 3, provide an adequate fit for the data in a majority of hourly cases. However, based on these results and the results presented in Table 4.7 a tolerance value of $\gamma = 30$ or 35 degrees will provide an adequate fit, given Model 3, for the largest number of hourly cases. Therefore, we proceed with the tolerance value set to 30 degrees for Models 3 and 4.

### 4.11 Model Estimation on Observed Data

Once all pertinent model specifications have been made, we proceed by estimating parameters for Models 3 and 4 for all hourly cases of observed data, excluding the cases where estimates based on pseudolikelihood are outside of the parameter space. We have 172 and 165 such cases for Model 3 and 4 respectively. In the following sections, we present results for Model 3 only, as results are similar for the directional dependence of Model 4. A short discussion of the estimated dependence parameter for the outer neighborhood of Model 4 can be found at the end of the section.

#### 4.11.1 Parameter Relationships with Observed Average Wind Speed

We begin by examining possible relationships between estimated parameters and the marginal observed wind speed for each case. Figure 4.25 gives plots of the estimated marginal mean parameter $\hat{\alpha}$ and the estimated constant conditional variance $\hat{\tau}^2$ against the observed mean wind speed. In the conditional Gaussian MRF model, the parameter $\alpha$ represents a quantity very close to the marginal mean. However, in this case, $\alpha$ is a weighted average of wind speeds

<table>
<thead>
<tr>
<th>$\gamma$</th>
<th>$\alpha^* = 0.01$</th>
<th>$\alpha^* = 0.05$</th>
<th>$\alpha^* = 0.10$</th>
<th>$\alpha^* = 0.20$</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>2</td>
<td>5</td>
<td>13</td>
<td>25</td>
</tr>
<tr>
<td>25</td>
<td>1</td>
<td>3</td>
<td>9</td>
<td>20</td>
</tr>
<tr>
<td>30</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td>35</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>12</td>
</tr>
<tr>
<td>40</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>25</td>
</tr>
<tr>
<td>45</td>
<td>5</td>
<td>8</td>
<td>14</td>
<td>37</td>
</tr>
</tbody>
</table>

Table 4.8 Number of hourly cases for which there is evidence of a lack-of-fit at varying levels of significance and all considered tolerance values.
across turbines due to the irregularity of the spatial domain and resulting unbalanced neighborhood structure. Since we have adapted our model to account for the unequal number of neighbors (by dividing each dependence parameter by the total number of neighbors for each pair of locations) we expect the estimated $\alpha$ values to be very similar to the observed marginal mean. As expected, Figure 4.25(a) shows that the estimated values of $\alpha$ are closely related to the observed marginal mean wind speed. Additionally, Figure 4.25 displays the estimated conditional variance parameter $\tau^2$ values against the observed marginal mean values. In general, as the mean wind speed increases the conditional variance also tends to increase. This result is not surprising, as from a physical standpoint we might expect wind speeds at the turbine-level to be more variable in higher wind speed conditions.

Figures 4.26(a) and 4.26(b) examine the relationship between the observed marginal mean wind speed and estimated dependence parameter values for $\eta_1$ and $\eta_2$ respectively. For both parameters there is no apparent relationship between their estimated values and the mean wind speed conditions. We might have expected the dependence parameter estimates to be larger on days with stronger wind conditions, however this is not the case.
Additionally, we examine the relationship between the observed marginal wind speed vari-
Figure 4.27 Estimated parameter values for $\alpha$ and $\tau^2$ plotted against observed marginal wind speed variance.

The values of estimated dependence parameters are of particular interest, as they may aid in the understanding of the spatial wind process at the farm-level. We begin by examining the estimates of $\eta_1$ in relation to $\eta_2$ estimates. Figure 4.29 displays the two dependence parameters plotted against one another for all 172 hourly cases. The dependence parameters are not independent of one another, as their sum must be within the parameter space for one dependence...
We further investigate the difference in magnitude of dependence parameter estimates by examining the absolute value of the estimates relative to one another. Table 4.9 looks at the multiplicative disparity in the absolute values of the parameter estimates by tabulating the number of cases for which the magnitude of the first parameter estimate is less than the magnitude of the second, two times less, etc. Additionally, selected summary statistics are given for the absolute value of $\hat{\eta}_1$; it is the case that the estimates are relatively small compared to the estimates of $\eta_2$ on average or for the middle of the estimates. In general, the magnitude of $\hat{\eta}_1$ is small relative to $\hat{\eta}_2$, regardless of the direction (positive/negative) of the dependence.
Figure 4.29  Estimated dependence parameter values for 172 hourly cases.

| $F^*$ | $|\eta_1| < |\eta_2|$ | Number of Cases | $|\hat{\eta}_1|$ |
|-------|-------------------|----------------|----------------|
|       |                   | Number of Cases | Mean | Median | Max. |
| $F = 1$ | 147 (85.5%) | 0.316 | 0.321 | 0.996 |
| $F = 2$ | 115 (66.9%) | 0.224 | 0.252 | 0.673 |
| $F = 3$ | 77 (44.8%) | 0.148 | 0.184 | 0.464 |
| $F = 5$ | 50 (29.1%) | 0.081 | 0.149 | 0.309 |
| $F = 10$ | 22 (12.8%) | 0.015 | 0.021 | 0.178 |

Table 4.9  Comparison of dependence parameter magnitudes for 172 estimated cases.
A large majority of $\eta_1$ estimates are positive, with only 24 of the 172 cases producing negative dependence parameter estimates in the direction of the wind. We examine these cases by plotting the observed wind vectors. Figure 4.30 shows the observed wind vectors for the cases in which neighbors in the direction of the wind are characterized by a negative dependence parameter. The speed of the wind is reflected in the length of the vector. Varying wind speeds occur across the 24 cases, yet most of the speeds tend to be strong. Most noticeably, the wind direction in nearly all of these cases is in the Northwest/Southeast direction. This phenomenon indicates that the wind speed process may change under these directional conditions compared to others. Further investigation shows that for cases when the observed wind direction is between 135 to 180 degrees or the opposite directional window, estimates for $\eta_1$ are almost exclusively negative or positive but relatively small in magnitude. More specifically, the maximum estimate for $\eta_1$ under these conditions is 0.238.

It should be noted that the tendencies found in dependence parameter estimates were similar across results for Model 4. In cases where the wind speed was weak or moderate in strength
the third dependence parameter was typically estimated to be very small, almost near zero in most instances. In general, estimates for \( \eta_3 \) were noticeably different from zero in cases where the wind speed was very strong.

4.11.4 Parameter Estimates for Consecutive Hours

The MRF model presented in this work has been able to adequately capture the structure of the wind speed process for this particular wind farm. One natural extension we might consider for future work would be to add a temporal component to the model. Ideally, we would like to take advantage of any information that is present in past observations. As described in Section 4.2, we have selected two periods of time for which we have consecutive observations. One stretch is comprised of 39 consecutive observations from June 26, 2009, 1100 (LST) to June 28, 2009, 0300 (LST), and the other stretch has 43 observations from August 23, 2009, 1000 (LST) to August 25, 2009, 0500 (LST).

To investigate possible dependences in parameter estimates we examine the sample autocorrelation. Preliminary investigation showed no significant findings for the autocorrelation of the marginal mean \( \alpha \) nor the conditional variance \( \tau^2 \) estimates. We further examine the sample autocorrelation of the dependence parameter estimates. Figure 4.31 displays times series plots of \( \hat{\eta}_1 \) for the two the cases of consecutive periods.

Figure 4.32 shows the sample autocorrelation plots for the differenced parameter estimate values in these two cases. There is some suggestion of dependence between the parameter estimates at a lag of one or two, however longer stretches of consecutive data are needed to further investigate dependencies. The estimates of \( \eta_2 \) yielded similar results to those presented here. All of this suggests, that if we were to build a temporal model, we might consider allowing for an ARMA dependence structure for dependence parameters from hour to hour.
(a) June 2009 Consecutive Period  
(b) August 2009 Consecutive Period

Figure 4.31  Time series plots of \( \hat{\eta}_1 \) for two stretches of consecutive hourly observations.

(a) June 2009 Consecutive Period  
(b) August 2009 Consecutive Period

Figure 4.32  Sample autocorrelation plots for the differenced \( \hat{\eta}_1 \) values for two stretches of consecutive hourly observations.
4.12 Conclusion

We have used a Gaussian MRF to model the wind speed process at the farm-level and have incorporated wind direction into the model provide additional useful information for the layout of this particular wind farm. Several practical issues in terms of statistical application of the model have been addressed. We proposed and demonstrated the validity of using the estimated negative log-pseudolikelihood value as a criterion for selecting the neighborhood radius. Additionally, wind direction was incorporated into the MRF model by defining neighborhood structures based on the observed wind direction, and additional flexibility was introduced by a hybrid model of the unidirectional and directional neighborhood structures. We introduced several criteria for assessing the adequacy of a proposed spatial model. A model’s adequacy in capturing large-scale structure of wind speed was evaluated based on a model’s ability to reflect appropriate marginal means and variances. Typically, the more difficult and important aspect of spatial modeling is describing the small-scale variability of the process of interest. This can be particularly difficult to assess due to the lack of independence in our observations. We used conditional expectations and simulated data as one method. Additionally, we assessed the generalized spatial residuals using methods for multiple concliques developed by Kaiser et al. (2012) and developed a basic algorithm for determining concliques of maximum size on an irregular spatial domain. Finally, we showed that incorporating wind direction into the neighborhood structure is beneficial to the MRF model for this application.

Once issues of model implementation and assessment were addressed, we were able to estimate the model and gain better understanding of some aspects of wind behavior at the farm-level. We determined that a Gaussian MRF was able to adequately capture key characteristics of the wind process. In general, a purely directional neighborhood structure is the preferred model. Most conditional information about the wind speed at a given turbine is found in turbines located in a direction orthogonal (with a certain level of tolerance) to the observed wind direction. Additionally, when wind speed conditions are at the extreme ends of fast and slow winds, the dependence structure changes, and in particular, the dependence range/distance
lengthens for strong winds. We have also seen that negative dependence in the direction of the wind occurs for wind directions in the Northwest/Southeast directions, likely due to the layout of turbines for this particular wind farm. We have seen that as the marginal mean wind speed increases, so do the conditional and marginal variability. However, the strength of dependence does not having any apparent relation to mean wind speeds.

We have made contributions to both the practical application of statistical models to wind speed spatially, and to the understanding of the wind speed process at a farm-level. However, much work remains to be done in these areas. Results regarding the wind speed process in space are most likely not representative of all wind farms given differences in terrain, wind farm geography, etc. Additional investigation for farms and areas of interest are needed. However, many of the statistical application methods laid out here will be helpful for future work. Additionally, several authors have noted differences in wind behavior from season to season (Schreck et al., 2008). Further investigation into seasonal effects and model choice, parameter estimates, etc. should be done. Finally, a natural extension to these methods would be to account for the temporal variability of wind in a model. We have done a bit of preliminary evaluation of time dependencies in model parameters, however, a more comprehensive investigation needs to be conducted before proceeding with a spatio-temporal model.
CHAPTER 5. GENERAL CONCLUSIONS

5.1 General discussion

A better understanding of wind, particularly for wind speed and direction, in terms of forecasts, relationships between wind attributes, and wind behavior over space and time is necessary if wind energy is to become a viable energy option. In Chapter 1, a model for the bias of wind speed (and the bias of wind vector components) was developed through the use of a Bayesian hierarchical model. The model was fit to data at a height relevant to wind turbines and the training and forecasting periods used are relevant to the energy trading industry. In Chapter 2, the hierarchical model developed in Chapter 1 was extended to allow for dependence between the bias of the vector component forecasts. Simulation studies evaluated and characterized properties of the model relevant to implementation on actual data and forecasts, and model comparisons were discussed. Finally, Chapter 3 investigated the behavior of wind spatially while addressing practical strategies for addressing modeling issues. Additionally, features of wind speed and direction in space were characterized in the context of an application to data from a wind farm in central Iowa.

5.2 Recommendation for future research

In this dissertation we focused on forecasting wind speed and direction and spatially modeling observed wind speed at a farm-level. Below we outline some directions we see for future research in this field.

1. The alternative method for combining bias corrected wind speed/vector forecasts through weights based on variability did not produce forecasts differing by any noticeable amount from unweighted combinations. This was largely due to the assumption that the variance
of model biases came from the same underlying distribution across meteorological models. It would be interesting to examine the effect of allowing the underlying distribution to vary from one meteorological model to another and assess whether forecasts can be improved.

2. The models in Chapters 1 and 2 used covariates of wind shear and temperature difference from 120 to 40 m vertical heights. Although models were competitive with existing methodology, the use of other covariates in our models needs to be explored.

3. We have focused on the understanding of wind speed in space. An extension to this work is to allow for temporal dependence in our model and take advantage of any temporal dependence in model parameters or wind characteristics.
APPENDIX A. CONDITIONAL POSTERIOR DERIVATIONS FOR HIERARCHICAL MODEL

Distributions Implied by the Model

1. For \( t = 1, \ldots, T, y_{0,t} \) has density:

\[
f_{0}(y_{0,t} | \mu_t, \lambda_0) = \frac{\lambda_0^{1/2}}{(2\pi)^{1/2}} \exp \left\{ -\frac{\lambda_0}{2} (y_{0,t} - \mu_t)^2 \right\}.
\] (A.1)

2. For \( t = 1, \ldots, T \) and \( j = 1, \ldots, M, Y_{j,t} \) has density:

\[
f_{j,t}(y_{j,t} | \mu_t, b_{j,t}, \lambda_j) = \frac{\lambda_j^{1/2}}{(2\pi)^{1/2}} \exp \left\{ -\frac{\lambda_j}{2} (y_{j,t} - (\mu_t + b_{j,t}))^2 \right\}.
\] (A.2)

3. For \( j = 1, \ldots, M, \lambda_j \) has density:

\[
g_j(\lambda_j | \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} (\lambda_j)^{\alpha-1} \exp \{ -\beta \lambda_j \}.
\] (A.3)

4. For \( t = 1, \ldots, T \) and \( j = 1, \ldots, M, b_{j,t} \) has density:

\[
h_{j,t}(b_{j,t} | x_{1j,t}, x_{2j,t}, \alpha_{0j}, \alpha_{1j}, \alpha_{2j}, \tau) = \frac{\tau^{1/2}}{(2\pi)^{1/2}} \exp \left\{ -\frac{\tau}{2} (b_{j,t} - (\alpha_{0j} + \alpha_{1j} x_{1j,t} + \alpha_{2j} x_{2j,t}))^2 \right\}.
\] (A.4)

5. The joint density of \( Y_{j,1}, \ldots, Y_{j,T} \) for \( j = 1, \ldots, M \) is:

\[
f_j(y_{j,1}, \ldots, y_{j,T} | \mu_1, \ldots, \mu_T, b_{j,1}, \ldots, b_{j,T}, \lambda_j) = \frac{\lambda_j^{T/2}}{(2\pi)^{T/2}} \exp \left\{ -\frac{\lambda_j}{2} \sum_{t=1}^{T} (y_{j,t} - (\mu_t + b_{j,t}))^2 \right\}.
\] (A.5)

6. The joint density of \( Y_{1,t}, \ldots, Y_{M,t} \) for \( t = 1, \ldots, T \) is:

\[
f_t(y_{1,t}, \ldots, y_{M,t} | \mu_1, b_{1,t}, \ldots, b_{M,t}, \lambda_1, \ldots, \lambda_M) = \prod_{j=1}^{M} \frac{\lambda_j^{1/2}}{(2\pi)^{M/2}} \exp \left\{ \sum_{j=1}^{M} -\frac{\lambda_j}{2} (y_{j,t} - (\mu_t + b_{j,t}))^2 \right\}.
\] (A.6)
7. The joint density of $\lambda_1, \ldots, \lambda_M$ is:

$$g(\lambda_1, \ldots, \lambda_M \mid \alpha, \beta) = \frac{\beta^M \alpha}{\Gamma(\alpha)} \left( \prod_{j=1}^M \lambda_j \right)^{\alpha - 1} \exp \left\{ -\beta \sum_{j=1}^M \lambda_j \right\}. \quad (A.7)$$

8. The joint density of $b_{j,1}, \ldots, b_{j,T}$ for $j = 1, \ldots, M$ is:

$$h_j(b_{j,1}, \ldots, b_{j,T} \mid x_{1j,t}, x_{2j,t}, \alpha_{0j}, \alpha_{1j}, \alpha_{2j}, \tau) = \frac{\tau^{T/2}}{(2\pi)^{T/2}} \exp \left\{ -\frac{\tau}{2} \sum_{t=1}^T (b_{j,t} - (\alpha_{0j} + \alpha_{1j} x_{1j,t} + \alpha_{2j} x_{2j,t}))^2 \right\}. \quad (A.8)$$

9. The joint density of the entire set $\{b_{j,t} : t = 1, \ldots, T; j = 1, \ldots, M\}$, assuming the forecast models are independent, are:

$$h(\{b_{j,t} : t = 1, \ldots, T; j = 1, \ldots, M\} \mid x_{1j,t}, x_{2j,t}, \alpha_{0j}, \alpha_{1j}, \alpha_{2j}, \tau) = \frac{\tau^{MT/2}}{(2\pi)^{TM/2}} \exp \left\{ -\frac{\tau}{2} \sum_{t=1}^T \sum_{j=1}^M (b_{j,t} - (\alpha_{0j} + \alpha_{1j} x_{1j,t} + \alpha_{2j} x_{2j,t}))^2 \right\}. \quad (A.9)$$

**Conditional Posterior Distributions**

1. The full conditional posterior of $\tau$ is given as:

$$p(\tau \mid \cdot) \propto \Pi(\tau) h(\{b_{j,t} : t = 1, \ldots, T; j = 1, \ldots, M\} \mid x_{1j,t}, x_{2j,t}, \alpha_{0j}, \alpha_{1j}, \alpha_{2j}) \tau^c \exp \left\{ -\frac{\tau}{2} \sum_{t=1}^T \sum_{j=1}^M (b_{j,t} - (\alpha_{0j} + \alpha_{1j} x_{1j,t} + \alpha_{2j} x_{2j,t}))^2 \right\} \propto Gamma \left( c + MT/2, d + \frac{1}{2} \sum_{t=1}^T \sum_{j=1}^M (b_{j,t} - (\alpha_{0j} + \alpha_{1j} x_{1j,t} + \alpha_{2j} x_{2j,t}))^2 \right). \quad (A.10)$$

2. The full conditional posterior of $\alpha_{0j}$ for $j = 1, \ldots, M$ is given as:

$$p(\alpha_{0j} \mid \cdot) \propto \Pi(\alpha_{0j}) h_j(b_{j,1}, \ldots, b_{j,T} \mid x_{1j,t}, x_{2j,t}, \alpha_{0j}, \alpha_{1j}, \alpha_{2j}, \tau),$$

which can be written as:

$$p(\alpha_{0j} \mid \cdot) \propto \frac{\lambda_{\alpha}^{1/2}}{(2\pi)^{T/2}} \exp \left\{ -\frac{\lambda_{\alpha}}{2} \alpha_{0j}^2 \right\} \frac{\tau^{T/2}}{(2\pi)^{T/2}} \exp \left\{ -\frac{\tau}{2} \sum_{t=1}^T (b_{j,t} - (\alpha_{0j} + \alpha_{1j} x_{1j,t} + \alpha_{2j} x_{2j,t}))^2 \right\} \propto \exp \left\{ -\frac{\lambda_{\alpha}}{2} \alpha_{0j}^2 - \frac{\tau}{2} \sum_{t=1}^T (b_{j,t} - (\alpha_{0j} + \alpha_{1j} x_{1j,t} + \alpha_{2j} x_{2j,t}))^2 \right\}.$$
Note:

\[(b_{j,t} - (\alpha_{0j} + \alpha_{1j}x_{1j,t} + \alpha_{2j}x_{2j,t}))^2\]
\[= b^2_{j,t} - 2b_{j,t}(\alpha_{0j} + \alpha_{1j}x_{1j,t} + \alpha_{2j}x_{2j,t}) + (\alpha_{0j} + \alpha_{1j}x_{1j,t} + \alpha_{2j}x_{2j,t})^2\]
\[= b^2_{j,t} - 2b_{j,t}\alpha_{0j} - 2b_{j,t}\alpha_{1j}x_{1j,t} - 2b_{j,t}\alpha_{2j}x_{2j,t} + \alpha_{0j}^2 + \alpha_{1j}^2x_{1j,t}^2 + \alpha_{2j}^2x_{2j,t}^2 + 2\alpha_{0j}\alpha_{1j}x_{1j,t}\]
\[+ 2\alpha_{0j}\alpha_{2j}x_{2j,t} + 2\alpha_{1j}\alpha_{2j}x_{1j,t}x_{2j,t}\]  \(\text{(A.11)}\)

\[\propto \exp\left\{-\frac{\lambda}{2}\alpha_{0j}^2 - \frac{\tau}{2} \sum_{t=1}^{T} (-2b_{j,t}\alpha_{0j} + \alpha_{0j}^2 + 2\alpha_{0j}\alpha_{1j}x_{1j,t} + 2\alpha_{0j}\alpha_{2j}x_{2j,t})\right\}\]
\[\propto \exp\left\{-\frac{\lambda}{2}\alpha_{0j}^2 + \tau\alpha_{0j} \sum_{t=1}^{T} b_{j,t} - \frac{\tau}{2}T\alpha_{0j}^2 - \tau\alpha_{0j}\alpha_{1j} \sum_{t=1}^{T} x_{1j,t} - \tau\alpha_{0j}\alpha_{2j} \sum_{t=1}^{T} x_{2j,t}\right\}\]

\[\propto \exp\left\{-\frac{\lambda}{2}\alpha_{0j}^2 + \left(\frac{\tau}{\lambda + \tau T} \sum_{t=1}^{T} b_{j,t} - \tau\alpha_{1j} \sum_{t=1}^{T} x_{1j,t} - \tau\alpha_{2j} \sum_{t=1}^{T} x_{2j,t}\right)\alpha_{0j}\right\}\]
\[\propto \exp\left\{-\frac{\lambda}{2}\alpha_{0j}^2 + \left(\frac{\tau}{\lambda + \tau T} \sum_{t=1}^{T} b_{j,t} - \tau\alpha_{1j} \sum_{t=1}^{T} x_{1j,t} - \tau\alpha_{2j} \sum_{t=1}^{T} x_{2j,t}\right)^2\right\}\]

Thus, the conditional posterior of \(\alpha_{0j}\) is proportional to a Normal distribution with mean:

\[\frac{\tau \sum_{t=1}^{T} b_{j,t} - \tau\alpha_{1j} \sum_{t=1}^{T} x_{1j,t} - \tau\alpha_{2j} \sum_{t=1}^{T} x_{2j,t}}{\lambda + \tau T}\]  \(\text{(A.12)}\)

and standard deviation:

\[(\lambda + \tau T)^{-1/2}.\]  \(\text{(A.13)}\)

3. The full conditional posterior of \(\alpha_{1j}\) for \(j = 1, \ldots, M\) is given as:

\[p(\alpha_{1j} \mid \cdot) \propto \Pi(\alpha_{1j})h_j(b_{j,1}, \ldots, b_{j,T} \mid x_{1j,t}, x_{2j,t}, \alpha_{0j}, \alpha_{1j}, \alpha_{2j}, \tau),\]

which can be written as:
4. The full conditional posterior of $\alpha_{1j}$ and standard deviation:

$$p(\alpha_{1j} \mid \cdot) \propto \frac{\lambda_{\alpha}}{2\pi} \exp \left\{ -\frac{\lambda_{\alpha}}{2} \alpha_{1j}^2 \right\} \tau^{T/2} \exp \left\{ -\frac{\tau}{2} \sum_{t=1}^{T} \left( b_{j,t} - (\alpha_{0j} + \alpha_{1j}\alpha_{1j,t} + \alpha_{2j}\alpha_{2j,t}) \right)^2 \right\}$$

$$\propto \exp \left\{ -\frac{\lambda_{\alpha}}{2} \alpha_{1j}^2 - \frac{\tau}{2} \sum_{t=1}^{T} \left( -2b_{j,t}\alpha_{1j} + \alpha_{1j}^2 \alpha_{1j,t} + \alpha_{2j}^2 \alpha_{2j,t} + 2\alpha_{0j}\alpha_{1j}\alpha_{1j,t} + 2\alpha_{1j}\alpha_{2j}\alpha_{2j,t} \right) \right\}$$

$$\propto \exp \left\{ -\frac{\lambda_{\alpha}}{2} \alpha_{1j}^2 - \frac{\tau}{2} \sum_{t=1}^{T} \left( -2b_{j,t}\alpha_{1j} + \alpha_{1j}^2 \alpha_{1j,t} + \alpha_{2j}^2 \alpha_{2j,t} + 2\alpha_{0j}\alpha_{1j}\alpha_{1j,t} + 2\alpha_{1j}\alpha_{2j}\alpha_{2j,t} \right) \right\}$$

$$\propto \exp \left\{ -\frac{\lambda_{\alpha}}{2} \alpha_{1j}^2 + \frac{\tau}{2} \sum_{t=1}^{T} \left( b_{j,t}x_{1j,t} - \alpha_{1j} \sum_{t=1}^{T} x_{1j,t}^2 - \alpha_{2j} \sum_{t=1}^{T} x_{2j,t}^2 \right) \right\}$$

Thus, the conditional posterior of $\alpha_{1j}$ is proportional to a Normal distribution with mean:

$$\frac{\tau \sum_{t=1}^{T} b_{j,t}x_{1j,t} - \tau \alpha_{0j} \sum_{t=1}^{T} x_{1j,t} - \tau \alpha_{2j} \sum_{t=1}^{T} x_{2j,t}}{\lambda_{\alpha} + \tau \sum_{t=1}^{T} x_{1j,t}^2}$$

and standard deviation:

$$\left( \frac{\lambda_{\alpha} + \tau \sum_{t=1}^{T} x_{1j,t}^2}{2} \right)^{-1/2}$$

(A.14)

(A.15)

4. The full conditional posterior of $\alpha_{2j}$ for $j = 1, \ldots, M$ is given as:

$$p(\alpha_{2j} \mid \cdot) \propto \Pi(\alpha_{2j})b_{j}(b_{j,1}, \ldots, b_{j,T} \mid x_{1j,1}, x_{2j,1}, x_{1j,T}, x_{2j,T}, \alpha_{0j}, \alpha_{1j}, \alpha_{2j}, \tau),$$

which can be written as:

$$p(\alpha_{2j} \mid \cdot) \propto \frac{\lambda_{\alpha}}{2\pi} \exp \left\{ -\frac{\lambda_{\alpha}}{2} \alpha_{2j}^2 \right\} \tau^{T/2} \exp \left\{ -\frac{\tau}{2} \sum_{t=1}^{T} \left( b_{j,t} - (\alpha_{0j} + \alpha_{1j}\alpha_{1j,t} + \alpha_{2j}\alpha_{2j,t}) \right)^2 \right\}$$

$$\propto \exp \left\{ -\frac{\lambda_{\alpha}}{2} \alpha_{2j}^2 - \frac{\tau}{2} \sum_{t=1}^{T} \left( -2b_{j,t}\alpha_{2j} + \alpha_{2j}^2 \alpha_{2j,t} + 2\alpha_{0j}\alpha_{2j}\alpha_{2j,t} + 2\alpha_{1j}\alpha_{2j}\alpha_{2j,t} \right) \right\}$$

$$\propto \exp \left\{ -\frac{\lambda_{\alpha}}{2} \alpha_{2j}^2 + \frac{\tau}{2} \sum_{t=1}^{T} \left( b_{j,t}x_{2j,t} - \alpha_{2j} \sum_{t=1}^{T} x_{2j,t}^2 \right) \right\}$$

$$\propto \exp \left\{ -\frac{\lambda_{\alpha}}{2} \alpha_{2j}^2 + \frac{\tau}{2} \sum_{t=1}^{T} \left( \tau \sum_{t=1}^{T} b_{j,t}x_{2j,t} - \alpha_{2j} \sum_{t=1}^{T} x_{2j,t}^2 \right) \right\}$$

$$\propto \exp \left\{ -\frac{\lambda_{\alpha}}{2} \alpha_{2j}^2 + \frac{\tau}{2} \sum_{t=1}^{T} \left( \frac{\tau \sum_{t=1}^{T} b_{j,t}x_{2j,t} - \alpha_{2j} \sum_{t=1}^{T} x_{2j,t}^2}{\lambda_{\alpha} + \tau \sum_{t=1}^{T} x_{2j,t}^2} \right)^2 \right\}.$$
Thus, the conditional posterior of $\alpha_{2j}$ is proportional to a Normal distribution with mean:

$$
\frac{\tau \sum_{t=1}^{T} b_{j,t} x_{2j,t} - \tau \alpha_{0j} \sum_{t=1}^{T} x_{2j,t} - \tau \alpha_{1j} \sum_{t=1}^{T} x_{1j,t} x_{2j,t}}{\lambda_\alpha + \tau \sum_{t=1}^{T} x_{2j,t}^2},
$$

and standard deviation:

$$
\left( \lambda_\alpha + \tau \sum_{t=1}^{T} x_{2j,t}^2 \right)^{-1/2}
$$

(A.17)

5. The full conditional posteriors for $\lambda_j$ for $j = 1, \ldots, M$ are given as:

$$
p(\lambda_j \mid \cdot) \propto g_j(\lambda_j \mid \alpha, \beta) f_j(y_{j,1}, \ldots, y_{j,T} \mid \mu_1, \ldots, \mu_T, b_{j,1}, \ldots, b_{j,T}, \lambda_j),
$$

which can be written as:

$$
p(\lambda_j \mid \cdot) \propto \frac{\beta^\alpha}{F(\alpha)} (\lambda_j)^{\alpha-1} \exp\left\{-\beta \lambda_j\right\} \frac{\lambda_j^{T/2}}{(2\pi)^{T/2}} \exp\left\{-\frac{\lambda_j}{2} \sum_{t=1}^{T} (y_{j,t} - (\mu_t + b_{j,t}))^2\right\} \\
\propto (\lambda_j)^{\alpha-1} \exp\left\{-\beta \lambda_j\right\} \lambda_j^{T/2} \exp\left\{-\frac{\lambda_j}{2} \sum_{t=1}^{T} (y_{j,t} - (\mu_t + b_{j,t}))^2\right\} \\
\propto (\lambda_j)^{\alpha T/2 - 1} \exp\left\{-\lambda_j \left( \beta + \frac{1}{2} \sum_{t=1}^{T} (y_{j,t} - (\mu_t + b_{j,t}))^2 \right) \right\} \\
p(\lambda_j \mid \cdot) \propto \text{Gamma} \left( \alpha + T/2, \beta + \frac{1}{2} \sum_{t=1}^{T} (y_{j,t} - (\mu_t + b_{j,t}))^2 \right).
$$

(A.18)

6. The full conditional posteriors for $\mu_t$ for $t = 1, \ldots, T$ are given as:

$$
p(\mu_t \mid \cdot) \propto \Pi(\mu_t) f_0(y_{0,t} \mid \mu_t, \lambda_0) f_t(y_{1,t}, \ldots, y_{M,t} \mid \mu_t, b_{j,1}, \ldots, b_{j,T}, \lambda_1, \ldots, \lambda_j),
$$
which can be written as:

\[
p(\mu_j | \cdot) \propto \frac{\lambda_{\mu}^{1/2}}{(2\pi)^{1/2}} \exp \left\{ -\frac{\lambda_{\mu}}{2} (\mu_j - M_\mu)^2 \right\} \cdot \frac{\Pi_j^{1/2}}{(2\pi)^{1/2}} \exp \left\{ -\frac{\lambda_0}{2} (y_{j,t} - \mu_t)^2 \right\} \cdot \frac{\lambda_j^{1/2}}{(2\pi)^{1/2}} \exp \left\{ -\frac{\lambda_j}{2} (y_{j,t} - (\mu_t + b_{j,t}))^2 \right\}
\]

\[
\propto \exp \left\{ -\frac{\lambda_{\mu}}{2} (\mu_j - M_\mu)^2 - \frac{\lambda_0}{2} (y_{j,t} - \mu_j)^2 + \sum_{j=1}^{M} -\frac{\lambda_j}{2} (y_{j,t} - (\mu_t + b_{j,t}))^2 \right\}
\]

Thus, the conditional posterior of \( \mu_t \) for \( t = 1, \ldots, T \) is proportional to a Normal distribution with mean:

\[
\frac{\lambda_\mu M_\mu + \lambda_0 y_{0,t} + \sum_{j=1}^{M} \lambda_j y_{j,t} - \sum_{j=1}^{M} \lambda_j b_{j,t}}{\lambda_\mu + \lambda_0 + \sum_{j=1}^{M} \lambda_j},
\]

and standard deviation:

\[
\left( \lambda_\mu + \lambda_0 + \sum_{j=1}^{M} \lambda_j \right)^{-1/2}.
\]

7. The full conditional posteriors for \( b_{j,t} \) for \( t = 1, \ldots, T \) and \( j = 1, \ldots, M \) are given as:

\[
p(b_{j,t} | \cdot) \propto b_{j,t}(b_{j,t} | x_{1j,t}, x_{2j,t}, \alpha_{0j}, \alpha_{1j}, \alpha_{2j}, \tau) f_{j,t}(y_{j,t} | \mu_t, b_{j,t}, \lambda_j),
\]

which can be written as:

\[
p(b_{j,t} | \cdot) \propto \frac{\tau^{1/2}}{(2\pi)^{1/2}} \exp \left\{ -\frac{\tau}{2} (b_{j,t} - (\alpha_{0j} + \alpha_{1j} x_{1j,t} + \alpha_{2j} x_{2j,t}))^2 \right\} \cdot \frac{\lambda_j^{1/2}}{(2\pi)^{1/2}} \exp \left\{ -\frac{\lambda_j}{2} (y_{j,t} - (\mu_t + b_{j,t}))^2 \right\}
\]

\[
\propto \exp \left\{ -\frac{\tau}{2} (b_{j,t} - (\alpha_{0j} + \alpha_{1j} x_{1j,t} + \alpha_{2j} x_{2j,t}))^2 - \frac{\lambda_j}{2} (y_{j,t} - (\mu_t + b_{j,t}))^2 \right\}
\]

\[
\propto \exp \left\{ -\frac{\tau}{2} (b_{j,t}^2 - 2b_{j,t} (\alpha_{0j} + \alpha_{1j} x_{1j,t} + \alpha_{2j} x_{2j,t})) - \frac{\lambda_j}{2} (-2y_{j,t}(\mu_t + b_{j,t}) + (\mu_t + b_{j,t})^2) \right\}
\]

\[
\propto \exp \left\{ -\frac{\tau + \lambda_j}{2} b_{j,t}^2 + \tau \alpha_{0j} + \alpha_{1j} x_{1j,t} + \alpha_{2j} x_{2j,t} + \lambda_j y_{j,t} b_{j,t} - \frac{\lambda_j}{2} (2\mu_t b_{j,t} + b_{j,t}^2) \right\}
\]

\[
\propto \exp \left\{ -\frac{\tau + \lambda_j}{2} b_{j,t}^2 + (\tau \alpha_{0j} + \alpha_{1j} x_{1j,t} + \alpha_{2j} x_{2j,t} + \lambda_j y_{j,t} - \lambda_j \mu_t) b_{j,t} \right\}
\]

\[
\propto \exp \left\{ -\frac{\tau + \lambda_j}{2} \frac{\tau(\alpha_{0j} + \alpha_{1j} x_{1j,t} + \alpha_{2j} x_{2j,t} + \lambda_j y_{j,t} - \lambda_j \mu_t)}{\tau + \lambda_j} \right\}.
\]
Thus, the conditional posterior of $b_{j,t}$ for $t = 1, \ldots, T$ and $j = 1, \ldots, M$ is proportional to a Normal distribution with mean:

$$\tau \left( \alpha_{0j} + \alpha_{1j}x_{1j,t} + \alpha_{2j}x_{2j,t} \right) + \lambda_j y_{j,t} - \lambda_j \mu_t \over \tau + \lambda_j ,$$

and standard deviation:

$$(\tau + \lambda_j)^{-1/2}.$$
APPENDIX B. CONDITIONAL POSTERIOR DERIVATIONS FOR BIVARIATE HIERARCHICAL MODEL

Distributions Implied by the Model

1. For \( t = 1, \ldots, T \) the joint distribution of \( u_{0t} \) and \( v_{0t} \) is:

\[
f_0(u_{0t}, v_{0t} | \mu_{0t}, \nu_{0t}, \Sigma_0) = \frac{1}{2\pi |\Sigma_0|^{1/2}} \exp \left\{ -\frac{1}{2} \begin{pmatrix} u_{0t} - \mu_{0t} \\ v_{0t} - \nu_{0t} \end{pmatrix}' \Sigma_0^{-1} \begin{pmatrix} u_{0t} - \mu_{0t} \\ v_{0t} - \nu_{0t} \end{pmatrix} \right\}
\]

\[
= \frac{1}{2\pi \lambda_0 (1 - \rho^2_o)^{1/2}} \exp \left\{ -\frac{1}{2\lambda_0 (1 - \rho^2_o)} \left( (u_{0t} - \mu_{0t})^2 + (v_{0t} - \nu_{0t})^2 - 2\rho_o (u_{0t} - \mu_{0t}) (v_{0t} - \nu_{0t}) \right) \right\}
\]

(B.1)

2. For \( t = 1, \ldots, T \) and \( j = 1, \ldots, M \) the joint distribution of \( u_{j,t} \) and \( v_{j,t} \) is:

\[
f_{j,t}(u_{j,t}, v_{j,t} | \mu_{0t}, \nu_{0t}, b_{uj,t}, b_{vj,t}, \Sigma_{j,t}) = \frac{1}{2\pi |\Sigma_{j,t}|^{1/2}} \exp \left\{ -\frac{1}{2} \begin{pmatrix} u_{j,t} - (\mu_{0t} + b_{uj,t}) \\ v_{j,t} - (\nu_{0t} + b_{vj,t}) \end{pmatrix}' \Sigma_{j,t}^{-1} \begin{pmatrix} u_{j,t} - (\mu_{0t} + b_{uj,t}) \\ v_{j,t} - (\nu_{0t} + b_{vj,t}) \end{pmatrix} \right\}
\]

(B.2)

Note:

\[
| \Sigma_{j,t} |^{1/2} = (\lambda_j^2 - \rho_t^2 \lambda_j^2)^{1/2} = (\lambda_j^2 (1 - \rho_t^2))^{1/2} = \lambda_j (1 - \rho_t^2)^{1/2},
\]

(B.3)

and

\[
\Sigma_{j,t}^{-1} = \frac{1}{\lambda_j^2 - \rho_t^2 \lambda_j^2} \begin{pmatrix} \lambda_j & -\rho_t \lambda_j \\ -\rho_t \lambda_j & \lambda_j \end{pmatrix} = \frac{1}{\lambda_j^2 (1 - \rho_t^2)} \begin{pmatrix} \lambda_j & -\rho_t \lambda_j \\ -\rho_t \lambda_j & \lambda_j \end{pmatrix} = \frac{1}{\lambda_j (1 - \rho_t^2)} \begin{pmatrix} 1 & -\rho_t \\ -\rho_t & 1 \end{pmatrix}
\]

(B.4)
\[ f_{j,t} (u_{j,t}, v_{j,t} \mid \mu_{0t}, \nu_{0t}, b_{uj,t}, b_{vj,t}, \kappa, \alpha, \beta, \lambda_j, \theta_{0t}) = \]
\[= \frac{1}{2\pi \lambda_j (1 - \rho_t^2)^{1/2}} \exp \left\{ -\frac{1}{2\lambda_j (1 - \rho_t^2)} \begin{pmatrix}
  u_{j,t} - (\mu_{0t} + b_{uj,t}) \\
  v_{j,t} - (\nu_{0t} + b_{vj,t})
\end{pmatrix}^T \begin{pmatrix} \rho_t & -\rho_t \\
 -\rho_t & 1 \end{pmatrix} \begin{pmatrix}
  u_{j,t} - (\mu_{0t} + b_{uj,t}) \\
  v_{j,t} - (\nu_{0t} + b_{vj,t})
\end{pmatrix} \right\} \]
\[= \frac{1}{2\pi \lambda_j (1 - \rho_t^2)^{1/2}} \exp \left\{ -\frac{1}{2\lambda_j (1 - \rho_t^2)} \begin{pmatrix}
  u_{j,t} - (\mu_{0t} + b_{uj,t}) - \rho_t (v_{j,t} - (\nu_{0t} + b_{vj,t})) \\
  v_{j,t} - (\nu_{0t} + b_{vj,t})
\end{pmatrix}^T \begin{pmatrix} \rho_t & 0 \\
 0 & 1 \end{pmatrix} \begin{pmatrix}
  u_{j,t} - (\mu_{0t} + b_{uj,t}) - \rho_t (v_{j,t} - (\nu_{0t} + b_{vj,t})) \\
  v_{j,t} - (\nu_{0t} + b_{vj,t})
\end{pmatrix} \right\} \]
\[= \frac{1}{2\pi \lambda_j (1 - \rho_t^2)^{1/2}} \exp \left\{ -\frac{1}{2\lambda_j (1 - \rho_t^2)} \left[ (u_{j,t} - (\mu_{0t} + b_{uj,t}))^2 - 2\rho_t (u_{j,t} - (\mu_{0t} + b_{uj,t}))(v_{j,t} - (\nu_{0t} + b_{vj,t})) + (v_{j,t} - (\nu_{0t} + b_{vj,t}))^2 \right] \right\} \]
\[= \frac{1}{2\pi \lambda_j (1 - \rho_t^2)^{1/2}} \exp \left\{ -\frac{1}{2\lambda_j (1 - \rho_t^2)} \left[ (u_{j,t} - (\mu_{0t} + b_{uj,t}))^2 - 2\rho_t (u_{j,t} - (\mu_{0t} + b_{uj,t}))(v_{j,t} - (\nu_{0t} + b_{vj,t})) + (v_{j,t} - (\nu_{0t} + b_{vj,t}))^2 \right] \right\} \]

3. For \( j = 1, \ldots, M \), \( \lambda_j \) has density:
\[ g_j (\lambda_j \mid c, d) = \frac{\Gamma(c)}{\Gamma(c - 1)} \lambda_j^{c - 1 - 1} \exp \left\{ -\frac{d}{\lambda_j} \right\} \] (B.6)

4. For \( j = 1, \ldots, M \) and \( t = 1, \ldots, T \), \( b_{uj,t} \) has density:
\[ h_{j,t} (b_{uj,t} \mid \alpha_{0j}, \alpha_{1j}, \alpha_{2j}, x_{1j,t}, x_{2j,t}, \tau) = \left( \frac{T}{2\pi} \right)^{1/2} \exp \left\{ -\frac{T}{2} (b_{uj,t} - (\alpha_{0j} + \alpha_{1j} x_{1j,t} + \alpha_{2j} x_{2j,t}))^2 \right\} \] (B.7)

5. For \( j = 1, \ldots, M \) and \( t = 1, \ldots, T \), \( b_{vj,t} \) has density:
\[ h_{j,t} (b_{vj,t} \mid \eta_{0j}, \eta_{1j}, \eta_{2j}, x_{1j,t}, x_{2j,t}, \gamma) = \left( \frac{T}{2\pi} \right)^{1/2} \exp \left\{ -\frac{\gamma}{2} (b_{vj,t} - (\eta_{0j} + \eta_{1j} x_{1j,t} + \eta_{2j} x_{2j,t}))^2 \right\} \] (B.8)

6. For \( t = 1, \ldots, T \) the joint density of \( u_{j,t} \) and \( v_{j,t} \) is:
\[ f_t (u_{1,t}, \ldots, u_{M,t}, v_{1,t}, \ldots, v_{M,t} \mid \mu_{0t}, \nu_{0t}, b_{u1,t}, \ldots, b_{uM,t}, b_{v1,t}, \ldots, b_{vM,t}, \lambda_1, \ldots, \lambda_M, \kappa, \alpha, \beta, \theta_{0t}) = \]
\[= \prod_{j=1}^M \frac{1}{2\pi \lambda_j (1 - \rho_t^2)^{1/2}} \exp \left\{ -\frac{1}{2\lambda_j (1 - \rho_t^2)} \left[ (u_{j,t} - (\mu_{0t} + b_{uj,t}))^2 - 2(\rho_t (u_{j,t} - (\mu_{0t} + b_{uj,t}))(v_{j,t} - (\nu_{0t} + b_{vj,t})) + (v_{j,t} - (\nu_{0t} + b_{vj,t}))^2 \right] \right\} \]
\[= \left( \frac{1}{2\pi (1 - \rho_t^2)^{1/2}} \right)^M \prod_{j=1}^M \frac{1}{\lambda_j} \exp \left\{ -\sum_{j=1}^M \left[ \frac{1}{2\lambda_j (1 - \rho_t^2)} \left[ (u_{j,t} - (\mu_{0t} + b_{uj,t}))^2 - 2(\rho_t (u_{j,t} - (\mu_{0t} + b_{uj,t}))(v_{j,t} - (\nu_{0t} + b_{vj,t})) + (v_{j,t} - (\nu_{0t} + b_{vj,t}))^2 \right] \right) \right\} \] (B.9)
7. For $j = 1, \ldots, M$ the joint density of $u_{j,t}$ and $v_{j,t}$ is:

$$f_j(u_{j,1}, \ldots, u_{j,T}; v_{j,1}, \ldots, v_{j,T} | \mu_0, \ldots, \mu_{0T}, \nu_0, \ldots, \nu_{0T}, b_{u,j,1}, \ldots, b_{u,j,T}, b_{v,j,1}, \ldots, b_{v,j,T}, \lambda_j, \kappa, \alpha, \beta, \theta_{0t}) =$$

$$= \prod_{t=1}^{T} \frac{1}{2\pi \lambda_j (1 - [\kappa \sin(2\pi \theta_0 \alpha - \beta)]^2)^{1/2}} \exp \left\{ -\frac{1}{2\lambda_j (1 - [\kappa \sin(2\pi \theta_0 \alpha - \beta)]^2)} \left[ (u_{j,t} - (\mu_0 + b_{u,j,t}))^2 - 2(\kappa \sin(2\pi \theta_0 \alpha - \beta))(u_{j,t} - (\mu_0 + b_{u,j,t}))(v_{j,t} - (\nu_0 + b_{v,j,t})) + (v_{j,t} - (\nu_0 + b_{v,j,t}))^2 \right] \right\}$$

$$= \left( \frac{1}{2\pi \lambda_j} \right)^T \prod_{t=1}^{T} \frac{1}{2\pi \lambda_j (1 - [\kappa \sin(2\pi \theta_0 \alpha - \beta)]^2)^{1/2}} \exp \left\{ -\frac{1}{2\lambda_j \sum_{t=1}^{T} (1 - [\kappa \sin(2\pi \theta_0 \alpha - \beta)]^2)} \left[ (u_{j,t} - (\mu_0 + b_{u,j,t}))^2 - 2(\kappa \sin(2\pi \theta_0 \alpha - \beta))(u_{j,t} - (\mu_0 + b_{u,j,t}))(v_{j,t} - (\nu_0 + b_{v,j,t})) + (v_{j,t} - (\nu_0 + b_{v,j,t}))^2 \right] \right\}$$

(B.10)

8. The joint density of the entire set $\{u_{j,t}, v_{j,t} : t = 1, \ldots, T; j = 1, \ldots, M\}$ is:

$$f(\{u_{j,t}, v_{j,t} : t = 1, \ldots, T; j = 1, \ldots, M\} | \mu_0, \ldots, \mu_{0T}, \nu_0, \ldots, \nu_{0T}, \{b_{u,j,t}, b_{v,j,t} : t = 1, \ldots, T; j = 1, \ldots, M\}) =$$

$$= \prod_{t=1}^{T} \prod_{j=1}^{M} \frac{1}{2\pi \lambda_j (1 - [\kappa \sin(2\pi \theta_0 \alpha - \beta)]^2)^{1/2}} \exp \left\{ 1 \sum_{t=1}^{T} (1 - [\kappa \sin(2\pi \theta_0 \alpha - \beta)]^2) \right\} \exp \left\{ -\frac{1}{2\lambda_j (1 - [\kappa \sin(2\pi \theta_0 \alpha - \beta)]^2)} \left[ (u_{j,t} - (\mu_0 + b_{u,j,t}))^2 - 2(\kappa \sin(2\pi \theta_0 \alpha - \beta))(u_{j,t} - (\mu_0 + b_{u,j,t}))(v_{j,t} - (\nu_0 + b_{v,j,t})) + (v_{j,t} - (\nu_0 + b_{v,j,t}))^2 \right] \right\}$$

$$= \prod_{t=1}^{T} \prod_{j=1}^{M} \frac{1}{2\pi \lambda_j (1 - [\kappa \sin(2\pi \theta_0 \alpha - \beta)]^2)^{1/2}} \exp \left\{ \sum_{t=1}^{T} \sum_{j=1}^{M} -\frac{1}{2\lambda_j (1 - [\kappa \sin(2\pi \theta_0 \alpha - \beta)]^2)} \left[ (u_{j,t} - (\mu_0 + b_{u,j,t}))^2 - 2(\kappa \sin(2\pi \theta_0 \alpha - \beta))(u_{j,t} - (\mu_0 + b_{u,j,t}))(v_{j,t} - (\nu_0 + b_{v,j,t})) + (v_{j,t} - (\nu_0 + b_{v,j,t}))^2 \right] \right\}$$

(B.11)

9. The joint density of $\lambda_1, \ldots, \lambda_M$ is:

$$g(\lambda_1, \ldots, \lambda_M | c, d) = \prod_{j=1}^{M} \frac{d^\kappa}{\Gamma(c)} \lambda_j^{c-1} \exp \left\{ -\frac{d}{\lambda_j} \right\}$$

$$= \left( \frac{d^\kappa}{\Gamma(c)} \right)^M \left( \prod_{j=1}^{M} \lambda_j \right)^{c-1} \exp \left\{ -d \sum_{j=1}^{M} \frac{1}{\lambda_j} \right\}$$

(B.12)

10. For $j = 1, \ldots, M$, $b_{u,j,1}, \ldots, b_{u,j,T}$ have the joint density:

$$h_j(b_{u,j,1}, \ldots, b_{u,j,T} | \alpha_{0j}, \alpha_{1j}, \alpha_{2j}, x_{1j,1}, \ldots, x_{1j,T}, x_{2j,1}, \ldots, x_{2j,T}, T) =$$

$$= \prod_{t=1}^{T} \left( \frac{\tau}{2\pi} \right)^{1/2} \exp \left\{ -\frac{\tau}{2} \left( b_{u,j,t} - (\alpha_{0j} + \alpha_{1j} x_{1j,t} + \alpha_{2j} x_{2j,t}) \right)^2 \right\}$$

$$= \left( \frac{\tau}{2\pi} \right)^{T/2} \exp \left\{ -\frac{\tau}{2} \sum_{t=1}^{T} \left( b_{u,j,t} - (\alpha_{0j} + \alpha_{1j} x_{1j,t} + \alpha_{2j} x_{2j,t}) \right)^2 \right\}$$

(B.13)
11. For $j = 1, \ldots, M$, $b_{v,j,1}, \ldots, b_{v,j,T}$ has joint density:

$$h_j (b_{v,j,1}, \ldots, b_{v,j,T} \mid \eta_0, \eta_1, \eta_2, x_{1,j,1}, \ldots, x_{1,j,T}, x_{2,j,1}, \ldots, x_{2,j,T}, \gamma) =$$

$$= \prod_{t=1}^{T} \left( \frac{\gamma}{2\pi} \right)^{1/2} \exp \left\{ -\frac{\gamma}{2} (b_{v,j,t} - (\eta_0 + \eta_1 x_{1,j,t} + \eta_2 x_{2,j,t}))^2 \right\}$$

$$= \left( \frac{\gamma}{2\pi} \right)^{T/2} \exp \left\{ -\frac{\gamma}{2} \sum_{t=1}^{T} (b_{v,j,t} - (\eta_0 + \eta_1 x_{1,j,t} + \eta_2 x_{2,j,t}))^2 \right\} \quad (B.14)$$

12. For $t = 1, \ldots, T$, $b_{u1,t}, \ldots, b_{uM,t}$ has joint density:

$$h_t (b_{u1,t}, \ldots, b_{uM,t} \mid \alpha_0, 1, \alpha_0 M, \alpha_1, \ldots, \alpha_1 M, \alpha_2, \ldots, \alpha_2 M, x_{11,1}, \ldots, x_{1M}, x_{21,1}, \ldots, x_{2M}, \tau) =$$

$$= \prod_{j=1}^{M} \left( \frac{\tau}{2\pi} \right)^{1/2} \exp \left\{ -\frac{\tau}{2} (b_{u,j,t} - (\alpha_0 + \alpha_1 x_{1,j,t} + \alpha_2 x_{2,j,t}))^2 \right\}$$

$$= \left( \frac{\tau}{2\pi} \right)^{J/2} \exp \left\{ -\frac{\tau}{2} \sum_{j=1}^{M} (b_{u,j,t} - (\alpha_0 + \alpha_1 x_{1,j,t} + \alpha_2 x_{2,j,t}))^2 \right\} \quad (B.15)$$

13. For $t = 1, \ldots, T$, $b_{v1,t}, \ldots, b_{vM,t}$ has joint density:

$$h_t (b_{v1,t}, \ldots, b_{vM,t} \mid \eta_0, 1, \eta_0 M, \eta_1, \ldots, \eta_1 M, \eta_2, \ldots, \eta_2 M, x_{11,1}, \ldots, x_{1M}, x_{21,1}, \ldots, x_{2M}, \gamma) =$$

$$= \prod_{j=1}^{M} \left( \frac{\gamma}{2\pi} \right)^{1/2} \exp \left\{ -\frac{\gamma}{2} (b_{v,j,t} - (\eta_0 + \eta_1 x_{1,j,t} + \eta_2 x_{2,j,t}))^2 \right\}$$

$$= \left( \frac{\gamma}{2\pi} \right)^{J/2} \exp \left\{ -\frac{\gamma}{2} \sum_{j=1}^{M} (b_{v,j,t} - (\eta_0 + \eta_1 x_{1,j,t} + \eta_2 x_{2,j,t}))^2 \right\} \quad (B.16)$$

14. The joint density of the entire set $\{b_{u,j,t} : t = 1, \ldots, T; j = 1, \ldots, M\}$ is:

$$h (\{b_{u,j,t} : t = 1, \ldots, T; j = 1, \ldots, M\} \mid \alpha_0, \alpha_1, \alpha_2, x_{1,j,t}, x_{2,j,t}, \tau) =$$

$$= \prod_{j=1}^{M} \prod_{t=1}^{T} \left( \frac{\tau}{2\pi} \right)^{1/2} \exp \left\{ -\frac{\tau}{2} (b_{u,j,t} - (\alpha_0 + \alpha_1 x_{1,j,t} + \alpha_2 x_{2,j,t}))^2 \right\}$$

$$= \left( \frac{\tau}{2\pi} \right)^{MT/2} \exp \left\{ -\frac{\tau}{2} \sum_{j=1}^{M} \sum_{t=1}^{T} (b_{u,j,t} - (\alpha_0 + \alpha_1 x_{1,j,t} + \alpha_2 x_{2,j,t}))^2 \right\} \quad (B.17)$$

15. The joint density of the entire set $\{b_{v,j,t} : t = 1, \ldots, T; j = 1, \ldots, M\}$ is:

$$h (\{b_{v,j,t} : t = 1, \ldots, T; j = 1, \ldots, M\} \mid \eta_0, \eta_1, \eta_2, x_{1,j,t}, x_{2,j,t}, \gamma) =$$

= \prod_{j=1}^{M} \prod_{t=1}^{T} \left( \frac{\gamma}{2\pi} \right)^{1/2} \exp \left\{ -\frac{\gamma}{2} \left( \frac{\gamma}{\tau} \right) \right\} \exp \left\{ -\frac{\gamma}{2} \sum_{j=1}^{M} \sum_{t=1}^{T} \left( b_{uj,t} - \left( \eta_{0j} + \eta_{1j} x_{1j,t} + \eta_{2j} x_{2j,t} \right) \right)^2 \right\} \\
= \left( \frac{\gamma}{2\pi} \right)^{MT/2} \exp \left\{ -\frac{\gamma}{2} \sum_{j=1}^{M} \sum_{t=1}^{T} \left( b_{uj,t} - \left( \eta_{0j} + \eta_{1j} x_{1j,t} + \eta_{2j} x_{2j,t} \right) \right)^2 \right\}

\text{(B.18)}

**Conditional Posteriors**

1. The full conditional posterior of \( \tau \) is given as:

\[
p(\tau \mid \cdot) \propto \Pi(\tau) h \left( \{b_{uj,t} : j = 1, \ldots, M; t = 1, \ldots, T\} \mid \alpha_{0j}, \alpha_{1j}, \alpha_{2j}, x_{1j,t}, x_{2j,t}, \tau \right)
\]

\[
\propto \frac{b^a}{\Gamma(a)} \tau^{a-1} \exp \left\{ -b \tau \left( \frac{\gamma}{2\pi} \right)^{MT/2} \exp \left\{ -\frac{\tau}{2} \sum_{j=1}^{M} \sum_{t=1}^{T} \left( b_{uj,t} - \left( \alpha_{0j} + \alpha_{1j} x_{1j,t} + \alpha_{2j} x_{2j,t} \right) \right)^2 \right\} \right\}
\]

\[
\propto \tau^{a+MT/2-1} \exp \left\{ -\tau \left( b + \frac{1}{2} \sum_{j=1}^{M} \sum_{t=1}^{T} \left( b_{uj,t} - \left( \alpha_{0j} + \alpha_{1j} x_{1j,t} + \alpha_{2j} x_{2j,t} \right) \right)^2 \right) \right\}
\]

\[
\propto G \left( a + MT/2, b + \frac{1}{2} \sum_{j=1}^{M} \sum_{t=1}^{T} \left( b_{uj,t} - \left( \alpha_{0j} + \alpha_{1j} x_{1j,t} + \alpha_{2j} x_{2j,t} \right) \right)^2 \right)
\]

**B.19**

2. The full conditional posterior of \( \gamma \) is given as:

\[
p(\gamma \mid \cdot) \propto \Pi(\gamma) h \left( \{b_{uj,t} : j = 1, \ldots, M; t = 1, \ldots, T\} \mid \eta_{0j}, \eta_{1j}, \eta_{2j}, x_{1j,t}, x_{2j,t}, \gamma \right)
\]

\[
\propto \frac{b^a}{\Gamma(a)} \gamma^{a-1} \exp \left\{ -b \gamma \left( \frac{\gamma}{2\pi} \right)^{MT/2} \exp \left\{ \gamma \sum_{j=1}^{M} \sum_{t=1}^{T} \left( b_{uj,t} - \left( \eta_{0j} + \eta_{1j} x_{1j,t} + \eta_{2j} x_{2j,t} \right) \right)^2 \right\} \right\}
\]

\[
\propto \gamma^{a+MT/2-1} \exp \left\{ -\gamma \left( b + \frac{1}{2} \sum_{j=1}^{M} \sum_{t=1}^{T} \left( b_{uj,t} - \left( \eta_{0j} + \eta_{1j} x_{1j,t} + \eta_{2j} x_{2j,t} \right) \right)^2 \right) \right\}
\]

\[
\propto G \left( a + MT/2, b + \frac{1}{2} \sum_{j=1}^{M} \sum_{t=1}^{T} \left( b_{uj,t} - \left( \eta_{0j} + \eta_{1j} x_{1j,t} + \eta_{2j} x_{2j,t} \right) \right)^2 \right)
\]

**B.20**

3. The full conditional posterior of \( \alpha_{0j} \) is given as:

\[
p(\alpha_{0j} \mid \cdot) \propto \Pi(\alpha_{0j}) h_j \left( b_{uj,1}, \ldots, b_{uj,T} \mid \alpha_{0j}, \alpha_{1j}, \alpha_{2j}, x_{1j,1}, \ldots, x_{1j,T}, x_{2j,1}, \ldots, x_{2j,T}, \tau \right)
\]

\[
\propto \left( \frac{\lambda_r}{2\pi} \right)^{1/2} \exp \left\{ -\frac{\lambda_r}{2} \alpha_{0j}^2 \right\} \exp \left\{ -\frac{\tau}{2} \sum_{t=1}^{T} \left( b_{uj,t} - \left( \alpha_{0j} + \alpha_{1j} x_{1j,t} + \alpha_{2j} x_{2j,t} \right) \right)^2 \right\}
\]

\[
\propto \exp \left\{ -\frac{\lambda_r}{2} \alpha_{0j}^2 - \frac{\tau}{2} \sum_{t=1}^{T} \left( b_{uj,t} - 2b_{uj,t} \left( \alpha_{0j} + \alpha_{1j} x_{1j,t} + \alpha_{2j} x_{2j,t} \right) + \left( \alpha_{0j} + \alpha_{1j} x_{1j,t} + \alpha_{2j} x_{2j,t} \right)^2 \right) \right\}
\]
4. The full conditional posterior of $\alpha_{ij}$ is given as:

$$p(\alpha_{ij} \mid \cdot) \propto \Pi(\alpha_{ij}) h_j(b_{u,j,t}, \ldots, b_{u,j,T} \mid \alpha_{ij}, \alpha_{ij+1} \ldots, x_{1j,t}, x_{2j,t}, \ldots, x_{2j,T}, \tau)$$

$$\propto \left(\frac{\lambda_r}{2\pi}\right)^{1/2} \exp\left\{-\frac{\lambda_r}{2} \alpha_{ij}^2 \right\} \left(\frac{\tau}{2}\right)^{T/2} \exp\left\{-\frac{\tau}{2} \sum_{t=1}^{T} (b_{u,j,t} - (\alpha_{ij+1} x_{1j,t} + \alpha_{ij+2} x_{2j,t})^2) \right\}$$

$$\propto \left(\frac{\lambda_r}{2}\right)^{1/2} \alpha_{ij}^2 \sum_{t=1}^{T} b_{u,j,t} - 2b_{u,j,t} (\alpha_{ij} + \alpha_{ij} x_{1j,t} + \alpha_{ij+2} x_{2j,t}) + (\alpha_{ij} + \alpha_{ij} x_{1j,t} + \alpha_{ij+2} x_{2j,t})^2$$

$$= b_{u,j,t} - 2b_{u,j,t} \alpha_{ij} - 2b_{u,j,t} \alpha_{ij} x_{1j,t} - 2b_{u,j,t} \alpha_{ij+2} x_{2j,t} + \alpha_{ij} + \alpha_{ij} x_{1j,t} + \alpha_{ij+2} x_{2j,t}$$

$$= b_{u,j,t} + 2\alpha_{ij} x_{1j,t} + 2\alpha_{ij} x_{2j,t} + 2\alpha_{ij} x_{1j,t} x_{2j,t} + 2\alpha_{ij} x_{1j,t} x_{2j,t}$$

Thus, the conditional posterior of $\alpha_{ij}$ is proportional to a Normal distribution with mean:

$$\frac{\tau}{\lambda_r + \tau T} \sum_{t=1}^{T} b_{u,j,t} - \tau \alpha_{ij} \sum_{t=1}^{T} x_{1j,t} - \tau \alpha_{ij+2} \sum_{t=1}^{T} x_{2j,t}, \quad (B.21)$$

and variance:

$$(\lambda_r + \tau T)^{-1} \quad (B.22)$$
Thus, the conditional posterior of $\alpha_{1j}$ is proportional to a Normal distribution with mean:

$$\frac{\tau \sum_{t=1}^{T} b_{uj,t} x_{1j,t} - \tau \alpha_{0j} \sum_{t=1}^{T} x_{1j,t} - \tau \alpha_{2j} \sum_{t=1}^{T} x_{2j,t} - \lambda_{r} + \tau \sum_{t=1}^{T} x_{1j,t}^{2}}{\lambda_{r} + \tau \sum_{t=1}^{T} x_{1j,t}^{2}},$$  \hspace{1cm} (B.23)

and variance:

$$\left(\lambda_{r} + \tau \sum_{t=1}^{T} x_{1j,t}^{2}\right)^{-1}$$ \hspace{1cm} (B.24)

5. The full conditional posterior of $\alpha_{2j}$ is given as:

$$p(\alpha_{2j} \mid \cdot) \propto \Pi(\alpha_{2j})h_{j} \left( b_{uj,1}, \ldots, b_{uj,T} \mid \alpha_{0j}, \alpha_{1j}, \alpha_{2j}, x_{1j,1}, \ldots, x_{1j,T}, x_{2j,1}, \ldots, x_{2j,T}, \tau \right)$$

$$\propto \left( \frac{\lambda_{r}}{2\pi} \right)^{1/2} \exp \left\{ -\frac{\lambda_{r}}{2} \alpha_{2j}^{2} \right\} \frac{\tau^{T/2}}{2\pi} \exp \left\{ -\frac{\tau}{2} \sum_{t=1}^{T} \left( b_{uj,t} - \left( \alpha_{0j} + \alpha_{1j} x_{1j,t} + \alpha_{2j} x_{2j,t} \right) \right)^{2} \right\}$$

$$\propto \exp \left\{ -\frac{\lambda_{r}}{2} \alpha_{2j}^{2} - \frac{\tau}{2} \sum_{t=1}^{T} \left( -2 b_{uj,t} \alpha_{0j} - \alpha_{0j} \alpha_{1j} x_{1j,t} + \alpha_{1j} \alpha_{2j} x_{2j,t} + \alpha_{2j} x_{2j,t}^{2} \right) \right\}$$

Thus, the conditional posterior of $\alpha_{2j}$ is proportional to a Normal distribution with mean:

$$\frac{\tau \sum_{t=1}^{T} b_{uj,t} x_{2j,t} - \tau \alpha_{0j} \sum_{t=1}^{T} x_{2j,t} - \tau \alpha_{1j} \sum_{t=1}^{T} x_{1j,t} x_{2j,t} - \lambda_{r} + \tau \sum_{t=1}^{T} x_{2j,t}^{2}}{\lambda_{r} + \tau \sum_{t=1}^{T} x_{2j,t}^{2}},$$ \hspace{1cm} (B.25)

and variance:

$$\left(\lambda_{r} + \tau \sum_{t=1}^{T} x_{2j,t}^{2}\right)^{-1}$$ \hspace{1cm} (B.26)

6. The full conditional posterior of $\eta_{0j}$ is given as:

$$p(\eta_{0j} \mid \cdot) \propto \Pi(\eta_{0j})h_{j} \left( b_{uj,1}, \ldots, b_{uj,T} \mid \eta_{0j}, \eta_{1j}, \eta_{2j}, x_{1j,1}, \ldots, x_{1j,T}, x_{2j,1}, \ldots, x_{2j,T}, \tau, \gamma \right)$$

$$\propto \left( \frac{\lambda_{r}}{2\pi} \right)^{1/2} \exp \left\{ -\frac{\lambda_{r}}{2} \eta_{0j}^{2} \right\} \frac{\gamma^{T/2}}{2\pi} \exp \left\{ -\frac{\gamma}{2} \sum_{t=1}^{T} \left( b_{uj,t} - \left( \eta_{0j} + \eta_{1j} x_{1j,t} + \eta_{2j} x_{2j,t} \right) \right)^{2} \right\}$$

$$\propto \exp \left\{ -\frac{\lambda_{r}}{2} \eta_{0j}^{2} - \frac{\gamma}{2} \sum_{t=1}^{T} \left( -2 b_{uj,t} \eta_{0j} - \eta_{0j} \eta_{1j} x_{1j,t} + \eta_{1j} \eta_{2j} x_{2j,t} + \eta_{2j} x_{2j,t}^{2} \right) \right\}$$
Note:

\[
(b_{j,t} - (\eta_{0j} + \eta_{1j}x_{1j,t} + \eta_{2j}x_{2j,t}))^2 = b_{j,t}^2 - 2b_{j,t}(\eta_{0j} + \eta_{1j}x_{1j,t} + \eta_{2j}x_{2j,t}) + (\eta_{0j} + \eta_{1j}x_{1j,t} + \eta_{2j}x_{2j,t})^2
\]

\[
= b_{j,t}^2 - 2b_{j,t}x_{0j} - 2b_{j,t}\eta_{1j}x_{1j,t} - 2b_{j,t}\eta_{2j}x_{2j,t} + \eta_{0j}^2 + \eta_{1j}^2x_{1j,t}^2
\]

\[
+ \eta_{2j}^2x_{2j,t} + 2\eta_{0j}\eta_{1j}x_{1j,t} + 2\eta_{0j}\eta_{2j}x_{2j,t} + 2\eta_{1j}\eta_{2j}x_{1j,t}x_{2j,t}
\]

\[
p(\eta_{0j} | \cdot) \propto \exp\left\{ -\frac{\lambda_r}{2} \eta_{0j}^2 - \frac{\gamma}{2} \sum_{t=1}^{T} (-2b_{j,t}\eta_{0j} + \eta_{0j}^2) \right\}
\]

\[
\propto \exp\left\{ -\frac{\lambda_r}{2} \eta_{0j}^2 + \gamma \eta_{0j} \sum_{t=1}^{T} b_{j,t} - \frac{\gamma}{2} T \eta_{0j}^2 - \gamma \eta_{0j} \eta_{1j} \sum_{t=1}^{T} x_{1j,t} - \gamma \eta_{0j} \eta_{2j} \sum_{t=1}^{T} x_{2j,t} \right\}
\]

\[
\propto \exp\left\{ -\frac{\lambda_r + \gamma T}{2} \eta_{0j}^2 + \left( \gamma \sum_{t=1}^{T} b_{j,t} - \gamma \eta_{1j} \sum_{t=1}^{T} x_{1j,t} - \gamma \eta_{2j} \sum_{t=1}^{T} x_{2j,t} \right) \eta_{0j} \right\}
\]

\[
\propto \exp\left\{ -\frac{\lambda_r + \gamma T}{2} \left( \eta_{0j} - \frac{\gamma \sum_{t=1}^{T} b_{j,t} - \gamma \eta_{1j} \sum_{t=1}^{T} x_{1j,t} - \gamma \eta_{2j} \sum_{t=1}^{T} x_{2j,t}}{\lambda_r + \gamma T} \right)^2 \right\}
\]

Thus, the conditional posterior of \(\eta_{0j}\) is proportional to a Normal distribution with mean:

\[
\frac{\gamma \sum_{t=1}^{T} b_{j,t}}{\lambda_r + \gamma T} - \frac{\gamma \eta_{1j} \sum_{t=1}^{T} x_{1j,t}}{\lambda_r + \gamma T} - \frac{\gamma \eta_{2j} \sum_{t=1}^{T} x_{2j,t}}{\lambda_r + \gamma T},
\]

and variance:

\[
(\lambda_r + \gamma T)^{-1}
\]

7. The full conditional posterior of \(\eta_{1j}\) is given as:

\[
p(\eta_{1j} | \cdot) \propto \Pi(\eta_{1j})h_j(b_{j,1}, \ldots, b_{j,T} | \eta_{0j}, \eta_{1j}, \eta_{2j}, x_{1j,1}, \ldots, x_{1j,T}, x_{2j,1}, \ldots, x_{2j,T})
\]

\[
\propto \left( \frac{\lambda_r}{2\pi} \right)^{1/2} \exp\left\{ -\frac{\lambda_r}{2} \eta_{1j}^2 \right\} \left( \frac{\gamma}{2\pi} \right)^{T/2} \exp\left\{ -\frac{\gamma}{2} \sum_{t=1}^{T} (b_{j,t} - (\eta_{0j} + \eta_{1j}x_{1j,t} + \eta_{2j}x_{2j,t}))^2 \right\}
\]

\[
\propto \exp\left\{ -\frac{\lambda_r}{2} \eta_{1j}^2 - \frac{\gamma}{2} \sum_{t=1}^{T} [b_{j,t} - 2b_{j,t}(\eta_{0j} + \eta_{1j}x_{1j,t} + \eta_{2j}x_{2j,t}) + (\eta_{0j} + \eta_{1j}x_{1j,t} + \eta_{2j}x_{2j,t})^2] \right\}
\]

\[
\propto \exp\left\{ -\frac{\lambda_r}{2} \eta_{1j}^2 + \gamma \eta_{1j} \sum_{t=1}^{T} b_{j,t} - \gamma \eta_{1j} \sum_{t=1}^{T} x_{1j,t} - \gamma \eta_{2j} \sum_{t=1}^{T} x_{2j,t} \right\}
\]

\[
\propto \exp\left\{ -\frac{\lambda_r + \gamma T}{2} \eta_{1j}^2 + \left( \gamma \sum_{t=1}^{T} b_{j,t}x_{1j,t} - \gamma \eta_{0j} \eta_{1j} \sum_{t=1}^{T} x_{1j,t} - \gamma \eta_{2j} \eta_{1j} \sum_{t=1}^{T} x_{2j,t} \right) \eta_{1j} \right\}
\]

\[
\propto \exp\left\{ -\frac{\lambda_r + \gamma T}{2} \left( \eta_{1j} - \frac{\gamma \sum_{t=1}^{T} b_{j,t}x_{1j,t} - \gamma \eta_{0j} \sum_{t=1}^{T} x_{1j,t} - \gamma \eta_{2j} \sum_{t=1}^{T} x_{2j,t}}{\lambda_r + \gamma T} \right)^2 \right\}
\]
Thus, the conditional posterior of $\eta_{1j}$ is proportional to a Normal distribution with mean:

$$\frac{\gamma \sum_{t=1}^{T} b_{ij,t} x_{1j,t} - \gamma \eta_{0j} \sum_{t=1}^{T} x_{1j,t} - \gamma \eta_{2j} \sum_{t=1}^{T} x_{1j,t} x_{2j,t}}{\lambda_r + \gamma \sum_{t=1}^{T} x_{1j,t}^2},$$

(B.29)

and variance:

$$\left(\lambda_r + \gamma \sum_{t=1}^{T} x_{1j,t}^2\right)^{-1}$$

(B.30)

8. The full conditional posterior of $\eta_{2j}$ is given as:

$$p(\eta_{2j} \mid \cdot) \propto \Pi(\eta_{2j}) b_{ij} (b_{ij,1}, \ldots, b_{ij,T} \mid \eta_{0j}, \eta_{1j}, \eta_{2j}, x_{1j,1}, \ldots, x_{1j,T}, x_{2j,1}, \ldots, x_{2j,T}, \cdot, \gamma)$$

$$\propto \left(\frac{\lambda_r}{2\pi}\right)^{1/2} \exp \left\{ -\frac{\lambda_r}{2} \eta_{2j}^2 \right\} \left(\frac{\gamma}{2\pi}\right)^{T/2} \exp \left\{ -\frac{\gamma}{2} \sum_{t=1}^{T} \left[ b_{ij,t} - 2b_{ij,1} + \eta_{0j} + \eta_{1j} x_{1j,t} + \eta_{2j} x_{2j,t} \right]^2 \right\}$$

$$\propto \left(\frac{\lambda_r}{2}\right)^{T/2} \exp \left\{ -\frac{\lambda_r}{2} \eta_{2j}^2 - \frac{\gamma}{2} \sum_{t=1}^{T} \left[ b_{ij,t} - 2b_{ij,1} + \eta_{0j} + \eta_{1j} x_{1j,t} + \eta_{2j} x_{2j,t} \right]^2 \right\}$$

$$\propto \exp \left\{ -\frac{\lambda_r}{2} \eta_{2j}^2 + \gamma \eta_{2j} \sum_{t=1}^{T} b_{ij,t} x_{2j,t} = \gamma \sum_{t=1}^{T} x_{2j,t} - \gamma \eta_{0j} \sum_{t=1}^{T} x_{2j,t} - \gamma \eta_{1j} \sum_{t=1}^{T} x_{1j,t} x_{2j,t} \right\}$$

$$\propto \exp \left\{ -\lambda_r + \gamma \sum_{t=1}^{T} x_{2j,t}^2 \right\} \eta_{2j}^2$$

Thus, the conditional posterior of $\eta_{2j}$ is proportional to a Normal distribution with mean:

$$\frac{\gamma \sum_{t=1}^{T} b_{ij,t} x_{2j,t} - \gamma \eta_{0j} \sum_{t=1}^{T} x_{2j,t} - \gamma \eta_{1j} \sum_{t=1}^{T} x_{1j,t} x_{2j,t}}{\lambda_r + \gamma \sum_{t=1}^{T} x_{2j,t}^2},$$

(B.31)

and variance:

$$\left(\lambda_r + \gamma \sum_{t=1}^{T} x_{2j,t}^2\right)^{-1}$$

(B.32)
9. The full conditional posterior of $\mu_{0t}$ is given as:

$$
p(\mu_{0t}) \propto \Pi(\mu_{0t}) f(u_{0t}, v_{0t} | \mu_{0t}, v_{0t}, \Sigma_0) f(t(u_{0t}, \ldots, u_{Mt}, v_{0t}, \ldots, v_{Mt} | \mu_{0t}, v_{0t}, b_{u0t}, \ldots, b_{oM}, b_{v0t}, \ldots, b_{oM}, \lambda_1, \ldots)) \propto \left(\frac{\lambda_0}{2\pi}\right)^{1/2} \exp\left\{-\frac{\lambda_0}{2} (\mu_{0t} - M_0)^2\right\} \frac{1}{2\pi\lambda_0(1 - \rho_0^2)^{1/2}} \exp\left\{-\frac{1}{2\lambda_0(1 - \rho_0^2)} (u_{0t} - \mu_{0t})^2 + (v_{0t} - \mu_{0t})^2 - 2\rho_0 (u_{0t} - \mu_{0t}) (v_{0t} - \mu_{0t})\right\} \left(\frac{1}{2\pi(1 - \rho_t^2)^{1/2}}\right)^M \left(\prod_{j=1}^M \frac{1}{\lambda_j}\right) \exp\left\{-\sum_{j=1}^M \frac{1}{2\lambda_j(1 - \rho_t^2)} \left[(u_{j,t} - (\mu_{0t} + b_{u_{j,t}}))^2 + (v_{j,t} - (\mu_{0t} + b_{v_{j,t}}))^2 - 2\rho_t (u_{j,t} - (\mu_{0t} + b_{u_{j,t}})) (v_{j,t} - (\mu_{0t} + b_{v_{j,t}}))\right]\right\}

Note:

$$(\mu_{0t} - M_0)^2 = \mu_{0t}^2 - 2\mu_{0t}M_0 + M_0^2 \quad \text{(B.33)}$$

$$(u_{0t} - \mu_{0t})^2 = u_{0t}^2 - 2u_{0t}\mu_{0t} + \mu_{0t}^2 \quad \text{(B.34)}$$

$$(u_{0t} - \mu_{0t})(v_{0t} - \mu_{0t}) = u_{0t}v_{0t} - u_{0t}v_{0t} - \mu_{0t}v_{0t} + \mu_{0t}v_{0t} \quad \text{(B.35)}$$

$$(u_{j,t} - (\mu_{0t} + b_{u_{j,t}}))^2 = u_{j,t}^2 - 2u_{j,t}(\mu_{0t} + b_{u_{j,t}}) + (\mu_{0t} + b_{u_{j,t}})^2 = u_{j,t}^2 - 2u_{j,t}\mu_{0t} - 2u_{j,t}b_{u_{j,t}} + \mu_{0t}^2 + 2\mu_{0t}b_{u_{j,t}} + b_{u_{j,t}}^2 \quad \text{(B.36)}$$

$$(u_{j,t} - (\mu_{0t} + b_{u_{j,t}}))(v_{j,t} - (\mu_{0t} + b_{v_{j,t}})) = u_{j,t}v_{j,t} - u_{j,t}(v_{0t} + b_{v_{j,t}}) - v_{j,t}(\mu_{0t} + b_{u_{j,t}}) + (\mu_{0t} + b_{u_{j,t}})(v_{0t} + b_{v_{j,t}}) = u_{j,t}v_{j,t} - u_{j,t}v_{0t} - u_{j,t}b_{v_{j,t}} - v_{j,t}\mu_{0t} - v_{j,t}b_{u_{j,t}} + \mu_{0t}v_{0t} + \mu_{0t}b_{v_{j,t}} + b_{o_{j,t}v_{0t}} + b_{u_{j,t}}b_{v_{j,t}} \quad \text{(B.37)}$$
\[ p(\mu \mid \nu) \propto \exp \left\{ -\frac{\lambda_\mu}{2} \left( \mu_0^2 - 2\mu_0 \mu + \mu^2 \right) \right\} \frac{1}{2\lambda_\nu(1 - \rho_\nu^2)} \left( -2\mu_0 \mu_\nu + \mu_0^2 - 2\rho_\nu (\mu_\nu \nu - \mu_\nu \nu) \right) \]

\[- \frac{1}{2\lambda_\nu(1 - \rho_\nu^2)} \left( -2\mu_0 \mu_\nu + \mu_0^2 + 2\rho_\nu \mu_\nu - 2\rho_\nu \mu_\nu \nu - 2\rho_\nu \mu_\nu \nu \right) \}

\[ \exp \left\{ -\frac{\lambda_\mu}{2} \mu_0^2 + \lambda_\mu \mu_0 \mu + \frac{u_0 \mu_0 \nu_0 + \nu_0 \mu_0}{\lambda_0(1 - \rho_\nu^2)} \right\} \frac{1}{2\lambda_\nu(1 - \rho_\nu^2)} \mu_0 \frac{1}{\lambda_\nu(1 - \rho_\nu^2)} \sum_{j=1}^{M} \frac{1}{\lambda_j} \left( -\frac{\rho_\nu}{\lambda_j} \sum_{j=1}^{M} \frac{b_{j,t}}{\lambda_j} - \frac{\rho_\nu}{\lambda_j} \sum_{j=1}^{M} \frac{v_{j,t}}{\lambda_j} + \frac{\rho_\nu}{\lambda_j} \sum_{j=1}^{M} \frac{b_{j,t}}{\lambda_j} \right) \mu_0 \}

\[ \exp \left\{ -\left( \left( \frac{\mu_0}{2} + \frac{1}{2\lambda(1 - \rho_\nu^2)} \right) \mu_0 + \left( \frac{\lambda_\mu + \frac{u_0 \nu_0 + \nu_0 \mu_0}{\lambda_0(1 - \rho_\nu^2)} \right) \mu_0 \right) \frac{1}{\lambda_\nu(1 - \rho_\nu^2)} \right\} \frac{1}{\lambda_\nu(1 - \rho_\nu^2)} \sum_{j=1}^{M} \frac{1}{\lambda_j} \frac{1}{\lambda_j} \frac{1}{\lambda_j} \frac{1}{\lambda_j} \]

Upon completion of the square, it can be seen that the conditional posterior distribution of \( \mu_0 \) is proportional to a Normal distribution with mean:

\[ \lambda_\mu_0 \mu + \frac{u_0 \nu_0 + \nu_0 \mu_0}{\lambda_0(1 - \rho_\nu^2)} + \frac{\sum_{j=1}^{M} \frac{b_{j,t}}{\lambda_j} + \frac{\sum_{j=1}^{M} \frac{v_{j,t}}{\lambda_j}}{\lambda_j} + \frac{\rho_\nu}{\lambda_j} \sum_{j=1}^{M} \frac{b_{j,t}}{\lambda_j}}{\lambda_j} \]

\[ \lambda_\mu + \frac{1}{\lambda_0(1 - \rho_\nu^2)} + \frac{\sum_{j=1}^{M} \frac{1}{\lambda_j}}{\lambda_j} \]

and variance:

\[ \left( \lambda_\mu + \frac{1}{\lambda_0(1 - \rho_\nu^2)} + \frac{\sum_{j=1}^{M} \frac{1}{\lambda_j}}{\lambda_j} \right)^{-1} \].

(B.38)

10. The full conditional posterior of \( \nu_0 \) is given as:

\[ p(\nu_0 \mid \mu) \propto p(\nu_0 \mid \mu) f(u_0, v_0 \mid \mu_0, \nu_0, \xi_0) f(u_{1:1}, ..., u_{M}, v_{1:1}, ..., v_{M} \mid \mu_0, \nu_0, b_{u_1:t}, ..., b_{u_{M},t}, b_{v_1:t}, ..., b_{v_{M},t}, \lambda_1, \ldots) \]

\[ \times \left( \frac{\lambda_\nu}{2\pi} \right)^{1/2} \exp \left\{ -\frac{\lambda_\nu}{2} \left( \nu_0 - \mu_0 \right)^2 \right\} \frac{1}{2\lambda_\nu(1 - \rho^2)^{1/2}} \exp \left\{ -\frac{1}{2\lambda_\nu(1 - \rho^2)} \left( u_0 - \mu_0 \right)^2 \right\} \]

\[ + \left( \nu_0 - \nu_0 \right)^2 - 2\rho_\nu (u_0 - \mu_0) (v_0 - \nu_0) \}

\[ \exp \left\{ -\frac{\lambda_\nu}{2} \left( v_0 - \mu_0 \right)^2 - \frac{1}{2\lambda_\nu(1 - \rho^2)} \right\} \left( v_0 - \nu_0 \right)^2 - 2\rho_\nu (v_0 - \mu_0) (v_0 - \nu_0) \}

\[ \exp \left\{ -\frac{\lambda_\nu}{2} \left( v_0 - \mu_0 \right)^2 - \frac{1}{2\lambda_\nu(1 - \rho^2)} \right\} \left( v_0 - \nu_0 \right)^2 - 2\rho_\nu (v_0 - \mu_0) (v_0 - \nu_0) \}

\[ \left( \nu_0 - \nu_0 \right)^2 - 2\rho_\nu (u_0 - \mu_0) (v_0 - \nu_0) \}

\[ \exp \left\{ -\frac{\lambda_\nu}{2} \left( v_0 - \mu_0 \right)^2 - \frac{1}{2\lambda_\nu(1 - \rho^2)} \right\} \left( v_0 - \nu_0 \right)^2 - 2\rho_\nu (v_0 - \mu_0) (v_0 - \nu_0) \}

\[ \sum_{j=1}^{M} \left( v_0 - \nu_0 \right)^2 - 2\rho_\nu (v_0 - \mu_0) (v_0 - \nu_0) \}

\[ \left( \nu_0 - \nu_0 \right)^2 - 2\rho_\nu (v_0 - \mu_0) (v_0 - \nu_0) \}

\[ \exp \left\{ -\frac{\lambda_\nu}{2} \left( v_0 - \mu_0 \right)^2 - \frac{1}{2\lambda_\nu(1 - \rho^2)} \right\} \left( v_0 - \nu_0 \right)^2 - 2\rho_\nu (v_0 - \mu_0) (v_0 - \nu_0) \}

\[ \sum_{j=1}^{M} \left( v_0 - \nu_0 \right)^2 - 2\rho_\nu (v_0 - \mu_0) (v_0 - \nu_0) \}

\[ \left( \nu_0 - \nu_0 \right)^2 - 2\rho_\nu (v_0 - \mu_0) (v_0 - \nu_0) \} \]
Note:

\[(v_{0t} - M_\nu)^2 = v_{0t}^2 - 2v_{0t}M_\nu + M_\nu^2 \]  \hspace{1cm} (B.40)

\[(v_{0t} - v_{0t})^2 = v_{0t}^2 - 2v_{0t}v_{0t} + v_{0t}^2 \]  \hspace{1cm} (B.41)

\[(v_{j,t} - (v_{0t} + b_{v,j,t}))^2 = v_{j,t}^2 - 2v_{j,t}(v_{0t} + b_{v,j,t}) + (v_{0t} + b_{v,j,t})^2 \]

\[= v_{j,t}^2 - 2v_{j,t}v_{0t} - 2v_{j,t}b_{v,j,t} + v_{0t}^2 + 2v_{0t}b_{v,j,t} + b_{v,j,t}^2 \]  \hspace{1cm} (B.42)

\[p(v_{0t}) \propto \exp \left\{ -\frac{\lambda_0}{2} \left( v_{0t}^2 - 2v_{0t}M_\nu \right) - \frac{1}{2\lambda_0(1 - \rho_0^2)} \left( -2v_{0t}v_{0t} + v_{0t}^2 - 2\rho_0(-v_{0t}v_{0t} + \mu_0v_{0t}) \right) \right. \]

\[- \sum_{j=1}^{M} \frac{1}{2\lambda_j(1 - \rho_j^2)} \left( -2v_{j,t}v_{0t} + v_{j,t}^2 - 2v_{0t}b_{v,j,t} - 2\rho_j(-v_{j,t}v_{0t} + \mu_jv_{0t} + b_{v,j,t}v_{0t}) \right) \right\} \]

\[\propto \exp \left\{ -\frac{\lambda_0}{2} v_{0t}^2 + \lambda_0v_{0t}M_\nu + \frac{v_{0t}v_{0t}}{\lambda_0(1 - \rho_0^2)} - \frac{v_{0t}^2}{2\lambda_0(1 - \rho_0^2)} - \frac{\rho_0v_{0t}v_{0t}}{\lambda_0(1 - \rho_0^2)} + \frac{\rho_0\mu_0v_{0t}}{\lambda_0(1 - \rho_0^2)} + \frac{v_{0t}}{(1 - \rho_0^2)} \sum_{j=1}^{M} \frac{v_{j,t}}{\lambda_j} \right. \]

\[- \frac{2(1 - \rho_0^2)}{\lambda_0(1 - \rho_0^2)} \sum_{j=1}^{M} \frac{1}{\lambda_j} v_{j,t} - \frac{2v_{0t}v_{0t}}{(1 - \rho_0^2)} \sum_{j=1}^{M} \frac{b_{v,j,t}}{\lambda_j} - \frac{v_{0t}^2}{(1 - \rho_0^2)} \sum_{j=1}^{M} \frac{u_{j,t}}{\lambda_j} + \frac{\rho_0v_{0t}v_{0t}}{(1 - \rho_0^2)} \sum_{j=1}^{M} \frac{1}{\lambda_j} + \frac{\rho_0\mu_0v_{0t}}{(1 - \rho_0^2)} \sum_{j=1}^{M} \frac{b_{v,j,t}}{\lambda_j} \right\} \]

\[\propto \exp \left\{ -\left( \frac{\lambda_0}{2} + \frac{1}{2\lambda_0(1 - \rho_0^2)} + \frac{\sum_{j=1}^{M} \frac{v_{j,t}}{\lambda_j}}{\lambda_0(1 - \rho_0^2)} \right) \right. \]

\[+ \frac{\sum_{j=1}^{M} \frac{v_{j,t}}{\lambda_j}}{(1 - \rho_0^2)} - \sum_{j=1}^{M} \frac{v_{j,t}}{\lambda_j} - \frac{\rho_0}{\lambda_0} \sum_{j=1}^{M} \frac{v_{j,t}}{\lambda_j} + \frac{\rho_0\mu_0}{\lambda_0} \sum_{j=1}^{M} \frac{1}{\lambda_j} + \frac{\rho_0}{\lambda_0} \sum_{j=1}^{M} \frac{b_{v,j,t}}{\lambda_j} \right\} \]

\[\lambda_\nu M_\nu + \frac{v_{0t} - \rho_0u_{0t} + \rho_0\mu_{0t}}{\lambda_0(1 - \rho_0^2)} + \frac{\sum_{j=1}^{M} \frac{v_{j,t}}{\lambda_j} - \rho_0 \sum_{j=1}^{M} \frac{v_{j,t}}{\lambda_j} + \rho_0 \mu_{0t} \sum_{j=1}^{M} \frac{1}{\lambda_j} + \rho_0 \sum_{j=1}^{M} \frac{b_{v,j,t}}{\lambda_j}}{(1 - \rho_0^2)} \]

\[\lambda_\nu + \frac{1}{\lambda_0(1 - \rho_0^2)} + \frac{\sum_{j=1}^{M} \frac{1}{\lambda_j}}{(1 - \rho_0^2)} \]

\hspace{1cm} (B.43)
and variance:

\[
\left( \lambda_0 + \frac{1}{\lambda_0 (1 - \rho_0^2)} + \sum_{j=1}^{M} \frac{1}{\lambda_j} \right)^{-1}.
\]  

(B.44)

11. The full conditional posterior of \( b_{uj,t} \) is given as:

\[
p(b_{uj,t}) \propto h_{j,t}(b_{uj,t} | x_{1j,t}, x_{2j,t}, \alpha_{oj}, \alpha_{1j}, \alpha_{2j}, \gamma) f_j, (u_{j,t}, v_{j,t} | \mu_{ot}, \nu_{ot}, b_{uj,t}, b_{vj,t}, \rho_t, \lambda_j)
\]

\[
\propto \left( \frac{1}{2\pi} \right)^{1/2} \exp \left\{ -\frac{1}{2 \pi \lambda_j (1 - \rho_t^2)^{1/2}} \right\} \exp \left\{ \frac{1}{2\lambda_j (1 - \rho_t^2)} \right\} 
\]

\[
\left[ (u_{j,t} - (\mu_{ot} + b_{uj,t}))^2 - 2\rho_t (u_{j,t} - (\mu_{ot} + b_{uj,t})) (v_{j,t} - (\nu_{ot} + b_{vj,t})) + (v_{j,t} - (\nu_{ot} + b_{vj,t}))^2 \right] \right\}
\]

\[
\propto \exp \left\{ -\frac{\tau}{2} \left( b_{uj,t}^2 - 2b_{uj,t} (\alpha_{oj} + \alpha_{1j} x_{1j,t} + \alpha_{2j} x_{2j,t}) \right) - \frac{1}{2\lambda_j (1 - \rho_t^2)} \right\} 
\]

\[
-2\rho_t (u_{j,t} - (\mu_{ot} + b_{uj,t})) (v_{j,t} - (\nu_{ot} + b_{vj,t})) \right\}
\]

Note:

\[
(u_{j,t} - (\mu_{ot} + b_{uj,t}))^2 = u_{j,t}^2 - 2u_{j,t} (\mu_{ot} + b_{uj,t}) + (\mu_{ot} + b_{uj,t})^2 
\]

\[
= u_{j,t}^2 - 2u_{j,t} \mu_{ot} - 2u_{j,t} b_{uj,t} + \mu_{ot}^2 + 2\mu_{ot} b_{uj,t} + b_{uj,t}^2 
\]

(B.45)

\[
(u_{j,t} - (\mu_{ot} + b_{uj,t})) (v_{j,t} - (\nu_{ot} + b_{vj,t})) 
\]

\[
= u_{j,t} v_{j,t} - u_{j,t} \nu_{ot} - u_{j,t} b_{uj,t} - v_{j,t} (\mu_{ot} + b_{uj,t}) - v_{j,t} (\nu_{ot} + b_{vj,t}) + (\mu_{ot} + b_{uj,t}) (\nu_{ot} + b_{vj,t}) 
\]

\[
= u_{j,t} v_{j,t} - u_{j,t} \nu_{ot} - u_{j,t} b_{uj,t} - v_{j,t} \mu_{ot} - v_{j,t} b_{uj,t} + \mu_{ot} \nu_{ot} + \mu_{ot} b_{uj,t} + \nu_{ot} b_{uj,t} + b_{uj,t} b_{vj,t} 
\]

(B.46)

\[
p(b_{uj,t}) \propto \exp \left\{ -\frac{\tau}{2} b_{uj,t}^2 + \tau b_{uj,t} (\alpha_{oj} + \alpha_{1j} x_{1j,t} + \alpha_{2j} x_{2j,t}) \right\} 
\]

\[
- \frac{1}{2\lambda_j (1 - \rho_t^2)} \left[ -2u_{j,t} \mu_{ot} - 2u_{j,t} b_{uj,t} + \mu_{ot}^2 + 2\mu_{ot} b_{uj,t} + b_{uj,t}^2 
\]

\[
-2\rho_t (-v_{j,t} \mu_{ot} - v_{j,t} b_{uj,t} + b_{uj,t} \nu_{ot}) \right\} \right\}
\]

\[
\propto \exp \left\{ -\frac{\tau}{2} b_{uj,t}^2 + \tau b_{uj,t} (\alpha_{oj} + \alpha_{1j} x_{1j,t} + \alpha_{2j} x_{2j,t}) \right\} 
\]

\[
+ \left\{ \tau (\alpha_{oj} + \alpha_{1j} x_{1j,t} + \alpha_{2j} x_{2j,t}) + \frac{u_{j,t} - \mu_{ot} - \rho_t v_{j,t} + \rho_t \nu_{ot} + \rho_t b_{uj,t}}{\lambda_j (1 - \rho_t^2)} \right\} b_{uj,t} \right\}
\]

Upon completion of the square, the conditional posterior distribution of \( b_{uj,t} \) is proportional to a Normal distribution with mean:

\[
\frac{\tau (\alpha_{oj} + \alpha_{1j} x_{1j,t} + \alpha_{2j} x_{2j,t}) + \frac{u_{j,t} - \mu_{ot} - \rho_t v_{j,t} + \rho_t \nu_{ot} + \rho_t b_{uj,t}}{\lambda_j (1 - \rho_t^2)}}{\tau + \frac{1}{\lambda_j (1 - \rho_t^2)}},
\]

(B.47)
12. The full conditional posterior of $b_{v,j,t}$ is given as:

$$p(b_{v,j,t}) \propto h_j(b_{v,j,t} \mid x_{1,j,t}, x_{2,j,t}, \eta_{0j}, \eta_{1j}, \eta_{2j}, \gamma) f_j(u_{j,t} \mid \mu_0t, \nu_0t, b_{u,j,t}, b_{v,j,t}, \rho_t, \lambda_j)$$

$$\propto \left(\frac{\gamma}{2\pi}\right)^{1/2} \exp\left\{-\frac{\gamma}{2} (b_{v,j,t} - (\eta_{0j} + \eta_{1j} x_{1,j,t} + \eta_{2j} x_{2,j,t}))^2\right\} \frac{1}{2\lambda_j(1 - \rho_t^2)^{1/2}}$$

$$\exp\left\{-\frac{1}{2\lambda_j(1 - \rho_t^2)} \bigg\{ (u_{j,t} - (\mu_0t + b_{u,j,t}))^2 - 2\rho_t (u_{j,t} - (\mu_0t + b_{u,j,t})) (v_{j,t} - (\nu_0t + b_{v,j,t})) + (v_{j,t} - (\nu_0t + b_{v,j,t}))^2 \bigg\}\right\}$$

$$\propto \exp\left\{-\frac{\gamma}{2} \left( b_{v,j,t}^2 - 2b_{v,j,t} (\eta_{0j} + \eta_{1j} x_{1,j,t} + \eta_{2j} x_{2,j,t}) \right) - \frac{1}{2\lambda_j(1 - \rho_t^2)} \left( v_{j,t} - (\nu_0t + b_{v,j,t}) \right)^2 \right\}$$

Note:

$$(v_{j,t} - (\nu_0t + b_{v,j,t}))^2 = v_{j,t}^2 - 2v_{j,t} \nu_0t - 2v_{j,t} b_{v,j,t} + \nu_0^2 + 2\nu_0t b_{v,j,t} + b_{v,j,t}^2$$

(B.49)

(B.50)

Upon completion of the square, the conditional posterior distribution of $b_{v,j,t}$ is proportional to a Normal distribution with mean:

$$\gamma (\eta_{0j} + \eta_{1j} x_{1,j,t} + \eta_{2j} x_{2,j,t}) + \frac{v_{j,t} - \nu_0t - \rho_t u_{j,t} + \rho_t \mu_0t + \rho_t b_{u,j,t}}{\lambda_j(1 - \rho_t^2)} b_{v,j,t}$$

(B.51)

and variance:

$$\left(\gamma + \frac{1}{\lambda_j(1 - \rho_t^2)}\right)^{-1}.$$  

(B.52)
13. The full conditional posterior of \( \lambda_j \) is given as:

\[
p(\lambda_j) \propto g_j(\lambda_j \mid c, d) f_j(u_{j,1}, \ldots, u_{j,T}, v_{j,1}, \ldots, v_{j,T} \mid \mu_0, \ldots, \mu_{0T}, \nu_0, \ldots, \nu_{0T}, b_{uj,1}, \ldots, b_{uj,T}, d_{uj,1}, \ldots, d_{uj,T}, \lambda_j, \rho_t)
\]

\[
\propto \frac{d^c}{\Gamma(c)} \lambda_j^{c-1} \exp \left\{ -\frac{d}{\lambda_j} \right\} \left( \prod_{t=1}^{T} \frac{T}{(1 - \rho_t^2)^{1/2}} \right) \exp \left\{ \sum_{t=1}^{T} -\frac{1}{2\lambda_j(1 - \rho_t^2)} \left[ (u_{j,t} - (\mu_0 + b_{uj,t}))^2 - 2\rho_t (u_{j,t} - (\mu_0 + b_{uj,t}))(v_{j,t} - (\nu_0 + b_{uj,t})) + (v_{j,t} - (\nu_0 + b_{uj,t}))^2 \right] \right\}
\]

\[
\propto \lambda_j^{c-1} \exp \left\{ -\frac{d}{\lambda_j} \right\} \left( \frac{1}{\lambda_j} \right)^T \exp \left\{ \sum_{t=1}^{T} -\frac{1}{2\lambda_j(1 - \rho_t^2)} \left[ (u_{j,t} - (\mu_0 + b_{uj,t}))^2 - 2\rho_t (u_{j,t} - (\mu_0 + b_{uj,t}))(v_{j,t} - (\nu_0 + b_{uj,t})) + (v_{j,t} - (\nu_0 + b_{uj,t}))^2 \right] \right\}
\]

\[
\propto \lambda_j^{c-T-1} \exp \left\{ -\frac{d + \frac{1}{2} \sum_{t=1}^{T} \frac{1}{1 - \rho_t^2}}{2\lambda_j} \left[ (u_{j,t} - (\mu_0 + b_{uj,t}))^2 - 2\rho_t (u_{j,t} - (\mu_0 + b_{uj,t}))(v_{j,t} - (\nu_0 + b_{uj,t})) + (v_{j,t} - (\nu_0 + b_{uj,t}))^2 \right] \right\}
\]

Thus, the full conditional posterior distribution of \( \lambda_j \) is proportional to:

\[
IG(c + T,
\]

\[
d + \frac{1}{2} \sum_{t=1}^{T} \frac{1}{1 - \rho_t^2} \left[ (u_{j,t} - (\mu_0 + b_{uj,t}))^2 - 2\rho_t (u_{j,t} - (\mu_0 + b_{uj,t}))(v_{j,t} - (\nu_0 + b_{uj,t})) + (v_{j,t} - (\nu_0 + b_{uj,t}))^2 \right]
\]

(B.53)

14. The full conditional posterior distribution of \( \kappa \) is given as:

\[
p(\kappa) \propto \Pi(\kappa) f \{ (u_{j,t}, v_{j,t} : t = 1, \ldots, T; j = 1, \ldots, M) \mid \mu_0, \ldots, \mu_{0T}, \nu_0, \ldots, \nu_{0T}, \}
\]

\[
\{ b_{uj,1}, b_{uj,t} : t = 1, \ldots, T; j = 1, \ldots, M \}, \lambda_1, \ldots, \lambda_M, \kappa, \alpha, \beta, \theta_{0T}
\]

\[
\propto \frac{1}{\kappa^M} \prod_{j=1}^{M} \prod_{t=1}^{T} \left( \frac{1}{2\pi \lambda_j(1 - [\kappa \sin(2\pi \theta_{0T} \alpha - \beta)]^2)^{1/2}} \right) \exp \left\{ \sum_{t=1}^{T} \sum_{j=1}^{M} -\frac{1}{2\lambda_j(1 - [\kappa \sin(2\pi \theta_{0T} \alpha - \beta)]^2)} \left[ (u_{j,t} - (\mu_0 + b_{uj,t}))^2 - 2(\kappa \sin(2\pi \theta_{0T} \alpha - \beta))(u_{j,t} - (\mu_0 + b_{uj,t}))(v_{j,t} - (\nu_0 + b_{uj,t})) + (v_{j,t} - (\nu_0 + b_{uj,t}))^2 \right] \right\}
\]

(B.54)

The full conditional posterior of \( \kappa \) cannot be rewritten into the form of a well-known distribution. Therefore, a rejection sampling algorithm will need to be implemented.

15. The full conditional posterior distribution of \( \alpha \) is given as:

\[
p(\alpha) \propto \Pi(\alpha) f \{ (u_{j,t}, v_{j,t} : t = 1, \ldots, T; j = 1, \ldots, M) \mid \mu_0, \ldots, \mu_{0T}, \nu_0, \ldots, \nu_{0T}, \}
\]

\[
\{ b_{uj,1}, b_{uj,t} : t = 1, \ldots, T; j = 1, \ldots, M \}, \lambda_1, \ldots, \lambda_M, \kappa, \alpha, \beta, \theta_{0T}
\]

\[
\propto \frac{1}{\alpha^M} \prod_{j=1}^{M} \prod_{t=1}^{T} \left( \frac{1}{2\pi \lambda_j(1 - [\kappa \sin(2\pi \theta_{0T} \alpha - \beta)]^2)^{1/2}} \right) \exp \left\{ \sum_{t=1}^{T} \sum_{j=1}^{M} -\frac{1}{2\lambda_j(1 - [\kappa \sin(2\pi \theta_{0T} \alpha - \beta)]^2)} \left[ (u_{j,t} - (\mu_0 + b_{uj,t}))^2 - 2(\kappa \sin(2\pi \theta_{0T} \alpha - \beta))(u_{j,t} - (\mu_0 + b_{uj,t}))(v_{j,t} - (\nu_0 + b_{uj,t})) + (v_{j,t} - (\nu_0 + b_{uj,t}))^2 \right] \right\}
\]

(B.55)
The full conditional posterior of $\alpha$ cannot be rewritten into the form of a well-known distribution. Therefore, a rejection sampling algorithm will need to be implemented.

The full conditional posterior distribution of $\beta$ is given as:

$$p(\beta) \propto \Pi(\beta)f(\{u_{j,t}, v_{j,t} : t = 1, \ldots, T; j = 1, \ldots, M\} | \mu_0, \ldots, \mu_0^T, \nu_0, \ldots, \nu_0^T, \{b_{uj,t}, b_{vj,t} : t = 1, \ldots, T; j = 1, \ldots, M\}, \lambda_1, \ldots, \lambda_M, \kappa, \alpha, \beta)$$

$$\propto \frac{1}{q-p} \prod_{j=1}^{M} \prod_{t=1}^{T} \left( \frac{1}{2\pi\lambda_j(1 - |\kappa \sin(2\pi\theta_0 \alpha - \beta)|^2)^{1/2}} \right) \exp \left\{ \sum_{t=1}^{T} \sum_{j=1}^{M} \frac{1}{2\lambda_j(1 - |\kappa \sin(2\pi\theta_0 \alpha - \beta)|^2)} \left[ (u_{j,t} - (\mu_0 + b_{uj,t}))^2 - 2(\kappa \sin(2\pi\theta_0 \alpha - \beta))(u_{j,t} - (\mu_0 + b_{uj,t}))(v_{j,t} - (\nu_0 + b_{vj,t})) + (v_{j,t} - (\nu_0 + b_{vj,t}))^2 \right] \right\}$$

(B.56)

The full conditional posterior of $\beta$ cannot be rewritten into the form of a well-known distribution. Therefore, a rejection sampling algorithm will need to be implemented.


