Advanced Statistical Modeling Constructs for Climate Extremes

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Lindsey Rose Dietz

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Snigdhansu Chatterjee, Adviser

May 2016

Acknowledgements

Thanks to my adviser, Ansu, for the hours of help and inspiration.

Thanks to my committee, Sandy, Scott, and Galin, for their essential contributions to my education.

Thanks to my colleagues and friends, especially Megan, for competing, collaborating, and commiserating.

Thanks to my parents, Jayne and John, and my sister Stacey. Their support throughout this adventure in academia will not be forgotten.

Finally, thanks to Todd for being my partner in statistics and life.

Dedication

For Jocelyn

You may encounter many defeats, but you must not be defeated. In fact, it may be necessary to encounter the defeats, so you can know who you are, what you can rise from, how you can still come out of it.

– Maya Angelou

Contents

List of Tables

List of Figures

Acronyms

ΔTT tropospheric temperature difference. 6, 128, 141, 143, 146, 153, 159

ACF autocorrelation function. 86, 112–114

AIC Akaike information criterion. 43, 56, 58, 64, 75, 96, 97, 99, 222, 223

amix asymmetric mixed. 220–223

AMO Atlantic Multidecadal Oscillation. 90–94, 96, 98, 99, 106, 110, 116

AR1 auto-regressive lag 1. 154, 166, 181

balog bivariate asymmetric logistic. 218, 222, 223

baneglog bivariate asymmetric negative logistic. 219, 222, 223

bbilog bivariate bilogistic. 218, 222, 223

BIC Bayesian information criterion. 43, 56, 58, 75, 96, 97, 99

blog bivariate logistic. 217, 222, 223

bnegbilog bivariate negative bilogistic. 220, 223

bneglog bivariate negative logistic. 219, 222, 223

BUGS Bayesian inference Using Gibbs Sampling. 86, 104, 108, 133

BVEVD Bivariate Extreme Value Distribution. 66, 68–71, 77, 217, 221, 223

CI current intensity. 14, 16

CPC Climate Prediction Center. 90, 91, 141

CPHC Central Pacific Hurricane Center. 16

ct Coles-Tawn. 221–223

DMI Indian Dipole Mode Index. 141, 146

DVT Dvorak Technique. 14–18, 20

ENSO the El Niño-Southern Oscillation. 6, 127, 141, 146

ERSST extended reconstructed sea surface temperature. 91

ESRL Earth System Research Laboratory. 90, 91

ESS effective sample size. 109, 110

GCS Generalized Chi-Square. 163

GEE generalized estimating equations. 154, 157

GEV Generalized Extreme Value. 3, 12, 54, 55, 57–63, 65, 66, 82, 217

GLM generalized linear model. 95, 96, 144, 146, 154, 157

GLMM generalized linear mixed model. 5–7, 127–132, 138–140, 142–145, 149, 153, 155–157, 163, 235

hr Husler-Reiss. 221–223

HURDAT2 revised Atlantic hurricane database. 4, 51, 88

IBTrACs International Best Track Archive for Climate Stewardship. 2, 12, 30–32, 36, 51, 72, 88

ICAT insurance catastrophe database. 72

ICOADS International Comprehensive Ocean-Atmosphere Data Set. 91

IOD Indian Ocean Dipole. 6, 127, 141

JAGS Just Another Gibbs Sampler. 108–110

JAMSTEC Japan Agency for Marine-Earth Science and Technology. 141

JMA Japan Meteorological Agency. 16

JTWC Joint Typhoon Warning Center. 20

kt knot. 16, 20, 22, 30, 36, 37, 39, 51, 52, 57

KZ Knaff-Zehr. 21, 26, 29, 51–53

LNMM logit-normal mixed model. 128, 133, 143, 146–148, 150–152, 154, 155, 163

LRT likelihood ratio test. 55, 56, 58, 66, 73–75, 96–99, 101, 144–146, 169, 221, 223

MAPE mean absolute prediction error. 43, 44, 58

MaxWS maximum wind speed. 2, 3, 11–15, 20, 21, 31, 33, 35, 36, 52, 54–56, 64, 66–70, 72–76, 78, 82, 217

mb millibars. 15, 20, 35, 36, 141, 142, 159, 195

MCMC Markov chain Monte Carlo. 4, 86, 87, 104, 109, 110, 112–115, 123, 132, 133, 229

MinCP minimum central pressure. 2, 3, 11–16, 20–22, 31, 33, 35, 36, 43, 52, 54, 56, 64, 66–70, 72–75, 78, 82, 217

MLE maximum likelihood estimate. 54, 56, 64, 66, 73, 108, 229, 230

MooN M-out-of-N. 3, 12, 33–35, 44–46, 56, 58–61, 67, 71, 74, 77

MSIM method of simulated moments. 132, 134–136, 138, 149

MSPE mean squared prediction error. 43, 44, 58

NAO North Atlantic Oscillation. 84, 85, 90–94, 96, 98

NCDC National Climatic Data Center. 140

NCEP National Centers for Environmental Prediction. 90, 91, 141

NEF natural exponential family. 129, 235

NHC National Hurricane Center. 16

NOAA National Oceanic and Atmospheric Administration. 30, 31, 89–91, 140

OLS ordinary least squares. 36

pdf probability density function. 54, 104, 129

PIRLS Penalized iteratively reweighted least squares. 131, 137, 225, 227

PL pseudo-likelihood. 139, 155

pmf probability mass function. 95, 96, 129

PQL Penalized quasi-likelihood. 131, 132, 142, 225, 226

PSD Physical Sciences Division. 91

ROCI radius of the outer closed isobar. 14, 18

RSMC Regional Specialized Meteorological Center. 13, 16, 20, 30

SAI synoptic activity index. 127

SE standard error. 52, 86, 109, 110, 112

SOI Southern Oscillation Index. 84, 85, 90–94, 96, 98

SSN sunspot number. 84, 90–94, 96, 98

SST sea surface temperature. 84, 85, 90–93, 141

STI Space-Time Index. 7, 8, 165, 166, 170–173, 179, 183–189, 192, 196–199

STrIde spatio-temporal identification. 7, 8, 165, 166, 171, 176, 177, 180–189, 192, 196–199

TCWC Tropical Cyclone Warning Center. 30

TPC Tropical Prediction Center. 16

WLS weighted least squares. 36, 37, 39–41, 43–53, 56

WMO World Meteorological Organization. 13

Part I

Advanced Statistical Analysis in Climate Data

Modeling any climate phenomena should consider relevant physics, however, it should not ignore approximations commonly made to simplify these constructs. The validity of these approximations must be verified in the presence of data. Physics provide a starting point for our research, but there are gaps in the physics knowledge relating to (i) use of extreme values, (ii) lack of random effects and latent variables, and (iii) spatial-temporal dependence. This thesis focuses on addressing the statistical challenges in properly modeling (i), (ii) and (iii) using climate data. Specifically, Part II focuses on tropical cyclones and Part III will address Indian monsoon precipitation.

Part II addresses the statistical study of intensity, frequency, and damages in tropical cyclones. In Chapter 1, a physical approximation relating maximum wind speed (MaxWS) and minimum central pressure (MinCP) is discussed and investigated at length in a global context. The MaxWS - MinCP relationship in tropical cyclones has been studied for several decades in the atmospheric sciences and related fields. However, there is now adequate data to statistically evaluate different approaches used in the literature to study tropical cyclone physics, and propose alternative statistical models that may be more consistent with the features and properties of the data while retaining physical interpretability.

Previous methodologies focused on combinations of qualitative methods (Dvorak, 1975, 1984; Koba et al., 1990), physical approximations (Takahashi, 1952; Kraft, 1961; Atkinson and Holliday, 1977), and multiple regression techniques (Knaff and Zehr, 2007; Holland, 2008). Many of these analyses suffer from untenable assumptions, opaque methodologies, and lack of uncertainty estimates. We test the MaxWS-MinCP relationship using aspects of the previous methods via the International Best Track Archive for Climate Stewardship (IBTrACs) data for global tropical cyclones. Our analyses begin with linear models based on transformations of physical approximations. Results indicate typical linear model as-

sumptions are violated especially constant variance of residuals. As a result, a weighted linear model is implemented, however, normality of residuals remains untenable. Uncertainty estimates are produced using asymptotic standard errors and a parametric M-out-of-N (MooN) bootstrap and subsampling; asymptotic errors are much smaller than those of the bootstrap indicating the possible inaccuracy of the asymptotic uncertainty estimates.

Therefore, we turn to a more statistically valid methodology within this context: extreme value modeling. We fit conditional models for log-transformed MaxWS utilizing a well studied physical approximation. The conditional model estimates a non-stationary Generalized Extreme Value (GEV) distribution for log(MaxWS) with a location parameter as a function of MinCP. Uncertainty estimates produced using asymptotic standard errors and the parametric MooN techniques are much more in line with each other than in the linear models.

Finally, we extend the study of MaxWS and MinCP to a bivariate realm. MaxWS and MinCP require joint study because data uncertainty may lead to potential changes in the bivariate distribution over time that a conditional model would be unable to capture. We examined several bivariate extreme value parameterizations before choosing a bivariate bilogistic extreme value model. Using this framework, we then study the joint relationship of damages and intensity in the Atlantic basin where intensity is modeled using the MaxWS-MinCP relationship. Simulation results using the estimated model provide a reasonable way of predicting possible bivariate intensity-damage distributions for individual tropical cyclones.

Chapter 1 highlights the importance of understanding the intensity mechanisms in tropical cyclones because intensity is closely related to potential damage. In particular, Atlantic tropical cyclones (more commonly known as hurricanes) can cause billions of dollars in damages and loss of lives annually in the United States.

Chapter 2 furthers the statistical exposition of tropical cyclones by linking their frequency, intensity, and damage capabilities in a joint model.

Prediction and estimation of annual frequency of tropical cyclones has been well studied, and several covariates have been used via regression methods, namely, Poisson regression (Elsner et al., 2001; Elsner, 2003; Elsner and Jagger, 2004, 2005, 2006; Villarini et al., 2010). Investigation of damages from tropical cyclones has also been done in separate studies (Jagger et al., 2008, 2010). Our research differs from these previous modeling efforts by jointly studying annual Atlantic tropical cyclone frequency, landfall, and damages. We propose a Bayesian hierarchical model which addresses intensity and frequency while quantifying probabilities of landfall and subsequent damages. We separate more intense Category 3-5 tropical cyclones and less intense tropical storms up to Category 2 tropical cyclones in the model because of the differences they exhibit in frequency, landfall, and damage.

Data are acquired for the analysis via the revised Atlantic hurricane database (HURDAT2) as well as a unique damage database known as ICAT. Several climatic covariates are also included for study within the frequency portion of the model. We take an empirical Bayes approach to selecting prior distributions for each component of the model. Frequency, landfall, and damages data are carefully explored and reasonable heuristics are developed for prior selection and setting hyperparameters. Inclusion of tropical storms along with tropical cyclones necessitates a modification of the frequency distribution due to overdispersion. Thus, we use a gamma-Poisson mixture in the frequency component of the Bayesian model for low intensity storms to act like a negative binomial distribution.

Markov chain Monte Carlo (MCMC) is used to estimate the posterior distribution for inference. The posterior mean estimates for frequency correspond to 13.4 low intensity and 2.45 high intensity storms per year. Low and high intensity expected landfall proportions were estimated around 11.5% and 42% respectively.

Thus, on average, the U.S. should expect about 1.54 low intensity and 1.03 high intensity landfalling storms per year. Low and high intensity damages indicated lognormal means of 19.17 and 21.84 respectively corresponding to about 212 million and 3.055 billion dollars worth of damage. Overall, the U.S. should be prepared for about 3.267 billion dollars in damage each year.

The usefulness of the model in a predictive sense is tested using the posterior predictive distributions for 2011 and 2012 data. Posterior predictive results indicate the model provides estimates which correspond closely with observed data, and the model could feasibly be employed as a tool for disaster planning and mitigation.

Part III begins with Chapter 3, which focuses on statistical analysis of the summer Indian monsoon precipitation data to provide insight symbiotic with deterministic physics modeling. Explanation of summer Indian monsoon precipitation has been a challenging problem in physics as well as data analysis. Previous statistical analyses regarding precipitation in Indian monsoons have explored two main areas- identifying methodologies of data analysis and covariate selection.

The establishment of appropriate statistical methodology for explanation and prediction of precipitation, while simultaneously capturing underlying variability, is paramount. These methods are used in identification of trends for prediction, however, trends tend to be inconsistent across studies and may relate to linked variability on different temporal and spatial scales. Studies such as (Goswami et al., 2006; Ghosh et al., 2012, 2009; Singh et al., 2014) utilized parametric – regression, extreme value theory, time series methods – and nonparametric statistical techniques, yet failed to make unanimous conclusions regarding the trends in monsoon precipitation. This suggests important properties of the Indian monsoon remain partially misunderstood.

In view of previous results, we propose generalized linear mixed models (GLMMs)

as a potential framework for analysis of Indian monsoon precipitation data. This implies part of the signal is random, and changes from one set of circumstances to another. GLMM may be suitable for capturing local, instantaneous variability arising from cloud or other physical micro-properties.

The second principal focus of literature has been identifying relevant covariates for study of Indian monsoon precipitation. Certain oscillations are commonly useful predictors for precipitation. For instance, the Indian Ocean Dipole (IOD) studied in (Rajeevan et al., 2008) was shown to modulate inter-annual, inter-decadal and long-term trends of extreme rainfall events. Most commonly, the El Niño-Southern Oscillation (ENSO) (Kumar et al., 1999; Li et al., 2007b; Prell and Kutzback, 1992; Turner and Annamalai, 2012) is cited as a driver of the monsoon. Several other climatic predictors of monsoons have been proposed; unfortunately, none have been conclusively attributed for the monsoon rainfall which suggests an intricate relationship between some or all of these factors. Because explicit attribution to covariates may not be possible, GLMM is a logical model for studying Indian monsoon precipitation. It allows underlying randomness to drive observed data in a particular hierarchy while still accounting for hypothesized drivers of rainfall.

The applicability of GLMM in this setting was verified using two studies. Study I (Dietz and Chatterjee, 2014) fitted annual logit-normal mixed models to indicators determined by 3 categories of rainfall: light ($0 < x < 64.4$mm/day), moderate ($64.4 \leq x < 124.4$ mm/day), and extreme (≥ 124.4 mm/day). The study utilized several estimation methods including the method of moments, likelihood approximation, and Bayesian estimation techniques. The outcome of these annual models show maximum temperature and elevation are consistently significant, aligning with the physics of precipitation. Tropospheric temperature difference (ΔTT) is also significant for many of the models. The most meaningful finding is the random

intercept by weather station is non-negligible in many of the models.

Study II (Dietz and Chatterjee, 2015) extends the results of Study I to examine the use of GLMM to other previously studied types of thresholds in precipitation data. Station-defined percentile thresholds were used as in (Krishnamurty et al., 2009) and fixed level thresholds slightly modified from those in Study I were used in (Goswami et al., 2006) to explore trends in monsoon rainfall intensity. Study II focuses on the inclusion of relevant covariates based on Study I and uses both threshold definitions with distinct purposes. Fixed threshold models elicit a physical interpretation across rainfall levels, and percentile-based thresholds aid in understanding local predicted probabilities of threshold exceedances and possible cycles in their occurrence.

As described in Chapter 3, the Indian monsoon precipitation thresholds model is useful for determining risk of threshold exceedances and can be applied in civil planning. However, the model is not implemented in a traditional spatio-temporal modeling context. The logit-normal mixed model specifically accounts for the time component by including repeated measures by station and year, however, it does not explicitly model possible spatial signals in the error structure through additional covariance requirements. Thus, to assess the adequacy of the model structure, in Chapter 4, we turn to spatio-temporal dependence hypothesis testing.

Testing for space and time dependence is a well studied process in certain classes of problems. In testing residuals for spatio-temporal dependence, there is a choice of statistic and testing procedure to be made by the researcher. Our analysis assesses the commonly used Space-Time Index (STI) (Griffith, 1981) against our own novel weighted statistic for spatio-temporal identification (STrIde). We carry out the test via an approximation to permutation testing.

Several simulations comparing STI to STrIde are performed. The two methods are compared directly on the null hypothesis of (a) complete independence. How-

ever, STrIde can also be tested on null hypotheses of (b) temporal dependence only and (c) spatial dependence only. Our study finds that Type I errors for STrIde in hypotheses (a)-(c) are in line with the selected size of the test. We demonstrate empirical power differences between STrIde and STI with results indicating STrIde is more powerful than STI in nearly all areas of the parameter space for separable spatio-temporal data.

Finally, we apply STrIde and STI to residuals arising from the Indian monsoon precipitation model in Study II of Chapter 3. The results provide evidence that the logit-normal mixed model has successfully captured the relevant spatio-temporal dependence in higher thresholds even when using the more powerful STrIde statistic for hypothesis testing.

Part II

Statistical Analysis of the Intensity, Frequency, and Damage of Tropical Cyclones

Chapter 1

Quantifying Wind Speed-Pressure Relationships in Tropical Cyclones

1.1 Wind Speed-Pressure Relationships in Tropical Cyclones

The relationship between maximum wind speed (MaxWS) and minimum central pressure (MinCP) (denoted MaxWS - MinCP) in tropical cyclones has been studied for several decades. As (Knaff and Zehr, 2007) remarks, possibly the most accurate and reliable estimate of tropical cyclone intensity is the MinCP, while destructive potential is better related to MaxWS.

Consequently, there has been ample research in the atmospheric sciences and related fields on understanding the relationship between these two variables. We present a brief review of the related literature later in this chapter. Studying the physics of tropical cyclonic activity naturally leads to better understanding of the MaxWS - MinCP relationship, the stress on such relationships under a changing climate scenario, and of risks associated from tropical cyclones to humans life and property, ecology, biodiversity and various other vulnerabilities. A review of the physical structure of tropical cyclones is presented in (Wang and Wu, 2004). More details may be obtained from (Maclay et al., 2008; Mei et al., 2012) and other outlets.

In the last few decades, data collection on tropical cyclones has gathered pace using a variety of observational techniques. Consequently, there is a now adequate global data to (i) statistically evaluate the different approaches used in the literature to study tropical cyclone physics, and (ii) study alternative statistical models that may be more consistent with the features and properties of the data while retaining physical interpretability.

Previous methodologies focus primarily on qualitative methods as well as combinations of physical approximations with multiple regression techniques. In Sec-

tion 1.2, we review the literature studying the relationship between MaxWS and MinCP. In particular, we introduce the *Qualitative Visual Methods*, *Cyclostrophic Balance Wind Approximations*, and *Regression Methods* for quantifying the relationship. These analyses suffer from some untenable assumptions and opaque methodologies. Analyses regularly eliminate variability in data and provide little or no statistical uncertainty quantification.

To better understand the current methods effectiveness, we first approach the MaxWS-MinCP relationship using a combination of the previous methods. Section 1.3 introduces the comprehensive International Best Track Archive for Climate Stewardship (IBTrACs) data set for global tropical cyclones and presents initial linear models based on modifying the cyclostrophic approximation. Results indicate typical linear model assumptions are violated including constant variance of residuals. A weighted linear model is implemented to adjust for this, however, residuals still do not appear normal. Uncertainty estimates are produced using asymptotic standard errors, a parametric M-out-of-N (MooN) bootstrap, and a parametric MooN subsampling; results indicate the asymptotic errors are much smaller than those of the bootstrap.

Therefore, we turn to extreme value modeling in Section 1.4. The model is an intuitive fit for extreme values such as MaxWS and utilizes the cyclostrophic balance wind approximation. The model fits a Generalized Extreme Value (GEV) distribution to log(MaxWS) with a location parameter as a function of MinCP. Uncertainty estimates are again produced using asymptotics and the two parametric MooN methods. In this case, uncertainty results were much more in line with each other.

Finally, in Section 1.5, we extend the study of MaxWS and MinCP to a bivariate realm. MaxWS and MinCP require joint study because data uncertainty may lead to potential changes in the bivariate distribution over time that a conditional

model would be unable to capture. We examine several bivariate extreme value parameterizations before choosing a bivariate bilogistic extreme value model. Additionally, using this framework, we study the joint relationship of damages and intensity in the Atlantic basin where intensity is modeled using the MaxWS-MinCP relationship. Simulation results using the estimated model provide a reasonable way of predicting possible bivariate intensity-damage distributions for individual tropical cyclones.

1.2 Current Practices in Modeling the MaxWS-MinCP Relationship

Existing study of the relationship between tropical cyclone intensity and potential for destruction, as quantified respectively by MaxWS and MinCP, can be classified into three broad categories. These categories are *Qualitative Visual Methods*, *Cyclostrophic Balance Wind Approximations*, and *Regression Methods*. In particular, these categories cover the current use of MaxWS-MinCP relationships at Regional Specialized Meteorological Centers (RSMCs) as part of the World Meteorological Organization (WMO) Tropical Cyclone Program[1]. RSMCs collect data and use analyses with the application of appropriate techniques to ensure accurate tropical cyclone forecasting and timely warnings.

[1] http://www.wmo.int/pages/prog/www/tcp/organization.html

1.2.1 Qualitative Visual Methods

1.2.1.1 Dvorak Technique (DVT)

One of the most popular ways of quantifying the MaxWS-MinCP relationship is the Dvorak Technique (Dvorak, 1975, 1984). This is a qualitative method that uses satellite imaging with wind speed and pressure specifications for different cloud formations. Essentially, analysts use images to identify the storm center, assign a pattern, take measurements, and estimate current intensity (CI). Table 1.1 indicates the CI values and their corresponding 1-minute sustained maximum wind speed values and pressures. These empirical values were obtained from Figure 9 of (Dvorak, 1984).

Because the DVT was only developed for the North Atlantic and West Pacific basins, regional modifications of the method have taken place in other basins. However, these modifications are somewhat ad-hoc, and as noted by (Velden et al., 2006), *"...some of these applications are simply rules of thumb, and are not rigorously proven or published results."*

In (Knaff et al., 2010), biases with respect to MaxWS estimates in the DVT were illuminated, and regression equations were generated to correct these biases in the Atlantic basin. Biases were quantified using aircraft reconnaissance data as compared to the DVT estimates. The verifications were stratified by intensity (V_{max}), latitude (ϕ), storm translation speed[2] ($1.5V_t^{0.63}$), intensity trend (ΔV_{max}), and radius of the outer closed isobar (ROCI). This correction was estimated in an unusual way where an initial bias was first estimated:

$$\widehat{\text{Bias}}_1 = 32.174 - 1.990V_{max} + \left(\frac{V_{max}}{5.070}\right)^2 + \left(\frac{V_{max}}{15.076}\right)^3 + \left(\frac{V_{max}}{34.351}\right)^4. \quad (1.1)$$

[2]This term is commonly used for translation speed as referenced in (Schwerdt et al., 1979).

Table 1.1: DVT and Koba Intensity Estimates. DVT 1-minute and Koba 10-minute value come directly from literature. 1-minute DVT values are converted using $C_1 = 0.88$ and $C_2 = 0.93$ to compare with Koba values. Values for MinCP come directly from literature.

	MaxWS (kt)				MinCP (millibars (mb))		
CI	DVT			Koba	DVT		Koba
	1-Min	10-Min (C_1)	10-Min (C_2)	10-Min	N. Atl.	W. Pac.	W. Pac.
1	25	22.00	23.25	22	-	-	1005
1.5	25	22.00	23.25	29	-	-	1002
2	30	26.40	27.90	36	1009	1000	998
2.5	35	30.80	32.55	43	1005	997	993
3	45	39.60	41.85	50	1000	991	987
3.5	55	48.40	51.15	57	994	984	981
4	65	57.20	60.45	64	987	976	973
4.5	77	67.76	71.61	71	979	966	965
5	90	79.20	83.70	78	970	954	956
5.5	102	89.76	94.86	85	960	941	947
6	115	101.20	106.95	93	948	927	937
6.5	127	111.76	118.11	100	935	914	926
7	140	123.20	130.20	107	921	898	914
7.5	155	136.40	144.15	115	906	879	901
8	170	149.60	158.10	122	890	858	888

Then,

$$\begin{aligned} \text{Bias} - \widehat{\text{Bias}}_1 = 12.385 - 22.639 \sin|\phi| + 0.859(1.5V_t^{0.63}) \\ + 0.016\text{ROCI} - 0.309\Delta V_{max}. \end{aligned} \tag{1.2}$$

The authors deterministically combined these regressions to get a final bias estimate of

$$\begin{aligned} \widehat{\text{Bias}} = 44.559 - 1.990V_{max} + \left(\frac{V_{max}}{5.070}\right)^2 + \left(\frac{V_{max}}{15.076}\right)^3 + \left(\frac{V_{max}}{34.351}\right)^4 \\ - 22.639 \sin|\phi| + 0.859(1.5V_t^{0.63}) + 0.016\text{ROCI} - 0.309\Delta V_{max}. \end{aligned} \tag{1.3}$$

The DVT is used for the Atlantic and East Pacific at the National Hurricane Center (NHC) Tropical Prediction Center (TPC) and for the Central Pacific at the Central Pacific Hurricane Center (CPHC).

1.2.1.2 Koba Technique

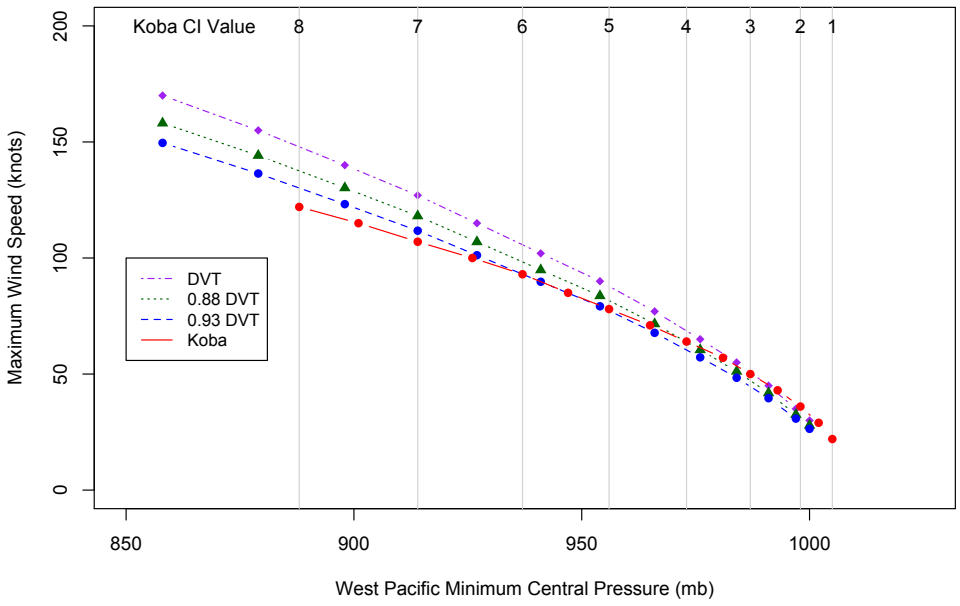
A method closely related to the DVT was proposed in (Koba et al., 1990)[3]. The Koba table came into operation in 1989 and incorporated a landfall rule due to typhoons passing over the Phillipines. Modifications to the DVT include cases where developing tropical cyclone makes landfall, thus, substantially decreasing intensity. This method is in use at RSMC Tokyo, after verification using the Japan Meteorological Agency (JMA) best-track data and CI numbers. Details of how exactly the verification was made, statistical procedures used, or uncertainties about the process are not apparent.

Values of the Koba table are seen in Table 1.1 for comparison with the DVT. To adequately compare wind speeds measured at different time lengths, we provide conversions of the DVT estimates in 1-minute knots (kts) to 10-minute kts as in Koba estimates. A common conversion from 1-minute wind speed (S_1) to 10-minute wind speed (S_{10}) is $S_{10} = C \cdot S_1$. This conversion comes from (Courtney and Knaff, 2009) and suggested values of C are 0.88 and 0.93.

As seen in Figure 1.1, in higher Koba CI values, the Koba technique estimates lower West Pacifica wind speeds than either of the converted the DVT values, while in lower CI values, Koba wind speeds are higher than converted the DVT values for similar pressures. The break even point occurs between CI 4 and 5. The unconverted the DVT values cross with Koba values between CI 2 and 3. We also note the DVT allows for much lower MinCP than the Koba technique.

[3]Though cited in many papers, this article only exists in Japanese.

Figure 1.1: Koba Table Comparisons to the DVT for the West Pacific. Unconverted 1-minute wind speed DVT estimates are provided along with two possible conversions to 10-minute wind speeds.



1.2.1.3 Critiques of the Qualitative Visual Methods

According to (Velden et al., 2006), the primary DVT limitation is that it must infer quantities from cloud patterns and features leading to two important sources of error. First, the technique is restricted by natural variability between the cloud patterns and observed wind speed. Second, the method is subjective. Maximum wind speeds can be overestimated especially for major tropical cyclones, and its North Atlantic application often underestimates the impact of translation on hurricane intensity. (Knaff et al., 2010) also find biases in the DVT noted in (1.1)-(1.3) in relation to latitude, translation speed, speed factor, ROCI, and intensity trend.

It is also not entirely clear how the DVT or Koba empirical values were obtained, or what uncertainty is associated with these values. It is also unclear if these values are expected to be valid irrespective of the the quality and quantity of satellite imagery, or advances in technology.

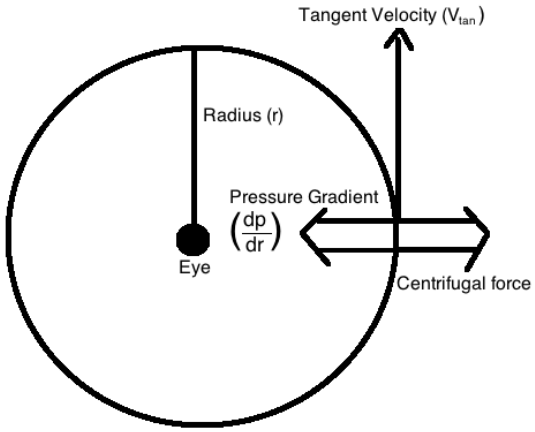
1.2.2 Cyclostrophic Balance Wind Approximations

Use of the cyclostrophic wind approximation seems to have originated from (Takahashi, 1952). This physical approximation comes from the cyclostrophic flow equation explained in detail in §3.2.4 of (Holton, 1992). Cyclostrophic flow can be used in situations where friction is ignorable and the Coriolis force is much smaller than the centrifugal force. The diagram in Figure 1.2 is a stylized depiction of the forces and terms in use for calculating the tangential wind speed (V_{tan}) given a storm radius (r) and pressure gradient ($\frac{\partial p}{\partial r}$). The equation for cyclostrophic flow is:

$$V_{tan} = \left(r \frac{\partial p}{\partial r} \right)^{0.5} . \tag{1.4}$$

To define the pressure gradient used in practice, let P_{ref} be a reference pressure

Figure 1.2: Cyclostrophic Balance in a Tropical Cyclone



for the basin in which the storm matriculates and P_{min} be the minimum sea level pressure over the tropical cyclones' lifetime. Then, $\frac{\partial p}{\partial r} \approx (P_{ref} - P_{min})$, assuming the minimum pressure occurs in the eye of the storm. Tangential wind speed is approximated by maximum sustained wind speed[4], V_{max}. Replacing the radius r and the exponent 0.5 with a generic constant C and exponent n, respectively, leads to the final nonlinear approximation:

$$V_{max} = C\big(P_{ref} - P_{min}\big)^n. \tag{1.5}$$

The relation in (1.5) is not invariant, and values of C and n depend on the units in which V_{max}, P_{ref}, and P_{min} are measured.

Researchers have considered the parameters (C, P_{ref}, n) to vary for different ocean basins, and several studies have empirically investigated (1.5). Reference pressure is not typically estimated, and can instead be fixed at a value determined to be the average steady state pressure over the ocean basin of interest. However, some research determines a reference pressure empirically during modeling.

[4]Sustained wind speed is defined as wind speed gusting for 1 minute or longer.

Some estimates for the North Atlantic basin using MaxWS in kts and MinCP in mb[5] include $(7, 1013, 0.5)$ in (Landsea, 1993) and $(14, 1013, 0.5)$ in (Kraft, 1961). The estimates in (Kraft, 1961) are used within the National Oceanic and Atmospheric Administration (NOAA) official documentation (Jarvinen et al., 1988) regarding tropical cyclone data for the North Atlantic record. A reference pressure of 1013 can be thought of an approximation of the pressure of 1 atmospheric unit corresponding to 1013.25 mb.

After converting wind speeds from 1-minute kts to meters per second, (Holland, 2008) stated the DVT North Atlantic values in Table 1.1 correspond closely to a cyclostrophic approximation with parameters $(3.92, 1015, 0.644)$. Details of their analysis were not readily apparent; we conducted a simple analysis in R, and obtained a nonlinear least squares fit of $(3.65, 1017.19, 0.655)$.

Estimates for the West Pacific basin using wind speed in kts are $(6, 1010, 0.5)$ in (Takahashi, 1952) and $(6.7, 1010, 0.667)$ in (Atkinson and Holliday, 1977). In (Atkinson and Holliday, 1977), observations from 1947-1974 were used but only 76 data points were kept based on data screening. Some details of this screening process were not apparent, but the authors did state the major limiting factor was requiring peak gusts winds (usually occurring in the eye wall cloud) to have been observed at a weather station using anemometers.

Versions of the nonlinear cyclostrophic balance approximation method are used at the RSMC on La Reunion island, RSMC Fiji, the Perth tropical cyclone center, and at the Joint Typhoon Warning Center (JTWC) located in Pearl Harbor, Hawaii.

[5]Note that some studies use hectopascals which are equivalent to mb

1.2.2.1 Critiques of the Cyclostrophic Balance Wind Approximations

As noted by the estimates above, (1.4) is not commonly used directly. The assumption of perfect balance would give the exponent of 0.5, but most empirical studies allow this to vary. The form of the approximation requires a nonlinear estimation technique which may be less robust to data problems than a transformation of the equation to linearity. This approximation assumes no other possibly useful covariates such as translation speed, latitude, or polynomial terms in MinCP are needed.

Binning of data by intensity is recommended in (Landsea et al., 2004) since the number of weaker cases often outnumber the stronger cases. This technique is reflected in some of the above estimates, however, is not consistent throughout the literature and leads to unclear uncertainty estimates which are reported in limited or no detail in the literature discussed above.

1.2.3 Regression Methods

Regression methods for the MaxWS-MinCP relationship combine physical characteristics of tropical cyclones with the results of several regressions. Two of the more notable studies are discussed in this section. Both methods have been developed specifically on North Atlantic data. Definitions of parameters and units used within the models are elucidated in Table 1.2.

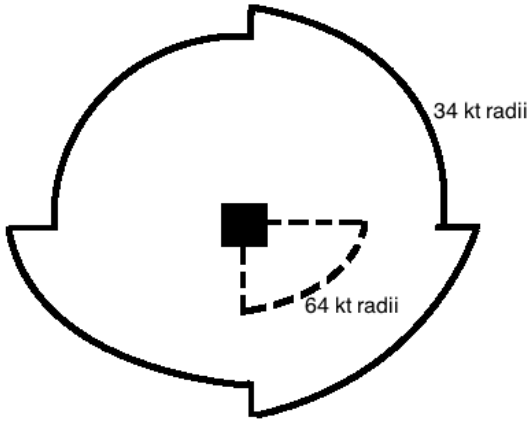
1.2.3.1 Knaff-Zehr (KZ) Model

The KZ model was developed in (Knaff and Zehr, 2007) using 15 years (1989-2004) of MinCP estimates, numerical analysis fields, and best-track intensities. Major findings indicated latitude, size of storm, and reference pressure are associated with MaxWS-MinCP relationships.

Table 1.2: Summary of Parameters used by (Knaff and Zehr, 2007) and (Holland, 2008)

Par.	Unit	Definition
ϕ	$^\circ$	Latitude
x	unitless	Modified Rankine vortex shape parameter
ρ	kg$\cdot$ m^{-3}	Air density
$\frac{\partial P_c}{\partial t}$	hPa $\cdot$ hour^{-1}	MinCP change per hour
P_{min}	hPa	MinCP
P_{ref}	hPa	Reference/environmental pressure
P_{rmw}	hPa	Estimated surface pressure in maximum wind region
R_{max}	km	Radius of maximum winds
S	unitless	Normalized storm size
T_{vs}	K	Virtual 10 meter Temperature
T_s	$^\circ$C	Surface Temperature
V_{500}	m$\cdot$s^{-1}	Average tangential winds from 400-600 km
V_{500c}	m$\cdot$s^{-1}	Tangential wind 500 km from the storm center
V_{srm}	m$\cdot$s^{-1}	Maximum surface wind speed
V_t	m$\cdot$s^{-1}	Storm translational speed
V_{st}	m$\cdot$s^{-1}	Asymmetry modified storm translational speed
V_{maxKZ}	kt	Maximum storm intensity (Knaff and Zehr, 2007)
V_{maxH}	m$\cdot$s^{-1}	Maximum storm intensity (Holland, 2008)

Figure 1.3: Radii Diagram of a Tropical Cyclone



In their model, reference (environmental) pressure, was estimated from data prior to fitting the model. This was done by taking the azimuthal mean pressure in an 800-1000 km annulus surrounding the cyclone center at each adjacent time in the data set.

To incorporate information on tropical cyclone size, the authors used average tangential wind data in the annulus of 400-600 km (V_{500}). However, to limit influence of intensity and latitude on the size estimate, V_{500} is normalized by the value by the climatological tangential wind 500 km from the center (V_{500c}). To do so, they used a modified Rankine vortex parametric model[6].

To model V_{500c} for each observation, the size of a storm is quantified using wind radii. Wind radii are the largest radii of that wind speed found in a quadrant. Quadrants are defined as northeast, southeast, southwest, and northwest. Figure 1.3 is a diagram of a tropical cyclone illuminating these type of radii commonly show on forecasting diagrams.

Typically, the modified Rankine model includes coefficients to incorporate the

[6]For a detailed explanation of the modified Rankine vortex, see (Holland, 1980; Mueller et al., 2006; Knaff et al., 2007)

asymmetric winds in a cyclone, however, the form used in (Knaff and Zehr, 2007) sets these asymmetry values to 0. Thus, the final form for wind speed (V_{Rc}) as a function of radius (R) from the center, latitude (ϕ), and maximum overall wind speed (V_{max}) is

$$V_{Rc} = V_{max}\left(\frac{R_{max}}{R}\right)^{x} \tag{1.6}$$

$$x = x_0 + x_1 V_{max} + x_2(\phi - 25) \tag{1.7}$$

$$R_{max} = m_0 + m_1 V_{max} + m_2(\phi - 25) \tag{1.8}$$

This model was previously estimated in (Knaff et al., 2007) using a constrained optimization technique over all the coefficients in an effort minimize the mean square error between predicted and actual radii values while constraining estimates to maintain physical consistency. Coefficients for the modified[7] Rankine vortex model are procured from Table 1 of (Knaff et al., 2007). The coefficients (x_0, x_1, x_2) come directly from the table, while coefficients (m_0, m_1, m_2) were converted from nautical miles to km. Plugging in 500 km as the radius of interest, the reported Rankine vortex model was:

$$V_{500c} = V_{max}\left(\frac{R_{max}}{500}\right)^{x} \tag{1.9}$$

$$x = 0.1147 + 0.0055 V_{max} - 0.001(\phi - 25) \tag{1.10}$$

$$R_{max} = 66.785 - 0.09102 V_{max} - 1.0619(\phi - 25) \tag{1.11}$$

[7]The term 'modified' refers to the fact that x is explicitly modeled. A true Rankine vortex would set $x = 1$.

Next, a size variable S was defined as:

$$S = V_{500}/V_{500c}, \tag{1.12}$$

and the approximation to the pressure gradient ΔP was defined as:

$$\Delta P = P_{ref} - P_{min}, \tag{1.13}$$

to create the following multiple regression model:

$$E[V_{srm}|\Delta P] = k_0 + k_1 S + k_2 \phi + k_3 \Delta P + k_4 \sqrt{|\Delta P|}. \tag{1.14}$$

The reported fitted model is:

$$\hat{V}_{srm} = 18.633 - 14.960S - 0.755\phi + 0.518\Delta P + 9.738\sqrt{|\Delta P|}. \tag{1.15}$$

Finally, a deterministic addition for translation speed (V_{ts}) is made based on results in (Schwerdt et al., 1979) to give:

$$V_{maxKZ} = \hat{V}_{srm} + 1.5V_{ts}^{0.63}. \tag{1.16}$$

This model is used with slight modification in Australia via (Courtney and Knaff, 2009). The equation defines V_{500} in terms of the radius of gales and uses 10-minute mean wind speed rather than 1-minute. A separate equation was also developed for low-latitude storms in Australia.

1.2.3.2 Holland Model

The second important model, the *Holland model*, is derived in (Holland, 2008). Again, the model uses (1.5) as its starting point, but adds structure to C. This research revises (Holland, 1980) and focuses on an additional parameter, b_s, within the previously constant C. It also explicitly accounts for air density (ρ). The Holland model fixes $P_{ref} = 1015$ to focus on modeling b_s which enables variation in the pressure gradient near the maximum winds and is an attempt to capture peakedness in the related wind profile.

The model for maximum storm intensity (V_{maxH}) is first defined in the form:

$$V_{maxH} = C(\Delta P)^{0.5} \tag{1.17}$$

$$C = \left(\frac{b_s}{\rho e}\right)^{0.5} \tag{1.18}$$

Though it looks similar to that in KZ model, the approach is to derive b_s directly from physics equations, and to then use that to estimate the maximum wind speed V_{maxH}. Since b_s is not directly observed, (1.17) is solved to derive b_s. Using the ideal gas law,

$$\rho = \frac{100 * P_{rmw}}{RT_{vs}}, \tag{1.19}$$

where P_{rmw} is the estimated surface pressure[8] in the maximum wind region, R is the universal gas constant[9], and T_{vs} is the virtual temperature at a 10 meter height in the maximum wind regime.

[8]Note the factor of 100 is necessary to put the units of the numerator into $n \cdot m^{-2}$.
[9] The gas constant used is $287.058 n \cdot m \cdot kg^{-1} \cdot K^{-1}$.

Thus, b_s can be represented as

$$b_s = \left(\frac{V_m^2 P_{rmw} e}{\Delta P R T_{vs}} \right), \tag{1.20}$$

and based on (Holland, 1980),

$$P_{rmw} = P_{min} + \frac{\Delta P}{3.7} \tag{1.21}$$

which assumes the radial pressure profile can be well approximated by rectangular hyperbole. To find values of T_{vs}, the following deterministic equations are employed:

$$T_{vs} = (T_s + 273.15)(1 + 0.81 q_m) \tag{1.22}$$

$$T_s = 28 - 3 \left(\frac{\phi - 10}{20} \right) \tag{1.23}$$

$$q_m = 0.9 \left(\frac{3.802}{P_{rmw}} \right) e^{17.67 T_s / (243.5 + T_s)} \tag{1.24}$$

where T_s is the surface air temperature (°C) and q_m is the vapor pressure at an assumed relative humidity of 90% modeled by a function approximating the Clausius-Claperyon equation (Emmanuel, 1991). The latitudinal variation of surface temperature was arrived using sea surface temperatures for the hurricane season. Finally, the regression of V_m on ΔP is fit:

$$E[V_m | \Delta P] = h_0 + h_1 \Delta P + h_2 (\Delta P)^2 \tag{1.25}$$

with the following result:

$$\hat{V}_m = 11 + 0.9 \Delta P - 0.0027 (\Delta P)^2. \tag{1.26}$$

Thus, V_m is substituted with $\hat{V}_m$ in (1.20) to calculate the 'actual' values of b_s. Additional variables are defined for use in modeling b_s including

$$V_{ts} = V_t^y \tag{1.27}$$

$$y = 0.6\left(1 - \frac{\Delta P}{215}\right) \tag{1.28}$$

where V_t is storm translation speed and V_{ts} is a modified translation speed based on the results regarding asymmetry in (Schwerdt et al., 1979). Unlike the KZ model, y is a function of pressure difference. They state that this exponent is *"arbitrary but captures the essence of the actual variation."*

Latitude (ϕ) and minimum pressure change over time ($\frac{\partial P_c}{\partial t}$) are also used in the following linear regression model:

$$E\left[b_s \middle| \Delta P, \frac{\partial P_c}{\partial t}, \phi, V_{ts}\right] = a_0 + a_1\Delta P + a_2(\Delta P)^2 + a_3\frac{\partial P_c}{\partial t} + a_4|\phi| + a_5 V_{ts}, \tag{1.29}$$

resulting in the estimate

$$\hat{b}_s = 1.0 - 4.4 * 10^{-5}(\Delta P)^2 + 0.01\Delta P + 0.03\frac{\partial P_c}{\partial t} - 0.014|\phi| + 0.15V_{ts}. \tag{1.30}$$

The estimate is then plugged back into (1.18) to provide estimates of V_{maxH}.

1.2.3.3 Critiques of the Regression Models

Both models provide useful insights on physical tropical cyclone modeling, however, fail to meet several statistical criteria. Regression techniques presented within either paper have not given estimates of parameter uncertainty, and no assessment of

models assumptions such as constant error variance are discussed in either article.

The methods are also opaque in some sense, relying on combinations of deterministic physics formulae, empirical estimates, and regression. The KZ model uses several empirically derived variables using formulae developed on other data such as the modified Rankine vortex. They ignore portions of the pre-derived Rankine model rather than re-estimating using their specified model form. They also utilize a regression plugged into a deterministic adjustment for translation speed as opposed to directly including translation speed as a regression variable.

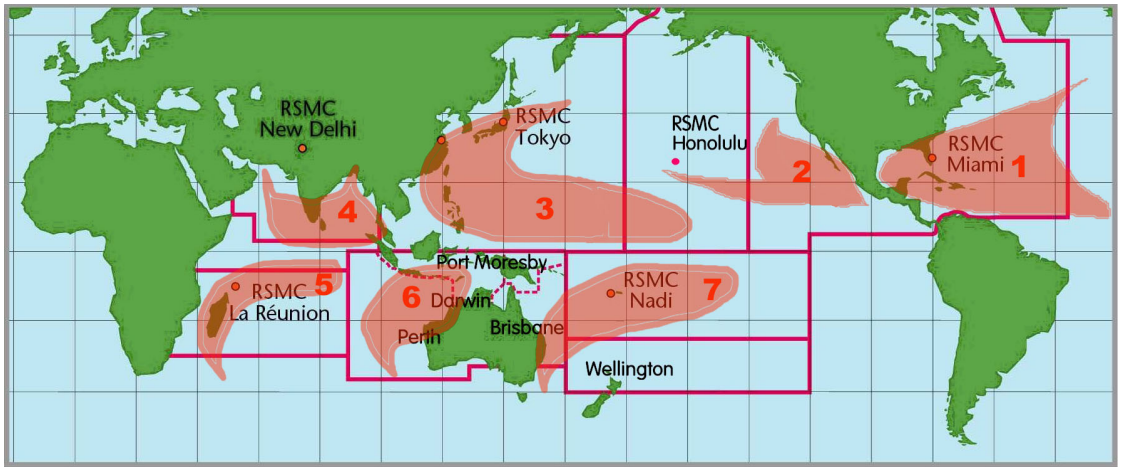
Holland calculates b_s using estimated values of maximum wind speed rather than the true values. They then plug estimates of b_s into a formula for maximum wind speed. This logic is somewhat circular in nature, and it is not clear what uncertainty propagates throughout the methodology.

As in the cyclostrophic models in Section 1.2.2, aggregation methods on data are used prior to modeling which essentially eliminates much of the data. Holland reports binning wind speed data into 5 meter/second groups to arrive at (1.25). However, doing so may mask extra variability seen in the weaker cases.

1.3 Tropical Cyclone Initial Investigation

In this section, we first present available data on the internal constructs of tropical cyclones, and discuss their interesting features. We also discuss limitations of the data resources, and caveats associated with analyses of such datasets. We then proceed to replicate and modify some of the data-centric approaches presented in Section 1.2.

Figure 1.4: NOAA Tropical Cyclone Basins



1.3.1 The IBTrACs Tropical Cyclone Data

National Oceanic and Atmospheric Administration (NOAA) defines tropical cyclone generating basins as shown in Figure 1.4[10]. Data from all basins has been acquired via version 3 release 6 of IBTrACs[11] (Knapp et al., 2010). This is a comprehensive data source on tropical cyclones which includes a collection of best track data from numerous tracking agencies including RSMCs and Tropical Cyclone Warning Centers (TCWCs). The data are objectively combined and the database does not use reanalysis of satellite or surface data.

However, Figure 1.5[12] and Table 1.3 shows IBTrACs basins are marginally different from the demarcation of storm tracking basins used by NOAA. In order to better align with NOAA definitions of basins, the South Indian basin within IBTrACs data was separated at the 100° longitude. A summary indicating the count of storms reaching tropical cyclone strength ($\geq$ 64 kts) within each of the

[10] http://www.prh.noaa.gov/cphc/pages/FAQ/Forecasting.php

[11] Upon completion of this work, v3r7 and v3r8 were released in Oct. 2015 and Jan. 2016, respectively. The only notable update was the inclusion of the 2014 season.

[12] Figure created in R using Table 1.3.

basins[13] between 1960-2013 is presented in Table 1.3.

Table 1.3: Tropical Cyclone Basin Demarcations and Storm Counts within NOAA and IBTrACs

		IBTrACs Lat.		IBTrACs Long.		
NOAA	IBTrACs	Min.	Max.	Min.	Max.	TCs
N. Atlantic	N. Atlantic (NA)	0	90	Varies	30	332
NW Pacific	W. Pacific (WP)	0	90	100	180	820
NE Pacific	E. Pacific (EP)	0	90	180	Varies	270
N. Indian	N. Indian (NI)	0	90	30	100	32
Australian/ SW Pacific	S. Pacific (SP)	-90	0	135	-70	222
SE Indian	S. Indian* (SEI)	-90	0	100	135	128
SW Indian	S. Indian* (SWI)	-90	0	10	100	205

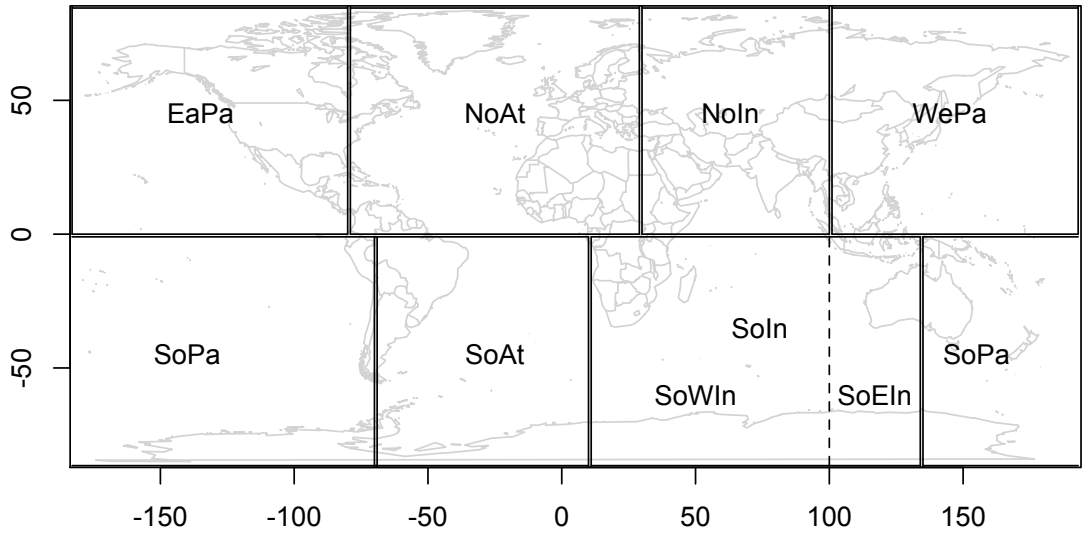
IBTrACs may have multiple streams of data for a each 6-hour time point of a storm if it is tracked by more than one agency. For storms where there were multiple values available at one time point, values are averaged for our analysis. All other variables are derived following this initial averaging.

Average latitude over an entire storm is derived by averaging all previously derived 6-hour time points available for a unique storm. IBTrACs also provides MinCP and MaxWS on a 6-hour basis when available. The minimum and maximum of each are taken for each storm, respectively.

Derivation of translation speed at each 6-hour time point is calculated using a finite differencing method. In order to calculate the distance between two geographical locations, we used the haversine distance, δ, recommended by (Sinnott, 1984) and implemented via `distHaversine{geosphere}` in R (Hijmans, 2015). For time ℓ, and all points other than the first and last, we calculate translation

[13] Although the IBTrACs database counts the South Atlantic as a possible area for storm development for completeness, there were no observations in this region.

Figure 1.5: IBTrACs Tropical Cyclone Basins



speed (V_{ts_ℓ}) in meters per second, as:

$$V_{ts_\ell} = \frac{\delta\big([\text{Lat}_{\ell-1}, \text{Long}_{\ell-1}], [\text{Lat}_{\ell+1}, \text{Long}_{\ell+1}]\big)}{12 * 60 * 60} \tag{1.31}$$

Translation speed at the first and last point of each storm were calculated using:

$$V_{ts_\ell} = \frac{\delta\big([\text{Lat}_{\ell}, \text{Long}_{\ell}], [\text{Lat}_{\ell+1}, \text{Long}_{\ell+1}]\big)}{6 * 60 * 60} \tag{1.32}$$

and

$$V_{ts_\ell} = \frac{\delta\big([\text{Lat}_{\ell-1}, \text{Long}_{\ell-1}], [\text{Lat}_{\ell}, \text{Long}_{\ell}]\big)}{6 * 60 * 60}, \tag{1.33}$$

respectively. To arrive at a storm average translation speed, another average was taken over these calculated 6-hour translation speeds.

Two caveats are associated with this data, which ought be considered when interpreting the results. First, wind speed is measured in different time units

depending on the basin. The common wind speed unit outside of the North Atlantic and Eastern Pacific is the maximum wind speed value over a 10-minute period. However, the U.S. bordered basins use a maximum wind speed value over a 1-minute period. There are several methods of converting the 1-minute wind speed to 10-minute wind speed that have been used in practice; for a comprehensive list of conversions, see (Knapp and Kruk, 2010; Harper et al., 2010). However, we choose not to modify the data and treat it as is. A second issue is that data records include those obtained using anemometers and other instruments as well as those processed from satellite imagery; consequently, data may not be of uniform quality. The data are also subjected to quantization in which only certain values of wind speed may be attained.

1.3.2 Standard Error Calculations in MaxWS-MinCP Models

Throughout the balance of the chapter, we fit several models for the MaxWS-MinCP relationship. In order to calculate standard errors for the basin estimates (linear combinations of the estimated parameters), we utilize two procedures: asymptotic standard error and parametric MooN bootstrapping/subsampling.

1.3.2.1 Asymptotic Standard Error

In cases where asymptotic standard errors are provided within estimation output, we use them directly to calculate the corresponding standard errors of linear combination (L) of parameters. The linear combinations in question are those which provide estimates by basin. For example, the estimate of intercept in the West Pacific basin is the intercept added to the difference of West Pacific from the reference basin which is the North Atlantic. Thus, L will be a vector of 0s and 1s

in the following sections. Using the estimated covariance matrix $\hat{\Sigma}$, the estimate of standard error will be calculated as $\sqrt{L^T \hat{\Sigma} L}$. In the least squares models, $\hat{\Sigma}$ is the corresponding least squares estimate. In the other models presented in this chapter, $\hat{\Sigma}$ results from maximum likelihood estimation and is the observed inverse Fisher information.

1.3.2.2 MooN Bootstrapping/Subsampling

As a benchmark to the standard asymptotics, we also use bootstrapping to provide standard errors. In particular, we utilize a parametric MooN bootstrap by sampling with replacement and parametric MooN subsampling by sampling without replacement. These methods provide checks underlying asymptotic assumptions as well as alternative uncertainty estimates when fitting extreme value models in Section 1.4 and Section 1.5. This can be especially useful in extreme value models as noted within (Bickel and Sakov, 2008).

To implement the methods, we first let N_B be the number of data points in each basin (B) and choose a percentage (p) of rows to sample. Define $M_B = \lfloor p * N_B \rfloor$. For each bootstrap sample, the algorithm is the following:

1. Sample M_B data points with replacement, $B \in \{1, 2, 3, 4, 5, 6, 7\}$; sample without replacement for subsampling. Join these points into a bootstrapped data set of $M = \sum_B M_B$ observations.

2. Simulate new responses using the original estimates in combination with the bootstrapped covariate data.

3. Refit the parameters of the model using the simulated data.

4. Save parameter estimates.

The following modeling sections use 1000 bootstrap samples to perform parametric MooN bootstrapping using $p \in \{0.66, 0.75, 1.00\}$ and parametric MooN subsampling using $p \in \{0.66, 0.75\}$.

1.3.3 Analysis Combining Previous Methods

As discussed previously, many of the published results on MaxWS-MinCP relationship do not contain details of the exact analysis procedure adopted. Consequently, we present several techniques that are plausible procedures for analyzing this relationship.

1.3.3.1 Comparisons to the Original Cyclostrophic Approximation

As a first step, it seems reasonable to consider a logarithmic transformation of both sides of (1.5). We also consider the reference pressure value P_{ref} to be fixed over all basins. However, since it is not transparent what this reference value should be, we initially test 4 separate settings: 1010 mb, 1013 mb, 1015 mb, 1017 mb.

Suppose (P_{min_i}, V_{Max_i}) is the (MinCP, MaxWS) pair for the i^{th} storm and let $\Delta P_i = P_{ref} - P_{min_i}$. Although we don't allow reference pressure to change by basin, we include a categorical variable for basin (B) to allow for different coefficient estimates by basin. Then, a statistical model that imitates the cyclostrophic approximation (1.5) is given by:

$$\textbf{Model 1:} \quad \log\big(V_{Max_i}\big) = \beta_{0B} + \beta_{1B} \log\big(\Delta P_i\big) + \epsilon_i \tag{1.34}$$

$$\mathbb{E}(\epsilon_i) = 0 \tag{1.35}$$

$$\mathbb{V}(\epsilon_i) = \sigma_1^2 \tag{1.36}$$

$$\epsilon_i \text{ independent of } \epsilon_j \text{ for all } i, j. \tag{1.37}$$

We fit the response separately in units of kts and meters/second for ease of comparison with previous literature. Results of the ordinary least squares (OLS) model indicated $\log(\Delta P)$ was inversely proportional to the residual variance. Thus, a weighted least squares (WLS) approach using $\log(\Delta P)$ as weights was implemented to estimate the original cyclostrophic balance parameters, $n_B = \beta_{1B}$ and $C_B = \exp(\beta_{0B})$ from (1.5), for each basin. An alternative to WLS would be to use the heteroskedastic variance estimate of (White, 1980) in conjunction with OLS mean estimates. However, because it is relatively clear the residual variability is a function of MinCP, we believe the efficiency loss using White's variance estimate is not warranted. Estimates and asymptotic standard errors are displayed in Table 1.4.

In general, n was estimated higher and C was estimated lower in southern hemisphere basins relative to the northern hemisphere basins. This was true for each fixed value of reference pressure. This makes sense due to storms in the southern hemisphere typically displaying slower wind speeds.

A simultaneous 95% family-wise confidence interval on North Atlantic parameters (using reference pressure of 1013 mb) indicated no statistical difference[14] to those provide by (Landsea, 1993) which estimates C as 7 using the response in meters per second and (Kraft, 1961) which estimates C as 14 using the response in kts. The 95% confidence intervals produced by the models were (6.46, 8.42) and (12.56, 16.37), respectively. We also find that n is not statistically different from the cyclostrophic balance exponent of 0.5 using either 1013 mb with 95% confidence interval (0.45, 0.52) or 1015 mb with confidence interval (0.48, 0.55) as the reference pressure. Thus, for clarity in the balance of the paper, we will set $P_{ref} = 1013$. We will also model MaxWS in kts since that is the reporting format of the IBTrACs database.

[14]Hypothesis test was conducted at an $\alpha = 0.05$ level.

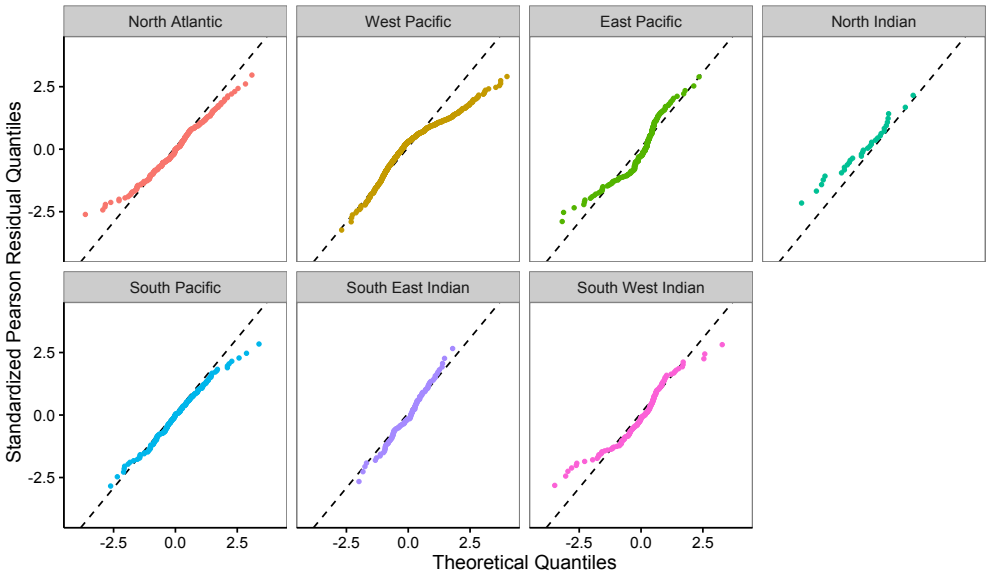
Table 1.4: WLS Estimates of Cyclostrophic Parameters for Tropical Cyclones with varying P_{ref} values. Asymptotic standard error estimates are in parentheses following parameter estimates. The dashed line divides Northern hemisphere from Southern hemisphere results.

P_{ref}	Basin	$\hat{C}$-kts	$\hat{C}$-m/s	$\hat{n}$
	NA	17.106 (0.770)	8.800 (0.396)	0.452 (0.012)
	EP	13.400 (0.784)	6.894 (0.404)	0.510 (0.015)
	WP	11.332 (0.441)	5.830 (0.227)	0.518 (0.009)
1010	NI	14.730 (2.480)	7.578 (1.276)	0.478 (0.042)
	SP	7.878 (0.711)	4.053 (0.366)	0.594 (0.022)
	SEI	8.119 (0.986)	4.177 (0.507)	0.588 (0.030)
	SWI	9.917 (0.894)	5.102 (0.460)	0.541 (0.022)
	NA	14.338 (0.698)	7.376 (0.359)	0.489 (0.013)
	EP	11.262 (0.706)	5.794 (0.363)	0.546 (0.016)
	WP	9.779 (0.402)	5.031 (0.207)	0.547 (0.010)
1013	NI	12.637 (2.268)	6.501 (1.167)	0.509 (0.045)
	SP	6.713 (0.640)	3.453 (0.329)	0.625 (0.023)
	SEI	6.952 (0.890)	3.576 (0.458)	0.618 (0.031)
	SWI	8.523 (0.811)	4.385 (0.417)	0.571 (0.023)
	NA	12.756 (0.652)	6.562 (0.335)	0.514 (0.013)
	EP	10.032 (0.658)	5.161 (0.338)	0.569 (0.016)
	WP	8.864 (0.378)	4.560 (0.194)	0.567 (0.010)
1015	NI	11.410 (2.133)	5.870 (1.098)	0.529 (0.046)
	SP	6.032 (0.595)	3.103 (0.306)	0.646 (0.024)
	SEI	6.266 (0.831)	3.224 (0.427)	0.639 (0.032)
	SWI	7.704 (0.760)	3.963 (0.391)	0.591 (0.023)
	NA	11.354 (0.608)	5.841 (0.313)	0.537 (0.014)
	EP	8.937 (0.611)	4.598 (0.314)	0.592 (0.017)
	WP	8.036 (0.354)	4.134 (0.182)	0.586 (0.010)
1017	NI	10.301 (2.004)	5.299 (1.031)	0.549 (0.047)
	SP	5.418 (0.553)	2.787 (0.285)	0.667 (0.024)
	SEI	5.647 (0.774)	2.905 (0.398)	0.659 (0.033)
	SWI	6.963 (0.711)	3.582 (0.366)	0.611 (0.024)

Figure 1.6 indicates that normality and constant variance assumptions for residuals are not satisfied for most of the basins with problem especially noticeable in the North Atlantic and West Pacific basins. This implies extensions of the model may provide more useful predictions.

Figure 1.6: Q-Q Plot and Residual vs. Fitted Value Plots by Basin for WLS Model 1 (Maximum Wind Speed in kts)

(a) Q-Q Plot



(b) Residual vs. Fitted Value Plot

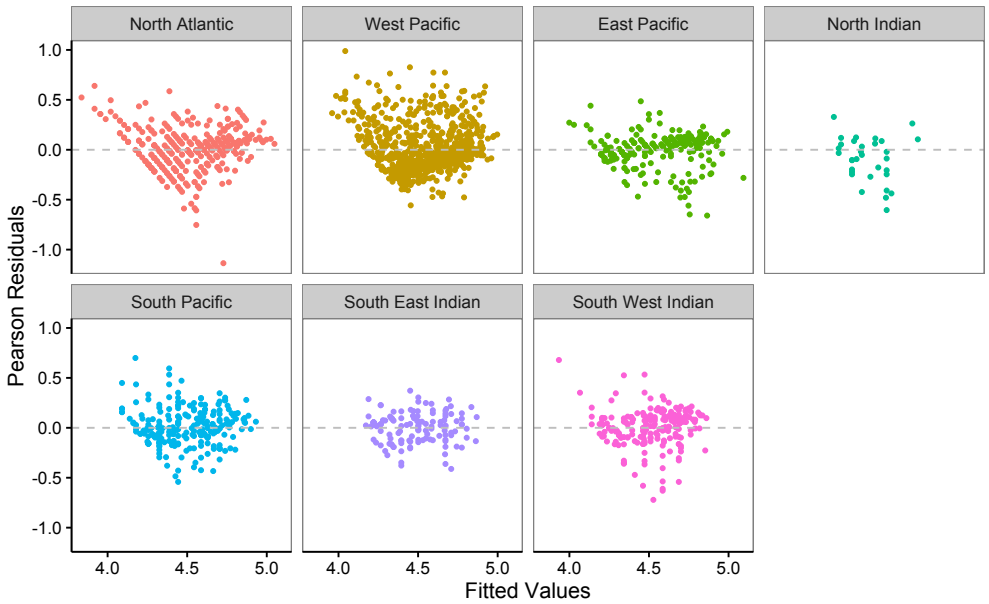
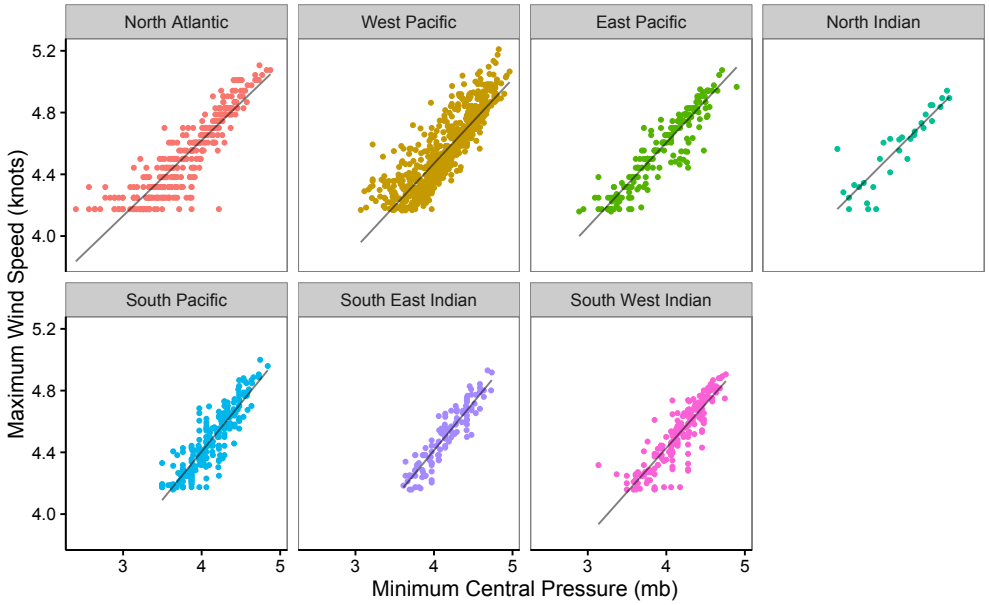
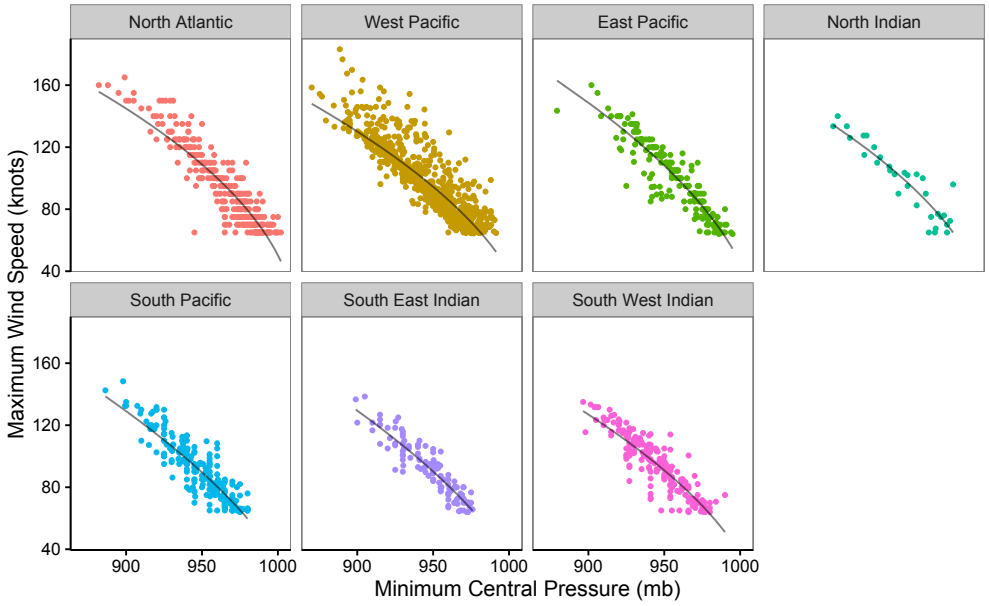


Figure 1.7: Predicted Wind Speeds on Log (Model) Scale and Original Scale in WLS Model 1

(a) Log Scale Predicted Wind Speeds



(b) Original Scale Predicted Wind Speeds



1.3.3.2 Extensions of the Transformed Cyclostrophic Approximation

Similar to (Holland, 2008), and to address the possible missing covariate issue, we move away from the cyclostrophic model and add a quadratic term. Following (Knaff and Zehr, 2007), we also add in other possibly important covariates of average latitude (ϕ_{Avg}) and average translation speed (V_{tsAvg}). These updated models are:

$$\textbf{Model 2:}\quad \log(V_{Max_i}) = \beta_{0B} + \beta_{1B}\log(\Delta P_i) + \beta_{2B}\log^2(\Delta P_i) + \epsilon_i \tag{1.38}$$

$$\mathbb{E}(\epsilon_i) = 0, \tag{1.39}$$

$$\mathbb{V}(\epsilon_i) = \sigma_2^2. \tag{1.40}$$

$$\epsilon_i \text{ independent of } \epsilon_j \text{ for all } i, j, \tag{1.41}$$

and

$$\textbf{Model 3:}\quad \log(V_{Max_i}) = \beta_{0B} + \beta_{1B}\log(\Delta P_i) + \beta_{2B}\log^2(\Delta P_i) + \beta_{3B}\phi_{Avg,i} + \beta_{4,B}V_{tsAvg,i} + \epsilon_i, \tag{1.42}$$

$$\mathbb{E}(\epsilon_i) = 0, \tag{1.43}$$

$$\mathbb{V}(\epsilon_i) = \sigma_3^2, \tag{1.44}$$

$$\epsilon_i \text{ independent of } \epsilon_j \text{ for all } i, j. \tag{1.45}$$

Once again, to address heteroskedascity of the residuals, we use WLS with the weight for the i^{th} observation, w_i, set equal to $\log(\Delta P_i)$. The WLS estimates of Models 1-3 are available in Table 1.5.

Table 1.5: WLS Model 1 - Model 3 Estimates

M	Basin	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
1	NA	2.663	0.489			
	EP	2.421	0.546			
	WP	2.280	0.547			
	NI	2.537	0.509			
	SP	1.904	0.625			
	SEI	1.939	0.618			
	SWI	2.143	0.571			
2	NA	5.241	-0.894	0.183		
	EP	2.664	0.420	0.016		
	WP	7.099	-1.816	0.288		
	NI	6.312	-1.417	0.243		
	SP	5.686	-1.213	0.222		
	SEI	2.859	0.173	0.054		
	SWI	5.816	-1.230	0.220		
3	NA	5.134	-0.778	0.166	-0.004	0.002
	EP	2.834	0.396	0.019	-0.006	-0.001
	WP	6.946	-1.738	0.278	-0.001	0.005
	NI	5.850	-1.217	0.215	0.007	0.005
	SP	5.466	-1.085	0.207	0.002	0.000
	SEI	2.885	0.159	0.055	0.000	0.001
	SWI	6.298	-1.442	0.245	0.000	-0.012

Table 1.6 provides several fit statistics and tests for each of the models. F-tests confirmed that Model 2 was preferred to the Model 1 ($p < 0.0001$) and similarly, Model 3, with the addition of translation speed and latitude, was preferred to either of the previous models ($p < 0.0001$, $p < 0.0001$). Model 3 is preferred in terms of adjusted R^2 and Akaike information criterion (AIC), however, Bayesian information criterion (BIC) favors the smaller Model 2.

Table 1.6: WLS Model 1 - Model 3 Fit Statistics. Bolded fit criteria indicate a preference for that model. Parentheses after F-test p-values indicate the preferred model.

M	$\hat{\sigma}_M$	R^2_{Adj}	AIC	BIC	S-W Test for Norm.	F-test v. M1	F-test v. M2
1	0.2075	0.7965	-3380	-3296	<0.0001	-	-
2	0.1940	0.8222	-3643	**-3520**	<0.0001	<0.0001 (2)	-
3	0.1914	**0.8268**	**-3682**	-3478	<0.0001	<0.0001 (3)	<0.0001 (3)

In Figure 1.8, Figure 1.9, and Figure 1.10, the WLS estimates ± 2 standard errors are displayed for each of the models and standard error types. Clearly, bootstrap estimates give much larger uncertainty values than that provided by the asymptotic standard errors. This indicates the asymptotics are not good estimates of the uncertainty. Based on the most conservative standard errors for Model 1 in Figure 1.8, all basins include the cyclostrophic exponent, 0.5, within a 2 standard error interval for β_1. Model 2 uncertainties in Figure 1.9 indicate the coefficient of the quadratic term of MinCP, β_2, is only significant in the North Atlantic and West Pacific which aligns with the comments from the previous section about the model possibly missing covariates.

In order to evaluate the model's predictive capabilities, we use both mean squared prediction error (MSPE) and mean absolute prediction error (MAPE). MAPE picks models with overall accuracy while MSPE favors models which limit

Table 1.7: Average Prediction Error For WLS MooN Parametric Bootstrapping. Bolded values indicate the smallest prediction error within error type.

Error	MooN p	Model 1		Model 2		Model 3	
		Sbsmp.	w/Rep.	Sbsmp.	w/Rep.	Sbsmp.	w/Rep.
MSPE	2/3	0.0152	0.0153	0.0135	0.0134	0.0499	0.0450
	3/4	0.0152	0.0151	0.0135	0.0135	0.0481	0.0455
	1	-	0.0153	-	**0.0133**	-	0.0457
MAPE	2/3	0.0918	0.0921	0.0859	0.0855	0.1221	0.1184
	3/4	0.0920	0.0917	0.0859	0.0858	0.1205	0.1183
	1	-	0.0921	-	**0.0854**	-	0.1184

the maximal prediction error. Table 1.7 gives the average over all bootstraps these metrics. Model 2 is preferred on both MSPE and MAPE. Despite the results of the F-tests, this result indicates translation speed and latitude coefficients are not useful in prediction. This is reflected in Figure 1.10 where bootstrapped standard errors indicate the coefficients are effectively 0.

All models reject the assumption of normal residuals using a Shapiro-Wilk test (Shapiro and Wilk, 1965) as reported in Table 1.6. Q-Q plots in Figure 1.11 and Figure 1.13 confirm this departure from normality which is still prominent in North Atlantic and West Pacific basins. The non-constant variance is somewhat resolved based on the residuals vs. fitted values plots, however, the West Pacific basin still shows a pattern.

Predictions in the log scale and back-transformed prediction to the original scale for Model 2 are shown in Figure 1.12. In the original scale, the curvature of the relationship for the basins is noticeably different and in some basins, there is almost no curvature. In Figure 1.14, actual wind speed vs. predicted wind speed in the log and original scale for Model 3. Generally speaking, the model fits well, however, the West Pacific continues to show a lack of fit.

Figure 1.8: Estimate +/- 2 Standard Errors for WLS Model 1. In all figures, the lines are defined as follows:
(1 - Purple Long Dashed): Asymptotic Standard Error;
(2 - Blue Solid): MooN bootstrap w/replacement, $p = 1.00$;
(3 - Green Short Dashed): MooN bootstrap subsampling, $p = 0.75$;
(4 - Green Solid): MooN bootstrap w/ replacement, $p = 0.75$;
(5 - Red Short Dashed): MooN bootstrap subsampling, $p = 0.66$;
(6 - Red Solid): MooN bootstrap w/ replacement, $p = 0.66$.

(a) β_0 (b) β_1

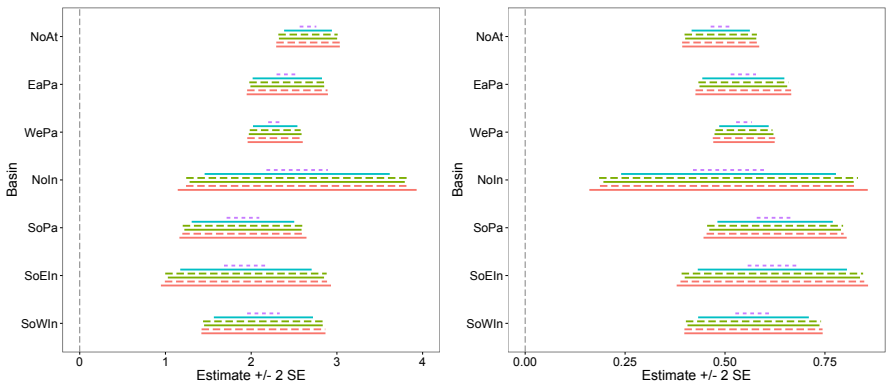


Figure 1.9: Estimate +/- 2 Standard Errors for WLS Model 2. In all figures, the lines are defined as follows:
(1 - Purple Long Dashed): Asymptotic Standard Error;
(2 - Blue Solid): MooN bootstrap w/replacement, $p = 1.00$;
(3 - Green Short Dashed): MooN bootstrap subsampling, $p = 0.75$;
(4 - Green Solid): MooN bootstrap w/ replacement, $p = 0.75$;
(5 - Red Short Dashed): MooN bootstrap subsampling, $p = 0.66$;
(6 - Red Solid): MooN bootstrap w/ replacement, $p = 0.66$.

(a) β_0 (b) β_1 (c) β_2

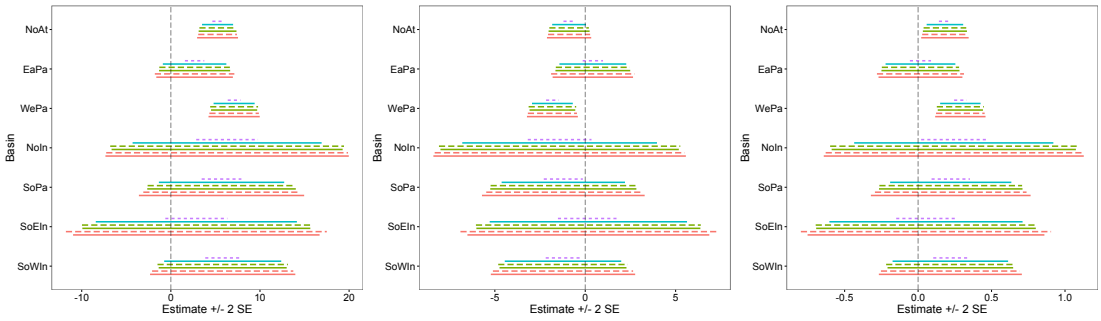


Figure 1.10: Estimate +/- 2 Standard Errors for WLS Model 3. In all figures, the lines are defined as follows:
(1 - Purple Long Dashed): Asymptotic Standard Error;
(2 - Blue Solid): MooN bootstrap w/replacement, $p = 1.00$;
(3 - Green Short Dashed): MooN bootstrap subsampling, $p = 0.75$;
(4 - Green Solid): MooN bootstrap w/ replacement, $p = 0.75$;
(5 - Red Short Dashed): MooN bootstrap subsampling, $p = 0.66$;
(6 - Red Solid): MooN bootstrap w/ replacement, $p = 0.66$.

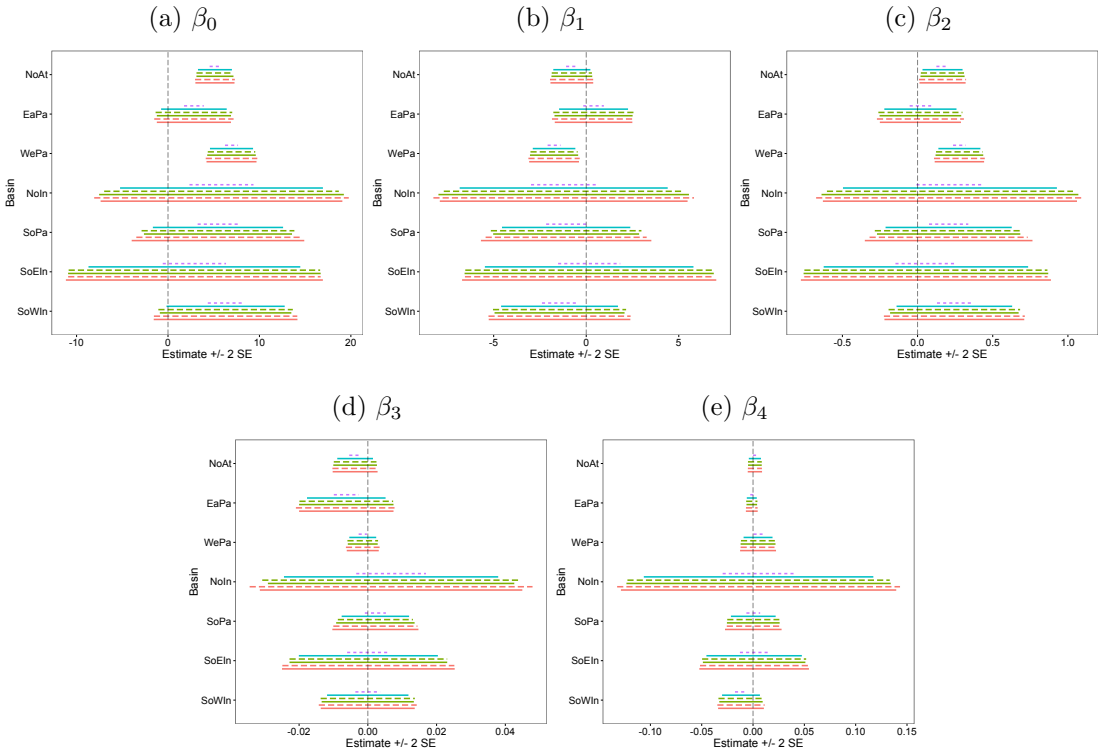
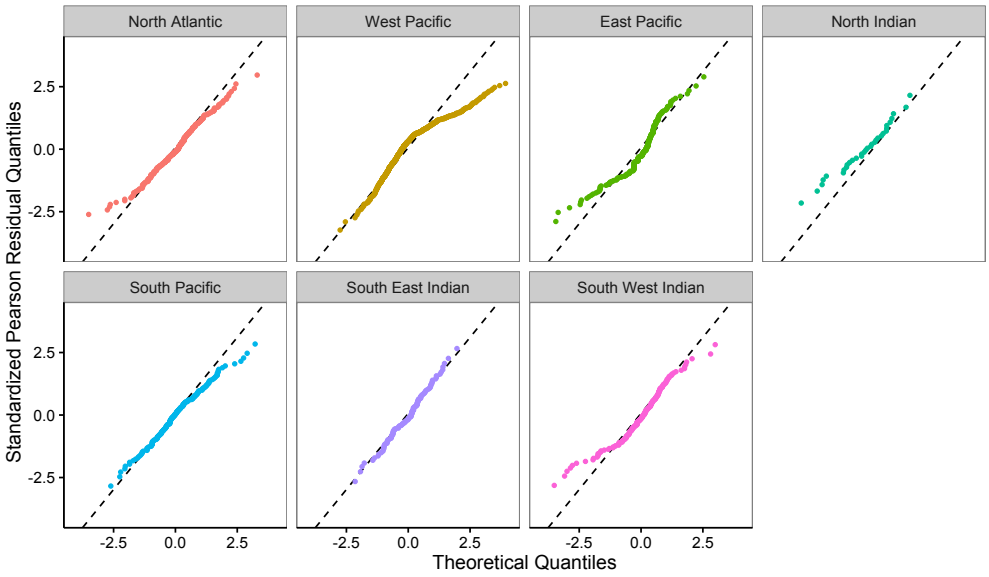


Figure 1.11: Q-Q Plot and Residual vs. Fitted Value Plots by Basin for WLS Model 2

(a) Q-Q Plot



(b) Residual vs. Fitted Value Plot

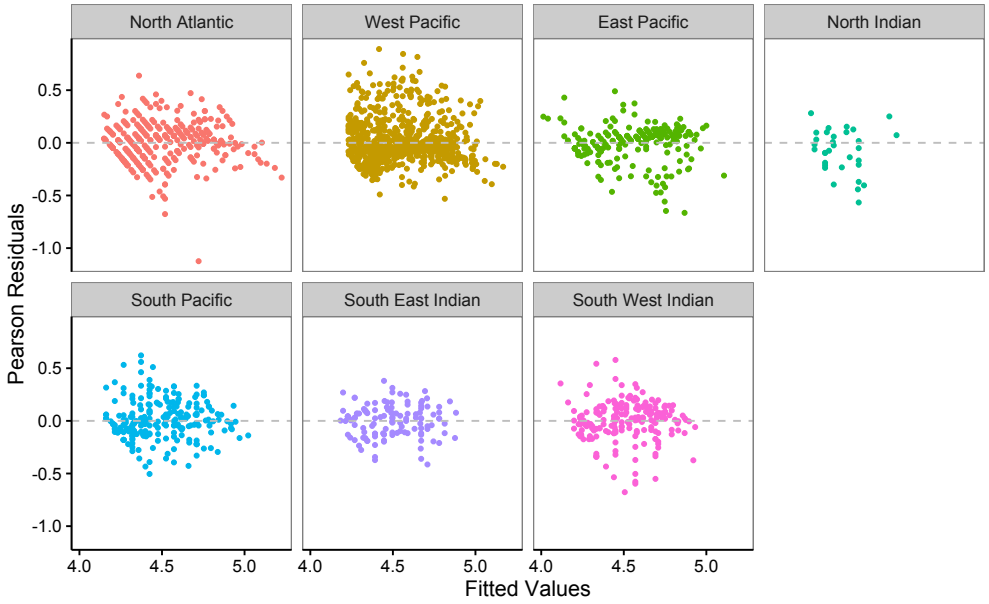
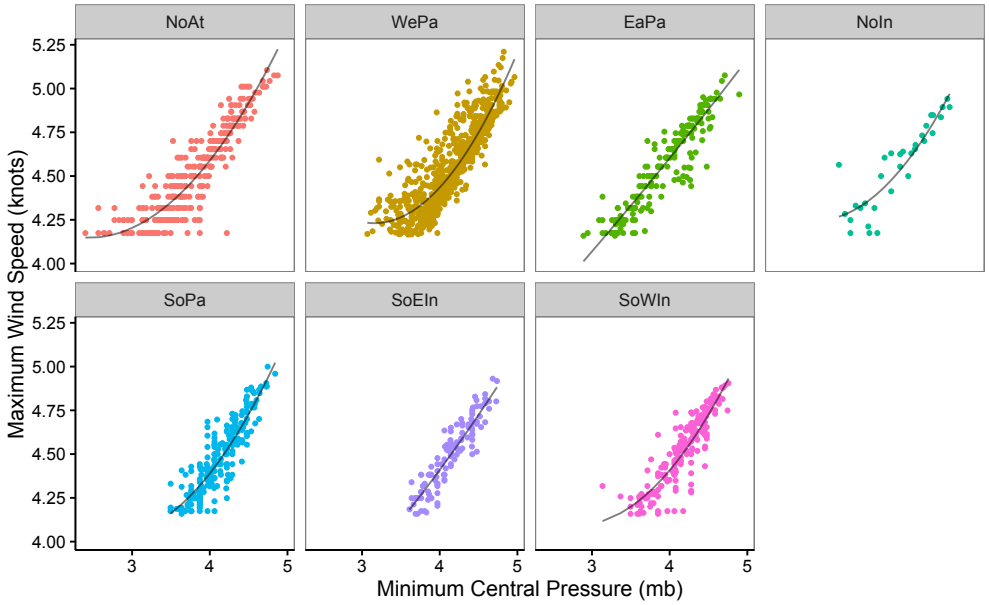


Figure 1.12: Predicted Wind Speeds on Log (Model) Scale and Original Scale in WLS Model 2

(a) Log Scale Predicted Wind Speeds



(b) Original Scale Predicted Wind Speeds

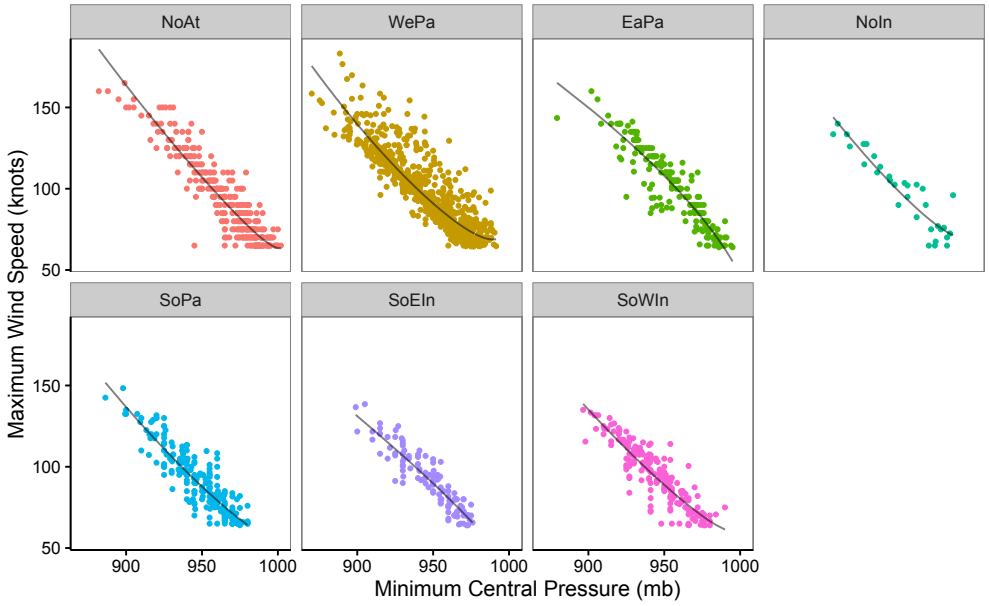
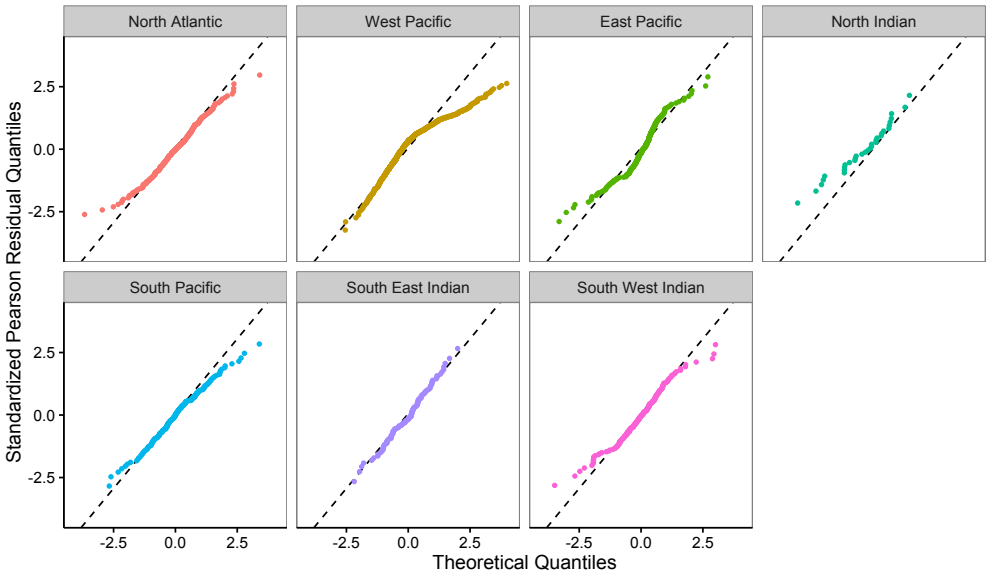


Figure 1.13: Q-Q Plot and Residual vs. Fitted Value Plots by Basin for WLS Model 3

(a) Q-Q Plot



(b) Residual vs. Fitted Value Plot

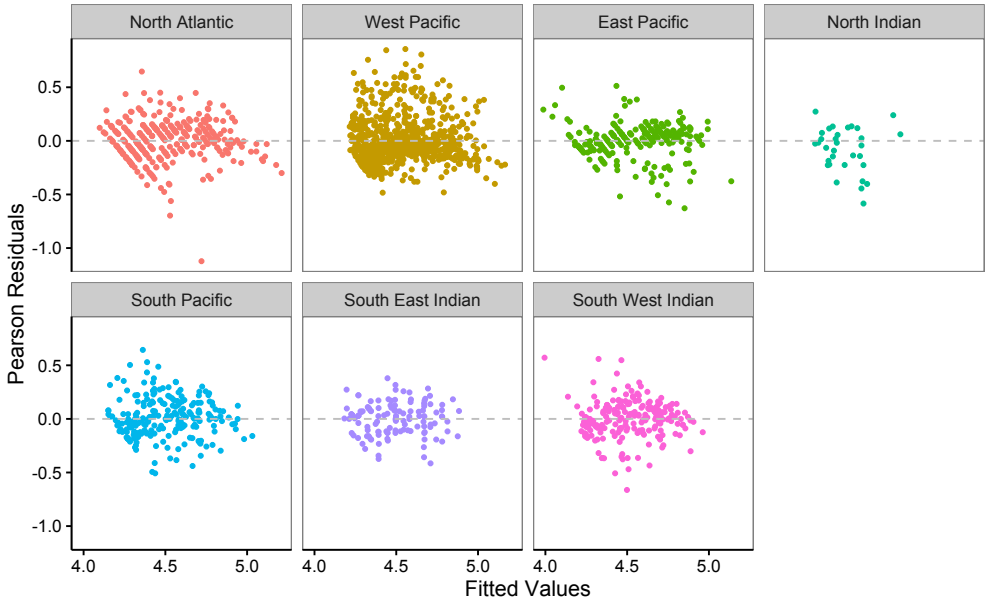
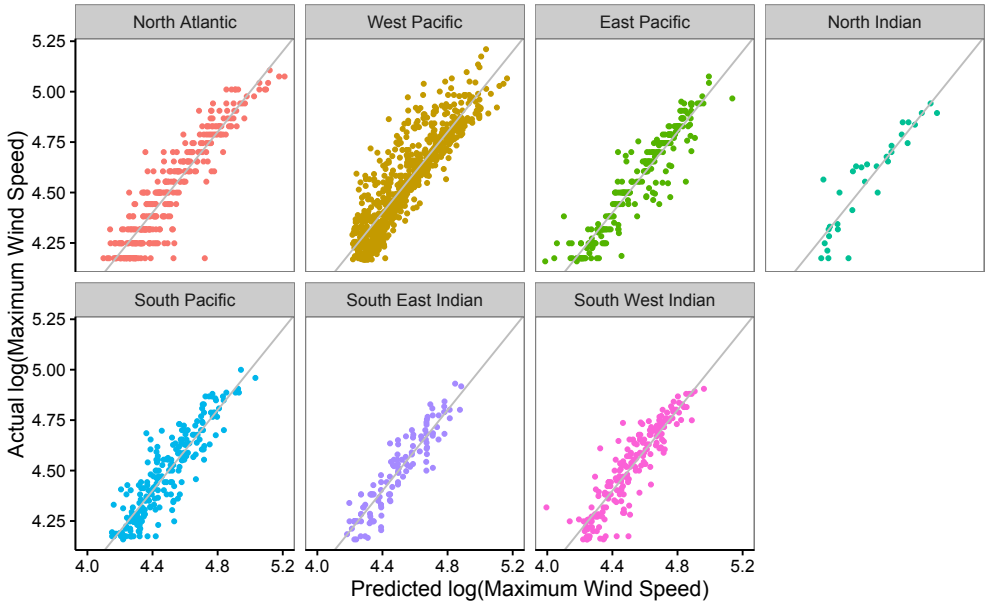
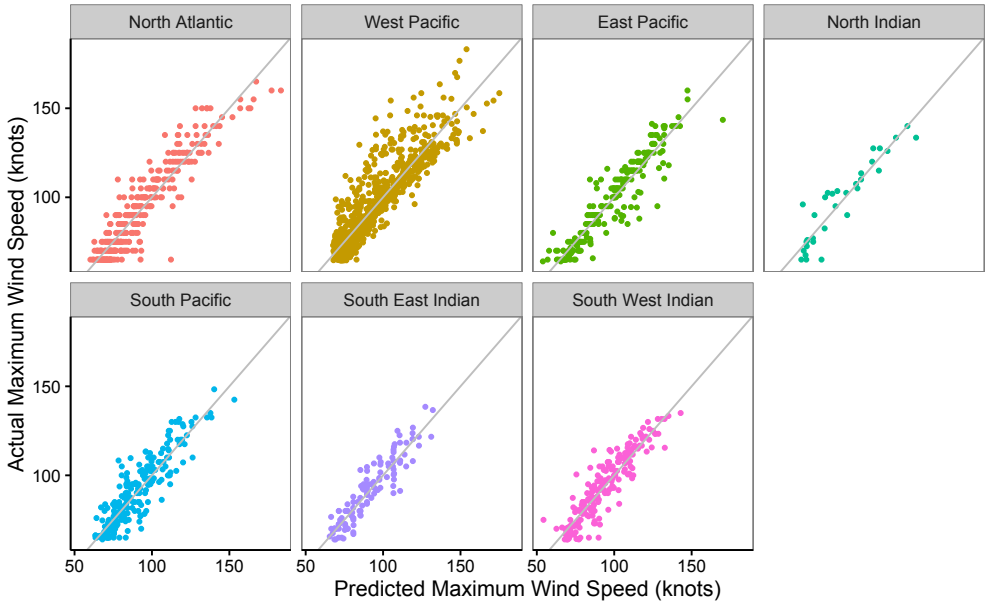


Figure 1.14: Actual by Predicted Wind Speed Plots by Basin for WLS Model 3. The 45° line indicates a perfect fit.

(a) Log Scale Wind Speeds



(b) Original Scale Wind Speeds



1.3.3.3 Comparison of WLS Models to Other Regression Methods for Atlantic Data

To ensure validity of our method, we benchmarked it against the models provided in Section 1.2.3. Since both the KZ and Holland models were developed on Atlantic data, we only utilized that basin. The necessary covariates for the KZ and Holland models were acquired by merging revised Atlantic hurricane database (HURDAT2)[15] data with IBTrACs data. Only the time period 2004-2013 contained sufficient data for modeling, thus, that is the period used for comparison. We refit Model 3 (WLS) from Section 1.3.3.2 on the available Atlantic data and include all 6-hour observations rather than taking storm maximums to be consistent with the previous methods. However, we recognize this violates an independence assumption in our WLS model, thus, we additionally fit a weighted linear mixed version of our model (WLS MM) which includes a random intercept by storm. After model fitting, we produced original scale predictions by back-transforming predictions[16] on the log scale into the wind speed scale of kts.

We used several metrics to quantify the appeal of our method over the KZ and Holland models. First, in Figure 1.15, we see graphically the actual versus predicted wind speeds and their line of best fit compared to a perfect fit. On average, both KZ and Holland models tend to overpredict more extreme wind speeds, while our models both slightly underpredict.

To quantitatively assess models, we used linear models of true wind speeds regressed on predicted wind speeds for each model setting. Our models perform best in terms of fitting an actual versus predicted linear regression, with root mean squared errors of 7.91 kts (WLS MM) and 9.42 kts (WLS), compared to 9.76 kts (KZ) and 10.77 kts (Holland).

[15] http://www.nhc.noaa.gov/data/hurdat/hurdat2-1851-2014-050515.txt

[16] Predictions in WLS MM were made using fixed effects only.

As a follow up, we implemented hypothesis tests in `linearHypothesis{car}`(Fox and Weisberg, 2011) which provides F-tests for the linear models of true wind speeds regressed on predicted wind speeds. We tested three separate hypotheses: (i) $H_0 : \beta_0 = 0$, (ii) $H_0 : \beta_1 = 1$, and (iii) (i) $H_0 : (\beta_0, \beta_1) = (0, 1)$. Hypothesis (i) identifies if there is an upward or downward bias in the prediction, hypothesis (ii) examines prediction correlation with observed data, and hypotheses (iii) more robustly tests (i) and (ii) simultaneously.

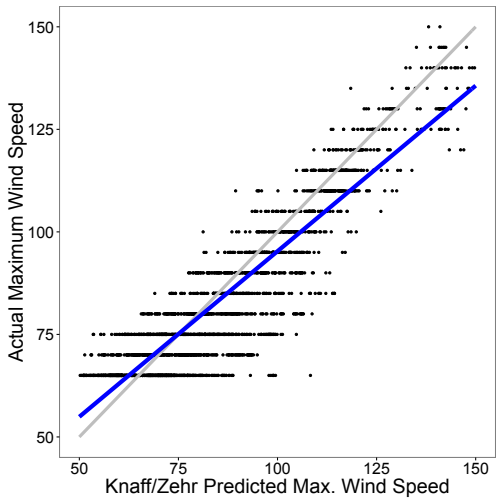
At the $\alpha = 0.05$ level, all true vs. predicted models rejected hypothesis (iii), however, both the WLS and WLS mixed model failed to reject hypothesis (i) indicating a lack of detectable bias. Based on the coefficients and standard errors (SEs) in Table 1.8, both KZ and Holland true vs. predicted models showed clear overprediction biases of 8 kts and 14.5 kts, respectively. Prediction accuracy, as defined by the slope of the regression, is statistically similar to 1 with the WLS mixed model. Even with failed assumptions in the initial model, the assessment of the WLS true vs. predicted model for hypothesis (ii) indicates a p-value of 0.0130, which is much less extreme than the p-values of <0.0001 produced by the corresponding Holland and KZ true vs. predicted models. Therefore, the form of our model shows promise in being a useful explanatory and predictive tool for understanding the MaxWS-MinCP relationship and, as shown here, surpasses some of the existing methodologies in prediction value.

Table 1.8: Actual Wind Speed vs. Predicted Wind Speed Regression Estimates

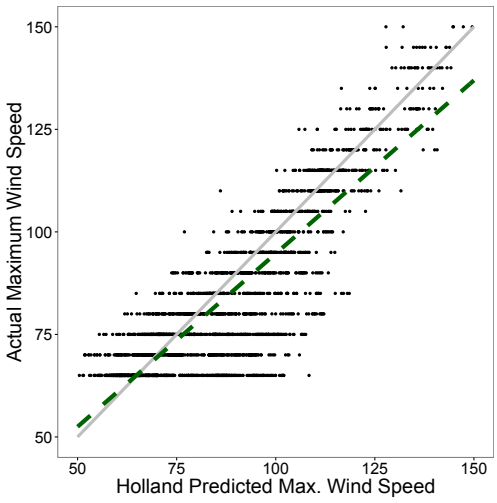
Method	$\hat{\beta}_0$ (SE $\hat{\beta}_0$)	$\hat{\beta}_1$ (SE $\hat{\beta}_1$)	$H_0 : \beta_0 = 0$	$H_0 : \beta_1 = 1$	$H_0 : (\beta_0, \beta_1) = (0, 1)$
KZ	14.542 (0.689)	0.807 (0.008)	< 0.0001	< 0.0001	< 0.0001
Holland	8.015 (0.951)	0.881 (0.011)	< 0.0001	< 0.0001	< 0.0001
WLS	-1.784 (0.952)	1.028 (0.011)	0.0609	0.0130	0.0008
WLS MM	-0.797 (0.760)	1.014 (0.009)	0.2948	0.1130	0.0155

Figure 1.15: Atlantic Wind Speed Model Comparison. The grey 45° line indicates a perfect fit of predicted to actual. The colored and bolded line indicates the result of a least squares fit of actual wind speeds regressed on predicted wind speeds.

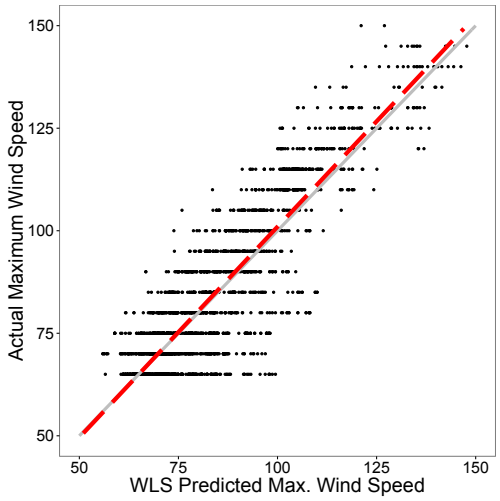
(a) KZ Model



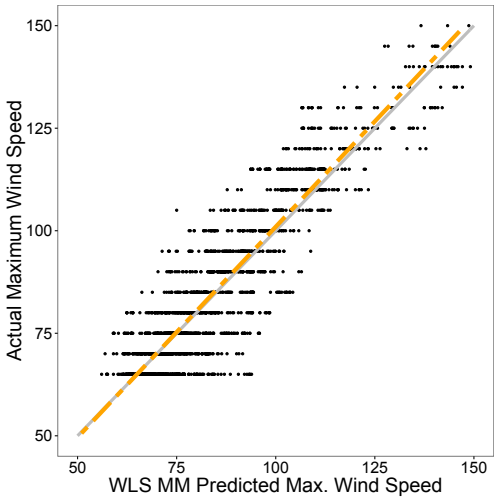
(b) Holland Model



(c) WLS Model



(d) WLS Mixed Model



1.4 Modeling the MaxWS-MinCP Relationship with non-stationary GEV Models

1.4.1 GEV Location Models using Cyclostrophic Assumptions

Theoretically, the assumption that maxima would follow a normal distribution is unlikely. This was suggested empirically in Section 1.3.3. However, we also saw there was prediction merit to using a transformed and augmented version cyclostrophic approximation. One way we could use this knowledge without implicit normality assumptions is modeling log(MaxWS) using a GEV model. Two versions of the GEV model, *stationary* and *non-stationary*, are described in (Coles, 2001). The stationary GEV combines the three types of extreme value distributions into a single density characterized by location parameter μ, scale parameter σ, and shape parameter ξ. A modification to the density allowing any of the parameters to become functions of available covariates rather than fixed constants. This is referred to as a non-stationary model.

We estimated a non-stationary GEV using maximum likelihood estimates (MLEs) generated by `fevd{extRemes}` in R (Gilleland and Katz, 2011). Let Y denote log(MaxWS) for a tropical cyclone. Then, the non-stationary GEV probability density function (pdf) is

$$f(y|\mu,\sigma,\xi) = \frac{1}{\sigma} t(y)^{\xi+1} e^{-t(y)} I(\sigma > 0); \tag{1.46}$$

$$t(y) = \begin{cases} \left[1 + \xi\left(\frac{y-\mu(\Delta P,\phi_{Avg},V_{tsAvg})}{\sigma}\right)\right]^{-1/\xi}; & \xi \neq 0 \\ \exp\left\{-\frac{y-\mu(\Delta P,\phi_{Avg},V_{tsAvg})}{\sigma}\right\}; & \xi = 0 \end{cases}. \tag{1.47}$$

The mean of this distribution is

$$\mathbb{E}(Y) = \begin{cases} \mu(\Delta P, \phi_{Avg}, V_{tsAvg}) + \sigma\xi; & \xi = 0 \\ \mu(\Delta P, \phi_{Avg}, V_{tsAvg}) + \sigma\frac{\Gamma(1-\xi)-1}{\xi}; & \xi \neq 0, \xi < 1 \\ \infty; & \xi \geq 1 \end{cases} . \tag{1.48}$$

The variance is

$$\mathbb{V}(Y) = \begin{cases} \sigma^2\frac{\pi^2}{6}; & \xi = 0 \\ \frac{\sigma^2}{\xi^2}\{\Gamma(1-2\xi) - [\Gamma(1-\xi)]^2\}; & \xi \neq 0, \xi < 0.5 \\ \infty; & \xi \geq 0.5 \end{cases} , \tag{1.49}$$

where $\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}dx$.

We initially tested a stationary GEV, i.e. no covariates, against a Gumbel distribution, effectively conducting the test of $H_0 : \xi = 0$ vs. $H_0 : \xi \neq 0$. This likelihood ratio test (LRT) produced a p-value of < 0.0001 indicating GEV was a better fit for log(MaxWS). Next, as in Section 1.3.3, we used data from all basins with a categorical variable for basin to fit linear location parameters. The three models considered were:

$$\textbf{Model 1: } \mu_1(\Delta P) = \beta_{0B} + \beta_{1B}\log(\Delta P), \tag{1.50}$$

$$\textbf{Model 2: } \mu_2(\Delta P) = \beta_{0B} + \beta_{1B}\log(\Delta P) + \beta_{2B}\log^2(\Delta P), \tag{1.51}$$

$$\begin{aligned} &\textbf{Model 3: } \mu_3(\Delta P, \phi_{Avg}, V_{tsAvg}) = \beta_{0B} + \beta_{1B}\log(\Delta P) \\ &+ \beta_{2B}\log^2(\Delta P) + \beta_{3B}\phi_{Avg} + \beta_{4B}V_{tsAvg}. \end{aligned} \tag{1.52}$$

1.4.2 GEV Location Model Results Discussion

Fitted model estimates are in Table 1.9. The estimated values of ξ are negative indicating a reversed Weibull distribution. This type of extreme value distribution

is bounded above which makes sense based on our use of log as well as wind speed limits given a certain set of climate conditions.

Empirical density is compared to each modeled density in Figure 1.16. This qualitative measure indicates that Model 1 seems to match the data distribution better than the more complex models. However, Table 1.10 indicates Model 2 is preferred based on BIC and Model 3 is preferred based on AIC and the LRTs. Table 1.11 also indicates that Model 3 has the best prediction errors.

Again, we used the asymptotic standard error and MooN bootstrapping and subsampling to calculate standard errors of basin-wise parameters. Standard asymptotic properties of MLEs hold if $\xi > -0.5$ (Smith, 1985) which we will assume is viable based on the estimates. Figure 1.17–Figure 1.19 provide standard errors for each of the models. As seen in these figures, the asymptotic standard error matches much more closely to the bootstrapping errors than they did in the WLS models. They also indicate many more significant effects in modeling the location parameter than were seen in the WLS. For instance, in Model 2, all basins other than the East Pacific and Southeast Indian show significant quadratic terms of MinCP. Similarly, in Model 3, the effect of translation speed and latitude becomes more pronounced in several basins.

Predicted returns for each model are back transformed to the original scale and displayed in Figure 1.20–Figure 1.22. A return level with a return period of $1/s$ storms is a threshold in the log(MaxWS) distribution whose probability of exceedance is s. For example, if $s = 0.01$, the corresponding return period is 100 storms. Each return level prediction is based on the covariates included which gives a distribution of return levels displayed by the boxplots. The Saffir-Simpson tropical cyclone categorization scale is overlaid on the plots to indicate the intensity of the predicted storms. Model 3 shows more high boxplot outliers than either Model 1 or 2. Based on predictions from each model, a Category 3 or higher

storm should be expected about 50% of the time within a 5 tropical cyclone return period regardless of basin. These returns may seem high, however, in a warming climate, faster wind speeds may be achievable. (Emmanuel, 2005) predicted that maximum wind speeds should increase about 5% for every 1°C increase in tropical ocean temperatures. A recent example of the possible climate change effect is the wind speed record of nearly 174 kts set by Hurricane Patricia near the coast of Mexico reached speeds on Oct. 23, 2015.

Table 1.9: Non-stationary GEV Location Estimates

M	Basin	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
1	NA	2.680	0.470			
	EP	2.436	0.532			
	WP	2.235	0.548			
	NI	2.382	0.537			
	SP	1.892	0.618			
	SEI	1.906	0.618			
	SWI	2.115	0.567			
2	NA	5.720	-1.163	0.217		
	EP	2.856	0.314	0.028		
	WP	7.260	-1.918	0.300		
	NI	6.990	-1.810	0.296		
	SP	6.872	-1.804	0.293		
	SEI	3.342	-0.077	0.084		
	SWI	7.063	-1.859	0.296		
3	NA	5.634	-1.059	0.202	-0.003	0.001
	EP	2.966	0.330	0.026	-0.007	-0.002
	WP	7.106	-1.832	0.290	-0.002	0.003
	NI	6.747	-1.696	0.279	0.005	-0.006
	SP	6.833	-1.773	0.289	0.001	0.000
	SEI	3.479	-0.148	0.092	-0.001	0.002
	SWI	7.453	-2.026	0.317	0.000	-0.015

Table 1.10: Non-stationary GEV Scale and Shape Estimates, Fit Statistics, and LRTs

M	$\hat{\sigma}_M$	$\hat{\xi}_M$	AIC	BIC	LRT v. M1	LRT v. M2
1	0.1018	-0.1615	-3292	-3202	-	-
2	0.0943	-0.1755	-3604	**-3475**	<0.0001	-
3	0.0915	-0.1670	**-3667**	-3456	<0.0001	<0.0001

Table 1.11: Average Prediction Error for GEV MooN Parametric Bootstrapping

Error	MooN p	Model 1		Model 2		Model 3	
		Sbsmp.	w/Rep.	Sbsmp.	w/Rep.	Sbsmp.	w/Rep.
MSPE	2/3	0.0130	0.0127	0.0108	0.0108	0.0108	0.0104
	3/4	0.0122	0.0124	0.0109	0.0105	**0.0102**	0.0103
	1	-	0.0124	-	0.0108	-	0.0108
MAPE	2/3	0.0858	0.0840	0.0767	0.0770	0.0762	**0.0736**
	3/4	0.0825	0.0823	0.0769	0.0755	0.0738	0.0747
	1	-	0.0828	-	0.0765	-	0.0769

Figure 1.16: Non-stationary GEV Fitted Density v. Empirical Density

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

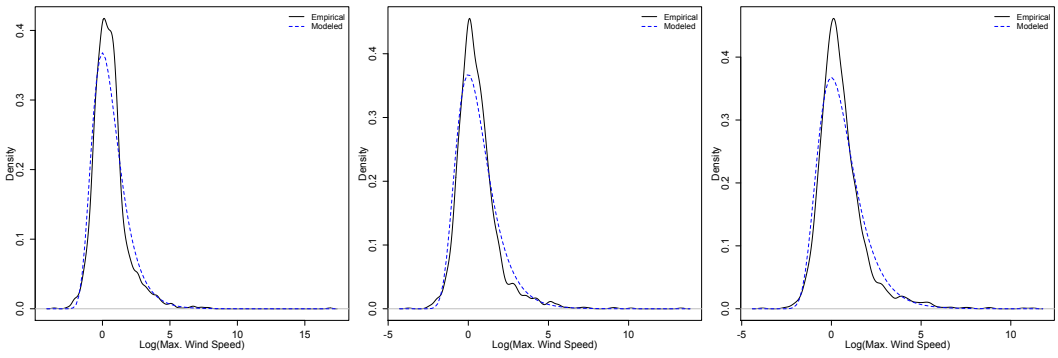


Figure 1.17: Estimate +/- 2 Standard Errors for Parameters in GEV Model 1. In all figures, the lines are defined as follows:
(1 - Purple Long Dashed): Asymptotic Standard Error;
(2 - Blue Solid): MooN bootstrap w/replacement, $p = 1.00$;
(3 - Green Short Dashed): MooN bootstrap subsampling, $p = 0.75$;
(4 - Green Solid): MooN bootstrap w/ replacement, $p = 0.75$;
(5 - Red Short Dashed): MooN bootstrap subsampling, $p = 0.66$;
(6 - Red Solid): MooN bootstrap w/ replacement, $p = 0.66$.

(a) β_0 (b) β_1 (c) σ and ξ

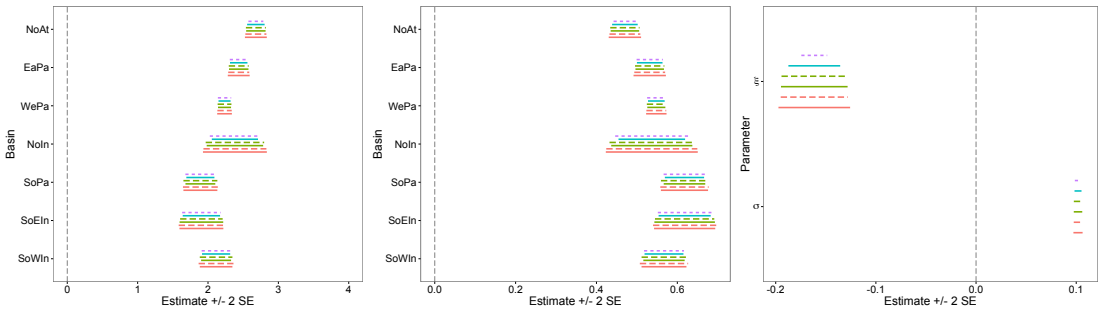
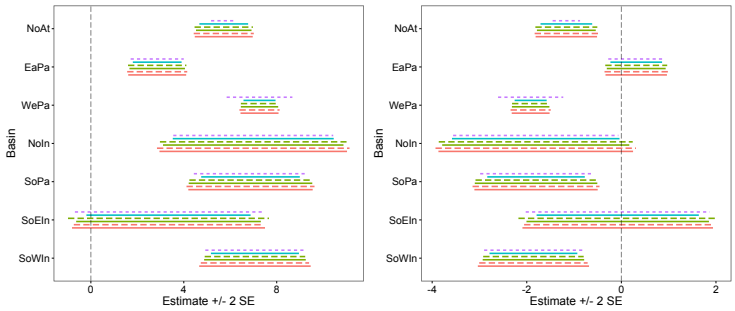


Figure 1.18: Estimate +/- 2 Standard Errors for Parameters in GEV Model 2. In all figures, the lines are defined as follows:
(1 - Purple Long Dashed): Asymptotic Standard Error;
(2 - Blue Solid): MooN bootstrap w/replacement, $p = 1.00$;
(3 - Green Short Dashed): MooN bootstrap subsampling, $p = 0.75$;
(4 - Green Solid): MooN bootstrap w/ replacement, $p = 0.75$;
(5 - Red Short Dashed): MooN bootstrap subsampling, $p = 0.66$;
(6 - Red Solid): MooN bootstrap w/ replacement, $p = 0.66$.

(a) β_0 (b) β_1



(c) β_2 (d) σ and ξ

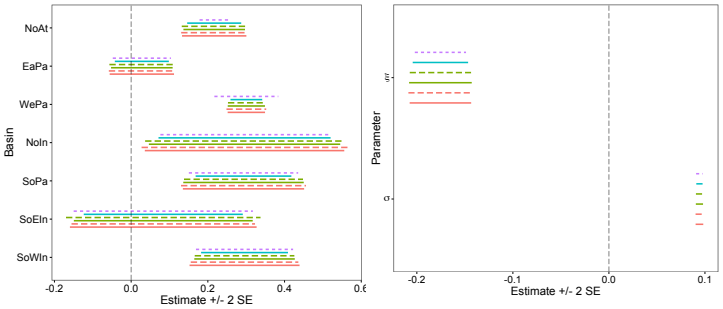


Figure 1.19: Estimate +/- 2 Standard Errors for Parameters in GEV Model 3. In all figures, the lines are defined as follows:
(1 - Purple Long Dashed): Asymptotic Standard Error;
(2 - Blue Solid): MooN bootstrap w/replacement, $p = 1.00$;
(3 - Green Short Dashed): MooN bootstrap subsampling, $p = 0.75$;
(4 - Green Solid): MooN bootstrap w/ replacement, $p = 0.75$;
(5 - Red Short Dashed): MooN bootstrap subsampling, $p = 0.66$;
(6 - Red Solid): MooN bootstrap w/ replacement, $p = 0.66$.

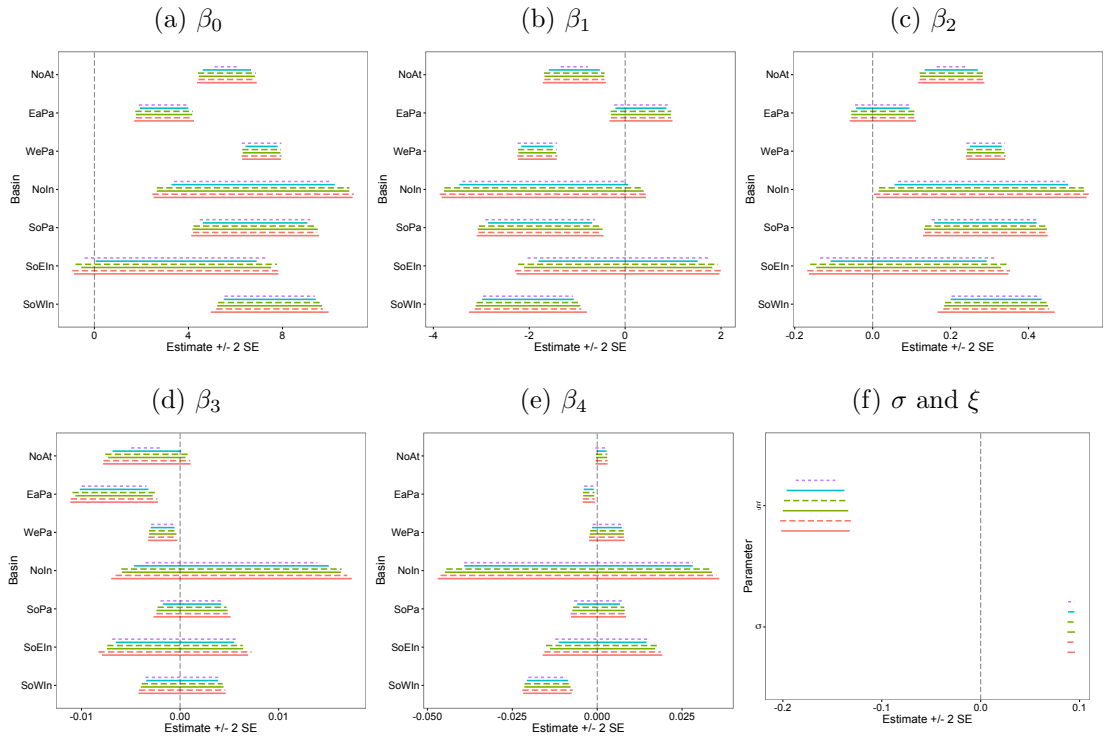


Figure 1.20: Predicted Returns for Non-stationary GEV Model 1

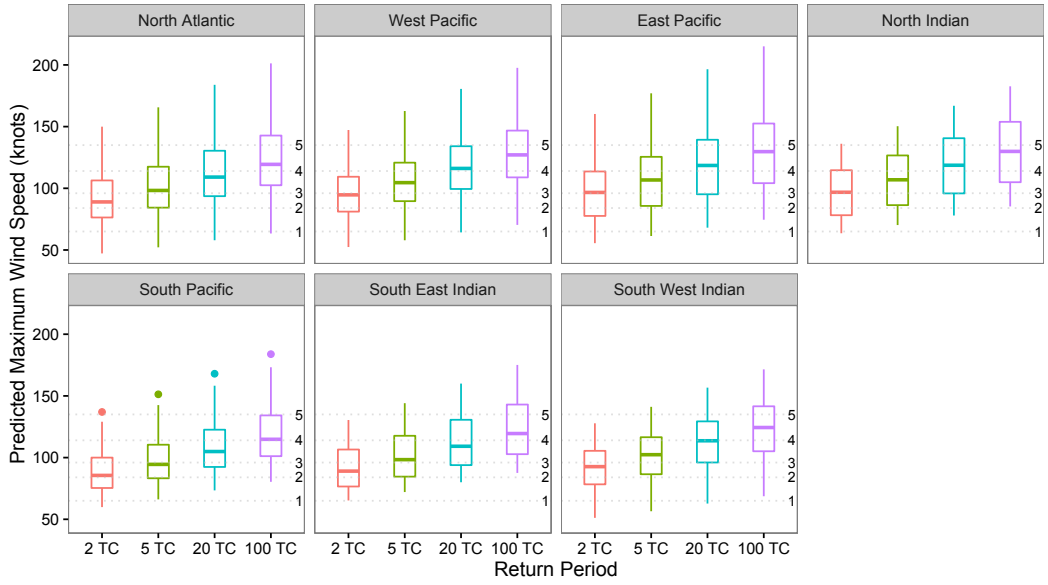


Figure 1.21: Predicted Returns for Non-stationary GEV Model 2

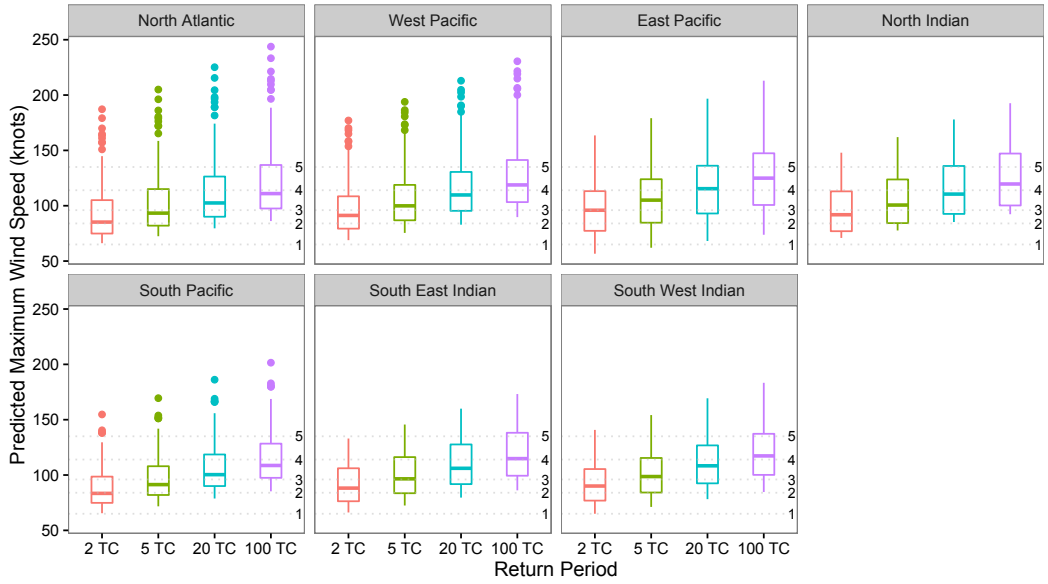
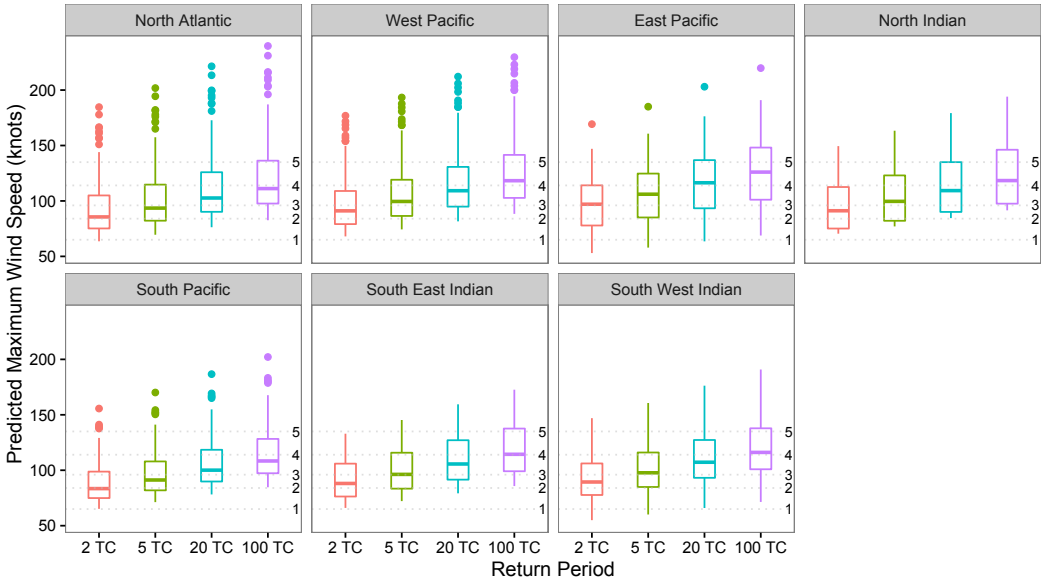


Figure 1.22: Predicted Returns for Non-stationary GEV Model 3



1.5 Bivariate Extreme Value Models for the MaxWS-MinCP Relationship

We propose that MaxWS and MinCP should be studied jointly, not just in a conditional form as we've done in the previous sections. The previous sections did not elucidate one important feature of the data. Intensity, as characterized by MaxWS and indirectly, MinCP, has a bimodal distribution (Holland and Bruyère, 2014). There may also be potential changes in the joint distribution over time that a conditional model would not be able to capture. To further understand the dynamics of this intensity relationship, the bivariate uncertainty of MaxWS and MinCP needs to be studied.

1.5.1 Bivariate Extreme Value Distributions for MaxWS-MinCP

Several forms of bivariate extreme value distributions are available. Models can be fit using MLEs within `fbvevd{evd}` in R (Stephenson, 2002). This software provides 9 possible functional forms of bivariate extreme value distributions. Preliminary analysis in Appendix A.1 based on qualitative comparisons to a nonparametric fit and AIC, indicated a bilogistic form to be the most relevant for this data. The bilogistic extreme value distribution function (Smith, 1990; Joe et al.,

1992) has the following parameterization:

$$G(y_1, y_2) = \exp\left\{-y_1 q^{1-\alpha} - y_2(1-q)^{1-\beta}\right\} \tag{1.53}$$

$$y_1(x_1) = \left\{1 + \frac{\xi_1(x_1 - \mu_1)}{\sigma_1}\right\}^{-1/\xi_1} I(\sigma_1 > 0) \tag{1.54}$$

$$y_2(x_2) = \left\{1 + \frac{\xi_2(x_2 - \mu_2)}{\sigma_1}\right\}^{-1/\xi_2} I(\sigma_2 > 0) \tag{1.55}$$

where $q = q(y_1, y_2; \alpha, \beta)$ is the root of the equation

$$(1-\alpha)y_1(1-q)^{\beta} - (1-\beta)y_2 q^{\alpha} = 0 \tag{1.56}$$

$$0 < \alpha, \beta < 1. \tag{1.57}$$

Each margin of the joint distribution is a GEV distribution with location (μ_1, μ_2), scale (σ_1, σ_2), and shape (ξ_1, ξ_2) parameters. If either shape parameter is 0, the marginal reduces to a Gumbel distribution defined by continuity. One useful aspect of this bivariate distribution is the dependence parameters, α and β. The combination of these values gives an estimate of asymmetry, $\alpha - \beta$, which can be used to test for a reduction to a logistic extreme value model with a single dependence parameter. They can also be used to get a total dependence estimate, $(\alpha + \beta)/2$, where 0 indicates complete dependence and 1 indicates independence. If $\alpha = \beta \to 1$, then

$$G(y_1, y_2) = \exp\{-y_1 - y_2\} = \exp\{-y_1\}\exp\{-y_2\} = G_1(y_1)G_2(y_2) \tag{1.58}$$

implying independence between the two margins. Similarly, if α fixed between 0

and 1 and $\beta \to 1$, (1.56) becomes

$$(1-\alpha)y_1(1-q) = 0 \tag{1.59}$$

implying $q = 1$, and again, gives the result in (1.58). The same argument holds for β fixed fixed between 0 and 1 and $\alpha \to 1$. Complete dependence is obtained in the limit as $\alpha = \beta \to 0$.

Similar to the univariate GEV models in Section 1.4, we used location parameter modeling. In order to maintain simplicity, yet still reflect basin differences, we fit separate location parameters by basin. Let x_{1B} be log(MaxWS) and x_{2B} be log(1013−MinCP) in basin B. The model is then:

$$G(y_{1B}, y_{2B}) = \exp\left\{-y_{1B}q^{1-\alpha} - y_{2B}(1-q)^{1-\beta}\right\} \tag{1.60}$$

$$y_{1B}(x_{1B}) = \left\{1 + \frac{\xi_1(x_{1B}-\mu_{1B})}{\sigma_1}\right\}^{-1/\xi_1} I(\sigma_1 > 0) \tag{1.61}$$

$$y_{2B}(x_{2B}) = \left\{1 + \frac{\xi_2(x_{2B}-\mu_{2B})}{\sigma_2}\right\}^{-1/\xi_2} I(\sigma_2 > 0) \tag{1.62}$$

MLEs are displayed in Table 1.12. Both marginal distributions are estimated as reversed Weibull GEVs indicated by their negative shape parameters. A LRT again a single location parameter model indicated location parameters by basin were necessary with $p < 0.0001$. A second LRT indicated α and β are significantly different ($p < 0.0001$) denoting an estimated asymmetry, $\hat{\alpha} - \hat{\beta}$, of 0.116 in the dependence structure of the Bivariate Extreme Value Distribution (BVEVD). The dependence estimate, $(\hat{\alpha} + \hat{\beta})/2$, is 0.273. As expected, this implies strong dependence between log(MaxWS) and log(1013-MinCP).

Figure 1.23, Figure 1.24, and Figure 1.25 show results of simulations from the bilogistic model by basin compared to the original data. These plots indicate

reasonable coverage of the data area, however, indicate that basin alone may not capture some aspects of the data. For example, some of the basins show bimodality in their actual density contours.

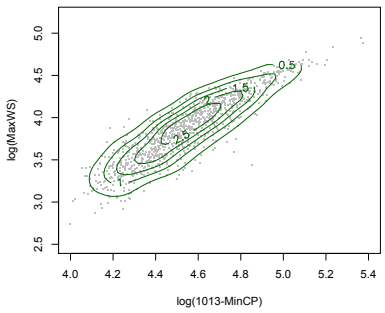
Uncertainty estimates based on the asymptotic standard error and MooN bootstrapping and subsampling are available in Figure 1.26. The bootstrapping and subsampling standard error estimates are slightly larger than those of the asymptotic standard errors. There is a fair amount of overlap in the intervals for the different basins in the log(MaxWS) margin although the estimates of wind speed tend to be higher in the northern hemisphere basins. The dependence parameters α and β indicate a large amount of uncertainty.

Table 1.12: Bilogistic MaxWS-MinCP Model Estimates

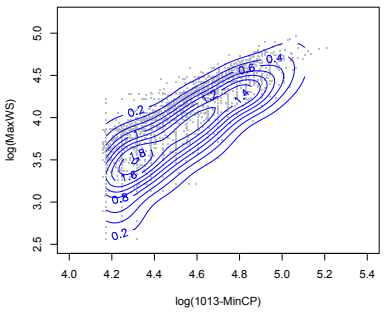
Basin	log(MaxWS)		log(1013-MinCP)	
	$\hat{\mu}_1$	$\hat{\sigma}_1$	$\hat{\mu}_2$	$\hat{\sigma}_2$
NA	4.438	0.217	3.702	0.383
EP	4.437	$\hat{\xi}_1$	3.755	$\hat{\xi}_2$
WP	4.462	-0.156	3.985	-0.266
NI	4.450		3.826	
SP	4.378	$\hat{\alpha}$	3.941	$\hat{\beta}$
SEI	4.390	0.331	3.945	0.215
SWI	4.416		3.971	

Figure 1.23: North Atlantic and East Pacific BVEVD MaxWS-MinCP Model Simulations. The left column shows 1000 simulated values based on estimated parameters overlaid with density contours. The right column shows the actual data overlaid with density contours.

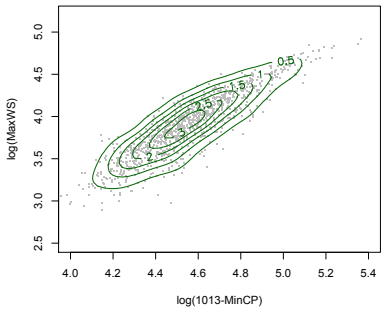
(a) North Atlantic Simulated



(b) North Atlantic Actual



(c) East Pacific Simulated



(d) East Pacific Actual

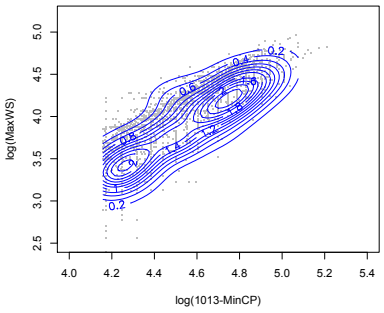
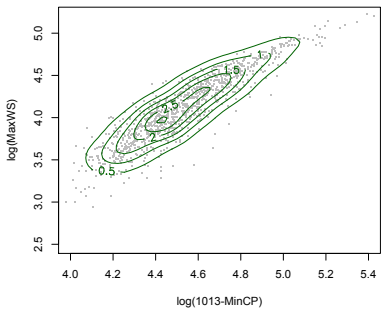
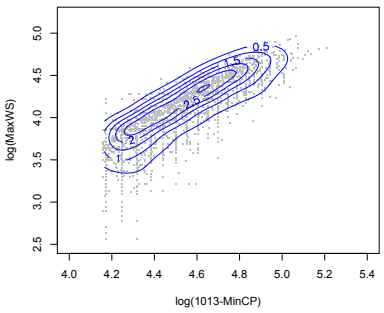


Figure 1.24: West Pacific and North Indian BVEVD MaxWS-MinCP Model Simulations. The left column shows 1000 simulated values based on estimated parameters overlaid with density contours. The right column shows the actual data overlaid with density contours.

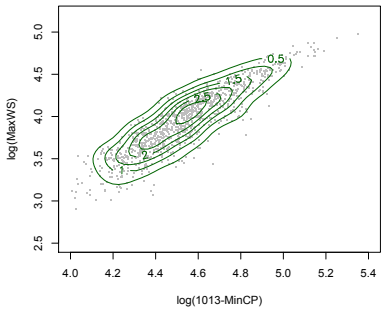
(a) West Pacific Simulated



(b) West Pacific Actual



(c) North Indian Simulated



(d) North Indian Actual

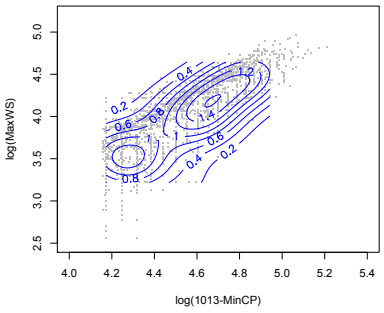
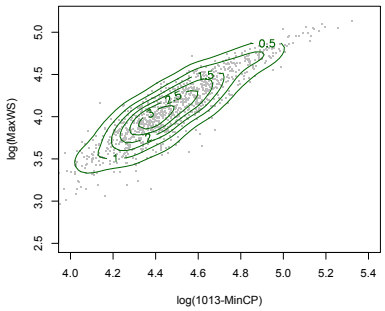
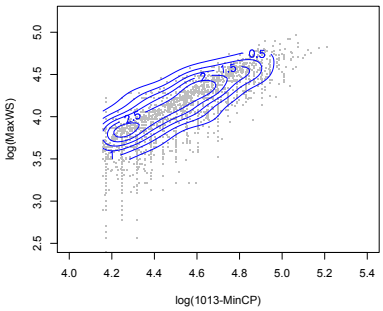


Figure 1.25: Southern Hemisphere BVEVD MaxWS-MinCP Model Simulations. The left column shows 1000 simulated values based on estimated parameters overlaid with density contours. The right column shows the actual data overlaid with density contours.

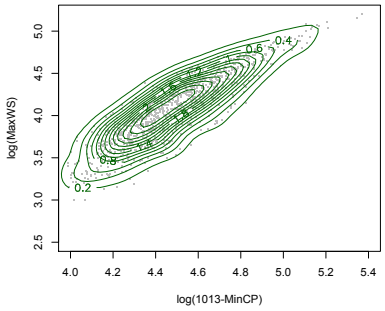
(a) South Pacific Simulated



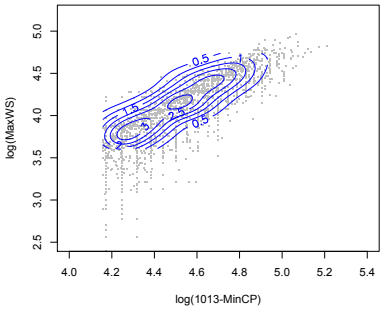
(b) South Pacific Actual



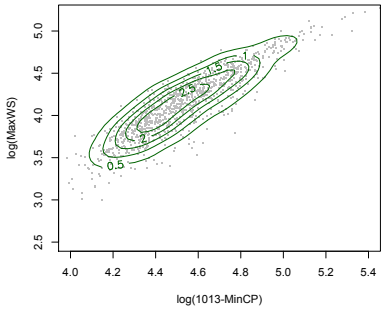
(c) South East Indian Simulated



(d) South East Indian Actual



(e) South West Indian Simulated



(f) South West Indian Actual

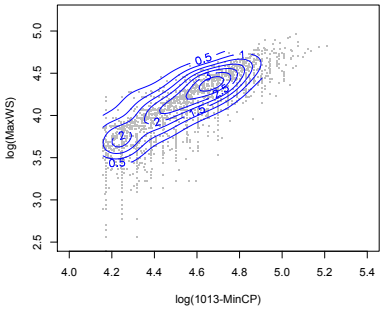
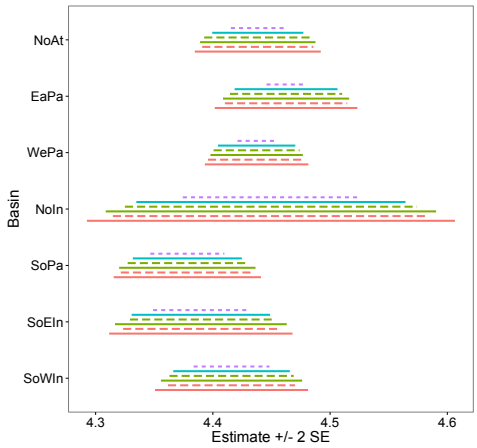
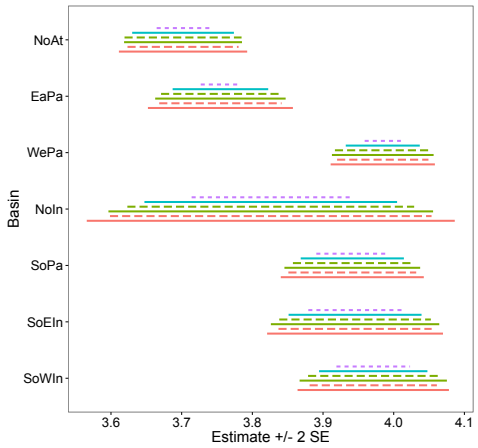


Figure 1.26: Estimated Standard Errors BVEVD Fit. In all figures, the lines are defined as follows:
(1 - Purple Long Dashed): Asymptotic Standard Error;
(2 - Blue Solid): MooN bootstrap w/ replacement, $p = 1.00$;
(3 - Green Short Dashed): MooN bootstrap subsampling, $p = 0.75$;
(4 - Green Solid): MooN bootstrap w/ replacement, $p = 0.75$;
(5 - Red Short Dashed): MooN bootstrap subsampling, $p = 0.66$;
(6 - Red Solid): MooN bootstrap w/ replacement, $p = 0.66$.

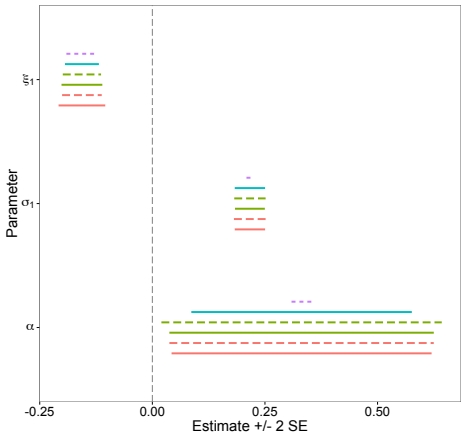
(a) log(MaxWS) Location (μ_1)



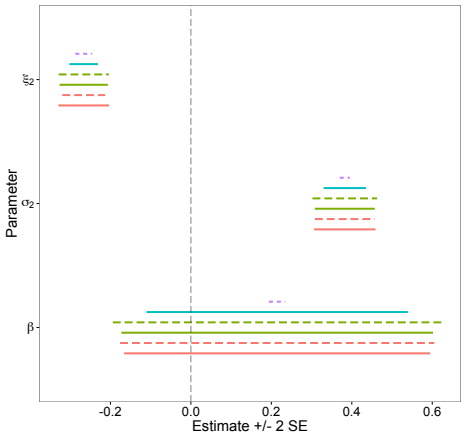
(b) log(1013-MinCP) Location (μ_2)



(c) log(MaxWS) Scale, Shape, and Dependence (ξ_1, σ_1, α)



(d) log(1013-MinCP) Scale, Shape, and Dependence (ξ_2, σ_2, β)



1.5.2 Using a Bivariate Extreme Value Model to Estimate Damages in the Atlantic Basin

One possible use for the bivariate extreme connection in the MaxWS and MinCP relationship is to use it in predicting tropical cyclone damages. To do so, we utilize damages data from the insurance catastrophe database (ICAT). The ICAT database[17] provides access to historical hurricane damage information. According to their site, all data is obtained from publicly available sources. This database provides fairly comprehensive damages for North Atlantic tropical cyclones from 1900-2014. The site notes that base damage values are retrieved from Monthly Weather Review and that data has also been normalized to 2014 dollars to reflect changes in inflation, wealth, and population in the storm area. Because this database only covers the North Atlantic, we restrict our attention to this basin only. We merge the damages data with the IBTrACs storm data from 1960-2014, and only utilize storms with a corresponding nonzero damage.

Again, we start with a bivariate bilogistic model based on its useful parameterization in the previous section. In order to incorporate the MaxWS-MinCP relationship, we model one margin as the transformed cyclostrophic relationship and the other margin as the damages. Let x_1 be log(MaxWS), z_1 be log(1013-MinCP), and x_2 as log(Damages). Then, our bivariate bilogistic model is:

$$G(y_1, y_2) = \exp\left\{-y_1 q^{1-\alpha} - y_2(1-q)^{1-\beta}\right\} \tag{1.63}$$

$$y_1(x_1, z_1) = \left\{1 + \frac{\xi_1(x_1 - \mu_1(z_1))}{\sigma_1}\right\}^{-1/\xi_1} I(\sigma_1 > 0) \tag{1.64}$$

$$y_2(x_2) = \left\{1 + \frac{\xi_2(x_2 - \mu_2)}{\sigma_2}\right\}^{-1/\xi_2} I(\sigma_2 > 0) \tag{1.65}$$

[17] http://www.icatdamageestimator.com/faq

with q as the root of (1.56). We test three models similar to what was done in prior sections:

$$\textbf{Model 0: } \mu_1(z_1) = \gamma_0 \tag{1.66}$$

$$\textbf{Model 1: } \mu_1(z_1) = \gamma_0 + \gamma_1 z_1 \tag{1.67}$$

$$\textbf{Model 2: } \mu_1(z_1) = \gamma_0 + \gamma_1 z_1 + \gamma_2 z_1^2 \tag{1.68}$$

Model 0 represents a true bivariate relationship for log(MaxWS) and log(Damage) while Models 1 and 2 include the MinCP relationship. Because of the complexity of this model and the questionable usefulness of the translation speed and latitude in the previous sections, we did not include them in the model.

1.5.2.1 Bivariate Results and Analysis

Table 1.13 displays the estimates for each of the models. Dependence estimates for Model 0, Model 1, and Model 2 are 0.597, 0.634, and 0.688, respectively. These values indicate decreasing dependence as polynomial terms of log(1013-MinCP) are added to the model. We also note that ξ_1 is estimated at -0.496 in Model 0 which is very close to -0.5 where MLEs become inconsistent.

The estimates of α and β are large indicating tail independence. As seen in Table 1.14, in Model 0 and Model 2, α was not significantly different from β with p-values of 0.1824 and 0.0524, respectively. Thus, we could effectively reduce the model to the logistic form in these cases. However, Model 1 indicates a significant difference between α and β with a p-value of 0.0095. LRTs indicate Model 1 as the best fit.

Figure 1.27 indicates the fitted versus empirical densities for the models. In Model 0, we see the bimodal shape of log(MaxWS) margin is not well accounted for by the model form. The graphics for Models 1 and 2 display data transformed

to stationarity using the parameters estimated for the location parameter. A separate model fit for the transformed data is used as the comparison density. The transformed value for Model 1 is

$$t_1(\text{MaxWS}) = \log(\text{MaxWS}) - \hat{\gamma}_{11}\log(1013 - \text{MinCP}) \tag{1.69}$$

and in Model 2, the transformation is

$$\begin{aligned} t_2(\text{MaxWS}) = \log(\text{MaxWS}) &- \hat{\gamma}_{11}\log(1013 - \text{MinCP}) \\ &- \hat{\gamma}_{12}\log^2(1013 - \text{MinCP}). \end{aligned} \tag{1.70}$$

Based on these marginal density plots, Model 1 appears to most closely fit the log(MaxWS) echoing the results of the LRTs. The log(Damage) marginal doesn't exhibit meaningful changes over the models but qualitatively appears to fit well in all three.

During model estimation, the optimization was not able to provide asymptotic standard errors, thus, we use only the results of MooN bootstrapping and subsampling to assess the standard errors. These results can be seen in Figure 1.28. Although α and β are restricted between 0 and 1, they varied quite a bit in the bootstrapping process, thus, they have fairly large standard errors. Model 1 standard errors indicate γ_{01} and γ_{11} are both larger than 0, however, in Model 2 only γ_{01} is nonzero. This agrees with the results of the LRT that choose Model 1 as the ideal fit and again shows that γ_{11} could feasibly be the ideal cyclostrophic exponent of 0.5. Variability about the log(MaxWS) scale parameter, σ_1, and the shape parameter, ξ_1, decreased as more terms of MinCP were added to the model. Standard errors in the log(Damage) margin were relatively consistent across models.

Table 1.13: Bilogistic MaxWS-MinCP Damage Model Estimates

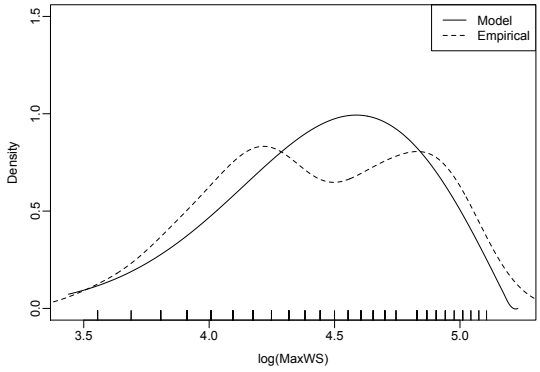
M	log(MaxWS)						log(Damage)			
	$\hat{\gamma}_0$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\sigma}_1$	$\hat{\xi}_1$	$\hat{\alpha}$	$\hat{\mu}_2$	$\hat{\sigma}_2$	$\hat{\xi}_2$	$\hat{\beta}$
0	4.336	-	-	0.431	-0.496	0.696	19.112	2.805	-0.377	0.498
1	2.561	0.494	-	0.154	-0.190	0.931	19.135	2.847	-0.371	0.416
2	2.998	0.230	0.038	0.151	-0.180	0.945	19.138	2.850	-0.372	0.431

Table 1.14: Bilogistic MaxWS-MinCP Damage Model Fit Statistics and Hypothesis Testing. Bolded fit criteria indicate a preference for that model. Parentheses after LRT p-values indicate the preferred model.

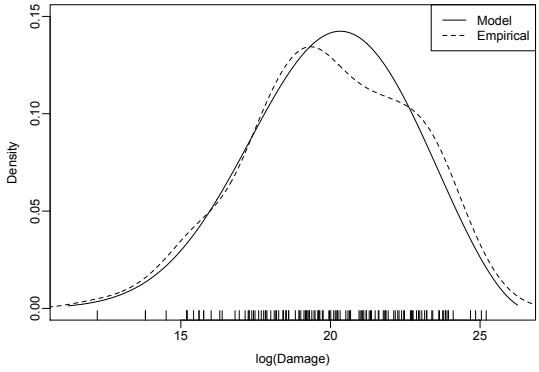
M	AIC	BIC	LRT v. M0	LRT v. M1	LRT $\alpha = \beta$
0	705.11	728.11	-	-	0.1824
1	517.99	**543.87**	<0.0001 (1)	-	0.0095
2	**517.59**	546.35	<0.0001 (2)	0.1220 (1)	0.0524

Figure 1.27: Fitted and Empirical Densities for the Marginals of the Bivariate Bilogistic Models. In Models 1 and 2, the marginal for log(MaxWS) shows the transformed data and the fit for that transformed data. The transformations are made using the estimated model parameters.

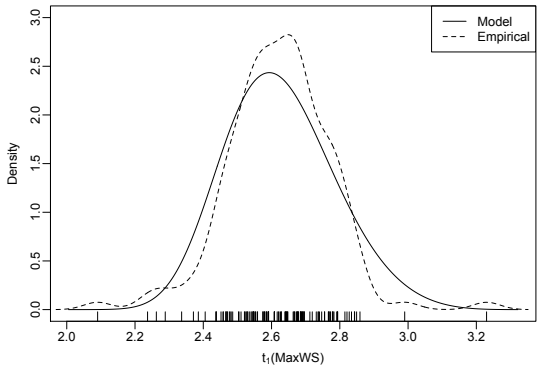
(a) Model 0: log(MaxWS)



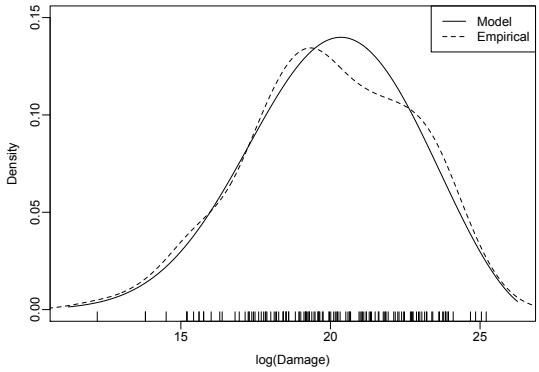
(b) Model 0: log(Damage)



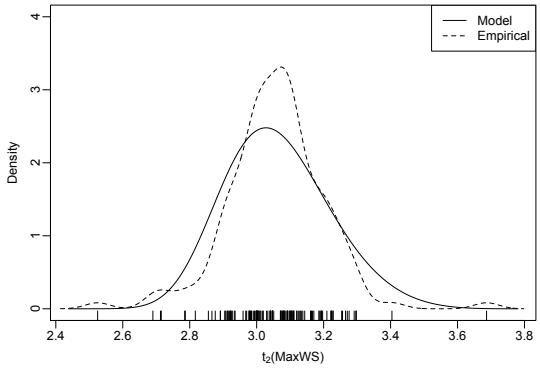
(c) Model 1: log(MaxWS)



(d) Model 1: log(Damage)



(e) Model 2: log(MaxWS)



(f) Model 2: log(Damage)

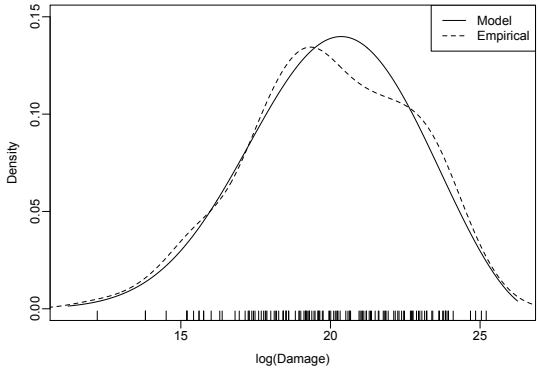
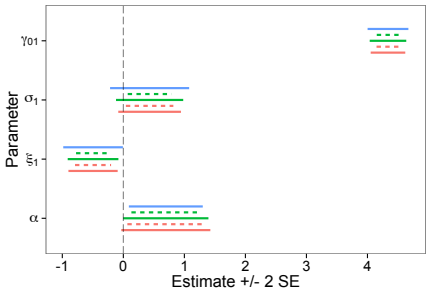
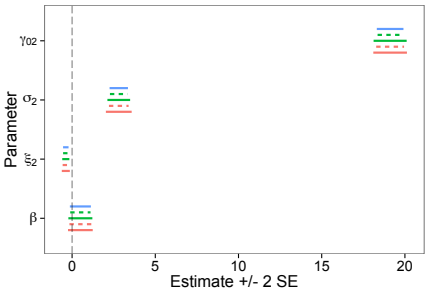


Figure 1.28: Estimate ± 2 Standard Errors for Parameters in BVEVD Damage Model Fitting. In all figures, the lines are defined as follows:
(1 - Blue Solid): MooN bootstrap w/replacement, $p = 1.00$;
(2 - Green Short Dashed): MooN bootstrap subsampling, $p = 0.75$;
(3 - Green Solid): MooN bootstrap w/ replacement, $p = 0.75$;
(4 - Red Short Dashed): MooN bootstrap subsampling, $p = 0.66$;
(5 - Red Solid): MooN bootstrap w/ replacement, $p = 0.66$.

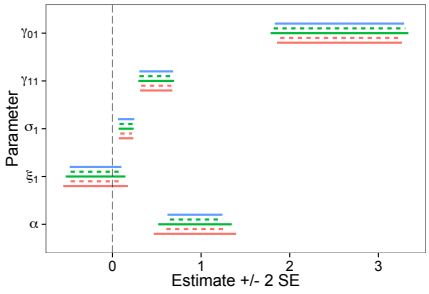
(a) Model 0: log(MaxWS)



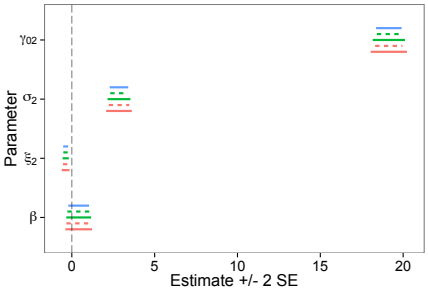
(b) Model 0: log(Damage)



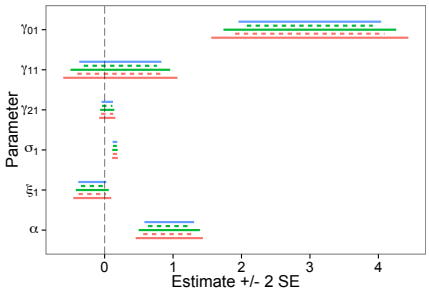
(c) Model 1: log(MaxWS)



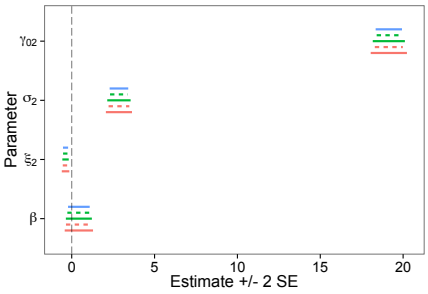
(d) Model 1: log(Damage)



(e) Model 2: log(MaxWS)



(f) Model 2: log(Damage)



1.5.2.2 Prediction Using the Bivariate Extreme Value Model to Estimate Damages in the Atlantic Basin

To assess the predictive power of this bivariate model, we fit models only using data from 1960-2008, then simulated from the fitted distributions for the 6 damage-causing Atlantic tropical cyclones in 2010-2012[18]. Since Model 0 was deemed a poor fit in the previous section, we use only Model 1 and Model 2. To specify the correct distribution, we input the corresponding storm minimum central pressure value to the model, and simulate from the resulting bivariate bilogistic model for log(MaxWS) and log(Damage).

One measure of how closely the observed values of log(MaxWS) and log(Damage) fit into the simulated distribution, is the Mahalonobis distance (Mahalanobis, 1936). We used the following algorithm for the 6 storms available in the data from 2010-2012:

1. Generate K bivariate bilogistic samples based on plugging in the true MinCP for the storm.

2. Calculate the Mahalonobis distance, $\sqrt{(x_i - \bar{x})^T S^{-1} (x_i - \bar{x})}$ where $\bar{x}$ is the sample mean and S is the sample covariance of the K simulated points, for all $K + 1$ points.

3. Order all the distances. Let the ordered index of the observed data point be i, and define the p-value as: $2 \min\left(\frac{i}{K+1}, \frac{K+1-i}{K+1}\right)$.

4. Repeat Steps 1-3 R times.

We used $K = 999$ and $R = 20$ in running this algorithm. Results of one simulation for each 2010 storm are seen in Figure 1.29 and 2011-12 storms are seen

[18]No damage causing storms occurred in 2009.

Table 1.15: Bilogistic Predictions for 2010-2012 Atlantic Hurricanes. Middle 95% Quantiles indicates the 0.025 and 0.975 quantiles of the generated p-values.

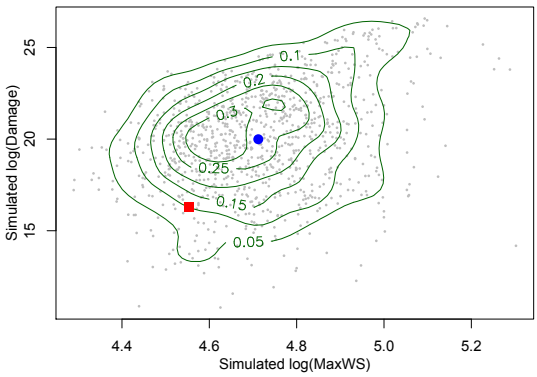
Year	Storm	Model 1		Model 2	
		Mid. 95% Qnt.	$\%p \leq 0.05$	Mid. 95% Qnt.	$\%p \leq 0.05$
2010	Alex	(0.061, 0.947)	5%	(0.083, 0.941)	5%
	Earl	(0.063, 0.922)	5%	(0.056, 0.752)	5%
	Hermine	(0.092, 0.950)	0%	(0.083, 0.887)	5%
2011	Irene	(0.025, 0.945)	10%	(0.147, 0.969)	0%
2012	Isaac	(0.045, 0.939)	5%	(0.057, 0.916)	5%
	Sandy	(0.062, 0.841)	5%	(0.076, 0.989)	0%

in Figure 1.30. Density contours are overlaid on each of the simulations using bivariate normal kernels implemented via `kde2d{MASS}`(Venables and Ripley, 2002). Table 1.15 summarizes the results of the 20 runs by providing the quantiles of the p-values and the percent of p-values less than or equal to 0.05. By random chance alone, we expect 5% of p-values to be 0.05 or below, however, if this percentage was much higher, it would indicate the observed data was an outlier and the model is not useful in prediction.

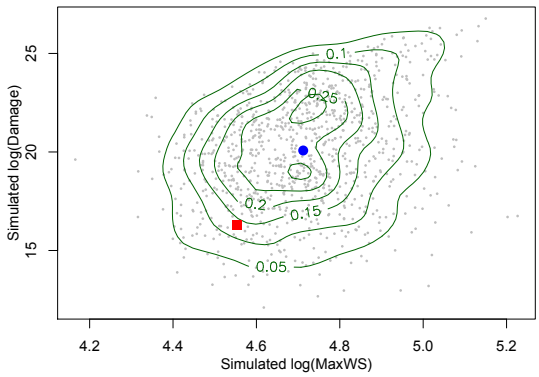
In general, both models do relatively well for many of the storms based on the metrics in Table 1.15. All of the 2010 hurricanes seem to fit into the simulated distribution in a reasonable way based on their p-value ranges. Model 2 seems to provide a better fit for the 2011-12 hurricanes in terms of the range of p-values. This is especially noticeable for Hurricane Irene when comparing the model simulations in Figure 1.30. Although Hurricane Sandy does well on the p-value metrics, it does seem to be an outlier in the log(Damage) space. However, it is still within a reasonable distance from the predicted average of the bivariate distribution.

Figure 1.29: Simulated Bivariate Bilogistic Models for 2010 Hurricanes. The blue circle is the sample mean of the simulated data and the red square is the observed value for the storm of interest.

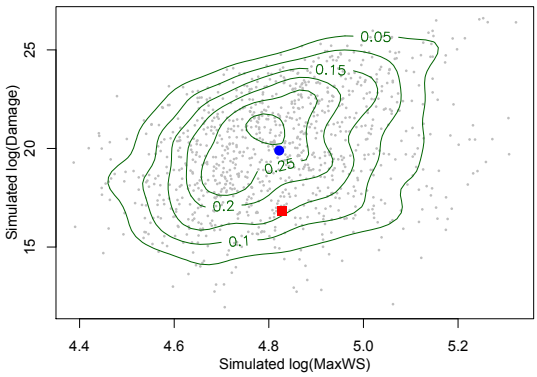
(a) Model 1: Hurricane Alex



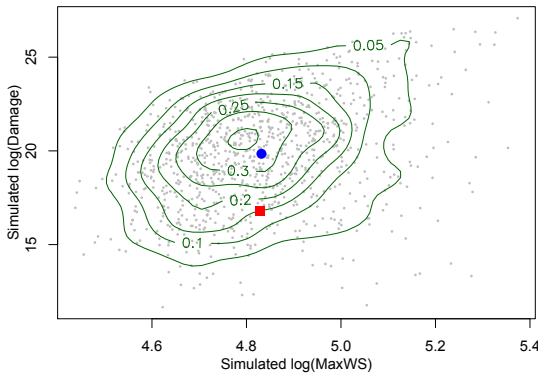
(b) Model 2: Hurricane Alex



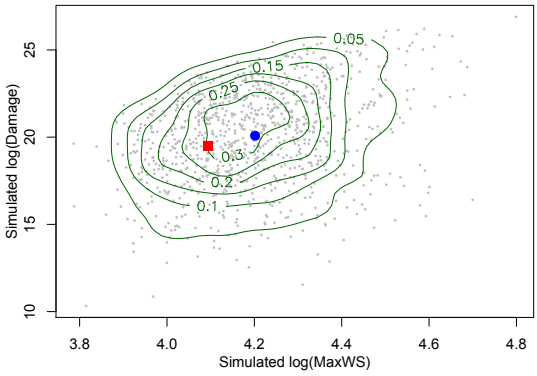
(c) Model 1: Hurricane Earl



(d) Model 2: Hurricane Earl



(e) Model 1: Tropical Storm Hermine



(f) Model 2: Tropical Storm Hermine

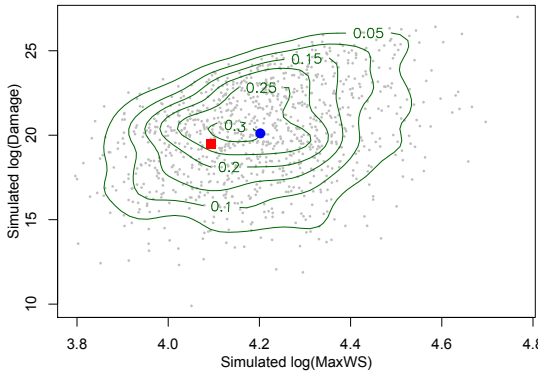
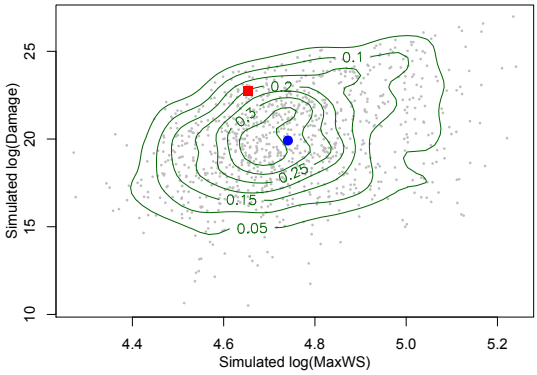
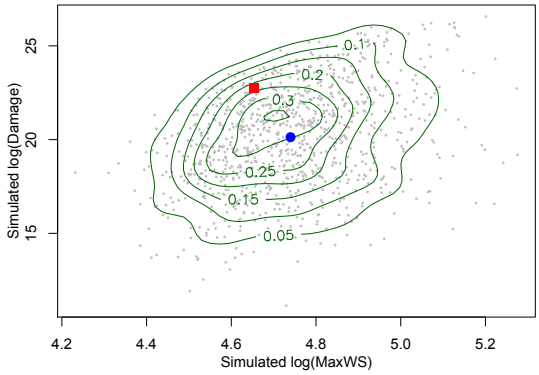


Figure 1.30: Simulated Bivariate Bilogistic Models for 2011-2012 Hurricanes. The blue circle is the sample mean of the simulated data and the red square is the observed value for the storm of interest.

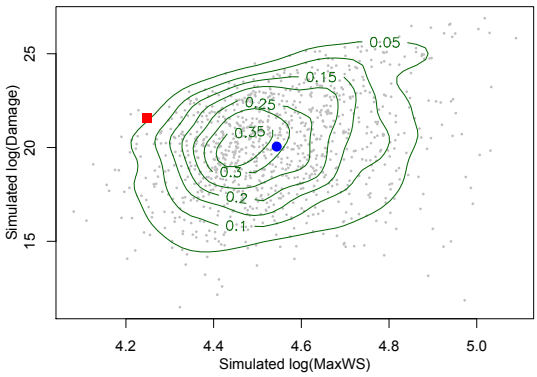
(a) Model 1: Hurricane Irene



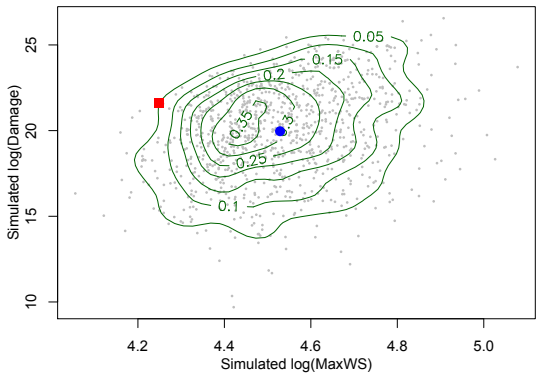
(b) Model 2: Hurricane Irene



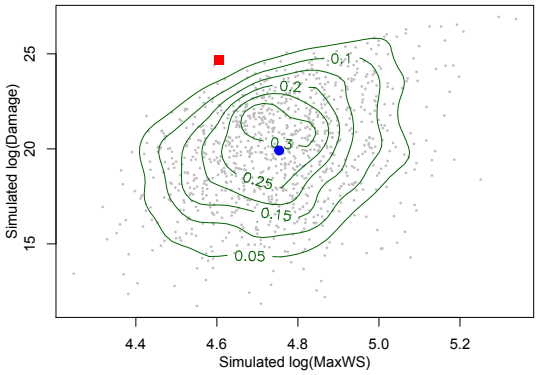
(c) Model 1: Hurricane Isaac



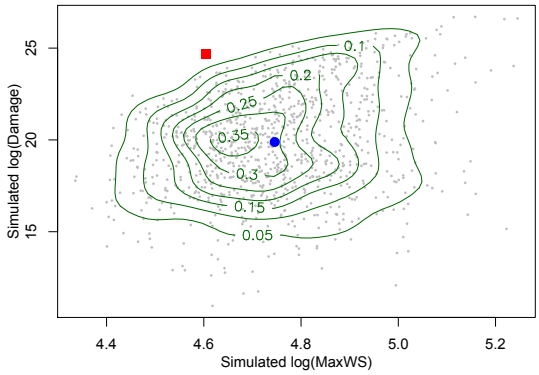
(d) Model 2: Hurricane Isaac



(e) Model 1: Hurricane Sandy



(f) Model 2: Hurricane Sandy



1.6 Conclusions on Modeling the MaxWS-MinCP Relationship

Overall, this chapter focused on understanding tropical cyclone intensity via modeling of the MaxWS-MinCP relationship. We first defined the current practices in modeling and highlighted many of their deficiencies including lack of systematic approaches and uncertainty estimates. To address these deficiencies, we fit weighted linear models of the transformed cyclostrophic wind approximation. This method was compared to the existing literature, and results indicated our models outperformed the most popular in current use.

Because of distributional assumption violations, we implemented a more intuitive model of the MaxWS-MinCP relationship using a univariate GEV for MaxWS conditional on MinCP. The uncertainties produced showed this was a better approach than using the linear model. The models also provide intuition on the return periods for each category of tropical cyclone.

Finally, we examined the MaxWS-MinCP relationship using a bivariate bilogistic extreme value model. This model indicated the large tail dependence between the two margins as one would expect. As a last step, we fit a bivariate bilogistic extreme value model of MaxWS and damage which included a conditional relationship for MinCP in the MaxWS margin. This model generated distributions which were reasonable for predictive use based on the distance between actual values and the simulated sample mean. To expand on the intensity and damage relationship, a Bayesian hierarchical model will be fit in Chapter 2 which simultaneously examines storm frequency, landfall, and damages.

Chapter 2

A Joint Bayesian Hierarchical Model of Occurrence, Landfall, and Damages for Atlantic Tropical Cyclones

2.1 The Study of Frequency, Intensity, and Damage in Tropical Cyclones

Chapter 1 highlighted the importance of understanding the intensity mechanisms in tropical cyclones because of the relationship with potential damage. In particular, Atlantic tropical cyclones (commonly known as hurricanes) can cause billions of dollars in damages and loss of lives annually in the United States. In extension of the previous chapter, we further the statistical exposition of tropical cyclones by linking their frequency, intensity, and damage capabilities in a joint model.

Prediction and estimation of annual frequency of tropical cyclones has been well studied, and several covariates have been used via regression methods. In (Elsner, 2003), Poisson regression is used to study the tropical cyclone counts with predictors Southern Oscillation Index (SOI) and North Atlantic Oscillation (NAO). Similarly, (Elsner et al., 2001) implemented Poisson regression with cold tongue index[1] as the predictor.

(Jagger and Elsner, 2010) utilized Bayesian model averaging of several Poisson regressions with covariates of the previous annual May-October monthly averages of the Atlantic SST, NAO, SOI, and sunspot number (SSN). They noted that September SSN parameter was negative correlated with the probability of a U.S. tropical cyclone in the following year.

Hierarchical Bayesian approaches for estimating Poisson regression coefficients to model tropical cyclone counts were also explored in (Elsner and Jagger, 2004, 2005, 2006). Poisson and negative binomial models for tropical cyclone counts are used in (Villarini et al., 2010), however, their work suggested no conclusive evidence to prefer a negative binomial fit. Covariates used in their study included

[1] Average sea surface temperature (SST) anomalies over 6°N-6°S, 180°W-90°W

NAO, SOI, Atlantic SST anomalies, and Tropical SST anomalies.

Investigation of damages from Atlantic tropical cyclones was done in (Jagger et al., 2008) where insurance losses were modeled using preseason averages of SOI, NAO, and Atlantic SST. In a follow-up, (Jagger et al., 2010), annual losses were modeled using a Poisson distribution on the annual number of loss events with a lognormal distribution for each loss amount. This did not take into account the probability of landfall explicitly, which we also purport to explore.

(Dailey et al., 2009) specifically addresses the issue of landfalling based on so-called 'cold' and 'warm' years of Atlantic SST, however, didn't find statistical differences in terms of landfalling probabilities for those designations. Their most interesting result involved breaking the United States into regions and assessing landfalling distributions based on 'cold' and 'warm' years. Similar patterns were observed in the regions but with different amounts of variability.

Our research differs from these previous modeling efforts by jointly studying annual Atlantic tropical cyclone frequency, landfall, and damages. We propose a Bayesian hierarchical model which addresses intensity and frequency while quantifying probabilities of landfall and subsequent damage estimates. We also separately model more intense Category 3-5 tropical cyclones and less intense tropical storm to Category 2 tropical cyclones based on their clearly distinguished differences in frequency, landfall probability, and damage capacity.

(Pielke, 2009) claims that model based prediction may not be able improve upon what is expected from long-term historical record of U.S. tropical cyclone landfalls and damages. However, findings using our model indicate otherwise as it does not rely on forecasting a time series structure, but rather, treats the distribution of years as random. By doing so, we inherently address the issue of 'hurricane droughts', i.e. several years without landfall, raised in (Hall and Hereid, 2015). This type of 'drought' would cause problems in a time series model, however,

should not affect our predictive power.

The following sections present exposition on data sources, preliminary analysis and prior specification, model implementation and results, and finally, posterior predictive distributions confirming the useful nature of our model. First, Section 2.2 provides the details of each data source from which we extract the frequency, landfalling nature, and damages for storms, along with the climatic covariates to be explored in the frequency model.

Next, we use extensive preliminary data analysis to select prior distributions for each component of the model in Section 2.3. Frequency, landfall, and damages are all carefully explored and reasonable heuristics are developed for prior selection and setting hyperparameters using empirical functions of the data. The full specification of the Bayesian hierarchical model is described in Section 2.4. The specification include a Poisson distribution for occurrence frequency, binomial distribution for landfall, and a mixture of a lognormal distribution and a mass at 0 for the annual damages. Inclusion of the tropical storms for low intensity storms necessitates a modification of the frequency distribution due to overdispersion. Thus, we use a Poisson-gamma mixture rather than a Poisson in the low intensity specification.

Markov chain Monte Carlo (MCMC) is used to estimate the posterior distribution for inference. The posterior is sampled using two similar MCMC implementations based on the Bayesian inference Using Gibbs Sampling (BUGS) language in R. Two types of MCMC standard errors (SEs) are reported, and the most conservative estimate is used to assess the quality of estimates. These include a time-series SE and a multivariate SE.

Along with SEs, trace, density, and autocorrelation function (ACF) plots are used to study characteristics of the posterior samples. All parameters show good mixing across iterations with the only possible issues of convergence appearing in

parameters constrained to be nonnegative, such as dispersion or variance parameters. However, even these look reasonable with posterior modes being close to the mean of the distribution, and density plots displaying nearly symmetric distributions. Both MCMC implementations produce similar estimates with the only notable disagreement produced in the dispersion parameter for low intensity storm frequency.

Posterior estimates for low and high intensity storms correspond to mean occurrences of 13.4 and 2.45 and landfall proportions of 11.5% and 42%, respectively. Low and high intensity damages indicated lognormal means of 19.17 and 21.84, respectively, corresponding to about 211 million and 3 billion dollars worth of damage. The synthesis of these estimates suggests that, on average, the U.S. should be prepared for about 3.267 billion dollars in annual damage.

Finally, in Section 2.5, the usefulness of the model in a predictive sense is examined via the posterior predictive distribution. Specifically, we fit separate models to assess the predictability of the 2011 and 2012 observations. We use two methods to simulate the posterior predictive distribution. The first method inputs posterior mean estimates into the the model specification and simulates a single realization of the posterior predictive distribution. The second method samples 100 parameter vectors from the posterior, simulates from the model specification for each, and averages over their output. Results for both 2011 and 2012 indicate the posterior predictive distribution matches well with the data. The two simulation methods provide similar output except the second method exhibits higher variability and provides smoother probability densities. This indicates the model specification is reasonable and could be employed as a tool in disaster mitigation.

2.2 Data Sources

2.2.1 Atlantic Tropical Cyclone Attributes Data

Data from the Atlantic tropical cyclone basin is available from 1851 up to 2014 in revised Atlantic hurricane database (HURDAT2)[2]. As in Chapter 1, data prior to 1960 are not used in this study due to possible inaccuracies. The data contains measurements on location, maximum winds, and central pressure taken every six hours during the storm lifetime (Landsea and Franklin, 2013). Alternatively, we could have used International Best Track Archive for Climate Stewardship (IBTrACs)[3] data as in Chapter 1. However, because the databases did not match exactly, and HURDAT2 is listed as the IBTrACs source for North Atlantic tropical cyclones, we chose to use the initial source for this analysis.

We define the maximum category a storm achieves by applying the Saffir-Simpson scale to the highest maximum wind speed[4] over the storm's lifetime. Tropical storms are included with tropical cyclones due to several damage events attributed to these less powerful storms. We note the wind speed data are subjected to quantization in which only multiples of five knots can be attained by maximum wind speed, however, this does not affect our planned analysis which is based on storm category rather than actual wind speed.

2.2.2 United States Tropical Cyclone Damage Data

A landfalling tropical storm is typically noted when the center (eye) of the cyclone crossing the coast with tropical storm or tropical cyclone strength. This definition

[2]Data were last updated in June 2015 and were retrieved from: http://www.aoml.noaa.gov/hrd/hurdat/Data_Storm.html

[3]https://www.ncdc.noaa.gov/ibtracs/

[4]Maximum wind speed is defined as maximum wind speed value over a 1-minute period

inherently assumes the center of the storm is tracked well and doesn't allow for ambiguity when tropical cyclone force winds make landfall, but the storm center does not. Different data sources can disagree on exactly when a storm makes landfall, or that a storm makes landfall at all. Using the original definition would cause our model to exclude storms that produce damage, but, technically, fail to make landfall. Rather than excluding this data, we use an alternate landfall definition to remove any possible ambiguity caused by different data sources. Thus, for the purposes of this research, we consider landfall to be equivalent to a nonzero damage value.

Damage data can be found within a variety of sources, however, a unique database has been developed by ICAT, a catastrophe insurance company. The ICAT database[5] provides access to historical hurricane damage information. According to their site, all data is pulled from publicly available sources. This database provide fairly comprehensive damages for North Atlantic tropical cyclones from 1900-2014. The site notes that base damage values are retrieved from Monthly Weather Review. Data in ICAT has also been normalized to 2014 dollars based on (Pielke et al., 2008) to reflect changes in inflation, wealth, and population in the storm area. ICAT is a useful resource because it is freely available and data are consistently reported in terms of time-based normalization.

Other less comprehensive sources of damage are available. One possible partial source is the National Oceanic and Atmospheric Administration (NOAA) top 30 costliest hurricanes technical report in (Blake et al., 2011). However, this only covers storms up to 2010 and many damage values do not match what is available in ICAT, thus, it would be difficult to determine the consistency with the rest of the storm damages.

[5] http://www.icatdamageestimator.com/faq

Weather Underground[6] also provides a record of damages, but, based on correspondence with technical representatives, their data is subject to error and exact sources are unknown. Finally, we note that insurance company AON[7] provides annual climate catastrophe insurance reporting dating back to 2008. The methodology proposed by this research should work with any internally consistent data source in the future.

2.2.3 Climate Covariate Data

Based on the literature review, we chose 6 covariates to examine for their predictive ability in Atlantic tropical storm/cyclone occurrence. These are Atlantic Multidecadal Oscillation (AMO), SOI, NAO, Atlantic SST, Niño 3.4 anomaly, and SSN. A short description and data source for each follows.

AMO is an ongoing series of long-duration changes in the sea surface temperature of the North Atlantic Ocean, with cool and warm phases that may last for 20-40 years at a time. It is the ten-year running mean of detrended Atlantic SST Anomalies north of the equator. Data[8] are retrieved from the NOAA Earth System Research Laboratory (ESRL) and are attributed to (Enfield et al., 2001).

SOI is defined as the normalized sea-level pressure difference between Tahiti and Darwin. Negative values of the SOI indicate an El Niño event. Monthly SOI values are obtained from National Centers for Environmental Prediction (NCEP) Climate Prediction Center (CPC)[9].

The Niño 3.4 anomaly series was also collected from NCEP CPC[10]. This series

[6] http://www.wunderground.com/hurricane/at2013.asp
[7] http://thoughtleadership.aonbenfield.com/Pages/home.aspx?Peril=Tropical%20Cyclone
[8] http://www.esrl.noaa.gov/psd/data/timeseries/AMO/
[9] ftp://ftp.cpc.ncep.noaa.gov/wd52dg/data/indices/soi
[10] http://www.cpc.ncep.noaa.gov/data/indices/ersst4.nino.mth.81-10.ascii

is an average of the SST from 5°S-5°N by 170°W-120°W with the 1951-2000 mean removed. Other Niño indices exist but are highly correlated with Niño 3.4, and are less commonly used in literature.

The NAO is characterized by fluctuations in sea level pressure differences. Strong positive phases of the NAO tend to be associated with above-average temperatures in the eastern United States and thus, provide a conducive environment for tropical cyclone development. Index values for the NAO are calculated as the difference in sea level pressure between Gibraltar and a station over southwest Iceland and are collected from NCEP CPC[11].

Atlantic SST averages gridded values over the region from 10-25°N by 80°W-20°W. Raw (unsmoothed and not detrended) monthly SST values are obtained via the NOAA ESRL Physical Sciences Division (PSD)[12]. This is version 3b of the data known as extended reconstructed sea surface temperature (ERSST) and was constructed using the most recently available International Comprehensive Ocean-Atmosphere Data Set (ICOADS) SST data (Smith et al., 2008).

Sunspots are magnetic disturbances of the sun surface having both dark and brighter regions. The brighter regions increase the intensity of the ultraviolet emissions. Increased sunspot numbers correspond to more magnetic disturbances. SSN are obtained from World Data Center-Sunspot Index and Long-term Solar Observations at the Royal Observatory of Belgium[13].

We use the monthly time series available for each covariate. In (Jagger et al., 2008), they suggested an average of the May and June values of the SOI, NAO, and Atlantic SST anomalies for prediction. We operate under this same premise for AMO and the Niño 3.4 Anomaly. However, with Sunspots, we use the average

[11] ftp://ftp.cpc.ncep.noaa.gov/wd52dg/data/indices/tele_index.nh
[12] http://www.esrl.noaa.gov/psd/
[13] http://www.sidc.be/silso/datafiles

of the monthly average sunspots for July to June of the predicting year, i.e. for 2015, we average July 2014 to June 2015 monthly values. We then looked at the correlations in these covariates for the 1960-2013 data to determine which to investigate as part of our modeling process.

As seen in Figure 2.1, AMO and Atlantic SST are highly correlated with a value of 0.87. Similarly, the Niño 3.4 Anomaly and SOI display nonnegligable correlation of -0.65. Thus, we remove Atlantic SST and the Niño 3.4 Anomaly from the analysis and move forward with four covariates, NAO, SOI, AMO, and SSN. The full available time series for each of the four covariates to be investigated are in Figure 2.2. Values used in our analysis correspond to the right column of charts in the graphic.

Figure 2.1: Pairwise Plots of May/June Averages of AMO, SOI, NAO, Atlantic SST, Niño 3.4 Anomaly, and July-June Average of SSN. Lines in plot represent a local smoother to show behavior of the data. Numbers in boxes above the diagonal are Pearson correlations.

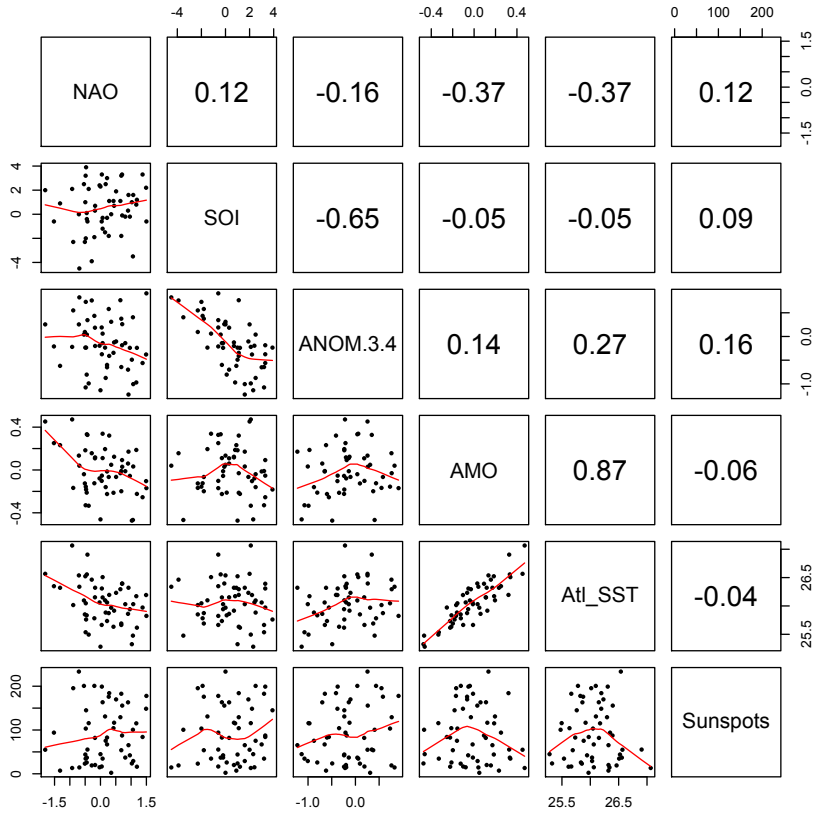
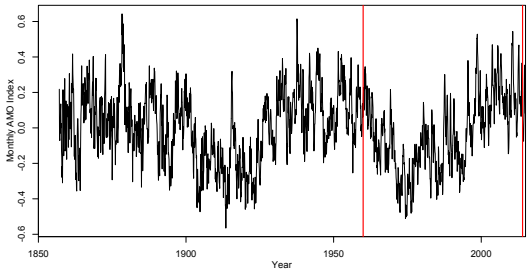
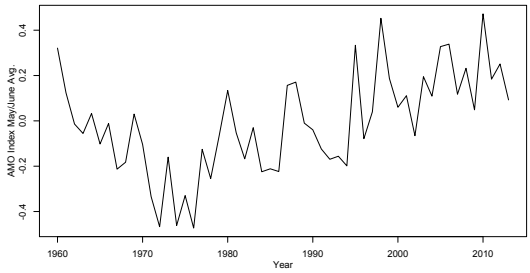


Figure 2.2: Time Series of AMO, SOI, NAO, and SSN. The left column gives the full available monthly time series with vertical lines indicating the portion used in deriving modeling variables. The right column are the May/June Averages used in the final model.

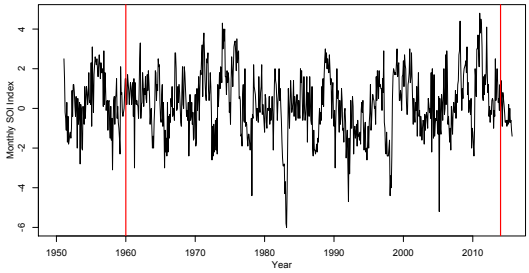
(a) Full AMO Monthly Series



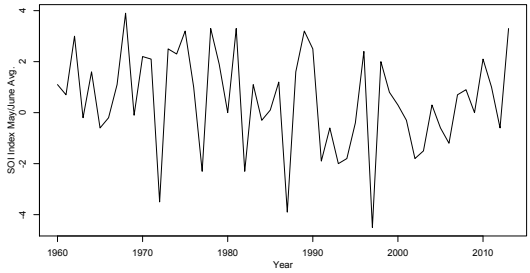
(b) 1960-2013 May/June Averages of AMO



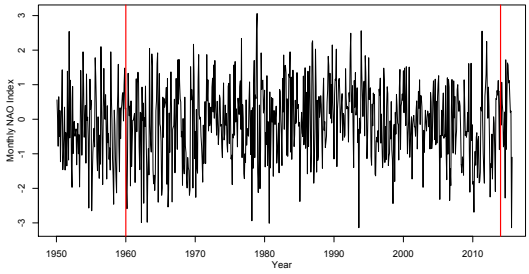
(c) Full SOI Monthly Series



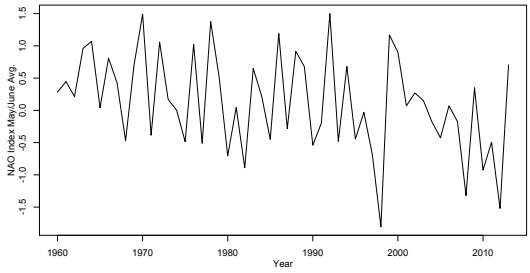
(d) 1960-2013 May/June Averages of SOI



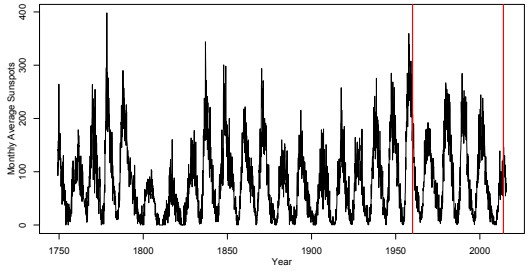
(e) Full NAO Monthly Series



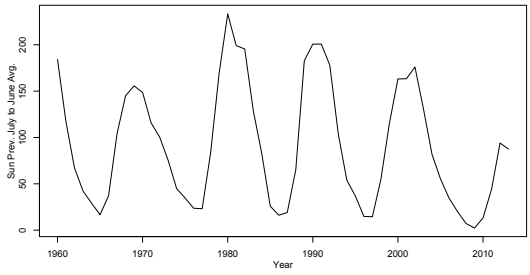
(f) 1960-2013 May/June Averages of NAO



(g) Full Monthly Average Sunspots Series



(h) 1960-2013 Lag. July-June Avg. Sunspots



2.3 Preliminary Empirical Analysis

This section rigorously examines that available data for each component of the model. Using this information, model and prior distributions are selected. Hyper-parameters are set based empirical functions of the data. This could be classified as an empirical Bayes method.

2.3.1 Storm Frequency Distribution

We first define two groups of storms. We refer to Group 1 as low intensity storms. This group contains tropical storms up to Category 2 tropical cyclones. Group 2, or high intensity storms, contains Category 3 through Category 5 tropical cyclones.

A simple model for count data is the Poisson generalized linear model (GLM). However, if data are overdispersed, one should consider more flexible count models to allow for variance to be larger than the mean. Thus, we consider two distributions for storm frequency: Poisson and negative binomial.

Given data for year i and annual covariates X_i, the Poisson model for Group C is the following:

$$[N_{C,i}|\mathbf{X}_i, \lambda_C] \sim \text{Poisson}(\lambda_C) \tag{2.1}$$

$$\log(\lambda_C) = \mathbf{x}_i^T \boldsymbol{\beta}_C \tag{2.2}$$

$$\mathbb{E}[N_{C,i}|\mathbf{X}_i, \lambda_C] = \mathbb{V}[N_{C,i}|\mathbf{X}_i, \lambda_C] = \lambda_C, \tag{2.3}$$

where the probability mass function (pmf) of the Poisson model is:

$$f(n) = \frac{e^{-\lambda_C}\lambda_C^n}{n!}. \tag{2.4}$$

The negative binomial formulation follows from integration of a Poisson-gamma

mixture to provide the following pseudo[14]-GLM:

$$[N_{C,i}|\mathbf{X}_i, \lambda_C, \rho_C] \sim \text{NB}(\lambda_C, \rho_C) \tag{2.5}$$

$$\log(\lambda_C) = \mathbf{x}_i^T \boldsymbol{\beta}_C \tag{2.6}$$

$$\mathbb{E}[N_{C,i}|\mathbf{X}_i, \lambda_C, \rho_C] = \lambda_C; \mathbb{V}[N_{C,i}|\mathbf{X}_i, \lambda_C, \rho_C] = \lambda_C + \frac{\lambda_C^2}{\rho_C}, \tag{2.7}$$

where the pmf of the negative binomial model (Zeileis et al., 2008) is:

$$f(n) = \frac{\Gamma(\rho_C + n)}{\Gamma(\rho_C) n!} \left(\frac{\rho_C}{\rho_C + \lambda_C}\right)^{\rho_C} \left(\frac{\lambda_C}{\rho_C + \lambda_C}\right)^{n}. \tag{2.8}$$

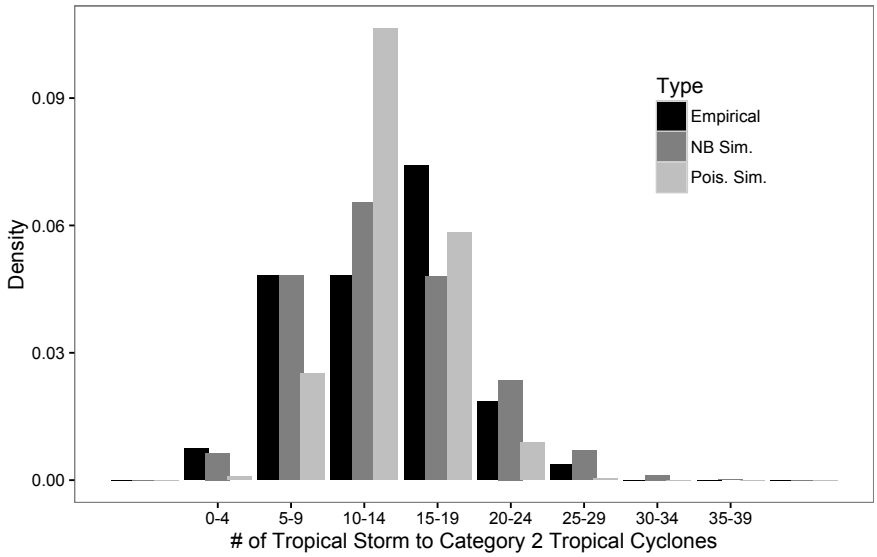
2.3.1.1 Low Intensity Storm Frequency

The plot of empirical density in Figure 2.3 indicates low intensity storms have a frequency distribution that is overdispersed compared to a Poisson distribution. The negative binomial model qualitatively appears to capture this overdispersion.

In Table 2.1, we show results of 4 models- 2 Poisson and 2 negative binomial. The full model (intercept, AMO, SOI, NAO, SSN) and an intercept-only model were fit. These two forms were selected based on a series of likelihood ratio tests (LRTs) not displayed here. The first noticeable result is the lack-of-fit in the Poisson models. Both models produced p-values less than 0.0001 indicating strong evidence of a lack-of-fit. Second, negative binomial models are preferred in all the relevant LRTs. Finally, note that none of the covariates are useful in predicting the number of expected storms per year as seen from the LRT comparing Model 3 versus Model 4 with a p-value of 0.2796. The Akaike information criterion (AIC) and Bayesian information criterion (BIC) both select Model 4 as the best fit to the data. Thus, we use Model 4 to specify the frequency distribution for the low

[14] If ρ_C is known, then this is a GLM.

Figure 2.3: Density of Annual Total Number of Low Intensity Storms. Poisson and negative binomial simulations were produced using models fit with only an intercept term. Simulated values are binned into groups of 5 to compare associated probability densities.

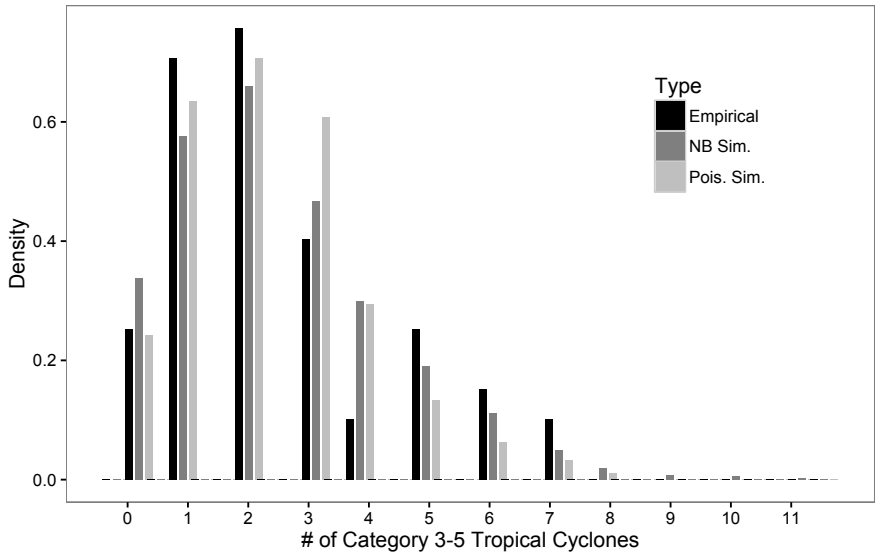


intensity storms. The prior mean for the intercept will be set at the estimated value from the frequentist negative binomial results. Similarly, the estimated value of ρ_1 will be used in specifying the gamma-Poisson mixture used in the final model.

Table 2.1: Low Intensity Storm Frequency Model Specification Testing. Numbers in parentheses indicate the preferred model based on the LRT. Bolded information criteria indicate the selected model by that criterion.

			LRT			Good.		
M	Fam.	Cov.	v. M1	v. M2	v. M3	of Fit	AIC	BIC
1	Poi.	All	-	-	-	<0.0001	355	365
2		Int.	0.0217 (1)	-	-	<0.0001	359	360
3	NB	All	<0.0001 (3)	<0.0001 (3)	-	0.3674	342	353
4		Int.	-	<0.0001 (4)	0.2796 (4)	0.5742	**339**	**343**

Figure 2.4: Density of Annual Total Number of High Intensity Tropical Cyclones. Poisson and Negative Binomial simulations were produced using models fit with only an intercept term.



2.3.1.2 High Intensity Storm Frequency

The empirical density in Figure 2.4 doesn't indicate overdispersion. The Poisson and negative binomial simulations qualitatively indicate a similar quality of fit.

To confirm our intuition, we again fit 4 models- 2 Poisson and 2 negative binomial with results displayed in Table 2.2. We fit the full model (intercept, AMO, SOI, NAO, SSN) and an intercept/AMO model selected based on a series of LRTs not displayed here.

Contrary to the less intense storms, the more intense storm frequencies appear to be well modeled by a Poisson fit. The LRT comparing both full models (Model 1 v. Model 3) and the intercept/AMO models (Model 2 v. Model 4) indicated statistical similarity with p-values of 1.0000 and 0.8840. Thus, although no model displayed a lack-of-fit, the more parsimonious Poisson models were preferred to their negative binomial counterparts. When comparing the 2 Poisson models

(Model 1 v. Model 2), the intercept/AMO model was deemed sufficient based on a p-value of 0.3452. This model also had the lowest AIC and BIC of the 4 models. Thus, we use Model 2 as the base distribution for the high intensity storms. The prior mean for the intercept and AMO coefficients will be set at the estimated values from the frequentist results.

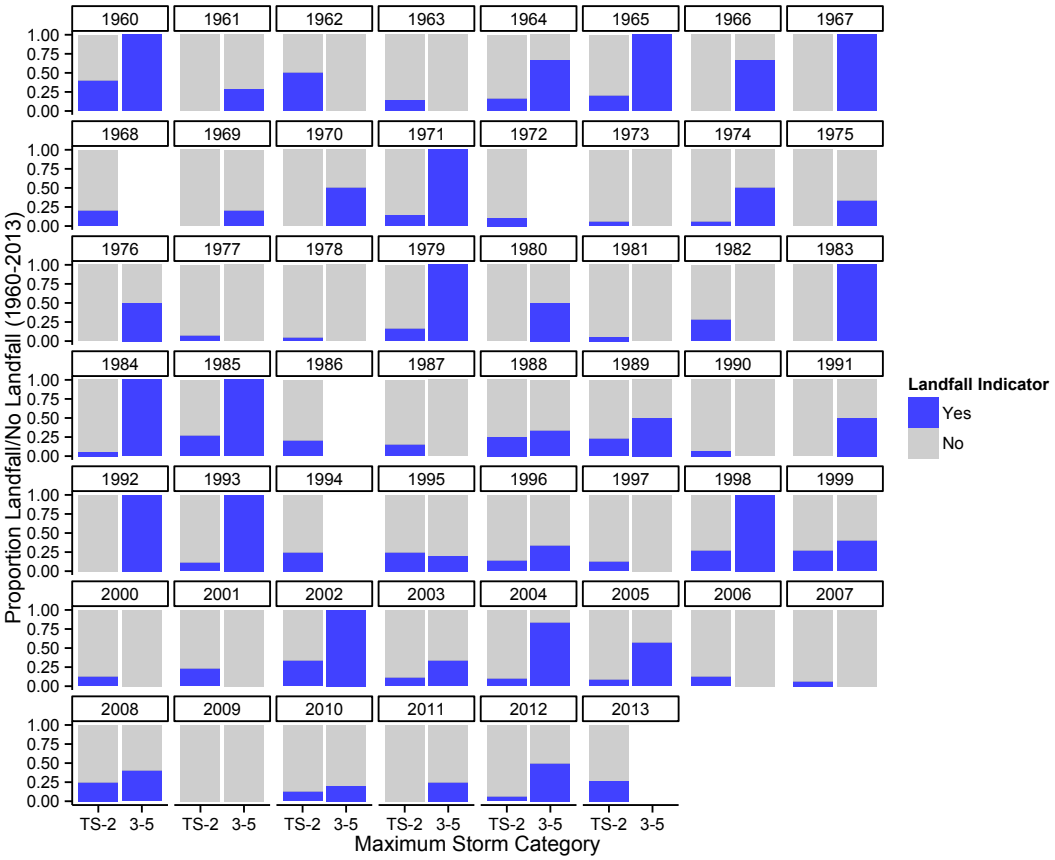
Table 2.2: High Intensity Storm Frequency Model Specification Testing. Numbers in parentheses indicate the preferred model based on the LRT. Bolded information criteria indicate the selected model by that criterion.

			LRT			Good.		
M	Fam.	Cov.	v. M1	v. M2	v. M3	of Fit	AIC	BIC
1	Poi.	All	-	-	-	0.4494	198	208
2		Int., AMO	0.3452 (2)	-	-	0.3845	**195**	**199**
3	NB	All	1.0000 (1)	0.3452 (2)	-	0.4498	200	212
4		Int., AMO	-	0.8840 (2)	0.3482 (4)	0.4369	197	203

2.3.2 Landfalling Storm Probabilities

As seen in Figure 2.5, the proportion of high intensity storms making landfall is usually higher than those of low intensity. To annualize the landfall observations, we assume a binomial distribution to describe the number of landfalling storms within each of the intensity groups. This implicitly assumes that within a season, landfall for a given storm and intensity group is an independent identically distributed Bernoulli trial. We make this assumption separately for both low and high intensity storms. However, because there are not enough high intensity storms within many seasons, we use all storms to give some verification of the assumption.

Figure 2.5: Annual Proportion of Storms from 1960-2013. A blank space indicates no storms occurred in that group in that year.



One possibility is to follow (Mailier et al., 2006), where a Poisson process model is used to determine the clustering in time (seriality) of extratropical cyclones in western Europe. However, they looked at monthly seriality rather than storm by storm seriality, thus, here we use a logit-normal mixed model for landfall occurrence (L_{ij}) with a random intercept (U_i) by year. The only fixed covariate was the landfall occurrence of the previous storm ($L_{i,j-1}$). Note for each year i, the data used starts from storm $j = 2$ so as not to falsely assign an observation from the previous years storms as a lagged landfall. The model is then:

$$\text{Level 1: } L_{ij}|\mathbf{U} = \mathbf{u} \sim \text{Bernoulli}(\theta_{ij}) \tag{2.9}$$

$$\text{logit}(\theta_{ij}) = \gamma_0 + \gamma_1 L_{i,j-1} + u_i \tag{2.10}$$

$$\text{Level 2: } U_i \overset{ind.}{\sim} \mathcal{N}(0, \sigma^2_{year}) \tag{2.11}$$

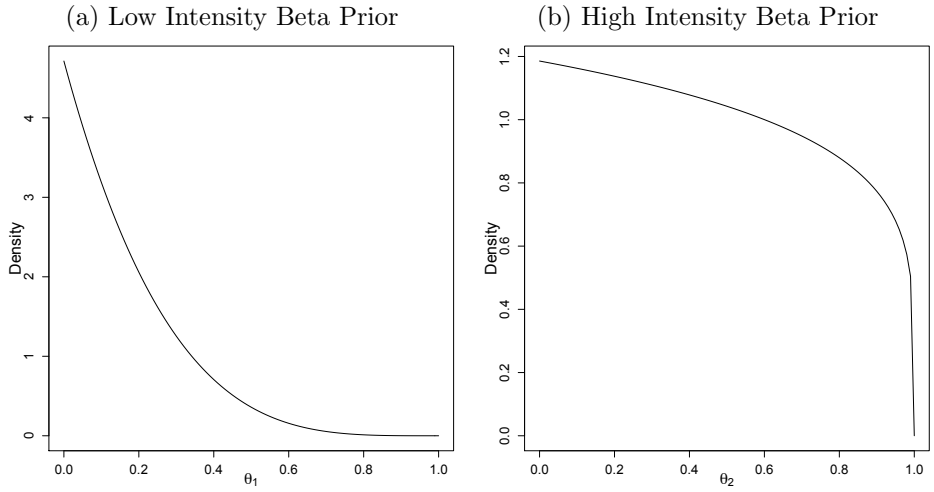
The estimates from the model are available in Table 2.3. A LRT indicated a lack of evidence ($p = 0.8219$) of a lag 1 dependence within storm season. However, the model estimates a large annual variability indicating the landfall probability may vary quite a bit by season. This fits well into our Bayesian hierarchy which assumes a distribution for landfall probability.

Table 2.3: Landfall Occurrence Logit-Normal Estimates

Parameter	Est.	Std. Error
γ_0	-1.7365	0.1364
γ_1	0.0604	0.2684
σ^2_{year}	0.3099	-

The natural prior densities for landfalling probabilities in each group of storms are beta distributions. In order to set the hyperparameters of the distributions, we first calculate raw estimates of landfalling probabilities, p_1 and p_2, using the indicators of landfall ($L_{C,ij}$) and the number of annual storms ($N_{C,i}$). One way to

Figure 2.6: Prior Densities of Landfall Probabilities for Low and High Intensity Storms



estimate p_1 and p_2 is to average the annual probabilities of landfall. Define

$$\hat{p}_{C,i} = \frac{\sum_{j=1}^{n_i} L_{C,ij}}{N_{C,i}} \tag{2.12}$$

and

$$\hat{p}_C = \frac{1}{n}\sum_{i=1}^{n} \hat{p}_{C,i}. \tag{2.13}$$

We use $\hat{p}_1$ and $\hat{p}_2$ to set the hyperparameters, α and κ, to arrive at the beta distribution priors in Figure 2.6 via the following definitions:

$$\alpha = 1 \tag{2.14}$$

$$\kappa = \frac{\alpha(1-\hat{p}_C) - \frac{1}{3}}{\hat{p}_C} + \frac{2}{3} \tag{2.15}$$

Based on (Kerman, 2011), the median (m) of a beta distribution can be approxi-

mated by:

$$m = \frac{\alpha - \frac{1}{3}}{\alpha + \kappa - \frac{2}{3}} \tag{2.16}$$

$$= \hat{p}_C \tag{2.17}$$

This gives a shape for each distribution that makes sense given the context of the problem, i.e. that the lower values of θ_1 and θ_2 contain a larger proportion of the density.

2.3.3 Damages By Storm Season

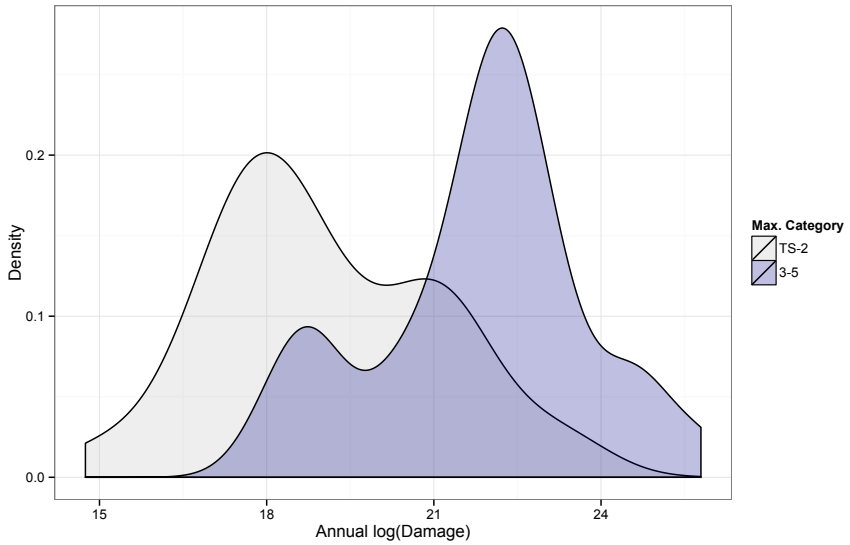
Section 2.3.2 shows that not all years have landfalls in each group. Thus, annual damage can be thought of a distribution with a substantial mass at 0. However, in years with damages, we do see considerable differences in the amount of damage by intensity group. In Figure 2.7, the log of nonzero damages are displayed. Modes on the log scales correspond approximately to damages around 66 million and 3.6 billion in 2014 dollars for the low and high intensity groups respectively. The difference is succinctly quantified by a single factor ANOVA on the annual log damages with a p-value less than 0.0001.

To set the prior for the lognormal mean, μ_C, we first define some data driven hyperparameters. Let m_C be the number of years with damages greater than 0 in Group C, and let $D^*_{C,i}$ be the total nonzero damages in year i. We calculate a sample mean of the log damages as:

$$\bar{x}_C = \frac{\sum_{i=1}^{m_C} \log(D^*_{C,i})}{m_C} \tag{2.18}$$

Thus, in order to provide a fairly flat prior for μ_C centered around $\bar{x}_C$, we use

Figure 2.7: Estimated Densities of Annual log(Damage) for Storms in Low and High Intensity Groups



a normal distribution with mean $\bar{x}_C$ and variance of 10^4. The variance chosen provided reasonable exploration of the parameter space and could be increased or decreased depending on the outcome of the MCMC. However, we feel it is large enough for the chain to sufficiently explore the parameter space.

The BUGS software used to run the model works more naturally with the inverse of a variance, also known as precision. Thus, we place a prior on this inverse rather than the variance itself. A suitable prior for this parameter is a gamma distribution. We use the parameterization with a shape parameter, v_1, and rate[15] parameter, v_2. The probability density function (pdf) is:

$$f(\tau) = \frac{{v_2}^{v_1}}{\Gamma(v_1)} \tau^{v_1 - 1} e^{-v_2 \tau} \tag{2.19}$$

[15] rate=1/scale

To create data driven hyperparameters, we use the following method:

$$\hat{s}_C^2 = \sum_{i=1}^{m_C} \frac{[\log(D^*_{C,i}) - \bar{x}_C]^2}{m_C - 1} \tag{2.20}$$

$$v_{C1}(\hat{s}_C) = \hat{s}_C^{-2} \tag{2.21}$$

$$v_{C2}(\hat{s}_C) = 1 \tag{2.22}$$

Thus, the prior for σ_C^{-2} is a gamma distribution with shape parameter $v_{C1}(\hat{s}_C)$ and rate parameter $v_{C2}(\hat{s}_C)$. This gives a gamma distribution with mean and variance equal to $\hat{s}_C^{-2}$.

2.4 Bayesian Hierarchical Modeling Implementation

2.4.1 Joint Frequency, Landfall, and Damage Model

Using the information from Section 2.3.1–Section 2.3.3, we now fully specify the hierarchical model. For a given year, i, define annual covariates as X_i. Let occurrence frequency of group C in year i be represented $N_{C,i}$, and let landfall frequency of be represented by $L_{C,i}$. Finally, let damages in year i be represented as $D_{C,i}$. Let the corresponding vector collections by group of all years for these variables be $\mathbf{N}_C$, $\mathbf{L}_C$, and $\mathbf{D}_C$.

The data indicated the distribution of annual low intensity storm frequencies is overdispersed. However, the chosen implementation software does not have the negative binomial parameterization needed for our model, thus, we substitute a Poisson-gamma mixture to act similarly. Therefore, the hierarchical specification for storm occurrence in (2.23) begins with a Poisson distribution with mean param-

eter, $\rho \cdot \lambda(\mathbf{x})$, as opposed to the usual Poisson regression mean parameterization, $\lambda(\mathbf{x})$. Then, let $\rho \sim \text{Gamma}(\phi, \phi)$ where the parameters are the shape and the rate. Note that $\mathbb{E}(\rho) = 1$ and $\mathbb{V}(\rho) = \phi^{-1}$, thus, the size of ϕ will be inversely related to the amount of overdispersion in the data. We also note that none of the covariates were statistically useful in the preliminary analysis, thus, $\log(\lambda(\mathbf{x}))$ only consists of an intercept term.

Low intensity landfall frequency, $\mathbf{L}_1$, is modeled using a binomial distribution with parameter θ_1 in (2.24) based on the elements in the vector of annual frequencies, $\mathbf{N}_1$. Finally, damages, $\mathbf{D}_1$, are modeled in (2.25) using a mixture distribution with a mass at 0 and nonzero damage modeled by a lognormal distribution with parameter (μ_1, σ_1^2). Hyperparameters are set using empirical functions of the data as described in Section 2.3.1. Priors are specified independently and employ normal distributions for $\boldsymbol{\beta}_1$, $\log(\phi)$, and μ_1 as in (2.26),(2.27), and (2.29), respectively. θ_1 is specified in (2.28) using a Beta distribution, and σ_1 is specified in (2.30) using a inverse transformation to a Gamma distribution.

The hierarchical specification in the high intensity storms is similar to that of low intensity save for the storm occurrence distribution. Storm occurrence in (2.31) utilizes a Poisson distribution with mean parameter, $\lambda(\mathbf{x})$, where $\log(\lambda(\mathbf{x})) = \mathbf{x}^T\boldsymbol{\beta}_2$. In this case, the initial analysis indicated AMO was predictive, thus, $\boldsymbol{\beta}_2$ is two-dimensional. As in the low intensity case, landfall frequency, $\mathbf{L}_2$, is modeled using a binomial distribution with parameter θ_2 in (2.32) based on the vector of annual frequencies, $\mathbf{N}_2$. Finally, damages, $\mathbf{D}_2$, are modeled in (2.33) using a mixture distribution with a mass at 0 and nonzero damage modeled by a lognormal distribution with parameter (μ_2, σ_2^2). Priors are specified independently and employ normal distributions for $\boldsymbol{\beta}_2$ and μ_1 as in (2.34) and (2.36), respectively. $\boldsymbol{\beta}_2$ is bivariate, but is specified with independent components. Similar to the low intensity case, θ_2 is specified using a Beta distribution in (2.35), and σ_2 is specified

in (2.37) using a inverse transformation to a Gamma distribution.

Low Intensity Storm Bayesian Specification

$$
\begin{aligned}
[N_{1,i}|\mathbf{X}_i, \phi, \boldsymbol{\beta}_1] &\sim \text{Poisson}(\rho_1 \cdot \lambda(\mathbf{x}_i)) \\
\log(\lambda(\mathbf{x}_i)) &= \mathbf{x}_i^T \boldsymbol{\beta}_1 \qquad (2.23) \\
[\rho_1|\phi] &\sim \text{Gamma}(\phi, \phi)
\end{aligned}
$$

$$
[L_{1,i}|N_{1,i}, \mathbf{X}_i, \phi, \boldsymbol{\beta}_1, \theta_1] \sim \text{Binomial}(n_{1,i}, \theta_1) \qquad (2.24)
$$

$$
\begin{aligned}
[D_{1,i}|L_{1,i}, N_{1,i}, \mathbf{X}_i, \phi, \boldsymbol{\beta}_1, \theta_1, \mu_1, \sigma_1] &\sim \\
(1-(1-\theta_1)^{n_{1,i}})\cdot\text{Lognormal}(\mu_1, \sigma_1^2) &+ (1-\theta_1)^{n_{1,i}} \cdot 0 \qquad (2.25)
\end{aligned}
$$

$$
[\boldsymbol{\beta}_1] \sim \mathcal{N}(\hat{\beta}_{01}, 10^4) \qquad (2.26)
$$

$$
[\log(\phi)] \sim \mathcal{N}(\log(\hat{\rho}_1), 1) \qquad (2.27)
$$

$$
[\theta_1] \sim \text{Beta}(\alpha(\hat{p}_1), \kappa(\hat{p}_1)) \qquad (2.28)
$$

$$
[\mu_1] \sim \mathcal{N}(\bar{x}_1, 10^4) \qquad (2.29)
$$

$$
\left[\sigma_1^{-2}\right] \sim \text{Gamma}(v_{11}(\hat{s}_1), v_{12}(\hat{s}_1)) \qquad (2.30)
$$

High Intensity Storm Bayesian Specification

$$\begin{aligned}[N_{2,i}|\mathbf{X}_i, \boldsymbol{\beta}_2] &\sim \text{Poisson}(\lambda(\mathbf{x}_i)) \\ \log(\lambda(\mathbf{x}_i)) &= \mathbf{x}_i^T \boldsymbol{\beta}_2\end{aligned} \tag{2.31}$$

$$[L_{2,i}|N_{2,i}, \mathbf{X}_i, \boldsymbol{\beta}_2, \theta_2] \sim \text{Binomial}(n_{2,i}, \theta_2) \tag{2.32}$$

$$\begin{aligned}[D_{2,i}|L_{2,i}, N_{2,i}, \mathbf{X}_i, \boldsymbol{\beta}_2, \theta_2, \mu_2, \sigma_2] \sim \\ (1-(1-\theta_2)^{n_{2,i}})\cdot\text{Lognormal}(\mu_2, \sigma_2^2) + (1-\theta_2)^{n_{2,i}} \cdot 0\end{aligned} \tag{2.33}$$

$$[\boldsymbol{\beta}_2] \sim \mathcal{N}_2([\hat{\beta}_{02}\ \hat{\beta}_{12}]^T, 10^4 I_2) \tag{2.34}$$

$$[\theta_2] \sim \text{Beta}(\alpha(\hat{p}_2), \kappa(\hat{p}_2)) \tag{2.35}$$

$$[\mu_2] \sim \mathcal{N}(\bar{x}_2, 10^4) \tag{2.36}$$

$$\left[\sigma_2^{-2}\right] \sim \text{Gamma}(v_{21}(\hat{s}_2), v_{22}(\hat{s}_2)) \tag{2.37}$$

2.4.2 Bayesian Estimation Implementation in R

To estimate the hierarchical Bayesian model, we used two separate implementations in R which are both based on the BUGS language. The first is *data cloning* (Lele et al., 2010). In essence, data cloning simply replicates the data set multiple times and estimates the posterior based on the cloned data. The usefulness of this is that the cloned posterior will converge more rapidly to a normal distribution centered around the maximum likelihood estimate (MLE). This method is discussed in detail within Appendix B.1.3 and is also used in Chapter 3. In R, the method is implemented using `dc.fit{dclone}`(Solymos, 2010). The second implementation is more standard and uses Just Another Gibbs Sampler (JAGS) (Plummer, 2015) to elicit dependent draws from the actual posterior.

Data cloning was implemented using 1, 2, and 5 clones. However, we only report the estimates of the the 5-clone chain. Each run consisted of 3 chains of length 1 million with adaptation on the first 100 iterations. JAGS was also implemented using 3 chains, each with 1 million iterations. The first 100 iterations of each chain were discarded.

When calculating MCMC SE, we use the results of a single chain. Two types of SEs are automatically provided within the summaries given by `dclone` and JAGS. The first is a naive SE which is equal to the standard deviation of the MCMC chain divided by the square root of the sample size. This SE ignores the autocorrelation in the chain and is likely anti-conservative. We don't report this SE due to its lack of theoretical foundation. The second is the time-series SE which is based on an estimate of the spectral density at 0 and incorporates autocorrelation, thus, likely provides a more accurate estimate of the SE. Based on data cloning theory, the reported SEs were multiplied by $\sqrt{5}$ to get an estimate of the error for the actual posterior rather than the cloned posterior.

We also use `mcse.multi{mcmcse}` (Flegal et al., 2015; Vats et al., 2015) to provide another SE type for JAGS since it has samples from the actual posterior. This method assesses the multivariate chain of posterior samples rather than separately looking at each parameter. This accounts for possible non-zero covariances between parameter estimates. We only report the diagonal of the covariance matrix which is essentially using batch means estimator for MCMC SE as derived in (Jones et al., 2006). This method also provides an effective sample size (ESS) for each parameter[16] as well as one for the total set of parameters based on the estimated parameter covariance matrix[17]. ESS is the sample size needed from an independent sample to provide the same standard error of the estimated posterior

[16]Using `ess{mcmcse}`
[17]Using `multiESS{mcmcse}`

mean as the correlated MCMC sample.

2.4.3 MCMC Inference

In Table 2.4 and Table 2.5, the posterior mean and standard deviation are reported to three digits beyond the decimal for completeness, however, based on the most conservative estimate for the MCMC SE (bolded in the tables), digits are underlined when not estimated with statistical precision. 95% credible intervals are only reported for JAGS because `dclone` provides the output of the cloned posterior. Overall, the most conservative single parameter ESS was 72,990 while the multivariate estimate of ESS was 403,601.

Both methods produce moment estimates which are quite similar. The only notable disagreement produced is in the dispersion parameter ϕ. The AMO slope coefficient, β_{21}, is no longer predictive in the Bayesian setting after being initially retained by frequentist model selection process.

Estimates for frequency in the low intensity situation correspond approximately to a mean occurrence of $e^{2.595} \approx 13.4$. Since the AMO coefficient doesn't appear useful, average frequency and variance in frequency for the high intensity storms is $e^{0.9} \approx 2.45$. Low and high intensity expected landfall proportions were estimated around 11.5% and 42%, respectively. This means that on average, the U.S. should expect about 1.54 low intensity and 1.03 high intensity landfalling storms per year. Low and high intensity damages indicated lognormal means of approximately 19.17 and 21.84, respectively corresponding to about 212 million and 3.055 billion dollars worth of damage. Thus, on average, the U.S. should be prepared for about 3.267 billion dollars in damage each year.

Table 2.4: Low Intensity Storm Posterior Estimation Results

Par.	Meth.	Posterior Mean	SD	95% Crd. Int.	Standard Error T.S.	Multi.	ESS
β_{01}	dclone	2.594	0.056	- -	**1.234e-04**	-	-
	JAGS	2.595	0.057	(2.483, 2.707)	**1.319e-04**	1.715e-05	188,687
ϕ	dclone	10.642	3.841	- -	**1.144e-02**	-	-
	JAGS	11.325	4.410	(5.588, 22.258)	1.337e-02	**1.797e-01**	108,299
θ_1	dclone	0.115	0.012	- -	**1.179e-05**	-	-
	JAGS	0.115	0.012	(0.093, 0.139)	**1.182e-05**	1.342e-07	1,042,285
μ_1	dclone	19.172	0.313	- -	**3.131e-04**	-	-
	JAGS	19.172	0.322	(18.537, 19.806)	**3.225e-04**	1.063e-04	978,632
σ_1^2	dclone	4.010	0.895	- -	**8.997e-04**	-	-
	JAGS	4.260	0.999	(2.735, 6.607)	**1.023e-03**	9.373e-04	1,064,779

Table 2.5: High Intensity Storm Posterior Estimation Results

Par.	Meth.	Posterior Mean	SD	95% Crd. Int.	Standard Error T.S.	Multi.	ESS
β_{02}	dclone	0.901	0.087	- -	**1.049e-04**	-	-
	JAGS	0.894	0.087	(0.721, 1.062)	**3.194e-04**	1.038e-04	72,990
β_{12}	dclone	0.259	0.394	- -	**4.680e-04**	-	-
	JAGS	0.255	0.395	(- 0.517, 1.028)	1.455e-03	**2.073e-03**	75,173
θ_2	dclone	0.421	0.043	- -	**4.271e-05**	-	-
	JAGS	0.422	0.042	(0.340, 0.506)	**4.234e-05**	1.948e-06	920,884
μ_2	dclone	21.837	0.360	- -	**3.599e-04**	-	-
	JAGS	21.836	0.374	(21.099, 22.572)	**3.737e-04**	1.361e-04	1,026,444
σ_2^2	dclone	4.396	1.081	- -	**1.086e-03**	-	-
	JAGS	4.732	1.234	(2.904, 7.678)	1.273e-03	**1.580e-03**	964,239

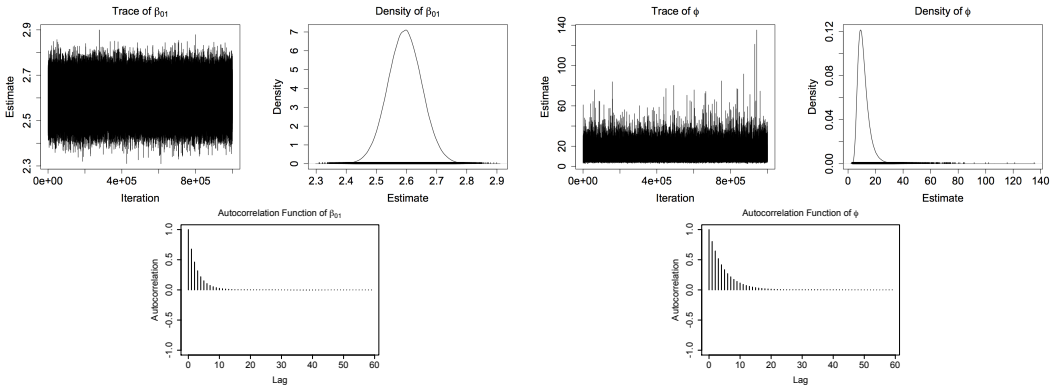
2.4.4 MCMC Diagnostics

Along with the SEs provided in the previous section, MCMC diagnostics are useful in telling whether the chain has converged, or at least that the sample space has been sufficiently explored and the chain is well mixed. Trace, density, and ACF plots are used to study these characteristics of the posterior samples.

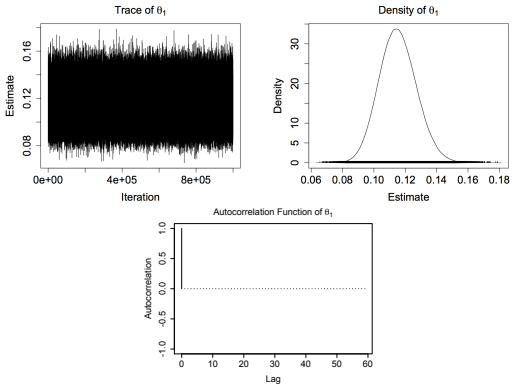
The low intensity storm parameters in Figure 2.8 show good mixing across the iterations. The only possible issues of convergence appear in the parameters which can only be positive, ϕ and σ_1^2. However, even these look reasonable with the posterior modes being close to the mean of the distribution. Similarly, the high intensity storm parameters in Figure 2.9 also appear to show good mixing in the trace plots and nearly normal symmetric posteriors except for σ_2^2.

Figure 2.8: MCMC Trace, Density, and ACF Plots For Low Intensity Storms

(a) Storm Occurrence Parameters



(b) Landfall Probability Parameter



(c) Log Damage Parameters

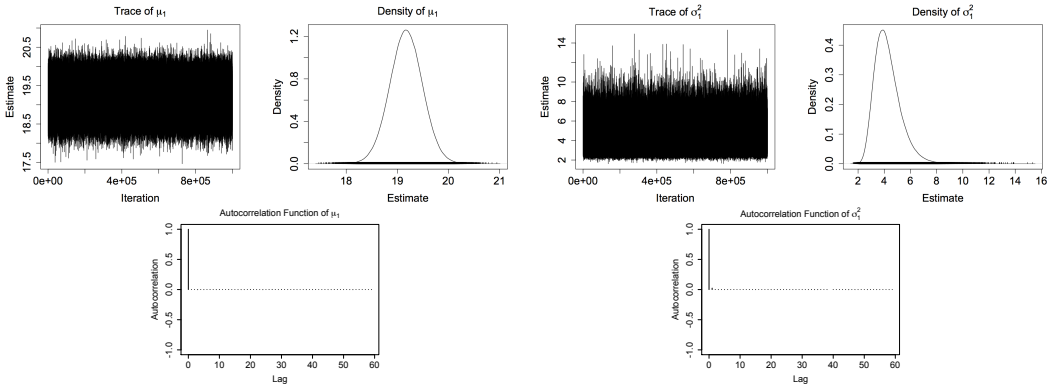
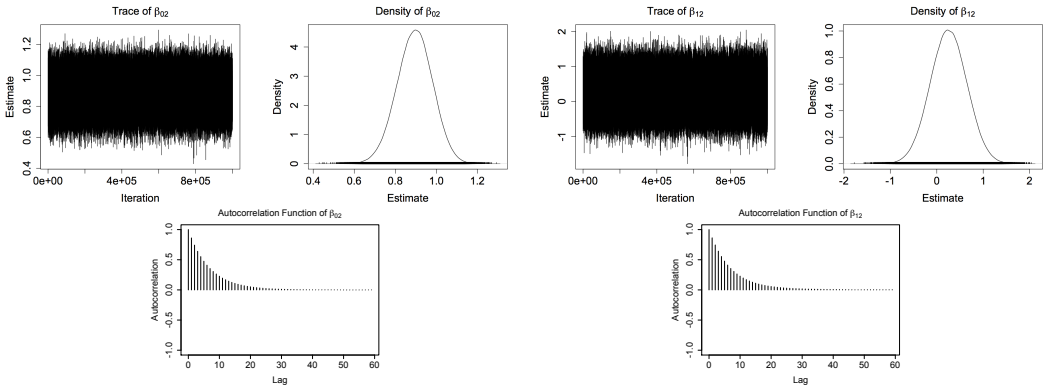
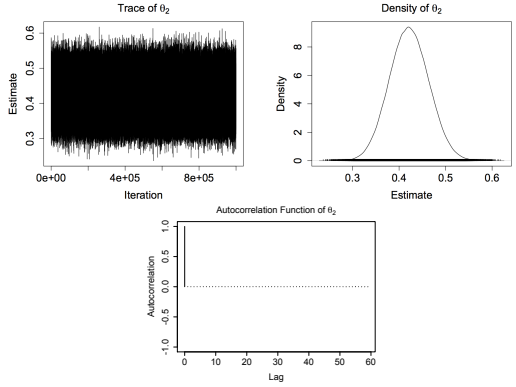


Figure 2.9: MCMC Trace, Density, and ACF Plots For High Intensity Storms

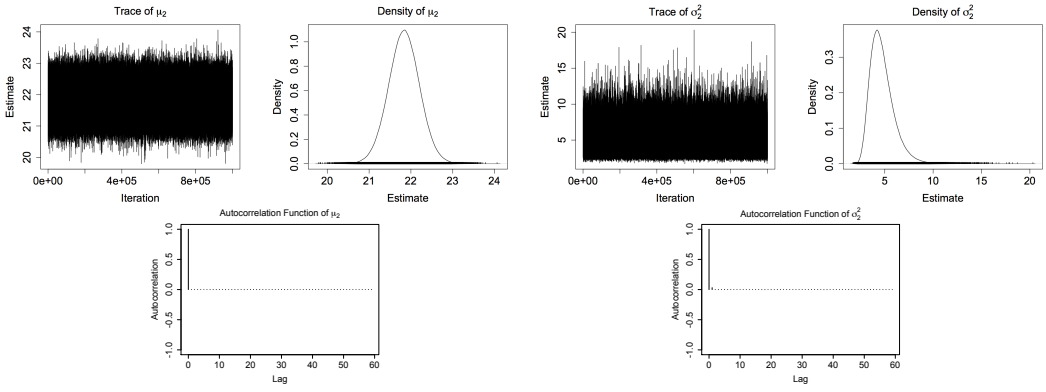
(a) Storm Occurrence Parameters



(b) Landfall Probability Parameter



(c) Log Damage Parameters



2.5 Posterior Predictive Analysis

The usefulness of this model lies in its ability to appropriately predict. Thus, we utilize the Bayesian specification to get an estimate of the posterior predictive distribution. Let the distributional specification of the a new data point, $\tilde{\mathbf{Y}}$, given data $\mathbf{Y} = [\mathbf{N}, \mathbf{L}, \mathbf{D}]$, parameters $\boldsymbol{\gamma} = [\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\sigma}, \phi]$, and hyperparameters $\boldsymbol{\alpha}$ be $p(\tilde{\mathbf{y}}|\mathbf{Y}, \boldsymbol{\gamma}, \boldsymbol{\alpha})$. Also, let the posterior distribution of $\boldsymbol{\gamma}$ given the data $\mathbf{Y}$ and hyperparameters be $p(\boldsymbol{\gamma}|\mathbf{Y}, \boldsymbol{\alpha})$. Then, the posterior predictive distribution for a new observation is

$$p(\tilde{\mathbf{y}}|\mathbf{Y}, \boldsymbol{\alpha}) = \int p(\tilde{\mathbf{y}}|\boldsymbol{\gamma}, \mathbf{Y}, \boldsymbol{\alpha}) p(\boldsymbol{\gamma}|\mathbf{Y}, \boldsymbol{\alpha}) d\boldsymbol{\gamma} \tag{2.38}$$

Because we cannot write down a simple analytical expression of the posterior predictive distribution in this case, we use two methods to simulate the distribution. First, let our hierarchical Bayesian model estimates, the posterior sample means, be $\hat{\boldsymbol{\gamma}}_{HB}$. We then sample $10,000$ times from model specification, $p(\tilde{\mathbf{y}}|\mathbf{Y}, \hat{\boldsymbol{\gamma}}_{HB}, \boldsymbol{\alpha})$. This provides a single realization of the posterior predictive distribution.

The second method randomly samples $\{\hat{\boldsymbol{\gamma}}_i; i = 1, ..., S\}$ from the MCMC generated set of posterior samples. In this case, we choose $S = 100$. Next, we sample $10,000$ times from $p(\tilde{\mathbf{y}}|\mathbf{Y}, \hat{\boldsymbol{\gamma}}_i, \boldsymbol{\alpha})$ for each i. Finally, we average over the 100 simulated distributions at each discrete value of $\mathbf{N}$ and $\mathbf{L}$. Since $\mathbf{D}$ is continuous, we first discretize nonzero damages into a grid of 150 equally spaced values. Probabilities are estimated at the 150 points using the `density{stats}` function in R with a bandwidth[18] of 0.25. Nonzero simulated values of log(Damage) for low (high) intensity storms were between 8.5 (9) and 29 (34), thus, we used a grid between 8

[18]Bandwidth is defined as the standard deviation of the smoothing kernel.

(8) and 30 (35) to allow the boundaries to decrease smoothly.

To test the model specification, we assess the usefulness of the model for years 2011 and 2012. Both were considered very active storm seasons, and are therefore good tests for the model prediction capabilities. Since a Bayesian specification corresponds to updating the model as new data become available, we use all available data up to the prediction year and include the preseason AMO covariate for high intensity storms.

For example, in predicting 2011, we use data from 1960-2010 to build the model including newly calculated empirical hyperparameters based on the same techniques used with the full data. We then sample the posterior predictive distribution using the two methods described above. The process for predicting 2012 was done in the same way using the 1960-2011 data. The posterior predictive distribution simulated output is shown in graphics which indicate observed values of occurrence, landfall, and damages by red dashed lines. For the annual damages, we display the log of positive damages and the \$0 damages on the same graph.

2.5.1 2011 Posterior Predictive Distribution

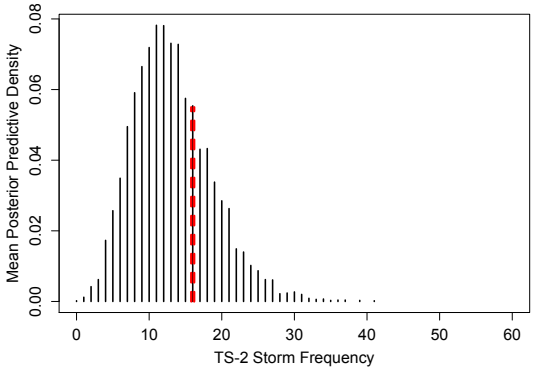
The results for 2011 indicate the model matches well with the data. The two posterior predictive methods provide similar output except that the second method has slightly more variability and provides a smoother density curve because of the averaging. As seen in Figure 2.10(a) and Figure 2.10(b), the actual number of storms was well within the predicted distribution for the low intensity group and had at least 30% of the probability higher than the observed value. The observed value of 16 storms had about a 5% occurrence in both versions of the posterior predictive. Similarly, in Figure 2.11(a) and Figure 2.11(b), high intensity storm occurrence was well within the limits of the predictive distribution. The observed

value of 4 storms accounted for nearly 15% of the samples and nearly 20% of the samples indicated more than 4 high intensity storms.

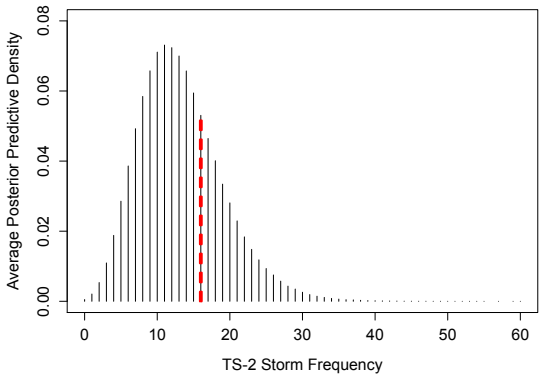
Simulated landfall values were also well within reason. The low intensity landfall predictive distribution in Figure 2.10(c) and Figure 2.10(d) indicated that not observing a landfall, as in 2011, was given a 24% and 25% chance respectively. Similarly, observing one high intensity landfall had the highest chance in Figure 2.11(c) and Figure 2.11(d) at 35%. There were \$0 in low intensity damages, thus, Figure 2.10(e) and Figure 2.10(f) reflect the chance of no landfall appropriately. The high intensity damages in Figure 2.11(e) and Figure 2.11(f) indicate the log of damage incurred in 2011 were well within the predictions of the model.

Figure 2.10: Posterior Predictive Distribution for 2011 Low Intensity Storms. In (a),(c), and (e), simulations are based on estimated mean of the posterior parameters. In the remaining plots, separate simulations using 100 samples of the posterior parameters are averaged. The actual values for 2012 are displayed with red dashed lines. The actual values for 2011 are displayed with red dashed lines.

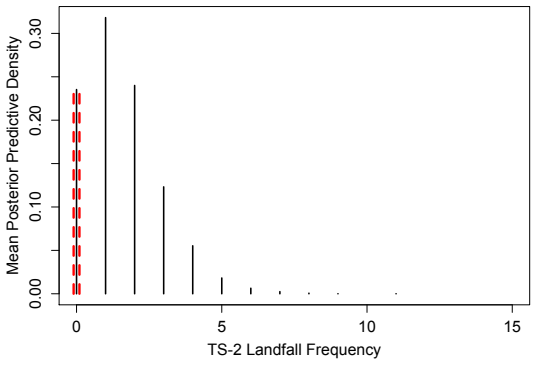
(a) Mean Est. Sim. Frequency



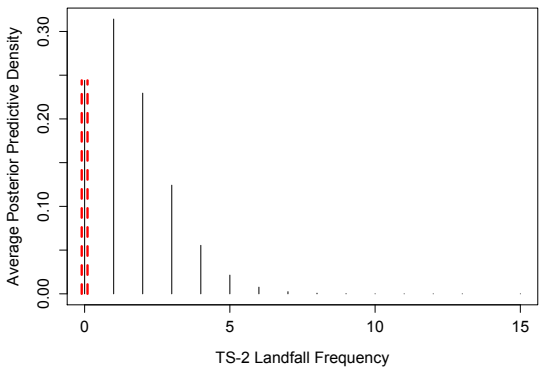
(b) Averaged Sim. Frequency



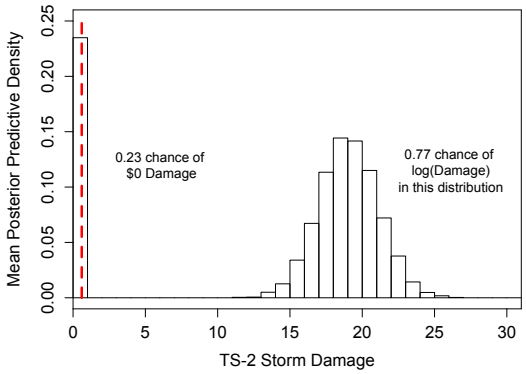
(c) Mean Est. Sim. Landfall



(d) Averaged Sim. Landfall



(e) Mean Est. Sim. Damages



(f) Averaged Sim. Damages

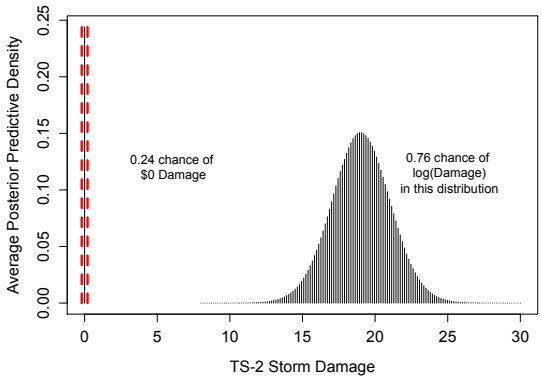
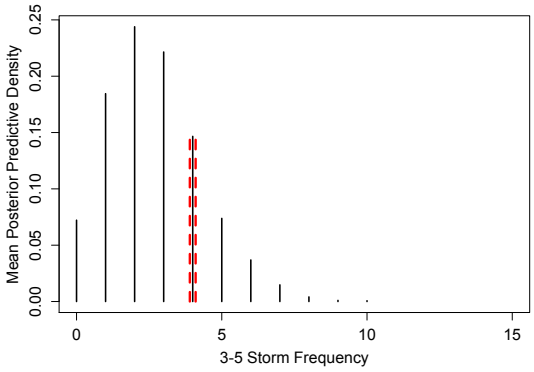
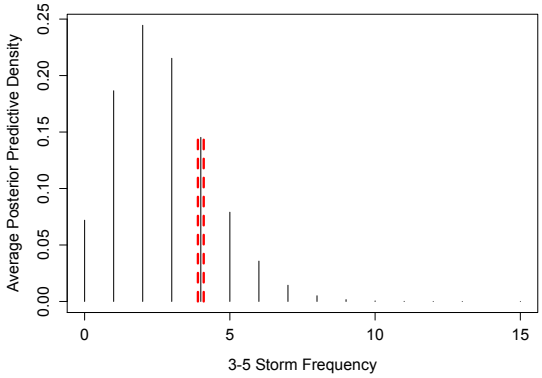


Figure 2.11: Posterior Predictive Distribution for 2011 High Intensity Storms. In (a),(c), and (e), simulations are based on estimated mean of the posterior parameters. In the remaining plots, separate simulations using 100 samples of the posterior parameters are averaged. The actual values for 2012 are displayed with red dashed lines. The actual values for 2011 are displayed with red dashed lines.

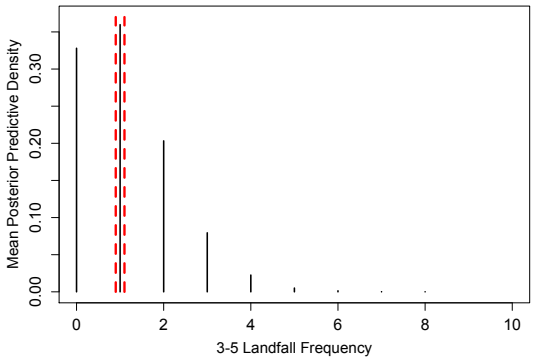
(a) Mean Est. Sim. Frequency



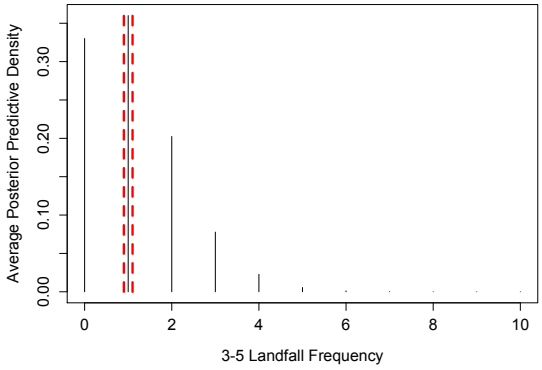
(b) Averaged Sim. Frequency



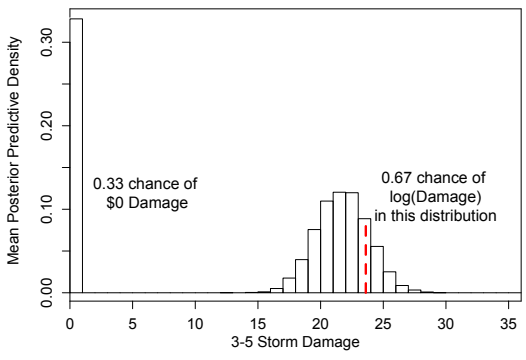
(c) Mean Est. Sim. Landfall



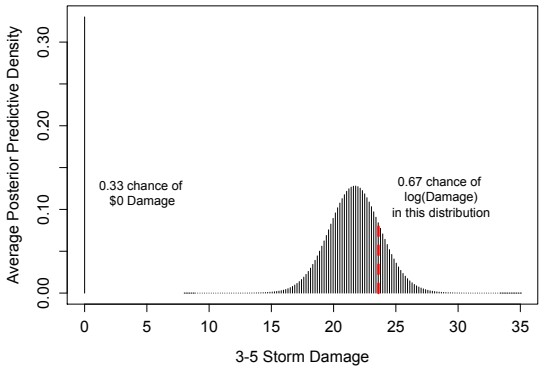
(d) Averaged Sim. Landfall



(e) Mean Est. Sim. Damages



(f) Averaged Sim. Damages



2.5.2 2012 Posterior Predictive Distribution

Similar to 2011, results for 2012 indicate the model matches well with the observations. Again, the two posterior predictive methods provide similar output with the second method showing more variability and a smoother density curve.

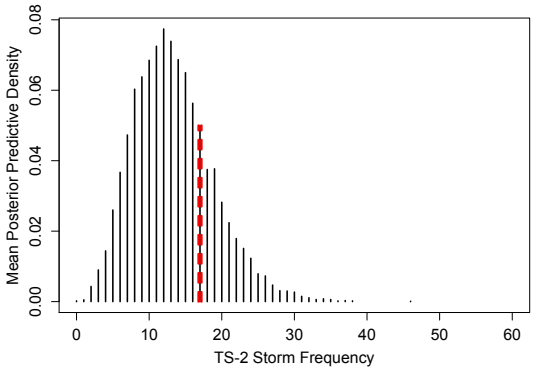
As seen in Figure 2.12(a) and Figure 2.12(b), the actual number of storms were within the predicted distribution for low intensity storms and at least 20% of the distribution was higher than the observed value. The observed value of 17 storms had about a 5% occurrence in the sample from the posterior predictive. Similarly, Figure 2.13(a), high intensity storm occurrence was well within the limits of the predictive distribution. The 2 observed storms accounted for about 25% of the samples with more than 50% of the samples indicating more high intensity storms.

Simulated landfall values were within reason. The low intensity landfall predictive distribution in Figure 2.12(c) and Figure 2.12(d) indicates observing a single landfall has a 33% chance. Similarly, observing one high intensity landfall has the highest chance in Figure 2.13(c) and Figure 2.13(d) at around 35%.

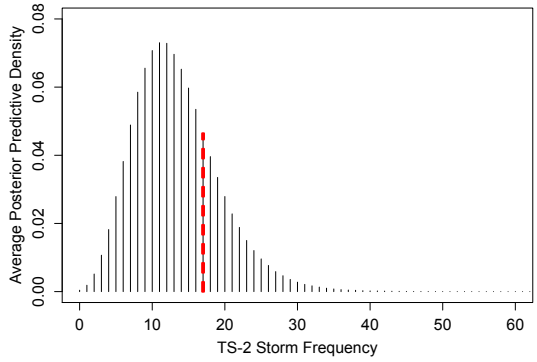
In both the low intensity scenario, Figure 2.12(e) and Figure 2.12(f), and the high intensity scenario, Figure 2.13(e) and Figure 2.13(f), the distribution indicate the damage incurred in 2012 was still within the predictions of the model. The observed values were both near the higher end of their respective damage distribution.

Figure 2.12: Posterior Predictive Distribution for 2012 Low Intensity Storms. In (a),(c), and (e), simulations are based on estimated mean of the posterior parameters. In the remaining plots, separate simulations using 100 samples of the posterior parameters are averaged. The actual values for 2012 are displayed with red dashed lines. The actual values for 2012 are displayed with red dashed lines.

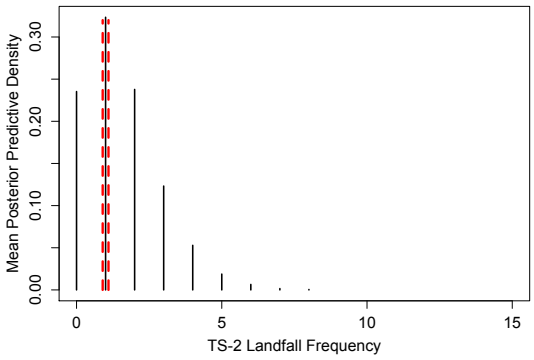
(a) Mean Est. Sim. Frequency



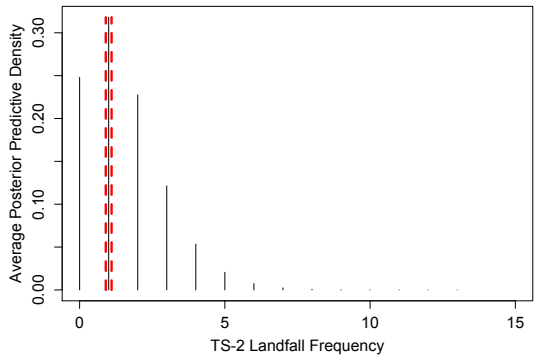
(b) Averaged Sim. Frequency



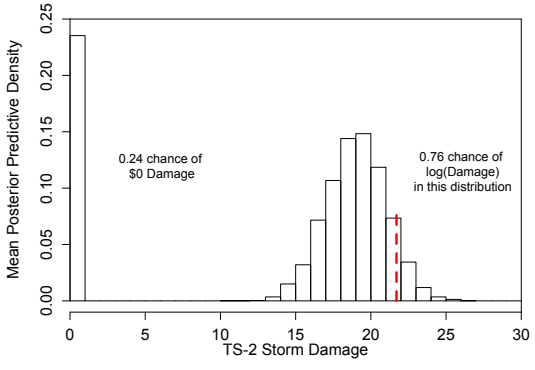
(c) Mean Est. Sim. Landfall



(d) Averaged Sim. Landfall



(e) Mean Est. Sim. Damages



(f) Averaged Sim. Damages

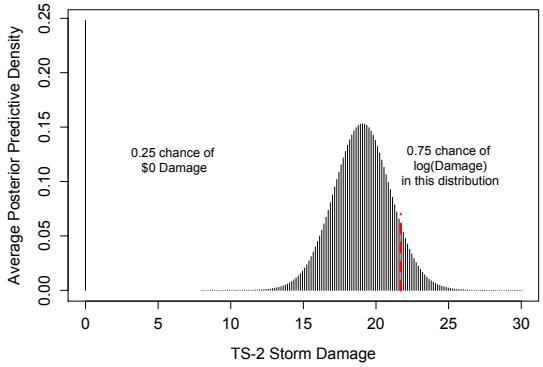
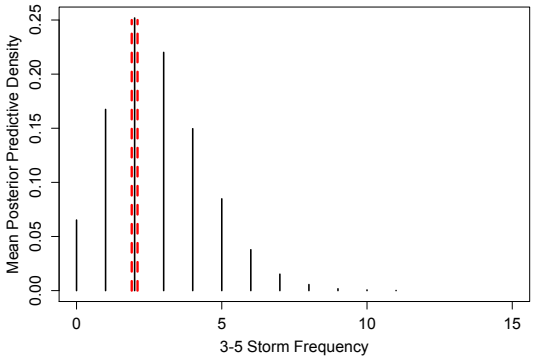
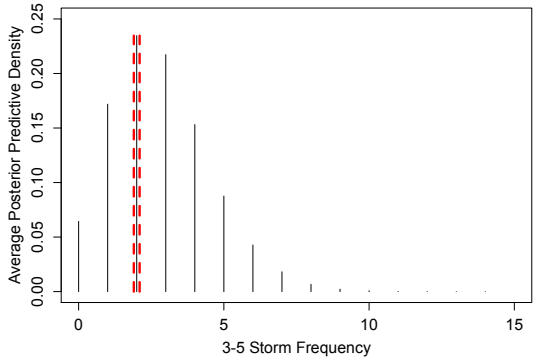


Figure 2.13: Posterior Predictive Distribution for 2012 High Intensity Storms. In (a),(c), and (e), simulations are based on estimated mean of the posterior parameters. In the remaining plots, separate simulations using 100 samples of the posterior parameters are averaged. The actual values for 2012 are displayed with red dashed lines.

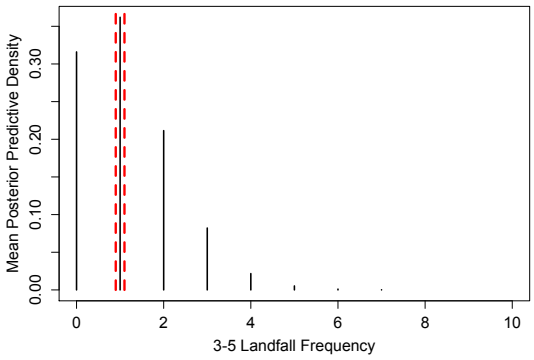
(a) Mean Est. Sim. Frequency



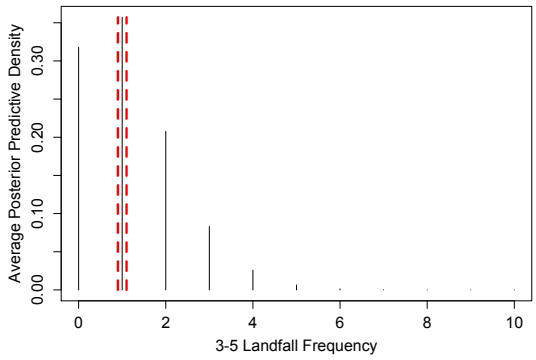
(b) Averaged Sim. Frequency



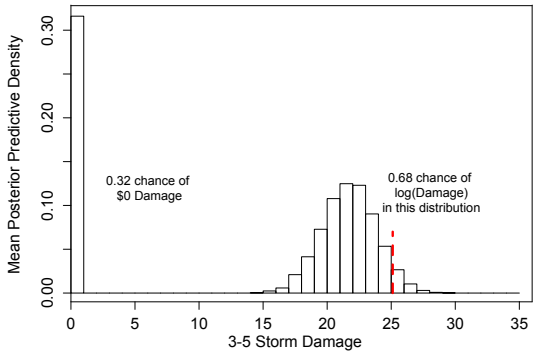
(c) Mean Est. Sim. Landfall



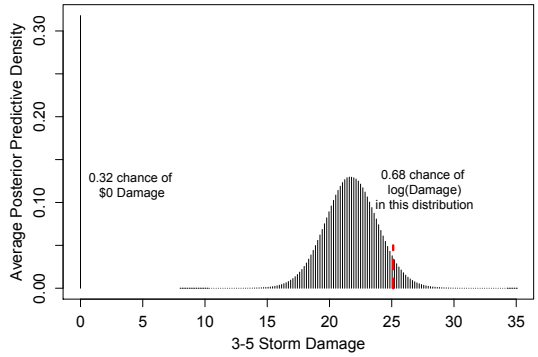
(d) Averaged Sim. Landfall



(e) Mean Est. Sim. Damages



(f) Averaged Sim. Damages



2.6 Conclusions on the Joint Bayesian Hierarchical Model

The Bayesian hierarchical model used in this chapter presents a relevant, intuitive, and insightful approach to the joint modeling of frequency, landfall, and damages related to tropical cyclones in the North Atlantic basin. When sufficient data are available, the framework could be implemented in other basins making only minor modifications to covariates or distributional assumptions.

The model separated the lower intensity storms from higher intensity in order to more accurately depict the bimodal nature of the damage distribution. We used an empirical Bayes approach to selecting prior distributions and developed reasonable heuristics to set hyperparameters. MCMC was used in estimating the posterior distribution and provided reasonable evidence for convergence of the chain. Estimates indicate that on average, the United States should prepare for approximately 3.267 billion dollars worth of damage per year.

Validation of this model comes from excellent performance in prediction, even in 2011 and 2012, years which exhibit higher frequencies than those seen in the recent past. The model provides a robust view of tropical cyclone seasons in terms of frequency, landfall, and damage prediction and allows risk to be mitigated based on the estimated parameters of the upcoming storm season.

Part III

Mixed Modeling and Dependence Analysis of Indian Monsoon Precipitation

Chapter 3

Analysis of Indian Monsoon Precipitation via Logit-Normal Mixed Models

3.1 Statistical Challenges of the Indian Monsoon

Explanation of summer Indian monsoon precipitation has been a challenging problem in physics as well as data analysis. In this chapter, we focus on statistical analysis of the summer monsoon precipitation data, to provide insight symbiotic with deterministic physics modeling. Previous statistical analysis studies regarding precipitation in Indian monsoons have explored two main areas- identifying methodology of data analysis and covariate selection.

The establishment of appropriate statistical methodology for explanation and prediction of precipitation, while simultaneously capturing underlying variability, is paramount. These methods are used in identification of trends for prediction, however, trends tend to be inconsistent across studies and may relate to linked variability on different temporal and spatial scales as noted by (Turner and Annamalai, 2012).

For instance, (Goswami et al., 2006) used daily central Indian rainfall and found rising trends in frequency and magnitude of extreme rain events along with decreasing light and moderate rainfall. While validating their 2006 study, (Ghosh et al., 2012) indicated increasing spatial variability in observed Indian rainfall extremes. They also found that moderate rainfall increased in central India despite a decreasing trend in occurrence of moderate rainfall. For high and extremely high rainfall, they noted a few locations experienced a significant upward or downward trend, however, most grid boxes showed a lack of trend.

A similar study conducted by (Ghosh et al., 2009) used a finer spatial scale and indicated a mixture of increases and decreases of extreme rainfall events dependent on location. An increasing trend in exceedances of 99^{th} (extreme) percentile daily rainfall was discovered by (Krishnamurty et al., 2009). On the other hand, they

stated many parts of India exhibited a decreasing trend for exceedances of the 90th (moderate to extreme) percentile. Increases in the frequency of both light and moderate to extreme rainfall events were observed in (Singh et al., 2014), along with decreasing probability of regional rainfall events and higher variability in the intensity of these events.

These studies utilized parametric techniques including regression, extreme value theory, and time series methods, and nonparametric statistical techniques, yet their lack of unanimity suggests important properties of the Indian monsoon remain at least partially misunderstood. In view of this disagreement, we propose adding the generalized linear mixed model (GLMM) as a potential framework for analysis of Indian monsoon precipitation data. The generalized part of the model allows for examination of threshold exceedances in daily rainfall, while use of a mixed model implies part of the signal is random, and changes from one set of circumstances to another. A mixed model could also capture local, instantaneous variability. Such local variability may arise from cloud and other physical micro-properties and year over year variations. When there is no such local variability, an appropriate variance component in the GLMM would be zero, thus, recovering the true underlying "fixed-effects" regression model.

The second principal focus of literature has been identifying relevant covariates for study of Indian monsoon precipitation. Certain oscillations are commonly useful predictors for precipitation. For instance, the synoptic activity index (SAI) developed in (Ajayamohan et al., 2008) correlated strongly with frequency of extreme rainfall. The Indian Ocean Dipole (IOD) studied in (Rajeevan et al., 2008) was shown to modulate inter-annual, inter-decadal and long-term trends of extreme rainfall events. Most commonly, the El Niño-Southern Oscillation (ENSO) (Kumar et al., 1999; Li et al., 2007b; Prell and Kutzback, 1992; Turner and Annamalai, 2012) is cited as a driver of the monsoon.

Several other climatic predictors of monsoons have been proposed in the literature including Himalayan/Eurasian snow extent (Kumar et al., 1999)), Pacific trade winds (Li and Yanai, 1996), atmospheric CO_2 concentration (Prell and Kutzback, 1992), and tropospheric temperature difference (ΔTT) (Xavier et al., 2007). Unfortunately, none have been conclusively attributed for the monsoon rainfall which suggests an intricate relationship between some or all of these factors. Because explicit attribution to covariates may not be possible, GLMM is a logical model for Indian monsoon precipitation. It allows underlying randomness to drive observed data in a particular hierarchy while still accounting for hypothesized drivers of rainfall.

The applicability of GLMM in this setting is verified using two studies. Study I (Dietz and Chatterjee, 2014) fits logit-normal mixed models (LNMMs) to indicators determined by 3 categories of rainfall: light ($0 < x < 64.4$mm/day), moderate ($64.4 \leq x < 124.4$ mm/day), and extreme (≥ 124.4 mm/day). The study utilizes several estimation methods including the method of moments, likelihood approximation, and Bayesian techniques. The outcome of annual models indicates maximum temperature and elevation are consistently important, aligning with the physics of precipitation. ΔTT is also significant in many of the models. The most meaningful finding is a random effect by weather station is non-negligible in many models.

Study II (Dietz and Chatterjee, 2015) extends the results of Study I to examine the use of GLMM with other previously studied thresholds in precipitation data. Station-defined percentile thresholds were used as in (Krishnamurty et al., 2009) and fixed level thresholds modified similar to those in Study I were used in (Goswami et al., 2006) to explore trends in monsoon rainfall intensity. Study II focuses on examining both threshold definitions with distinct purposes. Fixed threshold models elicit a physical interpretation across rainfall levels, and

percentile-based thresholds aid in understanding local predicted probabilities of threshold exceedances and possible cycles in their occurrence.

3.2 GLMM Preliminaries

3.2.1 Natural Exponential Families

The key component of a GLMM is the use of a natural exponential family (NEF) for the observed data. A NEF probability mass function (pmf) or probability density function (pdf) has several unique properties conducive to modeling. For further discussion of these properties, refer to Ch. 2 of (Keener, 2010). For simplicity, consider a univariate random variable Y distributed as a NEF. Define $a(\cdot)$ as a function of dispersion parameter ϕ, $r(\cdot, \cdot)$ as a function of Y and ϕ, and $c(\cdot)$ as a function of canonical parameter η known as the cumulant function.

The pmf/pdf then can be written:

$$f(y|\eta) = \exp\left\{\frac{y \cdot \eta - c(\eta)}{a(\phi)} - r(y, \phi)\right\}. \tag{3.1}$$

3.2.2 Mathematical Specifications

A GLMM is a probability model with a hierarchical structure; given the unobservable random effects, the observed data follows a NEF. Assuming observed data are independent conditional on the random effects, and there are $i = 1, \ldots, N$ observations, define the i^{th} response as Y_i.

Then, a two-layer GLMM can be written as:

$$\text{Level 1: } Y_i|\mathbf{U}=\mathbf{u} \overset{\text{ind.}}{\sim} f(y_i|\mathbf{u}) = \exp\left\{\frac{y_i\cdot\eta_i - c(\eta_i)}{a(\phi)} - r(y_i,\phi)\right\}, \tag{3.2}$$

$$\eta_i = \mathbf{x}_i^T\boldsymbol{\beta} + \mathbf{z}_i^{\mathrm{T}}\mathbf{u}, \tag{3.3}$$

$$\text{Level 2: } \mathbf{U} \sim \mathcal{N}_{\mathbf{M}}(\mathbf{0},\boldsymbol{\Sigma}). \tag{3.4}$$

The p-components of $\boldsymbol{\beta}$ are called *fixed effects*. The random effects covariance Σ is a function of the *variance components* $(\sigma_1,\ldots,\sigma_q)$. Fixed covariates are included via the $p\times 1$ vector $\mathbf{x}_i$. Random covariate vectors for the i^{th} data point and r^{th} variance component can be denoted by a $m_r\times 1$ vector $\mathbf{z}_{ir}$. Combining all variance component vectors forms the random covariate vector $\mathbf{z}_i = \left(\mathbf{z}_{i1}^{\mathrm{T}},\ldots,\mathbf{z}_{iq}^{\mathrm{T}}\right)^{\mathrm{T}}$ of length $M=\sum_{r=1}^{q} m_r$.

An additional assumption of conditional independence among observations implies the density of $\mathbf{Y}|\mathbf{U}=\mathbf{u}$ is

$$f(\mathbf{y}|\mathbf{u},\boldsymbol{\beta}) = \prod_{i=1}^{N} f(y_i|\mathbf{u},\boldsymbol{\beta}), \tag{3.5}$$

and the joint density of $(\mathbf{Y},\mathbf{U})$ is

$$f(\mathbf{y},\mathbf{u}|\boldsymbol{\beta},\Sigma) = f(\mathbf{y}|\mathbf{u},\boldsymbol{\beta})f(\mathbf{u}|\Sigma). \tag{3.6}$$

However, since random effects are unobserved, one must find the marginal distribution with respect to the observed data $\mathbf{Y}$ only. The log-likelihood is then:

$$\ell(\boldsymbol{\beta},\Sigma|\mathbf{Y}) = \log\int f(\mathbf{y},\mathbf{u}|\boldsymbol{\beta},\Sigma)d\mathbf{u}. \tag{3.7}$$

This integral is rarely analytically tractable. Thus, maximum likelihood esti-

mation is difficult. Many methods for inference have been proposed and several are discussed in Section 3.3.

3.3 GLMM Estimation Methods

3.3.1 Methods in R

This study focuses on a few methods implemented in R (R Core Team, 2014). The list is not comprehensive, but provides a reasonable cross-section of methodology available for GLMM estimation. These methods are implemented in Study I found in Section 3.5

3.3.1.1 Likelihood Approximation Estimation

Penalized quasi-likelihood (PQL) was proposed in (Breslow and Clayton, 1993) to approximate (3.7) using Laplace approximation. Further algorithmic details can be found in Section B.1.1.1. The R function which computes PQL estimates is `glmmPQL{MASS}` implemented by (Venables and Ripley, 2002). PQL is reasonably accurate when data are approximately normal and can be run quickly. However, (Lin and Breslow, 1996) and others have criticized this method for its bias in highly non-normal data. It can be especially poor in binomial data with a small sample size or true probabilities near zero or one. Reliance on the quadratic expansion of the log-likelihood is appropriate with normal random effects, yet it is difficult to assess normality of these unobserved effects.

Penalized iteratively reweighted least squares (PIRLS) is another approach to likelihood approximation in R implemented by (Bates, 2010). The main difference from PQL is that it attempts to approximate the true likelihood rather than the quasi-likelihood. This approximation may also be done by Laplace approximation

or Gauss-Hermite quadrature as discussed in (Bolker et al., 2009). Further algorithmic details can be found in Section B.1.1.2. The R function used to compute estimates is `glmer{lme4}` . This method can experience similar problems to PQL in cases where the random effects are non-normal. Gauss-Hermite quadrature can allay some of these issues, but is only computationally feasible for few random effects in limited model structures.

3.3.1.2 Method of Moments Estimation

Methodology known as the method of simulated moments (MSIM) is described in (Jiang, 1998). The method first derives a set of sufficient statistics. Then, estimating equations are obtained by equating sample moments of sufficient statistics to their expectations. Further algorithmic details can be found in Section B.1.2.1.

We implemented this method in a newly created R program. As shown in (Jiang, 1998), this method is consistent and potentially less computationally intensive than a Markov chain Monte Carlo (MCMC) method. However, it does require more computing time than any likelihood approximation method.

3.3.1.3 Bayesian Estimation

GLMM estimates can also be produced in a Bayesian framework. After choosing priors for the parameters of interest, the posterior distribution is calculated by multiplying the prior densities by the likelihood, $L(\boldsymbol{\beta}, \Sigma|\mathbf{Y})$, corresponding to (3.7). MCMC can then be used to generate a dependent sample from the posterior distribution from which estimates can be derived based on strong laws.

Data cloning is a method suggested by (Lele et al., 2010) to be used in conjunction with MCMC. Data cloning uses a 'k-cloned' version of the data and has a likelihood, $L^k(\boldsymbol{\beta}, \Sigma|\mathbf{Y})$ which exhibits quicker posterior convergence than MCMC

with the original likelihood. Further details are available in Appendix B.1.3. This method is implemented in R using `dclone{dclone}` . This method is discussed in (Solymos, 2010) and relies on the Bayesian inference Using Gibbs Sampling (BUGS) language for estimation of hierarchical models. The method is computationally intensive, and it may prove difficult to assess convergence, as with any MCMC implementation.

3.3.1.4 LNMM Simulations in R

For subject $i \in (1, \ldots, m)$ and observation $j \in (1, \ldots, n)$, 100 data sets are simulated from the model:

$$\text{Level 1: } Y_{ij}|\mathbf{u} \overset{ind.}{\sim} \text{Bernoulli}(\theta_{ij}) \tag{3.8}$$

$$\text{logit}(\theta_{ij}) = \eta_{ij} = \mu + u_i \tag{3.9}$$

$$\text{Level 2: } U_i \overset{\text{ind.}}{\sim} \mathcal{N}(0, \sigma^2). \tag{3.10}$$

After setting $(\mu, \sigma^2) = (2, 1)$ for $m \in \{10, 50, 200, 1000\}$ and $n \in \{2, 10, 50, 200\}$, means and standard errors over the estimates were calculated for each of the 16 settings. Estimation discrepancy between μ and $\hat{\mu}_{m,n}$, is assessed using squared error loss,

$$Q(\hat{\mu}_{m,n}) = (\hat{\mu}_{m,n} - \mu)^2, \tag{3.11}$$

while the difference between σ^2 and $\hat{\sigma}^2_{m,n}$ was quantified by Stein's loss,

$$S(\hat{\sigma}^2_{m,n}) = \frac{\hat{\sigma}^2_{m,n}}{\sigma^2} - 1 - \log \frac{\hat{\sigma}^2_{m,n}}{\sigma^2}. \tag{3.12}$$

Overall loss is defined as

$$G(\hat{\mu}_{m,n}, \hat{\sigma}^2_{m,n}) = Q(\hat{\mu}_{m,n}) + S(\hat{\sigma}^2_{m,n}). \tag{3.13}$$

A good estimation method should have $G(\hat{\mu}_{m,n}, \hat{\sigma}^2_{m,n}) \to 0$ as $m, n \to \infty$.

Subsequently, 4 of the 16 simulation settings were judged on computing speed. The settings used were combinations of (50, 200) subjects with (10, 200) observations. The `system.time()` command in R was used to record times. Simulations were independently run on 4 computers[1], and each estimation method was tested in sequence in one R script on a single core. Computing specifications can be found in Appendix B.2.1. MSIM was implemented in two ways for the speed test. In the intercept-only model (3.8)-(3.10), it is possible to use a simpler algorithm for estimation. However, a more general form of the algorithm is needed for problems including covariates. This form relies on matrices and does not work well with large data sets at this time. These methods are referred to MSIM Fast and MSIM Slow, respectively.

3.3.1.5 Simulation Analysis

An ideal method would provide high quality estimates in a short amount of time. However, this study indicated trade-offs between speed and accuracy in some methods.

Estimation results are displayed in Table 3.1–Table 3.4. All methods failed to reasonably estimate both μ and σ^2 in the minimal scenario with 10 subjects and 2 observations each. This was expected based on only 2 replications within subject. All other settings for MSIM, `dclone`, and `glmer` estimated μ within 2 standard errors. These methods provided reasonable estimates of σ^2 for settings

[1] See Appendix B.2.1 for details.

other than those with 10 subjects. The loss for both estimates went to 0 quickly for all three methods. In general, estimation by these three methods were unbiased. Estimates from `glmmPQL` did not converge to the true values of (μ, σ^2) as evidenced by combined loss greater than 0 for all settings. Further, this method displayed an underestimating bias in both parameters.

Speed testing was similar for each computer, therefore, only one set of results are shown in Table 3.5. Results indicated `glmmPQL` was fastest in the 50 subject cases and `glmer` was fastest in the 200 subject cases. These two likelihood methods are fast due to the nature of the approximations that they make. The Bayesian method, `dclone`, was slower ranging from 4 - 25 minutes to produce estimates. The simpler algorithm of MSIM Fast was somewhere in the middle, taking 3 - 6 seconds per run. The MSIM Slow method was much slower ranging from 1.5 minutes to nearly 4 hours. The case with 200 subjects 200 observations could not be handled by this method because the size of matrices and vectors exceeded the storage capacity allowed by R.

The `glmer` and `glmmPQL` methods were the fastest, but `glmmPQL` estimates were biased. Accurate estimates were produced by `dclone` despite being much slower than the approximation methods.

The intercept-only implementation of MSIM provided fast, accurate estimation. However, the matrix version was much slower and failed when too many observations were used. It should be noted that tuning parameters within each of the methods, such as convergence criteria for MSIM or number of MCMC samples in `dclone` impact computing time significantly[2].

Based on the output of these simulations, `glmmPQL` was not consistent. The other three methods- `glmer`, `dclone`, and MSIM- provided estimates with reasonable accuracy. Therefore, these estimation methods are used to fit models for

[2]See Appendix B.2.2 for detailed parameter specifications.

Study I in Section 3.5.

Table 3.1: Method of Simulated Moments (MSIM) Simulation Results. True values are $(\mu, \sigma^2) = (2, 1)$.

Par.	# of Sub.	Obs. per Subject 2	10	50	200
μ	10	17.41 (4.38)	2.11 (0.07)	2.05 (0.03)	2.00 (0.01)
	50	2.08 (0.05)	1.98 (0.02)	2.02 (0.01)	2.00 (0.00)
	200	2.01 (0.03)	1.98 (0.02)	1.99 (0.01)	1.99 (0.00)
	1000	2.00 (0.04)	1.99 (0.01)	2.01 (0.01)	2.00 (0.00)
σ^2	10	741.99 (302.51)	1.71 (0.33)	1.16 (0.09)	0.97 (0.04)
	50	1.02 (0.10)	0.98 (0.05)	0.98 (0.02)	0.99 (0.01)
	200	0.87 (0.05)	0.97 (0.03)	0.98 (0.02)	0.99 (0.01)
	1000	0.92 (0.06)	0.99 (0.02)	1.00 (0.02)	1.00 (0.01)
Loss	10	2885.74 (1016.73)	4.32 (0.71)	1.14 (0.33)	0.09 (0.01)
	50	3.29 (0.58)	0.19 (0.02)	0.04 (0.01)	0.01 (0.00)
	200	0.33 (0.04)	0.07 (0.01)	0.02 (0.00)	0.00 (0.00)
	1000	0.50 (0.16)	0.05 (0.00)	0.02 (0.00)	0.00 (0.00)

Table 3.2: Data Cloning (`dclone`) Simulation Results.
True values are $(\mu, \sigma^2) = (2, 1)$.

Par.	# of Sub.	Obs. per Subject 2	10	50	200
μ	10	13.18 (2.65)	2.12 (0.05)	2.03 (0.03)	2.02 (0.04)
	50	2.11 (0.07)	1.99 (0.02)	1.99 (0.02)	1.99 (0.01)
	200	2.05 (0.03)	2.02 (0.01)	1.99 (0.01)	2.01 (0.01)
	1000	2.01 (0.01)	2.00 (0.00)	1.99 (0.00)	2.00 (0.00)
σ^2	10	7.79 (1.79)	1.18 (0.11)	0.95 (0.06)	0.98 (0.05)
	50	1.67 (0.33)	1.00 (0.05)	0.99 (0.03)	1.00 (0.02)
	200	1.16 (0.08)	0.99 (0.02)	0.98 (0.01)	1.00 (0.01)
	1000	0.98 (0.04)	0.99 (0.01)	1.00 (0.01)	0.99 (0.00)
Loss	10	131.65 (86.2)	1.26 (0.14)	0.32 (0.04)	0.31 (0.05)
	50	1.58 (0.44)	0.19 (0.02)	0.06 (0.01)	0.04 (0.00)
	200	0.40 (0.06)	0.04 (0.00)	0.02 (0.00)	0.01 (0.00)
	1000	0.09 (0.01)	0.01 (0.00)	0.00 (0.00)	0.00 (0.00)

Table 3.3: PIRLS (`glmer`) Simulation Results. True values are $(\mu, \sigma^2) = (2, 1)$.

Par.	# of Sub.	Obs. per Subject 2	10	50	200
μ	10	6.02 (0.70)	2.77 (0.18)	2.33 (0.09)	2.10 (0.02)
	50	2.18 (0.09)	1.99 (0.02)	2.02 (0.01)	2.00 (0.00)
	200	2.03 (0.04)	1.99 (0.02)	1.99 (0.01)	1.99 (0.00)
	1000	2.02 (0.04)	1.99 (0.01)	2.01 (0.01)	2.00 (0.00)
σ^2	10	198.73 (81.39)	7.48 (1.36)	3.07 (0.76)	1.19 (0.04)
	50	1.66 (0.54)	0.95 (0.05)	0.94 (0.02)	0.94 (0.01)
	200	0.93 (0.06)	0.97 (0.03)	0.97 (0.01)	0.98 (0.01)
	1000	0.97 (0.05)	1.00 (0.02)	1.00 (0.01)	0.99 (0.00)
Loss	10	270.68 (84.19)	13.35 (1.91)	3.79 (1.12)	0.11 (0.02)
	50	4.96 (1.15)	0.19 (0.02)	0.04 (0.00)	0.01 (0.00)
	200	0.32 (0.04)	0.06 (0.01)	0.02 (0.00)	0.00 (0.00)
	1000	0.47 (0.20)	0.04 (0.00)	0.01 (0.00)	0.00 (0.00)

Table 3.4: Penalized Quasi-Likelihood (`glmmPQL`) Simulation Results. True values are $(\mu, \sigma^2) = (2, 1)$.

Par.	# of Sub.	Obs. per Subject 2	10	50	200
μ	10	3.10 (0.17)	1.92 (0.16)	1.34 (0.14)	0.68 (0.08)
	50	1.83 (0.06)	1.61 (0.03)	1.60 (0.02)	1.54 (0.01)
	200	1.81 (0.04)	1.71 (0.02)	1.73 (0.01)	1.72 (0.01)
	1000	1.81 (0.04)	1.81 (0.02)	1.81 (0.01)	1.79 (0.01)
σ^2	10	1.71 (0.13)	1.26 (0.11)	0.81 (0.11)	0.26 (0.06)
	50	0.52 (0.06)	0.25 (0.04)	0.15 (0.04)	0.01 (0.01)
	200	0.51 (0.04)	0.54 (0.03)	0.67 (0.02)	0.68 (0.01)
	1000	0.48 (0.04)	0.72 (0.03)	0.75 (0.01)	0.74 (0.01)
Loss	10	6.04 (0.50)	5.95 (0.47)	7.80 (0.50)	10.21 (0.47)
	50	4.81 (0.44)	5.61 (0.38)	6.26 (0.31)	7.89 (0.12)
	200	3.40 (0.48)	1.77 (0.33)	0.24 (0.06)	0.15 (0.01)
	1000	3.95 (0.47)	0.75 (0.25)	0.10 (0.01)	0.10 (0.00)

Table 3.5: Computing Time (in seconds) Results for GLMM Methods on nokomis

Method	(# of Subjects, Obs. per Subject) (50,10)	(50,200)	(200,10)	(200,200)
`glmer`	0.09	0.05	0.08	0.07
`glmmPQL`	0.29	0.23	0.38	0.39
MSIM Fast	2.58	3.42	2.48	2.48
`dclone`	10.03	11.36	38.07	40.00
MSIM Slow	94.73	9363.85	1069.47	NA

3.3.2 Methods in SAS

SAS also provides several options for fitting a GLMM. Using SAS/STAT® 9.3 for the Windows® operating system, these methods are implemented in Study II found in Section 3.6. A short simulation study for these methods is described in Section 3.6.3.

3.3.2.1 Likelihood Approximation Estimation

Exploring multiple estimation options is always wise due to the approximate nature of the estimation techniques in GLMM. SAS provides well-developed GLMM estimation procedures, specifically, PROC GLIMMIX and PROC NLMIXED. Several approximate likelihood estimation methods are available, however, method of moments and Bayesian estimation are not currently proceduralized within SAS.

Many methods exist within PROC GLIMMIX; the fastest computationally are the pseudo-likelihood (PL) methods. Four implementations of PL exist in SAS based on the combination of restricted or maximum likelihoods and the use of either subject-specific or population-averaged expansions for the linearization of the optimization function. These methods involve a doubly iterative optimization of the marginal log-likelihood determined by an approximate linear mixed model. These methods are flexible, however, they don't attempt to approximate the true log-likelihood.

PROC GLIMMIX can also use quadrature to approximate likelihood for some GLMMs. The quadrature option using more than 1 point failed in our implementation, thus, Laplace approximation (quadrature with 1 point) was used. For more details on the PL and quadrature methods of PROC GLIMMIX, refer to pages 2216-2231 of (SAS Institute Inc., 2011).

PROC NLMIXED can only fit certain structures of random effects. The

NLMIXED procedure approximates the marginal log-likelihood by integral approximation through Gaussian quadrature. Although this approximation may be more accurate, the number of random effects that can be practically managed is limited and computational time is considerably increased.

3.4 Indian Monsoon Precipitation Data

Multiple data sets of different temporal and spatial granularity have been used to study Indian monsoon precipitation in the literature. However, the initial goal was to test methods on widely and freely available data for the purpose of understanding the usefulness of GLMM in this context. This led to the selection of the data sources and covariates described in this section. A summary of covariates included in each of our studies can be found in Table 3.6.

Daily data for station level covariates of minimum and maximum temperature (°C), latitude and longitude (°), and elevation (meters) were collected from the National Climatic Data Center (NCDC)[3] in the National Oceanic and Atmospheric Administration (NOAA). Data were queried for all available Indian stations in the database. This data source was developed for a wide variety of potential applications, including climate analysis and monitoring studies that require data at a daily resolution. Quality assurance checks are routinely applied to the full data set according to (Menne et al., 2012).

Data were collected from January 1, 1973 to December 31, 2013. Unfortunately, this data has missing observations. In Study I (Section 3.5), only stations with ≥ 5 observations were included in analysis. One year in particular, 1975, did not contain enough data to be included in the analysis. In general, stations included in the data may change over time. In Study II (Section 3.6), only stations with $\geq$

[3] http://www.ncdc.noaa.gov/

40 days were included in the analysis. Two years, 1975-1976, were excluded due to not meeting criteria. The processed data included 36 weather stations.

Data for other climatic covariates of interest were also gathered. In Study I, the Indian Dipole Mode Index (DMI) was included. DMI takes the difference between sea surface temperature (SST) anomalies in the western and eastern Indian Ocean. The IOD was also included. IOD is an irregular oscillation in the Indian Ocean associated with Non-ENSO drought years. This index was available for 1973-2010 and data were retrieved from the Japan Agency for Marine-Earth Science and Technology (JAMSTEC) site[4].

Both studies included the Niño-3.4 anomaly series collected from the National Centers for Environmental Prediction (NCEP) Climate Prediction Center (CPC)[5]. As stated in (Wang, 2006) and other literature, Indian rainfall is strongly associated with ENSO, and onset of discharge in Niño-3.4 region can lead to drought in India. The occurrences of precipitation extremes are thought to be fewer in drought years. This index is a measure of the sea surface temperature which is known to be an important global climate driver. Since these data are gridded, they were aligned with the station closest in Euclidean distance by latitude and longitude.

Another covariate used in both studies is ΔTT. This is defined as the air temperature (°C) averaged between pressure levels 600 millibars (mb) and 200 mb. The hypothesis that Indian ocean warming leads to reduction in ΔTT which in turn reduces monsoon circulation is noted in (Xavier et al., 2007). Thus, the inclusion of this covariate in the models was relevant. Reanalysis data (Kalnay et al. (1996)) were collected from the NCEP Reanalysis site[6]. In Study II, further reanalysis data were collected including u-winds from 200 and 850 mb levels and

[4] http://www.jamstec.go.jp/e/

[5] http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/detrend.nino34.ascii.txt

[6] http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.derived.html

v-winds from 200 and 850 mb levels. Positive u-winds move west to east (westerlies), and positive v-winds move south to north (southerlies) and thus, both provide important tropospheric circulation information. Again, since these data are gridded, they were assigned to stations in the same method as the previous gridded covariates.

Table 3.6: Summary of Fixed Covariate Inclusion for Studies I and II

Covariate	**Study I**	**Study II**
Elevation	X	X
Max. Temp.	X	X
Min. Temp.	X	X
Latitude	X	X
Longitude	X	X
DMI	X	
Nino34	X	X
ΔTT	X	X
u-wind 200 mb		X
u-wind 850 mb		X
v-wind 200 mb		X
v-wind 850 mb		X

3.5 Study I: Annual GLMMs Using Indian Monsoon Precipitation Data

3.5.1 Goals of Study I

This study provides an introduction on extending GLMMs to climate applications. Because simulation findings in Section 3.3.1 showed that PQL was not accurate for the given application, the three remaining methods were used to fit logit-normal models with random intercepts by weather station for Indian summer rainfall data

in light, moderate, and extreme rainfall classifications. Maximum temperature and elevation were consistently significant in the models aligning with the physics of precipitation. ΔTT was also significant for many of the models. The most meaningful finding was a random effect by weather station was non-negligible in many of the models. This provides further credibility to the methodology in applications to climate. Overall, GLMMs could be a noteworthy addition to data analytics in climate applications.

3.5.2 Annual LNMMs

Fixed thresholds for India were defined in (Attri and Tyagi, 2010). This report used 3 categories of rainfall: light rainfall ($0 < x < 64.4$mm/day), moderate rainfall ($64.4 \leq x < 124.4$ mm/day), and extreme rainfall (≥ 124.4 mm/day).

Let weather station $i \in \{1, ..., m\}$ and day $j \in \{1, ..., n_i\}$. Define τ as one of the fixed thresholds above, and let a single precipitation event be defined as Z_{ij}. Finally, let $Y_{ij} = I(Z_{ij} > \tau)$. Let $\mathbf{x}_{ij}$ be a vector of covariates and $\mathbf{U}$ be a vector of random intercepts by station. Then,

$$\text{Level 1: } Y_{ij}|\mathbf{U} = \mathbf{u} \overset{\text{ind.}}{\sim} \text{Bernoulli}(\theta_{ij}), \tag{3.14}$$

$$\text{logit}(\theta_{ij}) = \mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + u_i, \tag{3.15}$$

$$\text{Level 2: } U_i \overset{\text{ind.}}{\sim} \mathcal{N}(0, \sigma^2_{station}). \tag{3.16}$$

Indicators (Y_{ij}) for light, moderate, and extreme rainfall events were derived to be used in the modeling. The typical inclusion period for the Indian summer monsoon is June 1 to September 30 (122 days), however, with the large degree of missingness, this didn't provide enough observations per station for annual models. Thus, the time frame was extended to range from May 1 to October 31 (184 days)

each year. Annual models (excluding 1975) were fit from 1973-2013.

3.5.3 Results of GLMMs

To aid interpretation and provide a basis for comparison among models, tests of significance were performed for both fixed and random parameter estimates. Results from a goodness-of-fit test for each year's model are also given.

To provide tests of significance for fixed effect coefficients, we propose the following procedure:

1. Run a generalized linear model (GLM) with all eligible fixed covariates.
2. Run a GLM with all eligible fixed covariates except one.
3. Perform a likelihood ratio test (LRT) to compare these models and get a p-value from the asymptotic χ^2_1 distribution.

The LRT for GLM described above gives an idea about the relative importance of fixed effects covariates that may be influential for light, moderate or heavy precipitation. The above procedure is supplemented by a Benjami-Hochberg (Benjamini and Hochberg, 1995) multiple testing correction procedure for 40 tests using the `p.adjust{stats}` function in R. This part of the analysis did not include random effects, owing to a lack of viable and theoretically justifiable testing procedure when a random effect is present. Inclusion of random effects are likely to reduce variance attributed to noise, thus, typically increasing significance levels.

A GLM with all fixed covariates to provide a test of goodness-of-fit based on residual deviance being asymptotically χ^2. For details, refer to (Faraway, 2006). This compares the fitted model to the saturated model which contains one parameter for each observation. Failure to reject the null hypothesis of this test indicates a lack-of-fit.

Finally, a test of the variance component in the GLMM fit by `glmer` is done using a LRT with a nonstandard asymptotic distribution. Because the models have a single variance component, the asymptotic distribution for the LRT corresponds half of the p-value obtained from the χ^2_1 distribution as noted by (Zhang and Lin, 2008). This test is only performed using `glmer` since the other two methods do not use maximum likelihood, although as noted before, the likelihood produced by `glmer` is approximate. Overall, we take a cautious view on the interpretation of these tests.

3.5.4 Discussion of Rainfall Models

Results of significance testing for each of the models can be found in Table 3.7. All covariates were significant in the light rainfall model in the majority of the years. However, 38% of the years showed lack-of-fit based on the deviance test. The moderate and extreme rainfall models showed no lack-of-fit, but had far fewer significant covariates over the years.

Clearly, maximum daily temperature was important in all three levels of rainfall aligning with the Clausius-Clapeyron equation regarding water vapor capacity of the atmosphere. Minimum temperature was significant in most years for light and moderate rainfall, but not in the majority of extreme rainfall models.

Elevation was also significant in many years across all rainfall levels. This aligns with the physical explanations of warm moist air cooling at higher altitudes to produce precipitation.

Latitude and longitude were both significant in most light rainfall years. Moderate and extreme rainfall did not show the same trend for latitude. Longitude was significant in over half of the extreme models, however, corrected p-values indicate these may mostly be false positives. Coefficient estimates for latitude indicated

Table 3.7: Proportion of p-values <0.05 for Annual LNMMs. p-values for fixed coefficients and goodness-of-fit test are from LRTs on GLM fits. p-values for the variance components are from LRTs that compare GLM and `glmer` fits. Proportions in parentheses represent corrected p-values based on the Benjamini-Hochberg correction for 40 tests.

	Variable	**Light**	**Moderate**	**Extreme**
Fixed	DMI	0.84 (0.78)	0.30 (0.22)	0.14 (0.03)
	Niño 3.4	0.68 (0.60)	0.20 (0.00)	0.13 (0.00)
	ΔTT	0.98 (0.98)	0.95 (0.95)	0.70 (0.70)
	Elevation	0.95 (0.95)	0.98 (0.98)	0.95 (0.93)
	Max. Temp.	1.00 (1.00)	0.98 (0.98)	1.00 (1.00)
	Min. Temp.	0.75 (0.75)	1.00 (1.00)	0.40 (0.33)
	Latitude	0.90 (0.75)	0.30 (0.00)	0.08 (0.00)
	Longitude	0.90 (0.78)	0.33 (0.03)	0.55 (0.15)
Random	Station	0.93 (0.51)	0.53 (0.53)	0.28 (0.20)
	Lack-of-Fit?	0.38 (0.38)	0.00 (0.00)	0.00 (0.00)

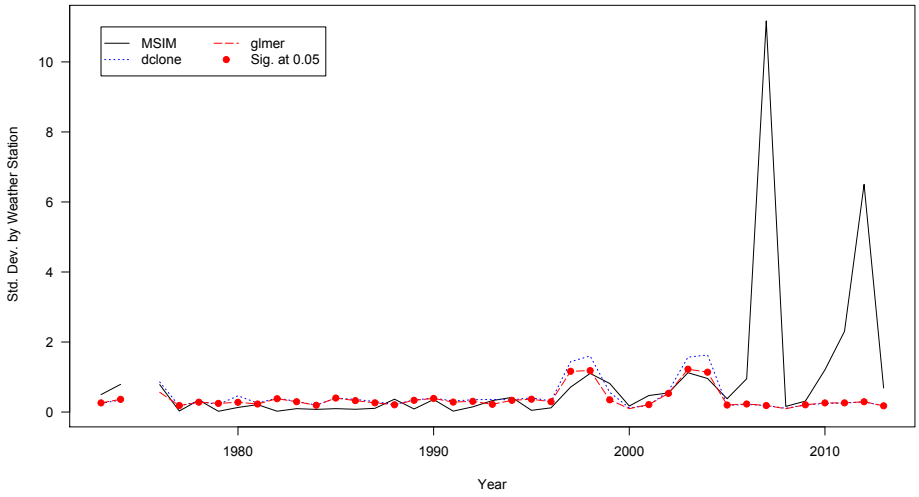
the probabilities of rain increasing going south to north. Longitude estimates were mostly negative indicating a decreasing probability of rainfall going west to east.

DMI was significant for the majority of light rainfall models, but not in moderate and extreme models. This corresponds to the DMI influence in non-ENSO drought years as hypothesized in previous literature. Similarly, ΔTT was significant in most years for all three rainfall levels as suggested by the hypothesis that it is instrumental in monsoon circulation.

The Niño 3.4 anomaly index was significant in the majority of light rainfall models, but in few of the moderate and extreme models. This could be related to the weakening of the relationship between ENSO and the Indian monsoon as discussed in (Chang et al., 2001), but may also be a function of the other covariates included in the modeling.

The station variance component was significant in many light rainfall models. In Figure 3.1, one can see there is less variability in general in light rainfall even

Figure 3.1: Weather station random effect standard deviation estimates for LN-MMs with light Indian rainfall ($0 < x \leq 64.4$ mm/day) as the response from 1973-2013. Estimates over time indicate variability near 0, however, most of the `glmer` estimates are significant at the 0.05 level.



though more years are significant. The variance was significant in about half of the moderate and one quarter of the extreme rainfall models. As seen in Figure 3.2 and Figure 3.3, these models tended to have higher magnitude variability than light rainfall even though fewer years were significant. The variance component does provide additional explanation for the rainfall variability and thus, vets the methodology use in this application. This verifies the thesis of this study- a meaningful portion of the variability in any precipitation category is a random component distinguishable from random noise variability.

Figure 3.2: Weather station random effect standard deviation estimates for LN-MMs with moderate Indian rainfall ($64.4 \leq x < 124.4$ mm/day) as the response from 1973-2013. Approximately half of the `glmer` estimates are significant at the 0.05 level.

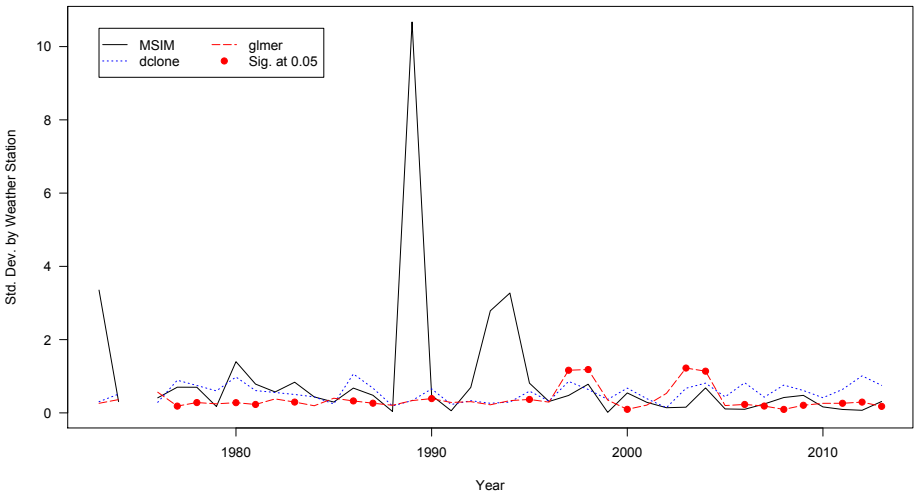
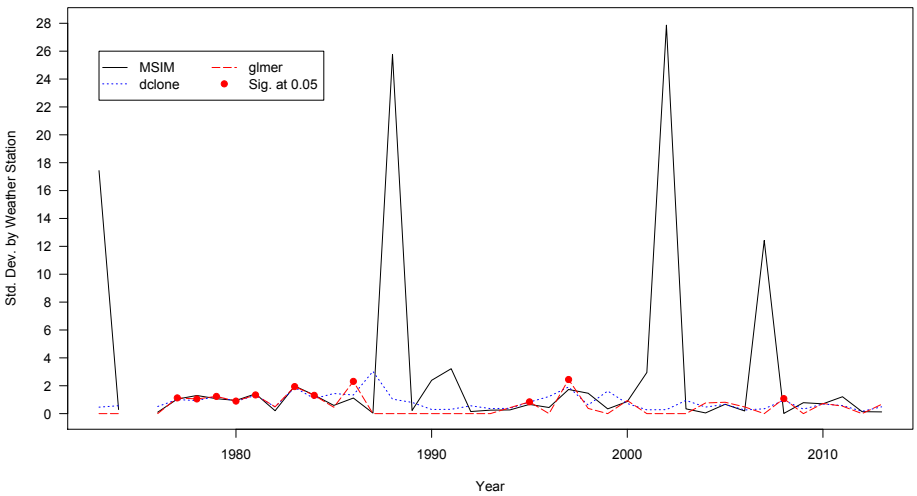


Figure 3.3: Weather station random effect standard deviation estimates for LN-MMs with extreme Indian rainfall (≥ 124.4 mm/day) as the response from 1973-2013. Approximately one quarter of the `glmer` estimates are significant at the 0.05 level.



3.5.5 Estimation Method Performance

Fixed coefficient estimates at each level of rainfall are depicted in Figure 3.4, Figure 3.5, and Figure 3.6. The three estimation methods produced different answers on at least some coefficients in each of the models. The best agreement amongst methods occurred for maximum temperature (moderate, extreme), longitude (light, extreme), and the variance components (all). Estimates for `glmer` and MSIM agreed more often than either agreed with `dclone`. However, they were did not always trend with each other. Light rainfall models displayed more disagreement among methods than moderate or extreme rainfall.

Standard deviations from `dclone` estimates indicated that the algorithm may have not converged for all parameters in the 10,000 samples taken from the posterior. One of the issues with using this method is difficulty in assessing convergence. It may require a much larger sample to provide suitable answers in this application. Based on this, `dclone` results were mostly inconclusive in this application.

The outcome of this application indicated `glmer` and MSIM provided more reasonable estimates, however, a longer run of `dclone` could also be useful. The three methods are representative of distinct statistical paradigms of estimation including approximate likelihood, Bayesian, and method of moments. Each of the methods has a distinct algorithm and set of assumptions, thus, it is wise to try multiple methods when using GLMM in an application.

Figure 3.4: Fixed coefficient estimates for LNMMs with light Indian rainfall ($0 < x < 64.4$ mm/day) response from 1973-2013.

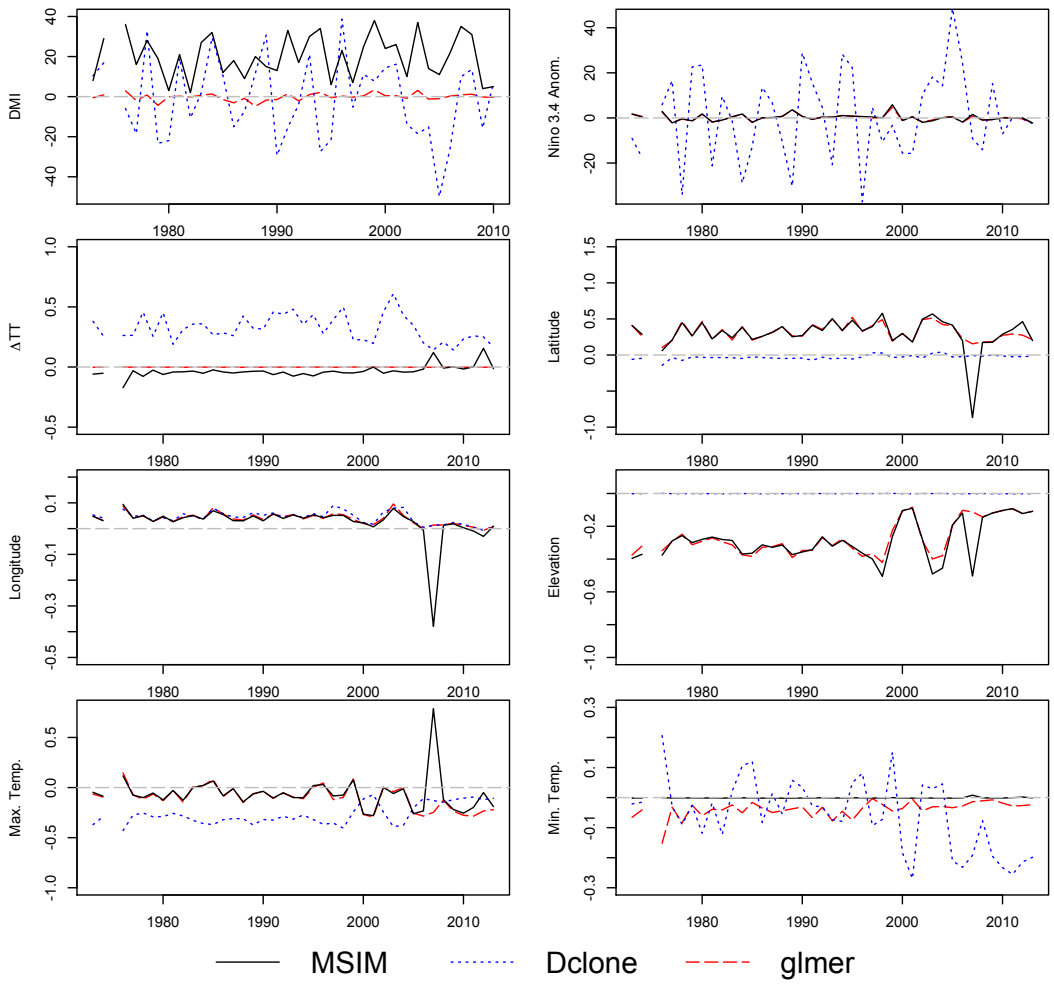


Figure 3.5: Fixed coefficient estimates for LNMMs with moderate Indian rainfall ($64.4 \leq x < 124.4$ mm/day) response from 1973-2013.

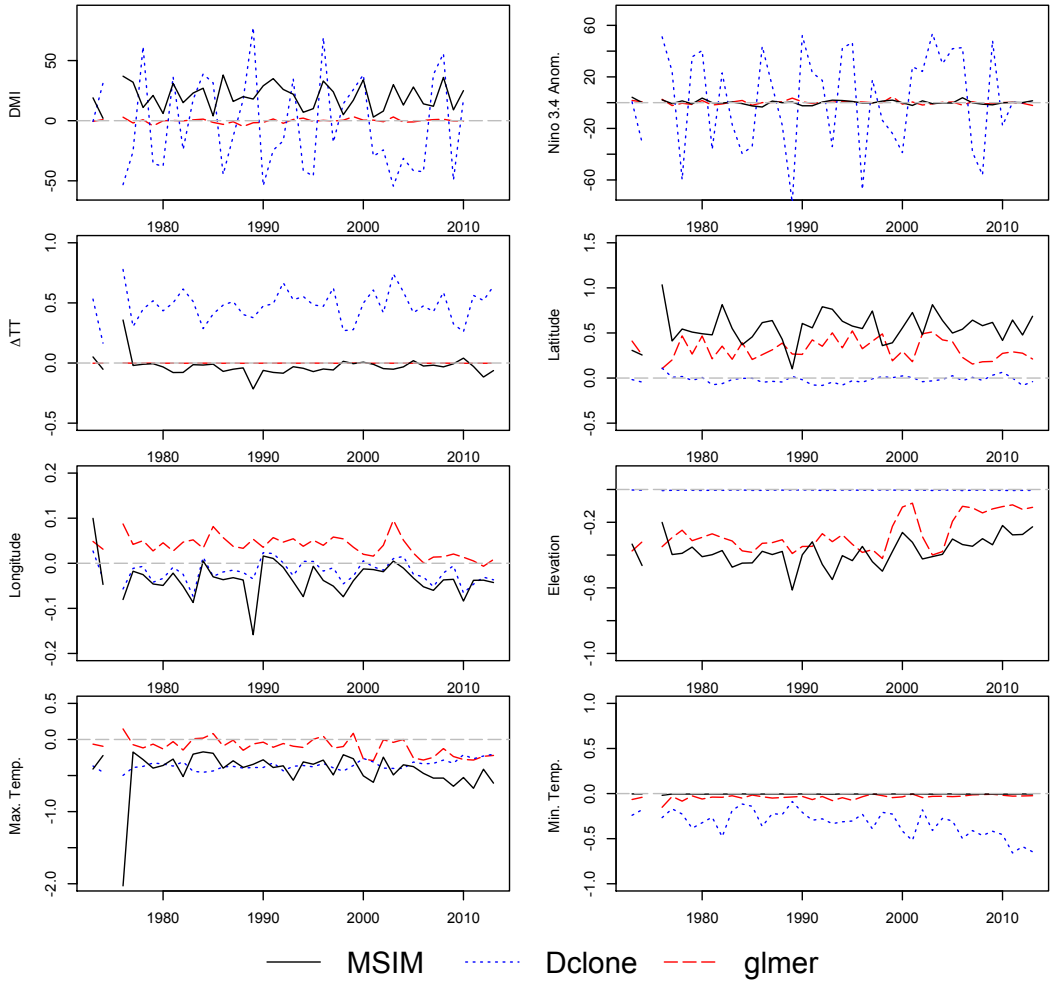
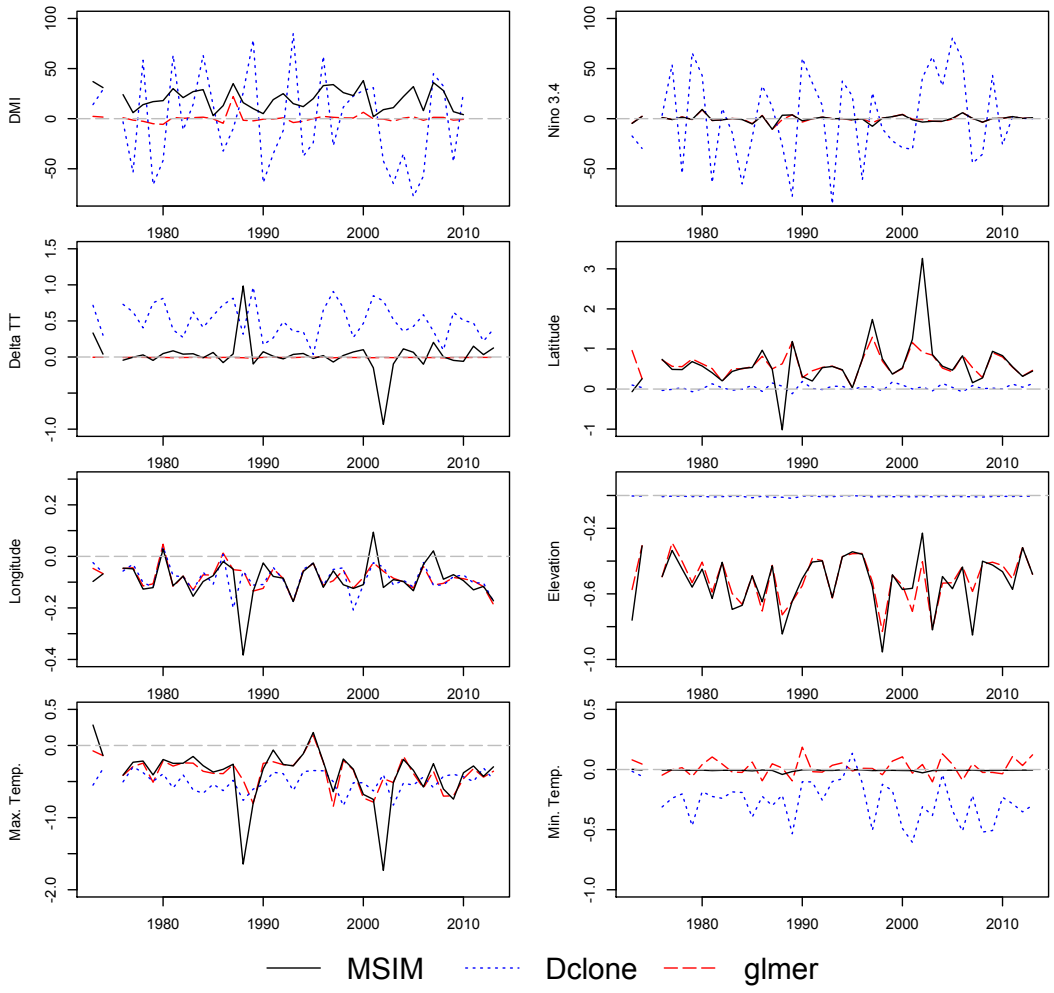


Figure 3.6: Fixed coefficient estimates for LNMMs with extreme Indian rainfall ($\geq$ 124.4 mm/day) response from 1973-2013.



3.5.6 Conclusions of Study I

Outcomes for Indian monsoon rainfall coincide with results in the Indian monsoon literature providing evidence of the usefulness of GLMM. Physical constructs are preserved by the models demonstrated by the importance of elevation, maximum temperature, and ΔTT in all levels of rainfall. Random effects are significant in several of the models indicating promise of modeling some of the unobservable and complicated interactions that underlie climate patterns.

The GLMM methods explored purposely included several styles of estimation including approximate likelihood, Bayesian, and method of moments type estimators. Each exhibits unique drawbacks, thus, use of multiple methods in the context of any application can verify reliability of estimates. This can provide researchers with higher confidence in results and will be more robust to limitations of any of the individual methods.

3.6 Study II: Studying Indian Monsoon Precipitation Thresholds using GLMMs

3.6.1 Goals of Study II

Study II extends the results of Study I to examine the use of GLMMs in other previously studied types of thresholds for precipitation data. Station-defined percentile thresholds were used in (Krishnamurty et al., 2009) and fixed level thresholds similar to those in Study I were used in (Goswami et al., 2006) to explore trends in monsoon rainfall intensity. This study includes relevant covariates based on Study I and uses each threshold definition with a distinct purpose. The fixed threshold model is used to elicit a physical interpretation across rainfall levels,

and percentile-based thresholds are used to quantify local predicted probabilities of threshold exceedances and potential cycles in their occurrence.

Rather than estimating annual LNMMs as in (Dietz and Chatterjee, 2014), a more robust approach is employed by using a single model for the entire time period. In addition, other relevant covariates are included, while maintaining the station random effect. A model with separate station and year random effects is also investigated. In this study, data are limited to a more accurate summer monsoon season, June 1 to September 30, for each year.

3.6.2 All Years LNMM

Let station $i \in \{1, ..., m\}$, day $d \in \{1, ..., n_i\}$, and year $k \in \{1, ..., K\}$. Given a threshold τ and precipitation event Z_{idk}, let $T_{idk} = I(Z_{idk} > \tau)$. Let $\mathbf{x}_{idk}$ be a vector of covariates and $\mathbf{U}$ and $\mathbf{W}$ be vectors of random effects for station and year, respectively. Then, the LNMM is:

$$\text{Level 1: } T_{idk}|\mathbf{U} = \mathbf{u}, \mathbf{W} = \mathbf{w} \overset{\text{ind.}}{\sim} \text{Bernoulli}(\theta_{idk}), \tag{3.17}$$

$$\text{logit}(\theta_{idk}) = \mathbf{x}_{idk}^{\mathrm{T}}\boldsymbol{\beta} + u_i + w_k, \tag{3.18}$$

$$\text{Level 2: } U_i \overset{\text{ind.}}{\sim} \mathcal{N}(0, \sigma^2_{station}), W_k \overset{\text{ind.}}{\sim} \mathcal{N}(0, \sigma^2_{year}) \tag{3.19}$$

$$U_i \text{ independent of } W_k \text{ for all } (i, k). \tag{3.20}$$

Two other types of models are used to provide benchmarks. These include GLM, which do not take into account repeated measures by station or year, and generalized estimating equations (GEE) models with auto-regressive lag 1 (AR1) structure for repeated events within weather station. Model selection is not used in this study; instead, models consisted of scientifically relevant covariates based on earlier literature.

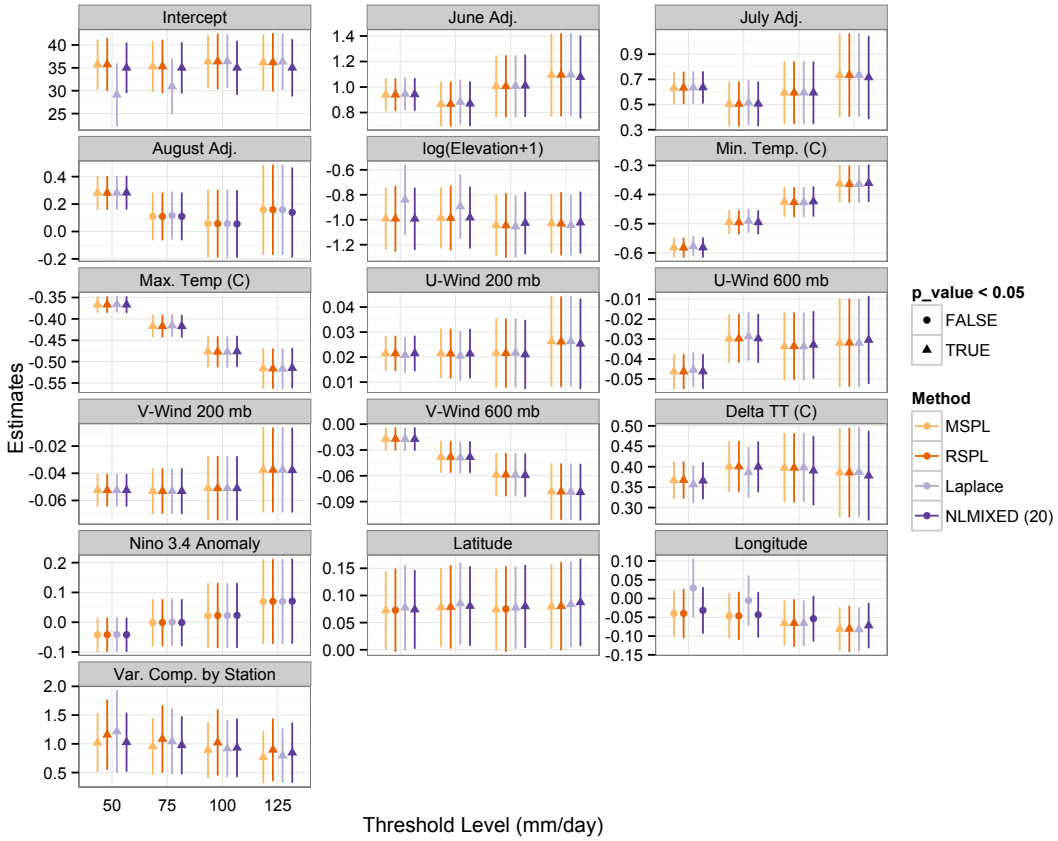
Within the rest of the study, discussion is provided on the fixed and percentile-based threshold models. Section 3.6.4 focuses on the interpretation of fixed threshold models in understanding covariates and variability at different threshold levels. Section 3.6.5 discusses the use of percentile-based threshold models to provide predicted probabilities on a local scale. Final commentary is highlighted in Section 3.6.6.

3.6.3 LNMM Estimation Comparison

SAS provides several methods of GLMM estimation as discussed in Section 3.3.2. Despite theoretical differences in the estimation techniques, PL methods in PROC GLIMMIX produced nearly the same results for the model in (3.17)-(3.20) as a 20 point quadrature estimated by PROC NLMIXED. Population-averaged expansions for random effects failed in both a maximum likelihood and a restricted maximum likelihood sense in cases with few Bernoulli successes. Thus, only subject-specific methods (MSPL[7] and RSPL[8]) estimates are reported for comparison. Laplace approximation was also implemented using PROC GLIMMIX. As seen in Figure 3.7, differences in the methods were quite minimal. Based on this, RSPL estimates were utilized in the study based on having the best computational speed with a large data set.

[7] *M*aximum *S*ubject-specific *P*seudo-*L*ikelihood
[8] *R*estricted maximum *S*ubject-specific *P*seudo-*L*ikelihood

Figure 3.7: Fixed threshold GLMM coefficients by Methodology. Statistical significance at the $\alpha = 0.05$ level is represented by the marker for each estimate. The reference level for the month adjustment is September, thus, statistical significance represents the significance of the difference from September. The bars represent 2 standard errors as estimated by each model using the default techniques within SAS.



GLMs and GEEs were estimated using PROC GENMOD. Uncertainty estimates within this study correspond to the default methods in these procedures. GLMM approximate standard errors for fixed effects are obtained by use of the delta method on the predicted population averaged probability estimates; variance component standard errors are based on asymptotic theory. GLM estimates use asymptotic normal standard errors while GEE provides empirically based standard errors. Detailed information on these procedures can be found in SAS Institute Inc. (2011).

3.6.4 Fixed Threshold Logit-Normal Models

Fixed thresholds of 50, 75, 100, and 125 mm/day were selected for modeling. 50 and 75 mm/day represent light and moderate thresholds. 100 mm/day was the high setting used by (Goswami et al., 2006), and 124.4 mm/day was the high setting in (Dietz and Chatterjee, 2014) based on (Attri and Tyagi, 2010). 125 mm/day is meant to approximate this.

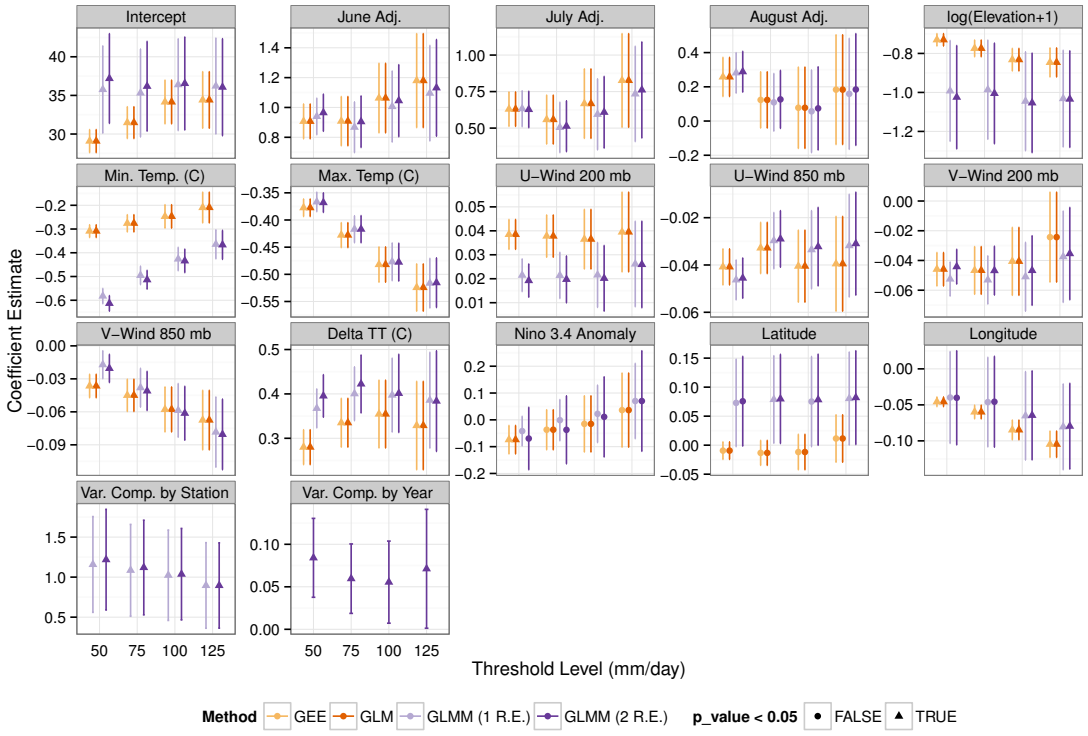
3.6.4.1 Fixed Threshold Fixed Effect Analysis

Coefficients for fixed thresholds are seen in Figure 3.8. Covariates are not scaled within the models to facilitate comparisons across different model types (GLM, GEE, GLMM). The Niño 3.4 anomaly, latitude, and longitude generally display non-significant estimates, although longitude is significant at higher thresholds.

Intercepts are higher in the GLMMs compared to GEE or GLM especially in lower thresholds. The intercept is constant over thresholds in the GLMMs, while GEE and GLM coefficients increase with threshold.

Monthly adjustments for June and July indicate a significant positive effect compared to September. August is not significantly different from September.

Figure 3.8: Fixed threshold fixed coefficient estimates. Statistical significance at $\alpha = 0.05$ level is represented by marker shape. The reference level for month is September, i.e., statistical significance indicates significant difference from September. Bars represent 2 standard errors.



June and July show a increasing trend as the threshold increases. This indicates a higher predicted probability of more extensive rainfall in June and July in comparison to September. This insight is consistent with earlier summer months typically containing more large rainfall events than September.

Western low elevation coastal areas and northeastern low lands receiving a large amount of rainfall may contribute to the significantly negative coefficient for log(Elevation+1). This estimate is relatively constant over threshold levels indicating a consistent effect. Both minimum and maximum temperatures coefficients are significantly negative. However, as the threshold increases, the magnitude of the minimum temperature coefficient decreases while the magnitude of the maximum temperature coefficient increases.

All monsoon circulation variables are significant in the models. The u-wind coefficients are positive at 200 mb and negative at 850 mb. Both are relatively constant as the threshold increased. The v-wind coefficients are negative at both pressure levels. The 850 mb coefficient decreased as threshold increased while the 200 mb is essentially constant as the threshold increased. The coefficient for ΔTT is significantly positive indicating higher probability of threshold exceedance as ΔTT increases.

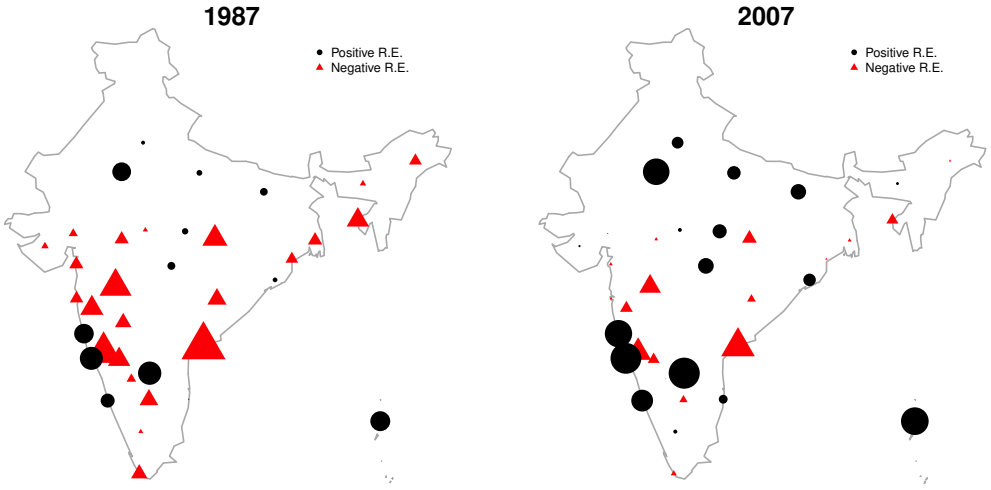
3.6.4.2 Fixed Threshold Random Effect Analysis

Testing for variance components[9] indicates both the intercept by station and intercept by year are significant over all threshold levels. However, the annual component makes up a much smaller proportion of the estimated variability. The station component decreases slightly as threshold increases.

In Figure 3.9, estimated random effects of the 125 mm/day exceedance model

[9]Note that this is the standard deviation (σ) of the random effects distribution

Figure 3.9: Estimated random effects for >125 mm/day rainfall. The magnitude is depicted by the relative size of the marker. Triangles (circles) indicate negative (positive) estimated random effects.



with station and year random effects are shown for two different years. Positive (negative) random effects correspond to a higher (lower) probability of rainfall than that estimated by the fixed effects alone. Stations tend to consistently indicate either positive or negative (of varying magnitudes by year) random effects.

In 1987, negative random effects were larger and mostly fell within the center of India. In 2007, the positive random effects were more pronounced especially along the west coast and northern areas of the subcontinent. The two years examined were compared with Indian Meteorological Society rainfall data[10]. This annual summer monsoon season data provides percentage deviations from average rainfall amounts for four geographic demarcations in India- northwest, central, northeast, and south peninsula. In 1987, all but northeast India indicated drought which agrees with the stronger negative random effects produced by the model. 2007 had higher than average rainfall in all but northwest India; again, this agrees with

[10] http://www.imd.gov.in/section/nhac/dynamic/data.htm

the stronger positive random effects and higher chances of a large precipitation even. Correspondence is not one-to-one because the model is fitting probabilities of exceedances rather than actual rainfall, but provides intuition for random effects.

3.6.5 Percentile Threshold Logit-Normal Models

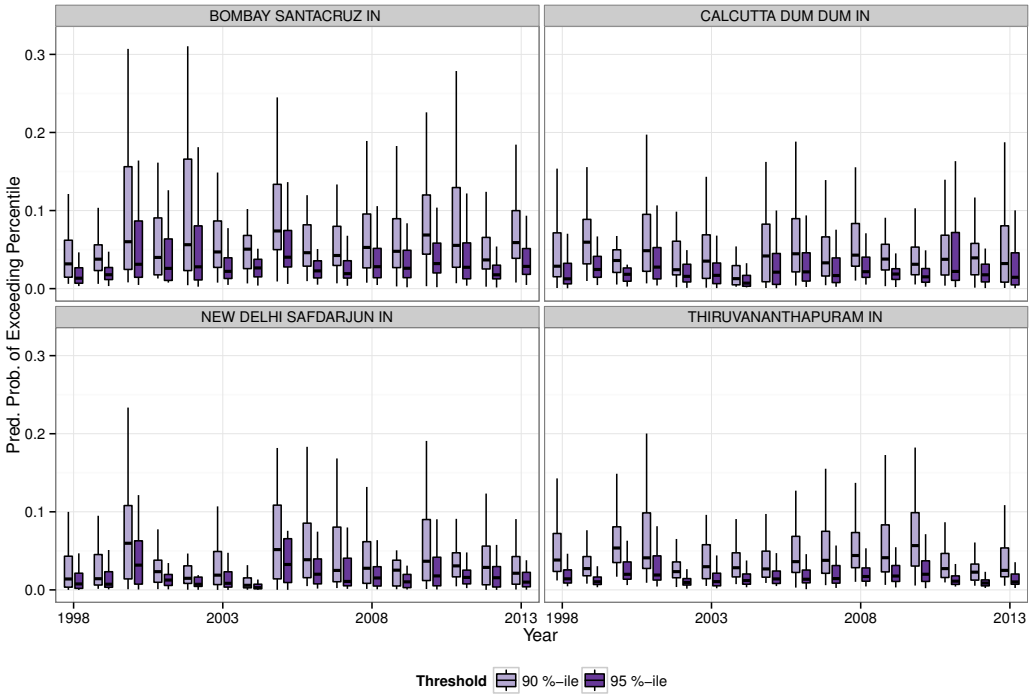
In (Krishnamurty et al., 2009), medians of the annual 90^{th} and 99^{th} percentiles were used as thresholds for examining station-level percentile exceedances. Because of missing data, thresholds were defined using direct 90^{th}, 95^{th}, and 99^{th} percentiles of the data. Models for the 99^{th} percentile failed to converge and are excluded.

Threshold exceedance predictions for 4 representative stations are displayed in Figure 3.10. Box plots indicate an expected pattern of decreasing probability as the threshold moves from the 90^{th} to the 95^{th} percentile. West coast stations, represented by Bombay (Mumbai), have markedly higher probabilities of exceeding their station thresholds. Bombay has station thresholds of 59.9 mm/day (90^{th}) and 92.9 mm/day (95^{th}). In comparison, more moderate exceedance probabilities were predicted for Calcutta (Kolkata) and New Delhi. Calcutta has thresholds of 39.9 mm/day (90^{th}) and 56.9 mm/day (95^{th}) and New Delhi thresholds are 34.0 mm/day (90^{th}) and 52.1 mm/day (95^{th}). Thiruvanthapuram, in the southernmost region of India, indicated low predicted probabilities of exceeding its extreme thresholds of 34.0 mm/day (90^{th}) and 49.0 mm/day (95^{th}). When compared with fixed thresholds analysis, percentile-based analysis suggests the use of much lower thresholds for understanding local monsoon behavior.

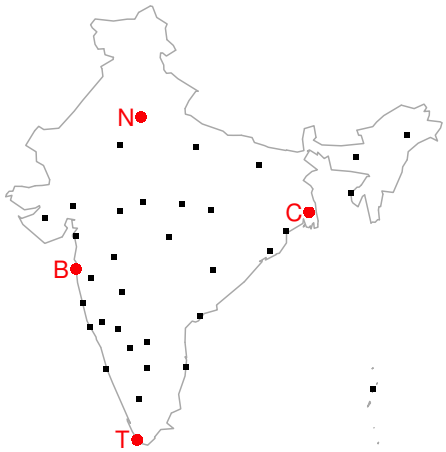
A final observation is the appearance of an irregular cycle in probability predictions in the stations from 1998-2013. The cycle is not consistent among all stations. One hypothesis is this may be due to the random effects for each station in each year which captures some of the idiosyncratic features of a location.

Figure 3.10: Percentile Threshold Predictions. Box plots show the distribution of daily predictions by year. Outliers are not shown for clarity of the graphics and consisted of <5 of yearly predictions.

(a) Box Plot Series for 4 Indian Cities



(b) Map of Highlighted Weather Stations



3.6.6 Conclusions of Study II

The analysis in this study serves as a starting point for climate scientists in exploring thresholds. These thresholds can be applied in an explicit context of risk assessment to civil structures or in an implicit context for further modeling. Specifically, fixed threshold analysis statistically examines relationships of climate covariates with rainfall probabilities in the context of increasing thresholds. This may be helpful in a large scale analysis of the Indian monsoon. Percentile-based thresholds are useful at a local scale for understanding risks of certain levels of rainfall.

Possible limitations of the approach include model fit and data issues. One measure of fit provided within SAS is a Generalized Chi-Square (GCS) statistic. This statistic should be around 1 if the model fits well. Fixed threshold models GCS ranged from 1.06 - 2.06, and increased with threshold, indicating a slight issue in fit at the higher thresholds. There were also outliers indicated by residual plots which may require a more robust fit in future. Missing data could be driving some of the results; several possibly important areas of India are not included in the data set based on availability. Unfortunately, the wet northeast as well as the central and northwest regions of India are poorly covered. Aggregating data may provide a different perspective and a more stable fit.

In general, the LNMM in this context provides valuable physical insights, such as the increasing importance of maximum temperature as threshold increases, as well as understanding of local predictions and their cycles. In Chapter 4, model residuals will be tested for spatio-temporal correlation to provide further evidence for the sufficiency of the GLMM in this context.

Chapter 4

Spatio-temporal Dependence Testing for the Indian Monsoon Logit-Normal Mixed Model

4.1 Hypothesis Testing for Space and Time Dependence

4.1.1 Problem Background

As seen in Chapter 3, the Indian monsoon precipitation thresholds model is useful for determining risk of threshold exceedances and can be applied in civil planning. The logit-normal mixed model specifically accounts for the time component by including repeated measures by station and year, however, it does not explicitly model possible spatial signals in the error structure through additional covariance requirements. Thus, to assess the adequacy of the model structure in this setting, we turn to spatio-temporal dependence hypothesis testing.

As discussed in Section 4.1.2, testing for space and time dependence is a well studied process in certain classes of problems. In testing residuals for spatio-temporal dependence, there is a choice of statistic and testing procedure to be made by the researcher. Our analysis assesses the commonly used Space-Time Index (STI) (Griffith, 1981) presented in Section 4.1.3. We also develop a new weighted statistic known as spatio-temporal identification (STrIde) in Section 4.1.4.

Several simulations comparing STI to STrIde are performed in Section 4.2. Asymptotic theory of the STrIde statistic will be assessed in future studies. Instead, we focus on a simulation study of the STrIde statistic using a variety of null hypotheses including (i) complete independence, (ii) temporal dependence only, and (iii) spatial dependence only. We examine Type I errors for (i)-(iii) in Section 4.2.5 and demonstrate empirical power differences between STrIde and STI in Section 4.2.6. The power study indicates STrIde is more powerful than STI in certain areas of the parameter space for separable spatio-temporal data.

Finally, in Section 4.3, we apply STrIde and STI to residuals arising from the logit-normal mixed model fit on the Indian monsoon precipitation data. The results provide evidence that the model has successfully captured the relevant spatio-temporal aspects of the data at higher thresholds even when using the more powerful STrIde statistic for hypothesis testing.

4.1.2 Literature Review

4.1.2.1 Temporal Dependence Testing

Temporal dependence in regression residuals was classically studied in (Durbin and Watson, 1950a,b). These papers develop the popular *Durbin-Watson Statistic* for testing for residual autocorrelation with an auto-regressive lag 1 (AR1) structure. This statistic is commonly used after a regression analysis to check for serial time correlation. However, the Durbin-Watson procedure tests the null hypothesis of independence versus a particular form of serial correlation, namely, the AR1 process. It is not informative about more general patterns of autocorrelation.

The *Box-Pierce* hypothesis test (Box and Pierce, 1970) is used on residuals of an autoregressive integrated moving average (ARIMA) model. One calculates the Q statistic, then tests independence up to k lags versus any lag of k or less exhibiting dependence, where k can be greater than or equal to 1. This test was refined by (Ljung and Box, 1978) to be more accurate for small samples and is applicable for univariate time series under the assumption of strictly exogenous regressors. The *Breusch-Godfrey* test (Breusch, 1978; Godfrey, 1978) extended the approach to test for autocorrelation in models with weakly exogenous regressors. Many other tests for specific types of time dependence can be found in econometrics literature.

4.1.2.2 Spatial Dependence Testing

A commonly used test for global spatial correlation between nearest neighbors in a two-dimensional stochastic scheme is *Moran's I* (Moran, 1950). Let $\{Y_i; i \in \{1, ..., n\}\}$ be a spatial series where i is a function of the coordinates, (i_1, i_2). Binary weights (w_{ij}) indicate a connection between location i and j. *Moran's I* is then:

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(Y_i - \bar{Y})(Y_j - \bar{Y})}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}, i \neq j. \tag{4.1}$$

Under either a randomization test implementation or a normality assumption for all Y_i, $E(I) = -\frac{1}{n-1}$. Thus, the null hypothesis of the test is $H_0 : \mu_I = -\frac{1}{n-1}$. Values of I significantly exceeding (less than) $-\frac{1}{n-1}$ indicate positive (negative) spatial autocorrelation. (Cliff and Ord, 1972) generalized *Moran's I* to derive a test for spatial correlation in a linear regression model.

Geary's C (also known as the contiguity ratio) is another classical measure of spatial correlation (Geary, 1954). It is considered to be more sensitive to local spatial autocorrelation than *Moran's I*. Using the same notation as in (4.1), the statistic is defined as:

$$C = \frac{(n-1)}{2\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(Y_i - Y_j)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}, i \neq j. \tag{4.2}$$

Again, for either a randomization test implementation or a normality assumption for all Y_i, $E(C) = 1$ and the null hypothesis is $H_0 : \mu_C = 1$. Values significantly lower (higher) than 1 indicate positive (negative) spatial correlation. *Moran's I*

and *Geary's C* are inversely linearly correlated but are not completely dependent.

However, (Li et al., 2007b) highlights some of the flaws of using *Moran's I* for spatial hypothesis testing within a spatial autoregressive model and develops an approximate profile-likelihood estimator (APLE) of the dependence parameter. The usefulness of APLE was highlighted using Monte Carlo testing as an approximation to permutation testing.

Other tests use the covariance function[1], C, of a spatial lattice process. First, define the two-dimensional spatial series by its coordinate values, $\{Y_{(i_1,i_2)}; i_1 \in \{1,...,n_1\}, i_2 \in \{1,...,n_2\}\}$. Then for a lag (g_1, g_2), the covariance function is:

$$C(g_1, g_2) = \text{cov}(Y_{(i_1,i_2)}, Y_{(i_1+g_1,i_2+g_2)}) = \mathbb{E}[(Y_{(i_1,i_2)} - \mu)(Y_{(i_1+g_1,i_2+g_2)} - \mu)] \tag{4.3}$$

Simplifying assumptions regarding the covariance can be tested through a spectral method developed in (Scaccia and Martin, 2005). Tests of axial symmetry, $H_0 : C(g_1, g_2) = C(g_1, -g_2)$, and separability, $H_0 : C(g_1, g_2) \propto C(g_1, 0)C(0, g_2)$, were developed to be performed sequentially since separability implies symmetry.

4.1.2.3 Spatio-temporal Dependence Testing

Testing for spatio-temporal dependence is not as straightforward as in the individual space or time settings. The goals of most tests in this area are to test separability or symmetry in spatio-temporal covariance functions. We present three such examples of tests, however, there exist other more specific tests in this field.

Define a spatio-temporal field as $\{Y(\mathbf{s}, t); \mathbf{s} \in \mathbb{R}^d, t \in \mathbb{Z}\}$, where $\mathbf{s}$ represents space and t represents time. The covariance function for space lag, $\mathbf{h}$, and time

[1] This is also known as the covariogram.

lag, u, is:

$$C(\mathbf{h}, u) = \text{cov}[Y(\mathbf{s}, t), Y(\mathbf{s} + \mathbf{h}, t + u)]. \tag{4.4}$$

The field is separable if

$$C(\mathbf{h}, u) \propto C_1(\mathbf{s}, \mathbf{s} + \mathbf{h})C_2(t, t + u) \tag{4.5}$$

for some spatial covariance, C_1 and temporal covariance, C_2.

In (Fuentes, 2006), a test of separability with H_0 identical to (4.5), is developed based on a nonparametric estimator of the spectral density. The test is carried out via ANOVA on a coherency function related to the spectral density estimate. Asymptotic theory related to the estimators is developed and an empirical power of 0.90 is demonstrated in selected simulations. However, the test is based on the spectral density estimate as a function of two filters which need to be carefully chosen by the practictioner.

A likelihood ratio test (LRT) for separability is derived in (Mitchell et al., 2006) in the context of multivariate repeated measures. In this case, $Y(\mathbf{s}, t)$ is thought of as $Y = XB + E\Sigma^{1/2}$ where X is a full rank covariate matrix, B is a parameter matrix, and E is an error matrix with independent rows of $N_m(0, I_m)$ distribution. Testing (4.5) equates to testing $H_0 : \Sigma = \Sigma_s \otimes \Sigma_t$ where Σ_s is a space covariance matrix and Σ_t is a time covariance matrix. Asymptotics based on the LRT were used to perform the test. However, empirical Type I error estimates indicated the asymptotic distribution failed to reject much more often than the desired size of the test even with relatively large samples.

Finally, (Li et al., 2007a) developed a test statistic for generalized hypothesis testing of covariance functions in spatio-temporal models. The test can be used

for separability, $H_0 : C(\mathbf{h}, u) \propto C(\mathbf{h}, 0)C(\mathbf{0}, u)$, and full symmetry, $H_0 : C(\mathbf{h}, u) = C(\mathbf{h}, -u)$ or $C(\mathbf{h}, u) = C(-\mathbf{h}, u)$. The test is formulated as a generic linear contrast and asymptotic results are used to perform the test. Simulations indicated good power for the test when there were more than 500 time points.

These tests implicitly assume that spatial-temporal dependence is present rather than testing for its existence. A review of the well-reputed texts in this area including (Cressie and Wikle, 2011), (Cressie, 1993), and (Banerjee et al., 2014) did not reveal hypothesis tests specifically for identifying spatio-temporal dependence among residuals as is needed for our analysis. However, (Cressie and Wikle, 2011) suggested the use of the STI in (Griffith, 1981) as a preliminary analysis tool for detection of spatio-temporal dependence before pursuing a spatio-temporal model. Thus, we move forward in assessing STI for our climate model context.

4.1.3 STI Statistic

The STI is an interpretable index for assessing the presence of spatial-temporal dependence in some contexts. It combines *Moran's I* and the *Durbin-Watson Statistic* and remains in current use as seen in (Kwan et al., 2015; Mitze, 2012; Griffith, 1988).

STI identifies spatio-temporal dependence under certain model constraints. Let $t \in \{1, ..., T\}$ represent discrete time and $i \in \{1, ..., n\}$ represent spatial location. Next, assume the spatio-temporal series, $\mathbf{Y}$, follows a vector autoregressive (VAR) process where each n-dimensional column vector, $\mathbf{Y}_t$ has the following form:

$$\mathbf{Y}_t = \mathbf{A}\mathbf{Y}_{t-1} + \boldsymbol{\phi}_t; \tag{4.6}$$

$$\boldsymbol{\phi}_t \overset{ind.}{\sim} \mathcal{N}_n(\mathbf{0}, \sigma^2\mathbf{I}), \tag{4.7}$$

with $\mathbf{A}$ being a fixed matrix of the spatial dependence in the previous time step. Note that in (Griffith, 1981), the error term, $\boldsymbol{\phi}_t$, was not set as a random variable, however, the implication was that it should be used in this sense. Define the overall mean as

$$\bar{Y} = \frac{1}{nT} \sum_{t=1}^{T} \sum_{i=1}^{n} Y_{i,t}. \tag{4.8}$$

and let $w_{ij,t-1}$=1 if stations i and j are neighbors at time $t-1$ and 0 otherwise. Then, STI is

$$STI = \frac{n(T-1)}{\sum_{t=2}^{T} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij,t-1}} \frac{\sum_{t=2}^{T} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij,t-1}(Y_{i,t} - \bar{Y})(Y_{j,t-1} - \bar{Y})}{\sum_{t=1}^{T} \sum_{i=1}^{n} (Y_{i,t} - \bar{Y})^2}. \tag{4.9}$$

The definition of $\bar{Y}$ hints at the possible detrimental effect of nonstationarity in data on STI. This was investigated within (Henebry, 1995) in terms of providing inference for spatial patterns. Model residuals are more likely to be stationary with mean 0, and should therefore, be a reasonable application for STI. The testing relies on a sampling distribution derived under independent, identically distributed normality assumptions of the elements of $\mathbf{Y}$. No central limit theorem is presented in (Griffith, 1981) to confirm the validity of the test under non-normal scenarios.

4.1.4 STrIde Statistic

Although STI provides a useful first effort in identifying spatio-temporal correlation in residuals, the hypothesis for testing is currently fairly restrictive. Thus, our goals in this study are to modify the null hypothesis to test presence of temporal

correlation or spatial correlation separately, as well as to modify the statement of the space-time process to include different forms of dependence.

We present a new statistic similar to (Li et al., 2007a), but most comparable to STI in its ease of use. Let $\{Z(\mathbf{s},t), \mathbf{s} \in \mathbb{R}^d, t \in \mathbb{Z}\}$ be a spatio-temporal residual series with locations $\mathcal{S} = \{\mathbf{s}_1, ..., \mathbf{s}_S\}$ and times $\mathcal{T} = \{t_1, ..., t_T\}$ where each element has mean 0 and a common variance σ^2. Then,

$$STrIde = \frac{1}{|\mathcal{S} \otimes \mathcal{T}|} \sum_{(\mathbf{s}_\ell, \mathbf{s}_m, t_i, t_j)} w(\mathbf{s}_\ell, \mathbf{s}_m, t_i, t_j) Z(\mathbf{s}_\ell, t_i) Z(\mathbf{s}_m, t_j). \tag{4.10}$$

Using the total elements, $|\mathcal{S} \otimes \mathcal{T}|$, as the denominator was based on the choice of testing procedure. We could sum the weights in the denominator, however, it will just be a multiple of the total elements. Our test relies on permutations and is data driven rather than relying on asymptotic theory. It also allows for a much longer time dependence to be incorporated than STI.

Choosing a weight, $w(\mathbf{s}_i, \mathbf{s}_j, t, u)$, could be done in many ways. We investigate 2 weights: one which is a function of distance in space and time and one binary version. Let $\mathbf{s}_m = [\text{lat}_m, \text{long}_m]$. Also, let $G(\mathbf{s}_m)$ indicate the neighborhood of $\mathbf{s}_m$. Points outside of this neighborhood are not considered in the calculation using the initial weighting scheme of

$$w(\mathbf{s}_\ell, \mathbf{s}_m, t_i, t_j) = C f(\mathbf{s}_\ell, \mathbf{s}_m, w_s) g(t_i, t_j, w_t). \tag{4.11}$$

This weight function allows for a space component, f, and a time component, g, as well as an overall modification, C.

The first weight we investigate is the following:

$$w_1(\mathbf{s}_\ell, \mathbf{s}_m, t_i, t_j) = I(\mathbf{s}_\ell \in G(\mathbf{s}_m)) \exp\left(-\frac{d_s(\mathbf{s}_\ell, \mathbf{s}_m)}{a_s}\right) \exp\left(-\frac{d_t(t_i, t_j)}{a_t}\right). \tag{4.12}$$

This weight function allows for a spatial modification, a_s, and a time modification, a_t. Distance functions, d_s and d_t can be chosen by the user, however, this study restricts the distance functions to the Euclidean norm.

As a secondary weight, we use a binary form based on the distance of a point in both time and space. This is similar to what is done in STI and is defined as:

$$w_2(\mathbf{s}_\ell, \mathbf{s}_m, t_i, t_j) = I(|t_i - t_j| \leq 1) I(\mathbf{s}_\ell \in G(\mathbf{s}_m)). \tag{4.13}$$

4.2 Simulation Study for the STrIde Statistic

4.2.1 Simulating a Separable Spatio-temporal Series

To generate data points indexed by space and time, we use a set of locations, $\mathcal{S} = \{\mathbf{s}_1, ..., \mathbf{s}_S\}$, where again, $\mathbf{s}_m = (\text{lat}_m, \text{long}_m)$. Also, define a set of times on n data points as $\mathcal{T} = \{t_1, ..., t_T\}$. Now, let the $S \times S$ space covariance matrix, Σ_S, be

$$\Sigma_S = \begin{bmatrix} \nu_{11} & \dots & \nu_{1S} \\ \vdots & \ddots & \vdots \\ \nu_{S1} & \dots & \nu_{SS} \end{bmatrix} \tag{4.14}$$

and let the $T \times T$ time covariance matrix, Σ_T, be

$$\Sigma_T = \begin{bmatrix} \phi_{11} & \dots & \phi_{1T} \\ \vdots & \ddots & \vdots \\ \phi_{T1} & \dots & \phi_{TT} \end{bmatrix}. \tag{4.15}$$

Because they are covariance matrices, Σ_S and Σ_T are symmetric and positive semi-definite. Let

$$\mathbf{Y}_{ST} \sim N_{ST}\left(\mathbf{0}_{ST}, \mathbf{I}_{ST}\right) \text{ (i.e. } \mathbf{Y}_{ST} \text{ is a vector of length } S * T). \tag{4.16}$$

Arrange $\mathbf{Y}_{ST}$ into an $S \times T$ matrix, $\mathbf{Y}_{S\times T}$, and let

$$\mathbf{Z}_{S\times T} = \Sigma_S^{1/2} \mathbf{Y}_{S\times T} \Sigma_T^{1/2}. \tag{4.17}$$

Then,

$$\mathbb{E}\left(\mathbf{Z}_{S\times T}\right) = \mathbf{0}_{S\times T}, \tag{4.18}$$

and by, Appendix C.1,

$$\mathbb{V}\left(\text{vec}\left(\mathbf{Z}_{S\times T}\right)\right) = \Sigma_S \otimes \Sigma_T \tag{4.19}$$

$$= \left(\begin{array}{c|c|c} \phi_{11}\Sigma_S & \dots & \phi_{1T}\Sigma_S \\ \hline \vdots & \ddots & \vdots \\ \hline \phi_{T1}\Sigma_S & \dots & \phi_{TT}\Sigma_S \end{array}\right) \tag{4.20}$$

$$= \left(\begin{array}{ccc|c|ccc} \phi_{11}\cdot\nu_{11} & \dots & \phi_{11}\cdot\nu_{1S} & & \phi_{1T}\cdot\nu_{11} & \dots & \phi_{1T}\cdot\nu_{1S} \\ \vdots & \ddots & \vdots & \dots & \vdots & \ddots & \vdots \\ \phi_{11}\cdot\nu_{S1} & \dots & \phi_{11}\cdot\nu_{SS} & & \phi_{1T}\cdot\nu_{S1} & \dots & \phi_{1T}\cdot\nu_{SS} \\ \hline \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \hline \phi_{T1}\cdot\nu_{11} & \dots & \phi_{T1}\cdot\nu_{1S} & & \phi_{TT}\cdot\nu_{11} & \dots & \phi_{TT}\cdot\nu_{1S} \\ \vdots & \ddots & \vdots & \dots & \vdots & \ddots & \vdots \\ \phi_{T1}\cdot\nu_{S1} & \dots & \phi_{T1}\cdot\nu_{SS} & & \phi_{TT}\cdot\nu_{S1} & \dots & \phi_{TT}\cdot\nu_{SS} \end{array}\right). \tag{4.21}$$

4.2.2 Specifying the Simulation Model in R

In order to implement the simulation in R, several settings were fixed while others were varied over simulations. Table 4.1 lists the various simulation specifications. Note that each experiment only used a portion of these settings, and they are explicitly stated within the results sections.

Table 4.1: Simulation Control Settings

Variable	Settings Used
Spatial grid size	4×4 or 7×7
Time points per grid point	30, 50, or 100
Number of random data sets	100
Time dependence (ρ_t)	-0.9, -0.5, -0.1, 0.1, 0.5, or 0.9
Time variance (σ_t)	1
Space dependence (ρ_s)	0.1, 0.3, 0.5, 0.7, or 0.9
Space variance (σ_s)	1

Properly defining covariances in (4.18) is essential in the following simulations. Time covariance between times i and j (ϕ_{ij}) was set using the following function:

$$\phi_{ij} = \sigma_t^2 \rho_t^{d_t(t_i,t_j)}; d_t(t_i,t_j) = |t_i - t_j|. \tag{4.22}$$

Spatial covariance between two-dimensional grid locations ℓ and m ($\nu_{\ell m}$) was simulated using `cov.sp{SpatialTools}` (French, 2014) in R. The function parameterizes using the variance[2] of the hidden process (σ_s^2) and the strength of dependence between locations (γ). To put this into parameters similar to the time covariance, we used the following equivalencies to transform γ:

$$\begin{aligned}
\nu_{lm} &= \sigma_s^2 \exp\left(\frac{-d_s(\mathbf{s}_\ell, \mathbf{s}_m)}{\gamma}\right) && (4.23)\\
&= \sigma_s^2 \exp\left(\frac{-d_s(\mathbf{s}_\ell, \mathbf{s}_m)}{-1/\log(\rho_s)}\right) && (4.24)\\
&= \sigma_s^2 \exp(d_s(\mathbf{s}_\ell, \mathbf{s}_m)\log(\rho_s)) && (4.25)\\
&= \sigma_s^2 \left[\exp(\log(\rho_s))\right]^{d_s(\mathbf{s}_\ell,\mathbf{s}_m)} && (4.26)\\
&= \sigma_s^2 \rho_s^{d_s(\mathbf{s}_\ell,\mathbf{s}_m)} && (4.27)
\end{aligned}$$

In this case, the distance function used was simply the Euclidean distance between two-dimensional spatial locations,

$$d_s(\mathbf{s}_\ell, \mathbf{s}_m) = \sqrt{(\text{lat}_\ell - \text{lat}_m)^2 + (\text{long}_\ell - \text{long}_m)^2}. \tag{4.28}$$

4.2.3 STrIde Statistical Settings

Several choices must be made in order to implement the statistical hypothesis testing using the STrIde statistic. The most pertinent of these are listed in Table 4.2

[2] This is sometimes called the partial sill.

for the simulation study.

Table 4.2: Statistical Control Settings for STrIde

Variable	Settings Used
Neighbors	1+, 3+
Number of bootstraps (B_P)	1000
Constant weight adjustment (C)	1
Time weight adjustment (a_t)	7
Space weight adjustment (a_s)	Maximum Euclidean Distance between neighbors
Null Hypothesis	Independent, Time, Space
Nominal significance level (α)	0.05

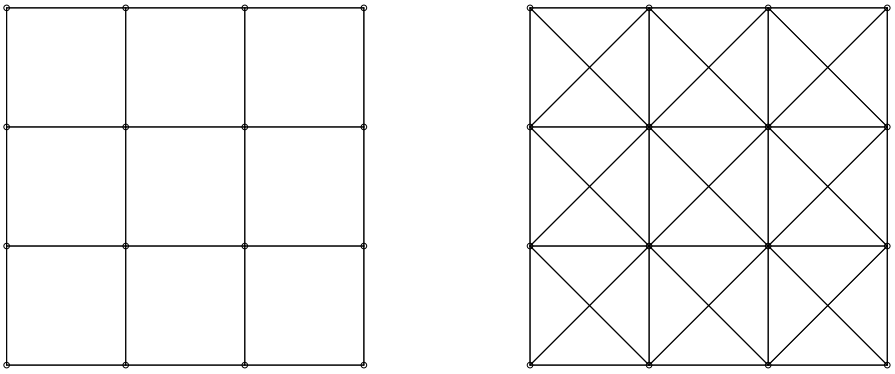
We also note that a missing data point elicits removal of any term using it from the summation as well as the denominator of the statistic. Additionally, the nominal significance level was chosen out of convention for a standard study to be evaluated at a 0.05 level. The number of bootstraps for permutation testing was fixed at 1000. The procedure for testing is addressed in Section 4.2.4.

In w_1, the choice of time and space modifications, a_t and a_s, was somewhat arbitrary and will elicit further study in the future. The current choices were made to put a lower bound on spatial dependence, but to have time correlation go to zero fairly quickly after 7 time periods. Considering the use for climate data, spatial dependence within a small area is fairly consistent, while time dependence drops quickly especially in daily data as in the application in Section 4.3.

One can specify a distance for spatial dependence to exist, or, implicitly specify distance by calculating neighbors such that all locations have at least n neighbors. For example, if each station is required to have at least 1 neighbor, this is equivalent to a distance setting anywhere between the maximum distance between two neighbors and the minimum distance at which all stations would have at least 2 neighbors. This was implemented in R using `dnearneigh{spdep}` (Bivand and

Figure 4.1: Neighbor alignment in a 4×4 location grid

(a) 1+ Neighbors (Rook Contiguity) (b) 3+ Neighbors (Queen Contiguity)



Piras, 2015). Figure 4.1 displays the 1+ and 3+ neighbor networks for the 4×4 grid scenario. Because the points are on an equally spaced grid, forcing each point to have 1 neighbor ensures all points in panel (a) have at least 2 neighbors. Interior points have 4 neighbors. In panel (b), requiring 3 neighbors provides a minimum of 3 for the corner points and a maximum of 8 for interior points. Further study on geographic layouts and the boundary effect could be tested in future research.

4.2.4 Nonparametric Bootstrapping Approximation to Permutation Testing

Testing based on permutations of the data allows for a more robust procedure and an extension of STI which doesn't rely on asymptotic normality to provide p-values. Because the number of possible permutations of the data is high for any meaningful data set, we bootstrapped samples of the set of possible permutations.

We allow sampling of permutations with replacement, but note that selection of the same permutation is of very low probability. For example, in permuting each element independently in a data set with $n = 10$, the number of permutations is 10! =3,628,800 indicating the chance of selecting a given permutation is 1/3,628,800. As noted in Section 4.2.3, in all simulations and applications, we selected $B_p = 1000$. Exposition on this type of sampling has been well studied in the literature and shown to work without much loss of power. Further details may be found in (Dwass, 1957; Marriott, 1979; Ernst, 2004; Jockel, 1986), § 3.6 of (Edgington and Onghena, 2007), or § 7.1.2.3 of (Thas, 2010). Simulation studies in (Guarte and Barrios, 2013) indicated this kind the procedure can correctly identify spatio-temporal dependence for reasonably sized data.

Considering a data matrix of size $S \times T$, let individual data elements be represented as (s_i, t_j) for all i, j. The testing procedure is as follows:

1. Permute $S \times T$ based on H_0 and refill the matrix with the permuted elements.
    - H_0 : Complete independence - (s_i, t_j) replaced by $(\pi(s_i, t_j))$.
    - H_0 : Spatially Independent Time Series - (s_i, t_j) replaced by $(\pi(s_i), t_j)$.
    - H_0 : Spatially Dependent; Temporally Independent Series - (s_i, t_j) replaced by $(s_i, \pi(t_j))$.

2. Calculate and save the STrIde statistic.

3. Repeat steps (1) and (2) B_P times.

4. Calculate the area to left of the 0.025 percentile and right of the 0.975 percentile of the permutation distribution to use as the 2-sided rejection region for H_0.

4.2.5 Type I Error Assessment through Simulations

To assess the STrIde statistic capabilities, we first examined Type I error testing under the 3 different null hypotheses. For a test of size $\alpha = 0.05$, a standard error on the Type I error for 100 tests is $\sqrt{\frac{0.05 \cdot 0.95}{100}} \approx 0.02$. Thus, in each testing scenario, we bolded values with at least a 0.04 positive difference from the nominal rate to highlight larger anti-conservative empirical deviations. We only utilize the distance scaled weight, w_1, in this setting as it is our primary goal to understand how this weight works in the testing.

4.2.5.0.1 H_0 : **Complete Independence** Independent standard normal variates were generated and arranged in an $S \times T$ matrix. In total, 36 settings were simulated:

1. 4×4, 7×7 grids of locations;

2. 30, 50, 100 time points;

3. 1+, 3+ neighbors;

4. No missing, 20% missing at random, 50% missing at random.

Table 4.3 indicates that as spatial grid size and temporal observations increase, empirical Type I error rates for the STrIde statistic are similar to the nominal

setting. Even small and missing data scenarios perform quite well in these tests of independence. The only meaningful difference from the nominal level occurred in the 50% missing data scenario. Both spatial neighbor settings yielded similar results.

Table 4.3: Simulated Type I Error of STrIde for Data Generated Under H_0: Complete Independence

		Grid Sizes					
		No Missing		**20% MAR**		**50% MAR**	
Times	**Ngh.**	4×4	7×7	4×4	7×7	4×4	7×7
30	1+	**0.09**	0.05	0.03	0.02	0.03	0.06
	3+	0.03	0.06	0.05	0.06	0.06	0.04
50	1+	0.05	0.06	0.08	0.07	0.04	0.02
	3+	0.06	0.04	0.06	0.04	0.04	0.04
100	1+	0.06	0.04	0.06	0.05	**0.10**	0.05
	3+	0.04	0.04	0.04	0.07	0.06	0.05

4.2.5.0.2 H_0 **: Spatially Independent Time Series** We generated spatially independent time series using AR1 variates with time parameter specified by the simulation specifics. The time series were generated using `arima.sim{stats}` in R. The standard deviation of innovations was set to 0.1. The time series were then set as the rows of an $S \times T$ matrix. In total, we simulated under 48 settings:

1. 4×4, 7×7 grids of locations;
2. 30, 100 time points;
3. 1+ and 3+ neighbors;
4. -0.90, -0.50, -0.10, 0.10, 0.50, 0.90 time dependence.

Table 4.4 again shows a propensity to approach the nominal error rate as spatial grid size and temporal observations increase. Low dependence settings of ± 0.10

had slightly higher Type I errors, however, they were within the 0.04 tolerance from 0.05 except in one scenario.

Table 4.4: Simulated Type I Error of STrIde for Data Generated Under H_0: Spatially Independent Time Series

			Time Dependence					
Grid Size	**Times**	**Ngh.**	-0.90	-0.50	-0.10	0.10	0.50	0.90
4×4	30	1+	0.06	0.05	0.07	0.06	0.06	0.03
		3+	0.06	0.01	0.04	0.03	0.04	0.03
	100	1+	0.08	0.03	**0.09**	0.05	0.03	0.03
		3+	0.05	0.05	0.07	0.08	0.04	0.04
7×7	30	1+	0.04	0.05	0.03	0.08	0.04	0.03
		3+	0.04	0.03	0.05	0.04	0.07	0.05
	100	1+	0.04	0.02	0.01	0.04	0.08	0.06
		3+	0.02	0.03	0.03	0.06	0.05	0.02

4.2.5.0.3 H_0 : Spatially Dependent; Temporally Independent Series

After creating a covariance matrix using (4.23), we simulated spatially dependent data from a multivariate normal with this covariance. Additionally, we generated an additional independent normal random variable with standard deviation 0.1 to add to each row (station). The data were arranged in an $S \times T$ matrix. In total, we simulated under 40 settings:

1. 4×4, 7×7 grids of locations;
2. 30, 100 time points;
3. 1+, 3+ neighbors;
4. 0.10, 0.30, 0.50, 0.70, 0.90 space dependence.

Table 4.5 indicates empirically reasonable Type I errors for all settings and only one tolerance exceedance. Lower dependence settings were still appropriately detected based on the empirical error rates.

Table 4.5: Simulated Type I Error of STrIde for Data Generated Under H_0: Spatially Dependent; Temporally Independent Series

			Space Dependence				
Grid Size	**Times**	**Ngh.**	0.10	0.30	0.50	0.70	0.90
4×4	30	1+	0.05	0.02	0.06	0.05	0.03
		3+	0.04	0.06	0.06	0.02	0.03
	100	1+	0.04	0.08	0.07	0.04	0.03
		3+	0.04	0.08	0.08	0.06	0.04
7×7	30	1+	0.06	0.04	0.04	0.00	0.03
		3+	0.01	0.06	0.03	0.02	0.07
	100	1+	**0.09**	0.05	0.07	0.08	0.04
		3+	0.06	0.08	0.05	0.02	0.04

4.2.6 Power Assessment through Simulations

Within each setting and under each null hypothesis, we generated 100 spatio-temporal series via the process described in Section 4.2.1. Empirical power is calculated using rejection at the $\alpha = 0.05$ level.

4.2.6.0.4 Power vs. H_0 : Complete Independence Spatio-temporal data were generated under 240 total settings over all combinations of:

1. 4×4, 7×7 grids of locations;
2. 30, 100 time points;
3. 1+, 3+ neighbors;
4. -0.90, -0.50, -0.10, 0.10, 0.50, 0.90 time parameters;
5. 0.10, 0.30, 0.50, 0.70, 0.90 space parameters.

STrIde was implemented with both weights, w_1 in (4.12) and w_2 in (4.13). Table 4.6-Table 4.9 show bolded power for the STI or STrIde method with at least

a 5% advantage over the other method. When neither method has an empirical advantage, but one of the STrIde weights is at least 5% lower, it is italicized.

The STrIde statistic displayed a noticeable power advantage when space or time parameters were weaker especially in the smallest data scenario. STrIde was also much more powerful at the smaller magnitude time parameters in the 3+ neighbor case.

The only advantage for STI appeared in the -0.50 time with 0.50 or 0.70 space in the smallest setting where there may have been a cancellation issue for the STrIde statistic. However, in climate scenarios for the use of this statistic, a negative time correlation would be a fairly unlikely scenario in a smaller spatial region, thus, we are confident in proceeding with it. Generally speaking, both statistics indicate power increases with sample size.

The STrIde weights show some differences, but are complementary to each other and, in combination, seem to provide a meaningful advantage over STI. On occasion, in the -0.50 time dependence setting, w_1 was much weaker than w_2. w_2 was weaker in the high, 0.50 and 0.90, time dependence settings crossed with the lowest space dependence.

Table 4.6: Empirical Power Comparison for Spatio-temporal Series at 16 (4×4) locations, 30 times/location vs. H_0: Complete Independence

		Space								
		0.10			0.30			0.50		
Ngh.	**Time**	STI	STrIde		STI	STrIde		STI	STrIde	
			w_1	w_2		w_1	w_2		w_1	w_2
1+	-0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	-0.50	0.95	**1.00**	**1.00**	0.99	0.98	**1.00**	1.00	*0.86*	0.98
	-0.10	0.01	0.09	**1.00**	0.12	0.14	**0.54**	0.36	0.19	**0.61**
	0.10	0.04	0.30	**0.96**	0.17	0.53	**0.55**	0.37	0.70	**0.88**
	0.50	0.96	1.00	*0.71*	0.98	**1.00**	0.85	1.00	1.00	1.00
	0.90	0.98	1.00	*0.82*	1.00	1.00	0.89	1.00	1.00	1.00
3+	-0.90	0.84	**1.00**	**1.00**	1.00	1.00	1.00	1.00	1.00	1.00
	-0.50	0.60	0.95	**1.00**	0.94	*0.67*	0.97	**1.00**	0.40	0.85
	-0.10	0.00	0.05	**0.94**	0.10	0.24	**0.70**	0.22	0.42	**0.91**
	0.10	0.00	0.34	**0.82**	0.05	0.73	**0.75**	0.10	0.82	**0.98**
	0.50	0.48	**0.96**	0.77	0.92	**1.00**	0.95	0.95	0.99	**1.00**
	0.90	0.84	**1.00**	0.85	0.97	1.00	*0.87*	1.00	1.00	1.00

		Space					
		0.70			0.90		
Ngh.	**Time**	STI	STrIde		STI	STrIde	
			w_1	w_2		w_1	w_2
1+	-0.90	1.00	1.00	1.00	1.00	1.00	1.00
	-0.50	**1.00**	0.71	0.92	**1.00**	0.66	0.90
	-0.10	0.40	0.30	**0.91**	0.66	0.37	**0.91**
	0.10	0.47	0.61	**1.00**	0.53	0.66	**1.00**
	0.50	1.00	0.99	1.00	0.99	0.99	1.00
	0.90	1.00	1.00	1.00	1.00	0.98	1.00
3+	-0.90	1.00	1.00	1.00	1.00	0.96	1.00
	-0.50	**0.98**	0.13	0.86	**1.00**	0.31	0.90
	-0.10	0.44	0.66	**0.94**	0.52	0.59	**0.99**
	0.10	0.31	0.90	**1.00**	0.51	0.82	**0.99**
	0.50	0.98	1.00	1.00	0.98	0.98	1.00
	0.90	1.00	1.00	1.00	1.00	1.00	1.00

Table 4.7: Empirical Power Comparison for Spatio-temporal Series at 16 (4×4) locations, 100 times/location vs. H_0: Complete Independence

Ngh.	**Time**	**Space** 0.10 STI	0.10 STrIde w_1	0.10 STrIde w_2	0.30 STI	0.30 STrIde w_1	0.30 STrIde w_2	0.50 STI	0.50 STrIde w_1	0.50 STrIde w_2
1+	-0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	-0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.98	1.00
	-0.10	0.13	0.19	**1.00**	0.45	0.25	**0.84**	**0.66**	0.56	0.60
	0.10	0.05	0.74	**1.00**	0.43	**0.95**	0.62	0.62	**1.00**	**1.00**
	0.50	1.00	1.00	*0.91*	1.00	1.00	1.00	1.00	1.00	1.00
	0.90	1.00	1.00	*0.84*	1.00	1.00	1.00	1.00	1.00	1.00
3+	-0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	-0.50	1.00	1.00	1.00	1.00	0.99	1.00	1.00	*0.55*	1.00
	-0.10	0.00	0.07	**1.00**	0.21	0.57	**0.69**	0.41	0.88	**1.00**
	0.10	0.01	0.67	**0.99**	0.13	**1.00**	0.91	0.46	**1.00**	**1.00**
	0.50	1.00	1.00	*0.71*	1.00	1.00	1.00	1.00	1.00	1.00
	0.90	1.00	1.00	*0.88*	1.00	1.00	1.00	1.00	1.00	1.00

Ngh.	**Time**	**Space** 0.70 STI	0.70 STrIde w_1	0.70 STrIde w_2	0.90 STI	0.90 STrIde w_1	0.90 STrIde w_2
1+	-0.90	1.00	1.00	1.00	1.00	1.00	1.00
	-0.50	1.00	*0.89*	0.99	1.00	*0.70*	0.99
	-0.10	0.73	0.69	**0.94**	0.75	0.75	**0.99**
	0.10	0.66	**1.00**	**1.00**	0.77	0.98	**1.00**
	0.50	1.00	1.00	1.00	1.00	1.00	1.00
	0.90	1.00	1.00	1.00	1.00	1.00	1.00
3+	-0.90	1.00	1.00	1.00	1.00	1.00	1.00
	-0.50	1.00	0.30	**0.97**	**1.00**	0.34	0.94
	-0.10	0.62	0.98	**1.00**	0.74	0.99	**1.00**
	0.10	0.51	**1.00**	**1.00**	0.62	**1.00**	**1.00**
	0.50	1.00	1.00	1.00	1.00	1.00	1.00
	0.90	1.00	1.00	1.00	1.00	1.00	1.00

Table 4.8: Empirical Power Comparison for Spatio-temporal Series at 49 (7×7) locations, 30 times/location vs. H_0: Complete Independence

Ngh.	**Time**	**Space** 0.10 STI	0.10 STrIde w_1	0.10 STrIde w_2	0.30 STI	0.30 STrIde w_1	0.30 STrIde w_2	0.50 STI	0.50 STrIde w_1	0.50 STrIde w_2
1+	-0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	-0.50	1.00	1.00	1.00	1.00	1.00	0.97	1.00	*0.90*	0.98
	-0.10	0.03	0.16	**1.00**	0.29	0.30	**0.97**	0.66	0.67	**1.00**
	0.10	0.03	0.83	**1.00**	0.25	0.93	**1.00**	0.51	0.99	**1.00**
	0.50	**1.00**	0.90	0.82	1.00	1.00	1.00	1.00	1.00	1.00
	0.90	1.00	1.00	*0.87*	1.00	1.00	1.00	1.00	1.00	1.00
3+	-0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	-0.50	0.96	*0.79*	1.00	1.00	1.00	1.00	1.00	*0.26*	0.95
	-0.10	0.00	0.66	**0.99**	0.06	0.04	**0.69**	0.37	0.97	**1.00**
	0.10	0.00	**1.00**	0.87	0.06	0.87	**1.00**	0.32	**1.00**	**1.00**
	0.50	0.97	1.00	*0.78*	1.00	1.00	1.00	1.00	1.00	1.00
	0.90	1.00	1.00	*0.92*	1.00	1.00	1.00	1.00	1.00	1.00

Ngh.	**Time**	**Space** 0.70 STI	0.70 STrIde w_1	0.70 STrIde w_2	0.90 STI	0.90 STrIde w_1	0.90 STrIde w_2
1+	-0.90	1.00	1.00	1.00	1.00	1.00	1.00
	-0.50	1.00	*0.79*	1.00	1.00	*0.82*	1.00
	-0.10	0.62	0.71	**0.85**	**0.80**	0.66	0.58
	0.10	0.60	0.98	**1.00**	0.63	**0.83**	0.76
	0.50	1.00	1.00	1.00	0.99	1.00	1.00
	0.90	1.00	1.00	1.00	1.00	1.00	0.99
3+	-0.90	1.00	1.00	1.00	1.00	1.00	1.00
	-0.50	1.00	*0.20*	0.95	**1.00**	0.47	0.94
	-0.10	0.60	**0.99**	**0.99**	0.73	0.95	**1.00**
	0.10	0.53	**1.00**	**1.00**	0.66	0.98	**1.00**
	0.50	1.00	1.00	1.00	0.96	1.00	1.00
	0.90	1.00	1.00	1.00	1.00	1.00	1.00

Table 4.9: Empirical Power Comparison for Spatio-temporal Series at 49 (7×7) locations, 100 times/location vs. H_0: Complete Independence

		Space								
		0.10			0.30			0.50		
Ngh.	**Time**	STI	STrIde		STI	STrIde		STI	STrIde	
			w_1	w_2		w_1	w_2		w_1	w_2
1+	-0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	-0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	-0.10	0.43	0.19	**1.00**	0.81	*0.57*	0.81	0.95	0.97	0.99
	0.10	0.48	**1.00**	**1.00**	0.77	**1.00**	0.96	0.88	**1.00**	**1.00**
	0.50	1.00	1.00	*0.94*	1.00	1.00	1.00	1.00	1.00	1.00
	0.90	1.00	1.00	*0.86*	1.00	1.00	1.00	1.00	1.00	1.00
3+	-0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	-0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99
	-0.10	0.00	0.04	**1.00**	0.50	**0.99**	0.91	0.80	0.46	**1.00**
	0.10	0.01	**1.00**	**1.00**	0.48	**1.00**	**1.00**	0.75	**1.00**	**1.00**
	0.50	1.00	1.00	*0.79*	1.00	1.00	1.00	1.00	1.00	1.00
	0.90	1.00	1.00	*0.89*	1.00	1.00	1.00	1.00	1.00	1.00

		Space					
		0.70			0.90		
Ngh.	**Time**	STI	STrIde		STI	STrIde	
			w_1	w_2		w_1	w_2
1+	-0.90	1.00	1.00	1.00	1.00	1.00	1.00
	-0.50	1.00	0.96	1.00	1.00	*0.81*	0.97
	-0.10	0.89	0.99	**1.00**	0.89	0.98	**1.00**
	0.10	0.90	**1.00**	**1.00**	0.83	1.00	1.00
	0.50	1.00	1.00	1.00	1.00	1.00	1.00
	0.90	1.00	1.00	1.00	1.00	1.00	1.00
3+	-0.90	1.00	1.00	1.00	1.00	1.00	1.00
	-0.50	1.00	1.00	1.00	1.00	1.00	0.95
	-0.10	0.87	0.35	**1.00**	0.83	0.74	**1.00**
	0.10	0.81	**1.00**	**1.00**	0.84	**1.00**	**1.00**
	0.50	1.00	1.00	1.00	1.00	1.00	1.00
	0.90	1.00	1.00	1.00	1.00	1.00	1.00

4.2.6.0.5 Power vs. Spatially Independent Time Series In this case, we'd expect H_0 to be rejected as the spatial parameter increases and the absolute value of the time parameter decreases. Since no comparison is available to STI in this test, we investigated the difference between the two STrIde weights, w_1 in (4.12) and w_2 in (4.13). We bolded power difference of 5% or more to highlight when a substantial empirical disparity was noted.

Based on the results in 4×4 spatial grid in Table 4.10 and Table 4.11, power increased for both weights when the number of times increased. Similarly, power tended to increase as spatial dependence grew due to the correct rejection of spatial independence in the null hypothesis. However, in the 7×7 spatial grid used for Table 4.12 and Table 4.13, power increased across all settings, and is mainly only lacking in the lowest spatial setting.

Differences in weights are quite noticeable in the tables. The binary weight, w_2, does much better in the very low spatial dependence scenarios although the difference decreases as the number of locations and time points increase. The more spatially and temporally adjusted weight, w_1, seems to produce higher power in the more intermediate spatial dependencies of 0.30 and 0.50. w_1 appears to be more in-line with the strength of the parameters than the binary w_2. w_1 may also more susceptible to low power with a small sample size and a weak form of the null hypothesis being true, however, it shows better alignment with the actual structure of the data in the case of medium to strong spatial dependence.

Table 4.10: Empirical Power for Spatio-temporal Series at 16 (4×4) locations, 30 times/location vs. H_0 : Spatially Independent Time Series. Differences of at least 5% are bolded.

Ngh.	**Time**	**Space**									
		0.10		0.30		0.50		0.70		0.90	
		w_1	w_2	w_1	w_2	w_1	w_2	w_1	w_2	w_1	w_2
1+	-0.90	0.22	**0.95**	**0.80**	0.69	**0.96**	0.63	**0.99**	0.91	0.99	0.96
	-0.50	0.17	**0.44**	**0.82**	0.62	**0.99**	0.63	**1.00**	0.84	0.96	0.92
	-0.10	0.18	**1.00**	0.73	0.69	**0.95**	0.79	0.98	0.96	0.92	**1.00**
	0.10	0.18	**1.00**	0.65	**0.75**	**0.94**	0.78	0.99	1.00	0.96	1.00
	0.50	0.11	**1.00**	0.52	**0.71**	0.83	0.80	0.95	1.00	0.87	**1.00**
	0.90	0.08	**0.94**	0.29	**0.66**	**0.66**	0.57	0.72	**0.90**	0.99	0.99
3+	-0.90	0.17	**0.88**	0.78	0.79	**0.93**	0.79	0.99	0.96	0.87	**0.98**
	-0.50	0.22	**0.56**	**0.72**	0.76	**0.98**	0.80	**0.99**	0.94	0.89	**0.96**
	-0.10	0.12	**0.96**	0.68	0.66	0.91	0.96	0.96	1.00	0.83	**1.00**
	0.10	0.08	**0.98**	0.61	**0.70**	**0.90**	0.95	0.87	**1.00**	0.78	**1.00**
	0.50	0.14	**0.97**	0.50	**0.68**	0.83	**0.97**	0.82	**1.00**	0.72	**1.00**
	0.90	0.10	**0.88**	0.31	**0.74**	0.57	**0.88**	0.65	**1.00**	0.46	**1.00**

Table 4.11: Empirical Power for Spatio-temporal Series at 16 (4×4) locations, 100 times/location vs. H_0 : Spatially Independent Time Series. Differences of at least 5% are bolded.

Ngh.	**Time**	**Space**									
		0.10		0.30		0.50		0.70		0.90	
		w_1	w_2	w_1	w_2	w_1	w_2	w_1	w_2	w_1	w_2
1+	-0.90	0.51	**1.00**	**1.00**	0.70	**1.00**	0.69	1.00	0.98	1.00	1.00
	-0.50	0.41	**0.53**	**1.00**	0.58	**1.00**	0.60	**1.00**	0.80	**1.00**	0.93
	-0.10	0.36	**1.00**	**1.00**	0.82	**1.00**	0.90	1.00	1.00	1.00	1.00
	0.10	0.28	**1.00**	**0.98**	0.88	**1.00**	0.92	1.00	1.00	1.00	1.00
	0.50	0.20	**1.00**	0.92	0.87	**1.00**	0.89	1.00	1.00	1.00	1.00
	0.90	0.11	**1.00**	**0.76**	0.60	**0.92**	0.72	0.98	1.00	0.95	1.00
3+	-0.90	0.44	**0.96**	**0.98**	0.72	**1.00**	0.93	1.00	1.00	1.00	1.00
	-0.50	0.30	**0.59**	**1.00**	0.70	**1.00**	0.79	1.00	0.95	0.99	0.97
	-0.10	0.39	**1.00**	**0.98**	0.79	1.00	1.00	1.00	1.00	1.00	1.00
	0.10	0.24	**1.00**	**0.97**	0.77	1.00	1.00	1.00	1.00	1.00	1.00
	0.50	0.22	**1.00**	0.86	0.82	1.00	1.00	0.99	1.00	0.94	**1.00**
	0.90	0.18	**0.97**	0.58	**0.78**	0.92	0.97	0.89	**1.00**	0.77	**1.00**

Table 4.12: Empirical Power for Spatio-temporal Series at 49 (7×7) locations, 30 times/location vs. H_0 : Spatially Independent Time Series. Differences of at least 5% are bolded.

Ngh.	**Time**	**Space**									
		0.10		0.30		0.50		0.70		0.90	
		w_1	w_2	w_1	w_2	w_1	w_2	w_1	w_2	w_1	w_2
1+	-0.90	0.62	**0.95**	**1.00**	0.57	1.00	0.98	1.00	0.99	1.00	1.00
	-0.50	**0.62**	0.52	**1.00**	0.50	**1.00**	0.70	**1.00**	0.88	**1.00**	0.92
	-0.10	0.50	**1.00**	**1.00**	0.59	1.00	1.00	1.00	1.00	1.00	1.00
	0.10	0.43	**1.00**	**1.00**	0.57	1.00	1.00	1.00	1.00	1.00	1.00
	0.50	0.42	**1.00**	**1.00**	0.56	1.00	1.00	1.00	1.00	1.00	1.00
	0.90	0.33	**1.00**	**0.79**	0.59	0.99	0.94	1.00	1.00	0.99	1.00
3+	-0.90	0.61	**0.85**	**1.00**	0.79	1.00	0.99	1.00	1.00	1.00	1.00
	-0.50	0.54	0.59	**1.00**	0.56	**1.00**	0.87	1.00	0.95	1.00	0.96
	-0.10	0.50	**0.98**	**1.00**	0.93	1.00	1.00	1.00	1.00	1.00	1.00
	0.10	0.49	**1.00**	0.99	0.95	1.00	1.00	1.00	1.00	0.99	1.00
	0.50	0.22	**1.00**	0.99	0.98	1.00	1.00	1.00	1.00	1.00	1.00
	0.90	0.24	**0.92**	0.83	0.84	0.98	0.98	1.00	1.00	0.97	1.00

Table 4.13: Empirical Power for Spatio-temporal Series at 49 (7×7) locations, 100 times/location vs. H_0 : Spatially Independent Time Series. Differences of at least 5% are bolded.

Ngh.	**Time**	**Space**									
		0.10		0.30		0.50		0.70		0.90	
		w_1	w_2	w_1	w_2	w_1	w_2	w_1	w_2	w_1	w_2
1+	-0.90	0.98	1.00	1.00	1.00	1.00	1.00	**1.00**	0.63	1.00	1.00
	-0.50	0.95	0.91	**1.00**	0.88	**1.00**	0.70	**1.00**	0.49	**1.00**	0.55
	-0.10	0.97	1.00	1.00	1.00	1.00	1.00	**1.00**	0.51	1.00	1.00
	0.10	0.90	**1.00**	1.00	1.00	1.00	1.00	**1.00**	0.57	1.00	1.00
	0.50	0.79	**1.00**	1.00	1.00	1.00	1.00	**1.00**	0.61	1.00	1.00
	0.90	0.53	**1.00**	1.00	1.00	1.00	1.00	**1.00**	0.58	1.00	1.00
3+	-0.90	1.00	0.99	**1.00**	0.92	1.00	1.00	1.00	1.00	1.00	1.00
	-0.50	**1.00**	0.61	**1.00**	0.72	**1.00**	0.84	**1.00**	0.94	1.00	0.98
	-0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.90	1.00	0.99	**0.98**	0.90	1.00	1.00	1.00	1.00	1.00	1.00

4.2.6.0.6 Power vs. Spatially Dependent; Temporally Independent In this case, we'd expect H_0 to be rejected as the absolute value of the time parameter increases and the spatial parameter decreases. Again, no comparison is available to STI, thus, the two STrIde weights, (4.12) and (4.13), are investigated. Bolded values indicate a power difference of 5% or more.

The 4×4 spatial grid results in Table 4.14 and Table 4.15 showed that power generally as spatial dependence grew. Lower power in the high spatial settings was expected because there is not enough evidence to reject the spatial dependence of the null hypothesis. Similarly, lower power is seen when time dependence is low and there is not enough evidence to reject that part of the null. Larger networks seemed to enhance this effect as well, showing lower power in the 3+ neighbor setting than its corresponding 1+ neighbor setting. In the 7×7 spatial grid results in Table 4.16 and Table 4.17, power increased across all settings although it remains low in the scenarios very close to the null hypothesis.

The differences between the weights are fairly noticeable in the smallest scenario in which w_2 is markedly better than w_1 in the small temporal dependence settings and in the highest spatial dependence settings. However, w_1 is consistently better at the 0.10 spatial dependence setting with high temporal dependence. w_1 is less able to distinguish settings very close to the null hypothesis, especially those with small time dependence.

Table 4.14: Empirical Power for Spatio-temporal Series at 16 (4×4) locations, 30 time points vs. H_0 : Spatially Dependent; Temporally Independent Series. Differences of at least 5% are bolded.

		Space									
		0.10		0.30		0.50		0.70		0.90	
Ngh.	**Time**	W_1	w_2	w_1	w_2	w_1	w_2	w_1	w_2	w_1	w_2
1+	-0.90	1.00	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	-0.50	1.00	1.00	0.97	1.00	0.94	**1.00**	0.85	**0.98**	0.48	**0.79**
	-0.10	0.07	**1.00**	0.14	**0.84**	0.13	**0.33**	0.09	0.13	0.10	**0.20**
	0.10	0.17	**0.98**	0.06	**0.42**	0.10	**0.23**	0.07	**0.41**	0.05	**0.52**
	0.50	**0.99**	0.88	0.89	0.55	0.81	**0.96**	0.66	**1.00**	0.35	**0.97**
	0.90	**1.00**	0.92	**1.00**	0.72	**0.99**	0.91	0.96	0.99	0.91	**1.00**
3+	-0.90	1.00	0.97	1.00	0.99	1.00	1.00	1.00	1.00	0.98	1.00
	-0.50	0.98	1.00	0.90	**1.00**	0.83	**0.94**	0.76	**0.93**	0.42	**0.67**
	-0.10	0.15	**0.97**	0.10	**0.48**	0.07	**0.29**	0.10	**0.42**	0.05	**0.48**
	0.10	0.11	**0.92**	0.04	**0.39**	0.08	**0.67**	0.07	**0.81**	0.06	**0.78**
	0.50	**0.98**	0.76	**1.00**	0.79	0.73	**1.00**	0.50	**1.00**	0.31	**1.00**
	0.90	**1.00**	0.90	**1.00**	0.88	0.99	0.97	0.95	**1.00**	0.90	**1.00**

Table 4.15: Empirical Power for Spatio-temporal Series at 16 (4×4) locations, 100 time points vs. H_0 : Spatially Dependent; Temporally Independent Series. Differences of at least 5% are bolded.

		Space									
		0.10		0.30		0.50		0.70		0.90	
Ngh.	**Time**	w_1	w_2	w_1	w_2	w_1	w_2	w_1	w_2	w_1	w_2
1+	-0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	-0.50	1.00	1.00	1.00	1.00	0.99	1.00	0.98	1.00	0.92	**1.00**
	-0.10	0.36	**1.00**	0.28	**1.00**	0.19	**0.59**	0.12	**0.17**	0.06	**0.25**
	0.10	0.41	**1.00**	0.23	**0.64**	0.16	**0.44**	0.10	**0.82**	0.09	**0.93**
	0.50	1.00	1.00	1.00	0.78	**1.00**	1.00	1.00	1.00	0.92	1.00
	0.90	**1.00**	0.88	**1.00**	0.92	1.00	1.00	1.00	1.00	1.00	1.00
3+	-0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	-0.50	1.00	1.00	0.99	1.00	0.99	1.00	0.96	1.00	0.88	**0.97**
	-0.10	0.25	**1.00**	0.19	**0.70**	0.14	**0.36**	0.11	**0.57**	0.06	**0.76**
	0.10	0.22	**1.00**	0.16	**0.45**	0.15	**0.96**	0.10	**1.00**	0.09	**1.00**
	0.50	**1.00**	0.88	1.00	0.99	1.00	1.00	0.98	1.00	0.91	**1.00**
	0.90	**1.00**	0.82	1.00	0.97	1.00	1.00	1.00	1.00	1.00	1.00

Table 4.16: Empirical Power for Spatio-temporal Series at 49 (7×7) locations, 30 time points vs. H_0 : Spatially Dependent; Temporally Independent Series. Differences of at least 5% are bolded.

		Space									
		0.10		0.30		0.50		0.70		0.90	
Ngh.	**Time**	w_1	w_2	w_1	w_2	w_1	w_2	w_1	w_2	w_1	w_2
1+	-0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	-0.50	1.00	1.00	1.00	1.00	1.00	1.00	0.97	1.00	0.66	**0.84**
	-0.10	0.32	**1.00**	0.20	**1.00**	0.15	**0.33**	0.19	0.17	0.07	**0.29**
	0.10	0.38	**1.00**	0.13	**0.41**	0.08	**0.65**	0.11	**0.84**	0.05	**0.74**
	0.50	1.00	0.97	**1.00**	0.91	1.00	1.00	0.92	**1.00**	0.48	**1.00**
	0.90	1.00	0.97	**1.00**	0.58	1.00	0.99	1.00	1.00	0.91	**1.00**
3+	-0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	-0.50	1.00	1.00	1.00	1.00	0.99	1.00	0.90	**0.97**	0.60	**0.74**
	-0.10	0.26	**1.00**	0.15	**0.40**	0.15	**0.58**	0.11	**0.72**	0.10	**0.67**
	0.10	0.18	**0.98**	0.07	**0.61**	0.13	**0.95**	0.04	**0.99**	0.07	**0.93**
	0.50	**1.00**	0.77	1.00	0.99	0.99	1.00	0.84	**1.00**	0.46	**1.00**
	0.90	**1.00**	0.90	**1.00**	0.95	1.00	1.00	1.00	1.00	0.91	1.00

Table 4.17: Empirical Power for Spatio-temporal Series at 49 (7×7) locations, 100 time points vs. H_0 : Spatially Dependent; Temporally Independent Series. Differences of at least 5% are bolded.

		Space									
		0.10		0.30		0.50		0.70		0.90	
Ngh.	**Time**	w_1	w_2	w_1	w_2	w_1	w_2	w_1	w_2	w_1	w_2
1+	-0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	-0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.97	1.00
	-0.10	0.80	**1.00**	0.61	**1.00**	0.40	**0.65**	0.24	0.24	0.07	**0.34**
	0.10	0.73	**1.00**	0.51	**0.60**	0.35	**0.93**	0.20	**1.00**	0.12	**0.98**
	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.94	**1.00**
	0.90	**1.00**	0.90	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00
3+	-0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	-0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.95	0.99
	-0.10	0.58	**1.00**	0.41	**0.78**	0.35	**0.71**	0.17	**0.90**	0.16	**0.94**
	0.10	0.59	**1.00**	0.39	**0.90**	0.24	**1.00**	0.15	1.00	0.07	**1.00**
	0.50	**1.00**	0.90	1.00	1.00	1.00	1.00	1.00	1.00	0.95	1.00
	0.90	**1.00**	0.86	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

4.3 Residual Testing for Indian Monsoon Precipitation Modeling

4.3.1 Logit-Normal Mixed Model for Indian Monsoon Precipitation

In Chapter 3, logit-normal mixed models were fit to Indian monsoon precipitation to discover relationships among covariates and the probability of exceeding precipitation thresholds. As previously stated, this model explicitly accounts for time dependence but only implicitly for spatial dependence, thus, we want to assess the need for further structure. The model uses daily data for summer monsoon seasons (1 June to 30 September) from 1973-2013.

Let station $i \in \{1, ..., m\}$, day $d \in \{1, ..., n_i\}$, and year $k \in \{1973, ..., 2013\}$. Given a precipitation (mm·day^{-1}) threshold $\tau \in \{50, 75, 100, 125\}$ and daily precipitation event Z_{idk}, let $T_{idk} = I(Z_{idk} > \tau)$. Let $\mathbf{x}_{ijk}$ be a vector of covariates and $\mathbf{U}$ and $\mathbf{W}$ be vectors of random intercepts for station and year, respectively. Recall the logit-normal mixed model specified in (3.17)-(3.20):

$$\text{Level 1: } T_{idk}|\mathbf{U} = \mathbf{u}, \mathbf{W} = \mathbf{w} \overset{\text{ind.}}{\sim} \text{Bernoulli}(\theta_{ijk}), \tag{4.29}$$

$$\text{logit}(\theta_{idk}) = \mathbf{x}_{idk}^{\mathrm{T}}\boldsymbol{\beta} + u_i + w_k, \tag{4.30}$$

$$\text{Level 2: } U_i \overset{\text{ind.}}{\sim} \mathcal{N}(0, \sigma^2_{station}), W_k \overset{\text{ind.}}{\sim} \mathcal{N}(0, \sigma^2_{year}) \tag{4.31}$$

$$U_i \text{ independent of } W_k \text{ for all } (i, k). \tag{4.32}$$

Covariates used in the model included a mix of climate variables - ΔTT, u-winds at 200 and 850 millibars (mb) levels, and v-winds at 200 and 850 mb levels, and Niño-3.4 anomaly- and station specific attributes- minimum temperature, maxi-

mum temperature, elevation, latitude, longitude. Data details can be found in Section 3.4.

We collected Pearson-like residuals (r_{idk}), defined in (4.33), during model estimation within SAS (SAS Institute Inc., 2011). Because of the model structure, linearized pseudo-data points ($\tilde{p}_{idk}$) are computed for comparison to logit($\hat{\theta}_{idk}$) for each observation during model fitting. See Appendix C.2 for the construction of $\tilde{\mathbf{p}}$. Empirical best linear unbiased predictors (eBLUPs) are used as estimates for random effects $\mathbf{U}$ and $\mathbf{W}$. Thus, the estimated residuals are:

$$\hat{r}_{idk} = \frac{\tilde{p}_{idk} - (\mathbf{x}_{idk}^{\mathrm{T}}\hat{\boldsymbol{\beta}} + \widehat{u_i} + \widehat{w_k})}{\mathrm{Var}(\tilde{p}_{idk}|\mathbf{U} = \hat{\mathbf{u}}, \mathbf{W} = \hat{\mathbf{w}})}. \tag{4.33}$$

Two distance settings- 500KM and 1000KM- are employed resulting in different neighborhoods seen in Figure 4.2. Great circle distances are utilized since the data represents coordinates on the Earth. Finally, note that STI was tested using the standard normal deviate suggested in (Griffith, 1981), while STrIde was tested based on a random sample of 1000 permutations.

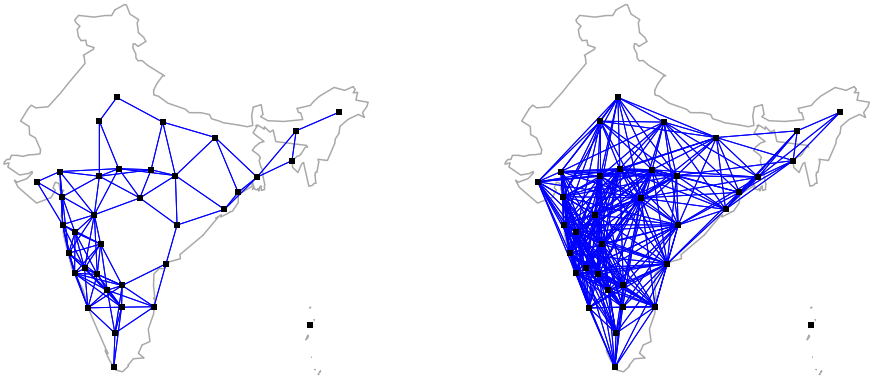
4.3.2 Results of Spatio-temporal Hypothesis Testing

Intuitively, we'd expect most extreme rainfall events to be a function of idiosyncratic station conditions. Thus, we'd conjecture at higher threshold event models we would not find evidence of residual dependence.

The test for complete independence gives distinct results based on the use of either STI or STrIde and is clearly influenced by the choices of neighbors. At the 500 KM spatial network, Table 4.18 displays higher rainfall thresholds do not show evidence of spatio-temporal dependence in model residuals for STI. However, the more powerful test, STrIde, detects dependence at all levels. After increasing

Figure 4.2: Neighbors Network for Indian Weather Stations. A line between two points indicates they are neighbors and their product has positive weight in the calculations.

(a) 500 KM Neighbors (b) 1000 KM Neighbors



the spatial network to include neighbors within 1000 KM, we find no evidence of dependence using STI, and significant dependence for STrIde is detected only in the light ($\geq$ 50 mm/day) and medium ($\geq$ 75 mm/day) rainfall thresholds. Taking into account the power of both tests, we conclude that if dependence exists at these higher levels, it is unlikely to be strong and may not require additional structure.

The significant results at lower thresholds are not surprising as rainfall in general has spatio-temporal dependence and exceedance of 50 or 75 mm/day is not unusual for all of India on a given day in the monsoon season.

Table 4.18: p-values for STI and STrIde Based on Model Residuals vs. H_0: Complete Independence. Results indicating dependence are bolded based on significance level $\alpha = 0.05$.

Rain in mm·day^{-1}	# KM in which Neighbors Exist					
	500			1000		
	STI	STrIde		STI	STrIde	
		w_1	w_2		w_1	w_2
≥ 50	**0.010**	**0.002**	**0.002**	0.250	**0.002**	**0.002**
≥ 75	0.160	**0.002**	**0.002**	0.670	**0.002**	**0.004**
≥ 100	0.580	**0.016**	**0.038**	0.900	0.084	0.156
≥ 125	0.840	**0.046**	0.092	0.970	0.104	0.242

We shift attention to the STrIde statistic versus the null hypotheses of spatially independent time series or spatially dependent, temporally independent series. The statistically significant results ifn the test against independence were reflected only at the lowest levels of rainfall threshold. The STrIde statistic did not show evidence of either spatial or temporal dependence at the 100 or 125 mm/day thresholds although p-values for the space hypotheses were generally lower than that for the time hypotheses. This indicates the remaining possible dependence in the residuals is likely to be spatial, however, it is not significant enough to be included in a further modeling effort for the more extreme thresholds.

Table 4.19: p-values for STrIde based on Model Residuals vs. H_0 : Spatially Dependent; Temporally Independent or H_0 : Spatially Independent Time Series. Results indicating dependence are bolded based on significance level $\alpha = 0.05$.

Rain in mm·day^{-1}	# KM in which Neighbors Exist							
	500				1000			
	w_1		w_2		w_1		w_2	
	Space	Time	Space	Time	Space	Time	Space	Time
≥ 50	**0.002**	**0.002**	**0.002**	**0.002**	**0.002**	**0.002**	**0.002**	**0.002**
≥ 75	**0.002**	**0.014**	**0.004**	**0.002**	**0.002**	**0.028**	**0.010**	**0.110**
≥ 100	0.058	0.412	0.076	0.290	0.172	0.899	0.194	0.851
≥ 125	0.118	0.466	0.148	0.308	0.178	0.697	0.356	0.999

4.4 Conclusions on Spatio-temporal Dependence Testing

This chapter explored detection of spatio-temporal dependence in model residuals, specifically in regard to the modeling presented in Chapter 3. We present a new statistic, STrIde, to compare to a commonly used interpretable measure, STI. The goals of our statistic were to assess a variety of dependence hypotheses through permutation tests rather than reliance on asymptotic normality.

The extensive simulation study in this chapter highlights the usefulness of the newly developed STrIde statistic and compares it to the STI. The study showed that STrIde produced reasonable empirical Type I errors and may be more powerful than STI in detecting dependence in spatio-temporal data. The choice of weights and neighborhood settings also appear to be quite important in the resulting power of the test.

In general, the results of the spatio-temporal testing indicate the Indian monsoon models provide valid inference for the probability of exceeding high precipitation thresholds. The possibly unmodeled portion of the dependence seems to be spatial in nature, but does not issue a strong enough signal to overturn our previous analysis.

However, we recognize that STrIde is currently an empirical heuristic method, and developing asymptotic theory regarding STrIde testing is relevant continuing work. Once asymptotic theory is developed, testing using asymptotic distributions can be compared to permutation testing. Another goal is to make the method more computationally feasible in computing permutations. Future simulations may also be performed to determine more specifically the effect of classes of weights on in the STrIde statistic.

References

Ajayamohan, R., Merryfield, W., and Kharin, V. (2008). Increasing Trend of Synoptic Activity and Its Relationship with Extreme Rain Events over Central India. *Journal of Climate*, 23(4):1004–1013.

Atkinson, G. D. and Holliday, C. R. (1977). Tropical Cyclone Minimum Sea Level Pressure/Maximum Sustained Wind Relationship for the Western North Pacific. *Monthy Weather Review*, 105(4):421–427.

Attri, S. D. and Tyagi, A. (2010). Climate Profile of India. Technical report, Government of India, Ministry of Earth Sciences, India Meteorological Department.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014). *Hierarchical Modeling and Analysis for Spatial Data*, volume 135 of *Monographs on Statistics and Applied Probability*. Chapman and Hall/CRC, 2nd edition.

Bates, D. (Preprint 2010). *lme4: Mixed-effects modeling with R*. Springer.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society, Series B*, 57:289–300.

Bickel, P. J. and Sakov, A. (2008). On the Choice of M in the M out of N Bootstrap and Confidence Bounds for Extrema. *Statistica Sinica*, 18:967–985.

Bivand, R. and Piras, G. (2015). Comparing Implementations of Estimation Methods for Spatial Econometrics. *Journal of Statistical Software*, 63(18):1–36.

Blake, E. S., Landsea, C. W., and Gibney, E. J. (2011). The Deadliest, Costliest, and Most Intense United States Tropical Cyclones From 1851 to 2010 (and Other Frequently Requested Hurricane Facts). Technical report, NOAA.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., and White, J.-S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology and Evolution*, 24(3):127–135.

Box, G. E. P. and Pierce, D. A. (1970). Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models. *Journal of the American Statistical Association*, 65(332):1509–1526.

Breslow, N. and Clayton, D. (1993). Approximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association*, 88(421):9–25.

Breusch, T. S. (1978). Testing for Autocorrelation in Dynamic Linear Models. *Australian Economic Papers*, 17(31):334–355.

Capéraà, P., Fougères, A. L., and Genest, C. (1997). A Non-parametric Estimation Procedure for Bivariate Extreme Value Copulas. *Biometrika*, 84(3):567–577.

Chang, C.-P., Harr, P., and Ju, J. (2001). Possible roles of atlantic circulations on the weakening indian monsoon rainfall– enso relationship. *Journal of Atmospheric and Oceanic Technology*, 14(11):2376–2380.

Cliff, A. and Ord, K. (1972). Testing for Spatial Autocorrelation Among Regression Residuals. *Geographical Analysis*, 4(3):267–284.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer, 1st edition.

Coles, S. G. and Tawn, J. A. (1991). Modelling Extreme Multivariate Events. *Journal of the Royal Statistical Society, Series B (Methodological)*, 53(2):377–392.

Coles, S. G. and Tawn, J. A. (1994). Statistical Methods for Multivariate Extremes: An Application to Structural Design. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 53(1):1–48.

Courtney, J. and Knaff, J. A. (2009). Adapting the Knaff and Zehr wind-pressure relationship for operational use in Tropical Cyclone Warning Centres. *Australian Meteorological and Oceanographic Journal*, 58:167–179.

Cressie, N. (1993). *Statistics for Spatial Data*. Wiley Series in Probability and Statistics. John Wiley and Sons, Inc., New York, New York.

Cressie, N. and Wikle, C. K. (2011). *Statistics for Spatio-Temporal Data*. Wiley Series in Probability and Statistics. John Wiley and Sons, Inc., Hoboken, New Jersey, 1st edition.

Dailey, P. S., Zuba, G., Ljung, G., Dima, I. M., and Guin, J. (2009). On the Relationship between North Atlantic Sea Surface Temperatures and U.S. Hurricane Landfall Risk. *Journal of Applied Meteorology and Climatology*, 48(1):111–129.

Dietz, L. R. and Chatterjee, S. (2014). Logit-Normal Mixed Model for Indian Monsoon Precipitation. *Nonlinear Processes in Geophysics*, 21:939–953.

Dietz, L. R. and Chatterjee, S. (2015). Investigation of Precipitation Thresholds in the Indian Monsoon Using Logit-Normal Mixed Models. In *Machine Learning*

and Data Mining Approaches to Climate Science: Proceedings of the Fourth International Workshop on Climate Informatics.

Durbin, J. and Watson, G. S. (1950a). Testing for Serial Correlation in Least Squares Regression: I. *Biometrika*, 37(3/4):409–428.

Durbin, J. and Watson, G. S. (1950b). Testing for Serial Correlation in Least Squares Regression: II. *Biometrika*, 38(1/2):159–177.

Dvorak, V. F. (1975). Tropical Cyclone Intensity Analysis and Forecasting from Satellite Imagery. *Monthly Weather Review*, 103(5):420–430.

Dvorak, V. F. (1984). Tropical Cyclone Intensity Analysis Using Satellite Data. Technical Report NESDIS 11, National Oceanic and Atmospheric Administration, Washington, D.C.

Dwass, M. (1957). Modified Randomization Tests for Nonparametric Hypotheses. *The Annals of Mathematical Statistics*, 28(1):181–187.

Edgington, E. and Onghena, P. (2007). *Randomization Tests*. Statistics: A Series of Textbooks and Monographs. Chapman & Hall/CRC, 4 edition.

Elsner, J. B. (2003). Tracking Hurricanes. *Bulletin of the American Meteorological Society*, 84(3):353–356.

Elsner, J. B., Bossak, B. H., and Niu, X. (2001). Secular Changes to the ENSO-U.S. Hurricane Relationship. *Geophysical Research Letters*, 28(21):4123–4126.

Elsner, J. B. and Jagger, T. H. (2004). A Hierarchical Bayesian Approach to Seasonal Hurricane Modeling. *Journal of Climate*, 17(14):2813–2827.

Elsner, J. B. and Jagger, T. H. (2005). Comparison of Hindcasts Anticipating the 2004 Florida Hurricane Season. *Weather and Forecasting*, 21(2):182–192.

Elsner, J. B. and Jagger, T. H. (2006). Prediction Models for Annual U.S. Hurricane Counts. *Journal of Climate*, 19(12):2935–2951.

Emmanuel, K. A. (1991). The Theory of Hurricanes. *Annual Review of Fluid Mechanics*, 23:179–196.

Emmanuel, K. A. (2005). Increasing Destructiveness of Tropical Cyclones Over the Past 30 years. *Nature*, 436:686–688.

Enfield, D. B., Mestas-Nunez, A. M., and Trimble, P. J. (2001). The Atlantic Multidecadal Oscillation and its relationship to rainfall and river flows in the continental U.S. *Geophysical Research Letters*, 28(10):2077–2080.

Ernst, M. D. (2004). Permutation Methods: A Basis for Exact Inference. *Statistical Science*, 19(4):676–685.

Faraway, J. J. (2006). *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*. Texts in Statistical Science. Taylor and Francis.

Flegal, J. M., Hughes, J., and Vats, D. (2015). *mcmcse: Monte Carlo Standard Errors for MCMC*. Riverside, CA and Minneapolis, MN. R package version 1.1-2.

Fox, J. and Weisberg, S. (2011). *An R Companion to Applied Regression*. Sage, Thousand Oaks, CA, 2nd edition.

French, J. (2014). *SpatialTools: Tools for spatial data analysis*. R package version 0.5.8.

Fuentes, M. (2006). Testing for Separability of Spatial-temporal Covariance Functions. *Journal of Statistical Planning and Inference*, 136:447–466.

Galambos, J. (1975). Order Statistics of Samples from Multivariate Distributions. *Journal of the American Statistical Association*, 70(351):674–680.

Geary, R. C. (1954). The Contiguity Ratio and Statistical Mapping. *The Incorporated Statistician*, 5(3):115–127+ 129–146.

Ghosh, S., Das, D., Kao, S.-C., and Ganguly, A. R. (2012). Lack of Uniform Trends but Increasing Spatial Variability in Observed Indian Rainfall Extremes. *Nature Climate Change*, 2(2):86–91.

Ghosh, S., Luniya, V., and Gupta, A. (2009). Trend Analysis of Indian Summer Monsoon Rainfall at Different Spatial Scales. *Atmospheric Science Letters*, 10:285–290.

Gilleland, E. and Katz, R. W. (2011). New Software to Analyze How Extremes Change Over Time. *Eos*, 92(2):13–14.

Godfrey, L. G. (1978). Testing Against General Autoregressive and Moving Average Error Models when the Regressors Include Lagged Dependent Variables. *Econometrica*, 46(6):1293–1301.

Goswami, B., Venugopal, V., Sengupta, D., Madhusoodanan, M. S., and Xavier, P. K. (2006). Increasing Trend of Extreme Rain Events Over India in a Warming Environment. *Science*, 314(5804):1442–1445.

Griffith, D. A. (1981). Interdependence in Space and Time: Numerical and interpretative considerations. In Griffith, D. A. and MacKinnon, R. D., editors, *Dynamic Spatial Models*, NATO Advanced Study Institute Series, pages 258–287, Alphen aan den Rijn. Sijthoff and Noordhoff.

Griffith, D. A. (1988). *Advanced Spatial Statistics*, volume 12 of *Advanced Studies in Theoretical and Applied Econometrics*. Springer, Netherlands.

Guarte, J. and Barrios, E. (2013). Nonparametric Hypothesis Testing in a Spatial-Temporal Model: A Simulation Study. *Communications in Statistics - Simulation and Computation*, 42(1):153–170.

Gumbel, E. J. (1960). Distributions des valeurs extremes en plusieurs dimensions. *Publications de l'Institut de statistique de l'Université de Paris*, 9:171–173.

Hall, T. and Hereid, K. (2015). The Frequency and Duration of U.S. Hurricane Droughts. *Geophysical Research Letters*, 42(9):3482–3485.

Harper, B., Kepert, J., and Ginger, J. (2010). *Guidelines For Converting Between Various Wind Averaging Period In Tropical Cyclone Conditions*. World Meteorological Organization.

Henebry, G. M. (1995). Spatial Model Error Analysis Using Autocorrelation Indices. *Ecological Modelling*, 82(1):75–91.

Hijmans, R. J. (2015). *geosphere: Spherical Trigonometry*. R package version 1.3-13.

Holland, G. J. (1980). An Analytic Model of the Wind and Pressure Profiles in Hurricanes. *Monthly Weather Review*, 108(8):1212–1218.

Holland, G. J. (2008). A Revised Hurricane Pressure–Wind Model. *Monthly Weather Review*, 136(9):3432–3445.

Holland, G. J. and Bruyère, C. L. (2014). Recent Intense Hurricane Response to Global Climate Change. *Climate Dynamics*, 42(3-4):617–627.

Holton, J. R. (1992). *An Introduction to Dynamic Meteorology*, volume 48 of *International Geophysics*. Academic Press, San Diego, 3rd edition.

Husler, J. and Reiss, R.-D. (1989). Maxima of Normal Random Vectors: Between Independence and Complete Dependence. *Statistics and Probability Letters*, 7(4):283–286.

Jagger, T. H. and Elsner, J. B. (2010). A consensus model for seasonal hurricane prediction. *Journal of Climate*, 23:6090–6099.

Jagger, T. H., Elsner, J. B., and Burch, R. K. (2010). Climate and solar signals in property damage losses from hurricanes affecting the United States. *Natural Hazards*, 58(1):541–557.

Jagger, T. H., Elsner, J. B., and Saunders, M. A. (2008). *Climate Extremes and Society*, chapter Forecasting U.S. Insured Hurricane Losses, pages 189–208. Cambridge University Press.

Jarvinen, B. R., Neumann, C. J., and Davis, M. A. (1988). A Tropical Cyclone Data Tape for the North Atlantic Basin, 1886-1982: Contents, Limitations, and Uses. NOAA Technical Memorandum NWS NHC 22, National Hurricane Center, Miami, Florida.

Jiang, J. (1998). Consistent Estimators in Generalized Linear Mixed Models. *Journal of the American Statistical Association*, 93(442):720–729.

Jockel, K.-H. (1986). Finite Sample Properties and Asymptotic Efficiency of Monte Carlo Tests. *The Annals of Statistics*, 14(1):336–347.

Joe, H. (1990). Families of Min-Stable Multivariate Exponential and Multivariate Extreme Value Distributions. *Statistics and Probability Letters*, 9(1):75–81.

Joe, H., Smith, R. L., and Weissman, I. (1992). Bivariate Threshold Methods for Extremes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 54(1):171–183.

Jones, G. L., Haran, M., Caffo, B. S., and Neath, R. (2006). Fixed-width Output Analysis for Markov chain Monte Carlo. *Journal of the American Statistical Association*, 101(476):1537–1547.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Jenne, R., and Joseph, D. (1996). The NCEP/NCAR 40-year Reanalysis Project. *Bulletin of the American Meteorological Society*, 77(3):437–470.

Keener, R. W. (2010). *Theoretical Statistics: Topics for a Core Course*. Springer Texts in Statistics. Springer.

Kerman, J. (2011). A Closed-Form Approximation for the Median of the Beta Distribution. arXiv:1111.0433 [math.ST].

Knaff, J. A., Brown, D. P., Courtney, J., Gallina, G. M., and II, J. L. B. (2010). An Evaluation of Dvorak Technique–Based Tropical Cyclone Intensity Estimates. *Weather and Forecasting*, 25(5):1362–1379.

Knaff, J. A., Sampson, C. R., DeMaria, M., Marchok, T. P., Gross, J. M., and McAdie, C. J. (2007). Statistical Tropical Cyclone Wind Radii Prediction Using Climatology and Persistence. *Weather and Forecasting*, 22(4):781–791.

Knaff, J. A. and Zehr, R. M. (2007). Reexamination of Tropical Cyclone Wind–Pressure Relationships. *Weather and Forecasting*, 22(1):71–88.

Knapp, K. R. and Kruk, M. C. (2010). Quantifying Interagency Differences in Tropical Cyclone Best-Track Wind Speed Estimates. *Monthly Weather Review*, 138(4):1459–1473.

Knapp, K. R., Kruk, M. C., Levinson, D. H., Diamond, H. J., and Neumann, C. J. (2010). The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying tropical cyclone best track data. The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying tropical cyclone best track data. *Bulletin of the American Meteorological Society*, 91(3):363–376.

Koba, H., Osano, S., Hagiwara, T., Akashi, S., and Kikuchi, T. (1990). Determination of Intensity of Typhoons Passing through the Philippine Islands (in Japanese). *Journal of Meteorological Research*, 41:157–162.

Kraft, R. H. (1961). The Hurricane's Central Pressure and Highest Wind. *Mariner's Weather Log*, 5:155.

Krishnamurty, C. K. B., Lall, U., and Kwon, H.-H. (2009). Changing Frequency and Intensity of Rainfall Extremes over India from 1951 to 2003. *Journal of Climate*, 22(18):4737–4746.

Kumar, K. K., Rajagopalan, B., and Cane, M. A. (1999). On the weakening relationship between the indian monsoon and enso. *Science*, 284(5423):2156–2159.

Kwan, M.-P., Richardson, D., Wang, D., and Zhou, C., editors (2015). *Space-Time Integration in Geography and GIScience*. Springer, Netherlands.

Landsea, C. W. (1993). A Climatology of Intense (or Major) Atlantic hurricanes. *Monthly Weather Review*, 121(6):1703–1713.

Landsea, C. W., Anderson, C., Charles, N., Gilbert Clark, J. D., Fernandez-Partagas, J., Hungerford, P., Neumann, C., and Zimmer, M. (2004). *Hurricanes and Typhoons*, chapter The Atlantic Hurricane Database Re-analysis

Project: Documentation for 1851-1910 Alterations and Additions to the HURDAT Database. Columbia University Press.

Landsea, C. W. and Franklin, J. L. (2013). Atlantic Hurricane Database Uncertainty and Presentation of a New Database Format. *Monthly Weather Review*, 141(10):3576–3592.

Lele, S. R., Nadeem, K., and Schumuland, B. (2010). Estimability and Likelihood Inference for Generalized Linear Mixed Models Using Data Cloning. *Journal of the American Statistical Association*, 105(492):1617–1625.

Li, B., Genton, M. G., and Sherman, M. (2007a). A Nonparametric Assessment of Properties of Space–Time Covariance Functions. *Journal of the American Statistical Association*, 102(478):736–744.

Li, C. and Yanai, M. (1996). The Onset and Interannual Variability of the Asian Summer Monsoon in Relation to Land Sea Thermal Contrast. *Journal of Climate*, 9(2):358–375.

Li, H., Calder, C. A., and Cressie, N. (2007b). Beyond Moran's I - Testing for Spatial Dependence Based on the Spatial Autoregressive Model. *Geographical Analysis*, 39(4):357–375.

Lin, X. and Breslow, N. E. (1996). Bias Correction in Generalized Linear Mixed Models With Multiple Components of Dispersion. *Journal of the American Statistical Association*, 91(435):1007–1016.

Ljung, G. M. and Box, G. E. P. (1978). On a Measure of Lack of Fit in Time Series Models. *Biometrika*, 65(2):297–303.

Maclay, K. S., DeMaria, M., and Haar, T. H. V. (2008). Tropical Cyclone Inner-Core Kinetic Energy Evolution. *Monthly Weather Review*, 136(12):4882–4898.

Mahalanobis, P. C. (1936). On the Generalised Distance in Statistics. *Proceedings of the National Institute of Sciences of India*, 2(1):49–55.

Mailier, P. J., Stephenson, D. B., Ferro, C. A. T., and Hodges, K. I. (2006). Serial Clustering of Extratropical Cyclones. *Monthly Weather Review*, 134(8):2224–2240.

Marriott, F. H. C. (1979). Barnard's Monte Carlo Tests: How Many Simulations? *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1):75–77.

McCulloch, C. E. and Searle, S. R. (2010). *Generalized, Linear, and Mixed Models*. Wiley Series in Probability and Statistics. John Wiley and Sons, Inc., Hoboken, New Jersey, 2nd edition.

Mei, W., Pasquero, C., and Primeau, F. (2012). The Effect of Translation Speed upon the Intensity of Tropical Cyclones over the Tropical Ocean. *Geophysical Research Letters*, 39(7):1–6.

Menne, M. J., Durre, I., Vose, R. S., Gleason, B. E., and Houston, T. G. (2012). An Overview of the Global Historical Climatology Network-Daily Database. *Journal of Atmospheric and Oceanic Technology*, 29(7):897–910.

Mitchell, M. W., Genton, M. G., and Gumpertz, M. L. (2006). A Likelihood Ratio Test for Separability of Covariances. *Journal of Multivariate Analysis*, 97(5):1025–1043.

Mitze, T. (2012). *Empirical Modelling in Regional Science*, volume 657 of *Lecture Notes in Economics and Mathematical Systems*. Springer-Verlag, Berlin Heidelberg.

Moran, P. A. P. (1950). Notes on Continuous Stochastic Phenomena. *Biometrika*, 37(1/2):17–23.

Mueller, K. J., DeMaria, M., Knaff, J. A., Kossin, J. P., and Haar, T. H. V. (2006). Objective Estimation of Tropical Cyclone Wind Structure from Infrared Satellite Data. *Weather and Forecasting*, 21(6):990–1005.

Pielke, R. A. (2009). United States Hurricane Landfalls and Damages: Can One- to Five-Year Predictions Beat Climatology? *Environmental Hazards*, 8(3):187–200.

Pielke, R. A., Gratz, J., Landsea, C. W., Collins, D., and Musulin, M. A. S. R. (2008). Normalized Hurricane Damage in the United States: 1900–2005. *Natural Hazards*, 9(1):29–42.

Plummer, M. (2015). *rjags: Bayesian Graphical Models using MCMC*. R package version 3-15.

Prell, W. L. and Kutzback, J. E. (1992). Sensitivity of the Indian Monsoon to Forcing Parameters and Implications for Its Evolution. *Nature*, 360(6405):647–652.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rajeevan, M., Bhate, J., and Jaswal, A. (2008). Analysis of Variability and Trends of Extreme Rainfall Events Over India Using 104 Years of Gridded Daily Rainfall Data. *Geophysical Research Letters*, 35(18):1–6.

SAS Institute Inc. (2011). *SAS/STAT 9.3 User's Guide*. SAS Institute Inc., Cary, NC.

Scaccia, L. and Martin, R. (2005). Testing Axial Symmetry and Separability of Lattice Processes. *Journal of Statistical Planning and Inference*, 131(1):19–39.

Schwerdt, R. W., Ho, F. P., and Watkins, R. R. (1979). Metorological Criteria for Standard Project Hurricane and Probable Maximum Hurricane Wind Fields—Gulf and East Coasts of the United States. Technical Report NWS 23, NOAA, National Hurricane/Tropical Prediction Center Library, 11691 S.W. 117th St., Miami, FL.

Shapiro, S. S. and Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, 52(3-4):591–611.

Singh, D., Tsiang, M., Rajaratnam, B., and Diffenbaugh, N. S. (2014). Observed Changes in Extreme Wet and Dry Spells During the South Asian Summer Monsoon Season. *Nature Climate Change*, 4(6):456–461.

Sinnott, R. (1984). Virtues of the Haversine. *Sky and Telescope*, 68(2):159.

Smith, R. L. (1985). Maximum Likelihood Estimation in a Class of Nonregular Cases. *Biometrika*, 72(1):67–90.

Smith, R. L. (1990). *Handbook of Applicable Mathematics*, volume 7, chapter Extreme Value Theory, pages 437–471. John Wiley.

Smith, T. M., Reynolds, R. W., Peterson, T. C., and Lawrimore, J. (2008). Improvements to NOAA's Historical Merged Land–Ocean Surface Temperature Analysis. *Journal of Climate*, 21(10):2283–2296.

Solymos, P. (2010). dclone: Data Cloning in R. *The R Journal*, 2(2):29–37.

Stephenson, A. G. (2002). evd: Extreme Value Distributions. *R News*, 2(2):31–32.

Takahashi, K. (1952). Techniques of the Typhoon Forecast. *The Geophysical Magazine*, 24:1–8.

Tawn, J. (1988). Bivariate Extreme Value Theory: Models and Estimation. *Biometrika*, 75(3):397–415.

Thas, O. (2010). *Comparing Distributions*. Springer Series in Statistics. Springer-Verlag, New York, 1st edition.

Tierney, L. and Kadane, J. B. (1986). Accurate Approximations for Posterior Moments and Marginal Densities. *Journal of the American Statistical Association*, 81(393):82–86.

Turner, A. G. and Annamalai, H. (2012). Climate Change and the South Asian Summer Monsoon. *Nature Climate Change*, 2(8):587–595.

Vats, D., Flegal, J. M., and Jones, G. L. (2015). Multivariate Output Analysis for Markov Chain Monte Carlo. arXiv:1512.07713v1 [math.ST].

Velden, C., Harper, B., Wells, F., II, J. L. B., Zehr, R., Olander, T., Mayfield, M., Guard, C., Lander, M., Edson, R., Avila, L., Burton, A., Turk, M., Kikuchi, A., Christian, A., Caroff, P., and McCrone, P. (2006). The Dvorak Tropical Cyclone Intensity Estimation Technique: A Satellite-Based Method that Has Endured for over 30 Years. *Bulletin of the American Meteorological Society*, 87(9):1195–1210.

Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer, New York, 4th edition.

Villarini, G., Vecchi, G. A., and Smith, J. A. (2010). Modeling the Dependence of Tropical Storm Counts in the North Atlantic Basin on Climate Indices. *American Meteorological Society*, 138(7):2681–2705.

Wang, B. (2006). *The Asian Monsoon*. Springer-Praxis Books in Environmental Sciences. Springer-Verlag, Berlin Heidelberg.

Wang, Y. and Wu, C. C. (2004). Current Understanding of Tropical Cyclone Structure and Intensity Changes – A Review. *Meteorology and Atmospheric Physics*, 87(4):257–278.

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, 48(4):817–838.

Wolfinger, R. and O'Connell, M. (1993). Generalized Linear Mixed Models a Pseudo-likelihood Approach. *Journal of Statistical Computation and Simulation*, 48(3-4):233–243.

Xavier, P. K., Marzina, C., and Goswami, B. N. (2007). An Objective Definition of the Indian Summer Monsoon Season and a New Perspective on the ENSO–Monsoon Relationship. *Quarterly Journal of the Royal Meteorological Society*, 133(624):749–764.

Zeileis, A., Kleiber, C., and Jackman, S. (2008). Regression Models for Count Data in R. *Journal of Statistical Software*, 27(8).

Zhang, D. and Lin, X. (2008). Variance Component Testing in Generalized Linear Mixed Models for Longitudinal/Clustered Data and other Related Topics. In Dunson, D., editor, *Random Effect and Latent Variable Model Selection*, volume 192 of *Lecture Notes in Statistics*, pages 19–36. Springer New York.

Appendix A

Supplementary Material for Chapter 1

A.1 Preliminary Analysis on Bivariate Extreme Value Models

Several forms of Bivariate Extreme Value Distributions (BVEVDs) are available. Within `fbvevd{evd}` in R (Stephenson, 2002), there are 9 possible parameterizations. We used location parameter modeling similar to the univariate conditional Generalized Extreme Value (GEV) models. In order to maintain simplicity, yet still reflect basin differences, we fit a separate intercept by basin in the location parameter. Let x_1 be log(maximum wind speed (MaxWS)) and x_2 be log(1013-minimum central pressure (MinCP)). Then,

$$y_1 = \left\{1 + \frac{\xi_1(x_1 - \mu_1)}{\sigma_1}\right\}^{-1/\xi_1} ; \sigma_1 > 0 \tag{A.1}$$

$$y_2 = \left\{1 + \frac{\xi_2(x_2 - \mu_2)}{\sigma_1}\right\}^{-1/\xi_2} ; \sigma_2 > 0. \tag{A.2}$$

A.2 Bivariate Extreme Value Distribution Parameterizations

A.2.1 Bivariate Logistic Distribution

The bivariate logistic (blog) distribution function (Gumbel, 1960) with parameter r is

$$G(y_1, y_2) = \exp\left\{-\left[y_1^{1/r} + y_2^{1/r}\right]^r\right\}, \tag{A.3}$$

where $0 < r \leq 1$. This is a special case of the bivariate asymmetric logistic model. Complete dependence is obtained in the limit as $r \to 0$. Independence is obtained

when $r = 1$.

A.2.2 Bivariate Asymmetric Logistic Distribution

The bivariate asymmetric logistic (balog) distribution function (Tawn, 1988) with parameters r and (t_1, t_2) is

$$G(y_1, y_2) = \exp\left\{-(1-t_1)y_1 - (1-t_2)y_2 - \left[(t_1 y_1)^{1/r} + (t_2 y_2)^{1/r}\right]^r\right\}, \quad \text{(A.4)}$$

where $0 < r \leq 1$ and $0 \leq t_1, t_2 \leq 1$. When $t_1 = t_2 = 1$ the asymmetric logistic model is equivalent to the logistic model. Independence is obtained when either $r = 1, t_1 = 0$ or $t_2 = 0$. Complete dependence is obtained in the limit when $t_1 = t_2 = 1$ and $r \to 0$. Different limits occur when t_1 and t_2 are fixed and $r \to 0$.

A.2.3 Bivariate Bilogistic Distribution

The bivariate bilogistic (bbilog) distribution (Smith, 1990; Joe et al., 1992) has the following parameterization:

$$G(y_1, y_2) = \exp\left\{-y_1 q^{1-\alpha} - y_2(1-q)^{1-\beta}\right\}, \quad \text{(A.5)}$$

where $q = q(y_1, y_2; \alpha, \beta)$ is the root of the equation

$$(1-\alpha)y_1(1-q)^{\beta} - (1-\beta)y_2 q^{\alpha} = 0, \quad \text{(A.6)}$$

with $0 < \alpha, \beta < 1$. Complete dependence is obtained in the limit as $\alpha = \beta \to 0$. Independence is obtained as $\alpha = \beta \to 1$, and when one of α, β is fixed and the other approaches 1.

A.2.4 Bivariate Negative Logistic Distribution

The bivariate negative logistic (bneglog) distribution function (Galambos, 1975) with parameter r is

$$G(y_1, y_2) = \exp\left\{-(y_1 + y_2) + \left[y_1^{-r} + y_2^{-r}\right]^{-1/r}\right\}, \tag{A.7}$$

where r > 0. This is a special case of the bivariate asymmetric negative logistic model. Independence is obtained in the limit as $r \to 0$. Complete dependence is obtained as $r \to \infty$.

A.2.5 Bivariate Asymmetric Negative Logistic Distribution

The bivariate asymmetric negative logistic (baneglog) distribution (Joe, 1990) function with parameters parameters r and (t_1, t_2) is

$$G(y_1, y_2) = \exp\left\{-(y_1 + y_2) + \left[(t_1 y_1)^{-r} + (t_2 y_2)^{-r}\right]^{-1/r}\right\}, \tag{A.8}$$

where $r > 0$ and $0 < t_1, t_2 \leq 1$. When $t_1 = t_2 = 1$ the asymmetric negative logistic model is equivalent to the negative logistic model. Independence is obtained in the limit as either r, t_1 or t_2 approaches zero. Complete dependence is obtained in the limit when $t_1 = t_2 = 1$ and $r \to \infty$. Different limits occur when t_1 and t_2 are fixed and $r \to \infty$.

A.2.6 Bivariate Negative Bilogistic Distribution

The bivariate negative bilogistic (bnegbilog) distribution function (Coles and Tawn, 1994) with parameters α and β is

$$G(y_1, y_2) = \exp\left\{-(y_1 + y_2) + y_1 q^{1+\alpha} + y_2(1-q)^{1+\beta}\right\}, \tag{A.9}$$

where $q = q(y_1, y_2; \alpha, \beta)$ is the root of the equation

$$(1+\alpha)y_1 q^{\alpha} - (1+\beta)y_2(1-q)^{\beta} = 0 \tag{A.10}$$

with $\alpha > 0$ and $\beta > 0$. When $\alpha = \beta$, the negative bilogistic model is equivalent to the negative logistic model with dependence parameter $1/\alpha = 1/\beta$. Complete dependence is obtained in the limit as $\alpha = \beta \to 0$. Independence is obtained as $\alpha = \beta \to \infty$, and when one of α, β is fixed and the other tends to ∞. Different limits occur when one of α, β is fixed and the other approaches 0.

A.2.7 Asymmetric Mixed Distribution

The asymmetric mixed (amix) distribution function (Tawn, 1988) with parameters α and β has a dependence function with the following cubic polynomial form:

$$A(t) = 1 - (\alpha + \beta)t + \alpha t^2 + \beta t^3, \tag{A.11}$$

where α and $\alpha + 3\beta$ are non-negative, and where $\alpha + \beta$ and $\alpha + 2\beta$ are less than or equal to one. These constraints imply that $\beta \in [-0.5, 0.5]$ and that $\alpha \in [0, 1.5]$, though α can only be greater than one if β is negative. The strength of dependence increases for increasing α (for fixed β). Complete dependence cannot be obtained. Independence is obtained when both parameters are 0.

A.2.8 Coles-Tawn Distribution

The Coles-Tawn (ct) distribution function (Coles and Tawn, 1991) with parameters $\alpha > 0$ and $\beta > 0$ is

$$G(y_1, y_2) = \exp\left\{-y_1\left[1 - Be(q; \alpha + 1, \beta)\right] - y_2 Be(q; \alpha, \beta + 1)\right\}, \tag{A.12}$$

where $q = \alpha y_2/(\alpha y_2 + \beta y_1)$ and $Be(q; \alpha, \beta)$ is the beta distribution function evaluated at q with shape parameters, $\xi_1 = \alpha$ and $\xi_2 = \beta$. Complete dependence is obtained in the limit as $\alpha = \beta \to \infty$. Independence is obtained as $\alpha = \beta \to 0$, and when one of α, β is fixed and the other approaches 0. Different limits occur when one of α, β is fixed and the other tends to ∞.

A.2.9 Husler-Reiss Distribution

The Husler-Reiss (hr) distribution function (Husler and Reiss, 1989) with parameter r is

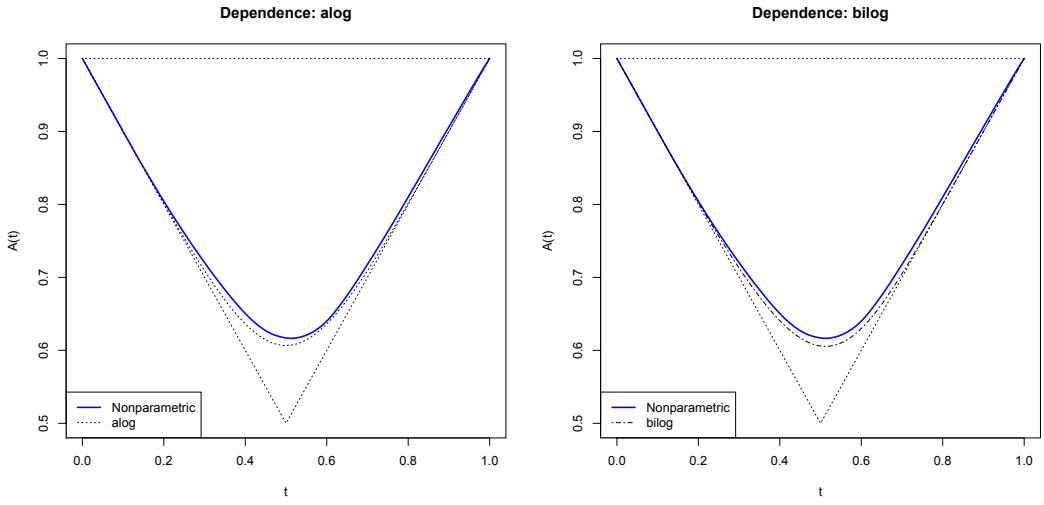
$$\begin{aligned} G(y_1, y_2) = \exp\Bigg\{ & -y_1 \Phi\left[r^{-1} + 0.5r \log\left(\frac{y_1}{y_2}\right)\right] \\ & -y_2 \Phi\left[r^{-1} + 0.5r \log\left(\frac{y_2}{y_1}\right)\right]\Bigg\}, \end{aligned} \tag{A.13}$$

where $\Phi()$ is the standard normal distribution function and $r > 0$. Independence is obtained in the limit as $r \to 0$. Complete dependence is obtained as $r \to \infty$.

A.3 BVEVD Model Selection Justification

The amix model could not converge without producing a singular information matrix, thus, it was excluded from further analysis. likelihood ratio tests (LRTs)

Figure A.1: Estimated Dependence Functions Compared to Nonparametric Estimation



are available for the nested models. The test of blog versus balog indicated a balog form was necessary with a p-value < 0.0001. The test of bneglog versus baneglog indicated a baneglog form was necessary with a p-value of < 0.0001. This immediately excluded amix, blog, and bneglog from further consideration.

Based on the Akaike information criterion (AIC) in Table A.1, the bbilog model provides the best fit. We also further excluded the hr and ct models due to their rather high comparative AIC. We also excluded the negative logistic forms since their AICs were not comparable to the logistic forms. Finally, using a nonparametric estimated dependence function (Capéraà et al., 1997), we compared the balog and bbilog as model candidates. The balog was slightly less symmetric than the nonparametric function, thus, based on AIC and this qualitative review, we choose the bbilog form as our final model.

Table A.1: Model Justification for BVEVD Forms

BVEVD Type	Name	AIC	Result
Logistic	blog	-1820	Excl. via LRT w/ balog
	balog	-1863	Excl. vis Qualitative Comparison
	bbilog	**-1866**	**Final Model**
Negative Logistic	bneglog	-1798	Excl. via LRT w/ baneglog
	baneglog	-1820	Excl. via AIC
	bnegbilog	-1843	Excl. via AIC
Other	hr	-1679	Excl. via AIC
	ct	-1735	Excl. via AIC
	amix	NA	Excl. via Nonconvergence

Appendix B

Supplementary Material for Chapter 3

B.1 GLMM Estimation Algorithms in R

B.1.1 Likelihood Approximation Estimation

The following methods, Penalized quasi-likelihood (PQL) and Penalized iteratively reweighted least squares (PIRLS), both make use of Laplace approximation to approximate an integral by a normal distribution (Tierney and Kadane, 1986)). First, note that (3.7) can be written as:

$$\log \int_{\mathbb{R}^M} f(\mathbf{y}|\mathbf{u},\boldsymbol{\beta}) f(\mathbf{u}|\Sigma) d\mathbf{u} = \tag{B.1}$$

$$\log \int_{\mathbb{R}^M} \exp\left\{\log f(\mathbf{y}|\mathbf{u},\boldsymbol{\beta}) + \log f(\mathbf{u}|\Sigma)\right\} d\mathbf{u}. \tag{B.2}$$

Let

$$h(\mathbf{u}) = \log f(\mathbf{y}|\mathbf{u},\boldsymbol{\beta}) + \log f(\mathbf{u}|\Sigma). \tag{B.3}$$

Then, express the log likelihood as follows:

$$\ell(\boldsymbol{\beta},\Sigma;\mathbf{y}) = \log \int_{\mathbb{R}^M} e^{h(\mathbf{u})} d\mathbf{u}. \tag{B.4}$$

This expression can now be approximated. To use the approximation, one first needs the maximizer of the integrand. Let $\mathbf{u}_0$ be the maximizer of $e^{h(\mathbf{u})}$. A Taylor expansion around $\mathbf{u}_0$ yields the approximation to the log-likelihood,

$$\ell(\boldsymbol{\beta},\Sigma;\mathbf{y}) \approx h(\mathbf{u}_0) + \frac{M}{2}\log 2\pi - \frac{1}{2}\log\left|-\frac{\partial^2 h(\mathbf{u})}{\partial \mathbf{u}, \partial \mathbf{u}^{\mathrm{T}}}\right|. \tag{B.5}$$

B.1.1.1 PQL

PQL approximates the high-dimensional integral in the likelihood using Laplace approximation as a method for obtaining $\mathbf{u_0}$ and $\partial^2 h(\mathbf{u})/\partial\mathbf{u}\partial\mathbf{u}^{\mathrm{T}}$. Assuming that $\mathbf{U} \sim \mathcal{N}_M(\mathbf{0}, \Sigma)$, then

$$\log f(\mathbf{u}|\Sigma) = -\frac{1}{2}\mathbf{u}^{\mathrm{T}}\Sigma^{-1}\mathbf{u} - \frac{M}{2}\log 2\pi - \frac{1}{2}\log|\Sigma|\,, \tag{B.6}$$

and thus, (B.3) becomes

$$h(\mathbf{u}) = \log f(\mathbf{y}|\mathbf{u}, \boldsymbol{\beta}) - \frac{1}{2}\mathbf{u}^{\mathrm{T}}\Sigma^{-1}\mathbf{u} - \frac{M}{2}\log 2\pi - \frac{1}{2}\log|\Sigma|\,. \tag{B.7}$$

This equation is differentiated with respect to $\mathbf{u}$ and $\boldsymbol{\beta}$ respectively. Further approximations are made within the derivatives because Σ is also unknown. The approximate derivatives are used to form estimating equations for the mean parameters. For more detailed discussion of these approximations, please refer to (McCulloch and Searle, 2010). The same estimating equations arise from jointly maximizing

$$\log f(\mathbf{y}|\mathbf{u}, \boldsymbol{\beta}) - \frac{1}{2}\mathbf{u}^{\mathrm{T}}\Sigma\mathbf{u}, \tag{B.8}$$

with respect to $\mathbf{u}$ and $\boldsymbol{\beta}$.

These equations are solved using Fisher scoring as an iterated re-weighted least squares (IRLS) problem. The quasi-likelihood, $\log f(\mathbf{y}|\mathbf{u}, \boldsymbol{\beta})$, is optimized taking into account the penalty, $\frac{1}{2}\mathbf{u}^{\mathrm{T}}\Sigma\mathbf{u}$. This penalty term has a shrinkage effect, i.e. forces values of $\mathbf{u}$ to be closer to zero.

Variance components in Σ are subsequently estimated using a restricted maximum likelihood approach. Further details on the estimation algorithm are found

in Section 2 of (Breslow and Clayton, 1993).

B.1.1.2 PIRLS Algorithm

Again, let $\mathbf{U} \sim \mathcal{N}_M(\mathbf{0}, \Sigma)$. Consider the decomposition of the random effects covariance matrix $\Sigma = \mathbf{\Gamma}\mathbf{\Gamma}^{\mathrm{T}}$. Then, $\mathbf{U} = \mathbf{\Gamma}\mathbf{V}$ where $\mathbf{V} \sim \mathcal{N}_q(\mathbf{0}, I_m)$. This implies that the canonical parameter in (3.3) can be written as

$$\eta_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + \mathbf{z}_i^{\mathrm{T}}\Gamma\mathbf{v}. \tag{B.9}$$

Substituting in $\mathbf{v}$ to (3.6), it can be noted that $f(\mathbf{y}, \mathbf{v})$ is proportional to $f(\mathbf{v}|\mathbf{y})$. Thus, $\mathbf{v}_0$ is found to maximize $f(\mathbf{v}|\mathbf{y})$.

The PIRLS algorithm is as follows.

1. Given starting values for $\boldsymbol{\beta}$, Σ, and $\mathbf{v}_0$, evaluate $\boldsymbol{\eta}$, $\mu_{\mathbf{Y}|\mathbf{v}}$, and $var_{\mathbf{Y}|\mathbf{v}}$. Let $\mathbf{W} = \mathrm{diag}\{var_{\mathbf{Y}|\mathbf{v}}^{-1}\}$.

2. Use a Gauss-Newton algorithm to solve

$$\mu_{\mathbf{V}|\mathbf{y}} = \arg\min_{\mathbf{v}} \left(||\mathbf{W}^{1/2}(\mathbf{y} - \mu_{\mathbf{Y}|\mathbf{v}})||^2 + ||\mathbf{v}||^{\mathbf{2}} \right). \tag{B.10}$$

3. Update $\mathbf{W}$, and check for convergence. If not converged, go to step 2.

Once the conditional mode $\tilde{\mathbf{v}}$ is determined, a Laplace approximation to the deviance $(-2\ell(\boldsymbol{\beta}, \Sigma; \mathbf{y}))$ is evaluated at $\tilde{\mathbf{v}}$. This evaluation may also be done by Gauss-Hermite quadrature in certain cases.

B.1.2 Method of Moments Estimation

B.1.2.1 Method of Simulated Moments Algorithm

Referring to the model elucidated in (3.17)-(3.20), let the dispersion function be $a(\phi) = \frac{w_i}{\phi}$ where w_i is a weight depending on the exponential family of the response. Let $\theta = (\beta, \sigma_{\mathbf{1}}, \ldots, \sigma_{\mathbf{q}})$. Restrict all elements of the $\mathbf{z}_i$ to be either 0 or 1. Represent

$$\mathbf{z}_i^{\mathrm{T}}\mathbf{u} = (\mathbf{z}_{i1}^{\mathrm{T}}\mathbf{u}_1, \ldots, \mathbf{z}_{iq}^{\mathrm{T}}\mathbf{u}_q) \tag{B.11}$$

$$= (\sigma_1\mathbf{z}_{i1}^{\mathrm{T}}\mathbf{v}_1, \ldots, \sigma_q\mathbf{z}_{iq}^{\mathrm{T}}\mathbf{v}_q), \tag{B.12}$$

where $\mathbf{V}_r \sim \mathcal{N}_{m_r}(0, I_{m_r})$.

Then,

$$f(y_i|\mathbf{v}) = C(y_i, \boldsymbol{\theta}, \phi) \exp\left\{ \left(\sum_{i=1}^{N} w_i \mathbf{x}_i y_i \right)^{\mathrm{T}} \frac{\beta}{\phi} + \sum_{r=1}^{q} \frac{\sigma_r}{\phi} \sum_{i=1}^{N} w_i y_i \mathbf{z}_{ir}^{\mathrm{T}} \mathbf{v}_r \right\}, \tag{B.13}$$

where $C(\cdot, \cdot, \cdot)$ represents the balance of the function.

This yields canonical parameters $(\boldsymbol{\beta}/\phi, \sigma_1/\phi, \ldots, \sigma_q/\phi)$ with corresponding sufficient statistics $\left(\left(\sum_{i=1}^{N} w_i \mathbf{x}_i y_i \right)^{\mathrm{T}}, \sum_{i=1}^{n} w_i y_i \mathbf{z}_{i1}, \ldots, \sum_{i=1}^{N} w_i y_i \mathbf{z}_{iq} \right)$.

Estimating equations are derived as

$$\sum_{i=1}^{n} w_i \mathbf{x}_i y_i \overset{set}{=} \sum_{i=1}^{n} w_i \mathbf{x}_i \mathbb{E}_\theta(y_i), \tag{B.14}$$

$$\sum_{l=1}^{m_r} \left(\sum_{i=1}^{n} w_i z_{irl} y_i \right)^2 \overset{set}{=} \sum_{l=1}^{m_r} \mathbb{E}_\theta \left(\sum_{i=1}^{n} w_i z_{irl} y_i \right)^2. \tag{B.15}$$

Note that the expectations on the right hand side are functions of the parameters while the formulae on the left hand sides are functions of data only. Since

the expectations are not available, they must be estimated by Monte Carlo simulation. The system of equations can then be solved for the parameters using the Newton-Raphson algorithm.

B.1.3 Bayesian Estimation

B.1.3.1 Data Cloning Algorithm

The algorithm can be summarized in the following three steps.

1. Create a k-cloned data set $\mathbf{y}_k$=($\mathbf{y}$,$\mathbf{y}$,...,$\mathbf{y}$) where the observed data vector is repeated k times. Choose a prior distribution $\pi(\boldsymbol{\beta}, \Sigma)$. The posterior distribution, $\pi_k(\boldsymbol{\beta}, \Sigma|\mathbf{Y})$, which corresponds to the k-cloned data is

$$\pi_k(\boldsymbol{\beta}, \Sigma|\mathbf{Y}) = \frac{[L(\boldsymbol{\beta}, \Sigma|\mathbf{Y})]^k \pi(\boldsymbol{\beta}, \Sigma)}{\int_{(\boldsymbol{\beta},\Sigma)} [L(\boldsymbol{\beta}, \Sigma|\mathbf{Y})]^k \pi(\boldsymbol{\beta}, \Sigma) d(\boldsymbol{\beta}, \Sigma)}. \tag{B.16}$$

   Under regularity conditions as $k \to \infty$,

$$\pi_k(\boldsymbol{\beta}, \Sigma|\mathbf{Y}) \to \mathcal{N}\left(\widehat{(\boldsymbol{\beta}, \Sigma)}, \frac{1}{k} S^{-1}\widehat{(\boldsymbol{\beta}, \Sigma)}\right), \tag{B.17}$$

   where $\widehat{(\boldsymbol{\beta}, \Sigma)}$ is the maximum likelihood estimate (MLE) of $(\boldsymbol{\beta}, \Sigma)$ and is the Fisher information matrix of the original data. Thus, large k means the posterior distribution is nearly degenerate at the MLE.

2. Use an appropriate Markov chain Monte Carlo (MCMC) algorithm, such as a Gibbs sampler or Metropolis-Hastings algorithm to generate a dependent sample from the posterior distribution $\pi_k(\boldsymbol{\beta}, \Sigma|\mathbf{Y})$.

3. Calculate the sample means and variances of the components of $(\boldsymbol{\beta}, \Sigma)$. Estimates of MLEs for $(\boldsymbol{\beta}, \Sigma)$ correspond to these sample means and approximate variances of estimated MLEs correspond to k times the posterior variance of the original data as seen in (B.17).

Refer to (Lele et al., 2010) for any further methodology details.

B.2 Settings for R Simulations and Application in Study I

B.2.1 Specifications of Computers for Speed Testing

- assawa: 2010 Frontier i7 8-core Intel i7 940 (2.93 GHz); 3 GB RAM
- geneva: 2011 Frontier i7 8-core Intel i7 950 (3.07 GHz); 6 GB RAM
- nokomis: 2012 Optiplex 7010 8-core Intel i7-3770 (3.40 GHz); 8 GB RAM
- tilde: 2013 Optiplex 7010 8-core Intel i7-3770 (3.40 GHz); 8 GB RAM

B.2.2 Estimation Methodology Specifications

MSIM Fast

-Number of Monte Carlo simulations: 100000

MSIM Slow

-Number of Monte Carlo simulations: 100

-Convergence criterion for Newton Raphson Method: Euclidean norm of iteration difference $\leq .01$

Dclone

-Clones: 5

-Prior for μ: N$(0, \frac{1}{0.0001})$

-Prior for $\frac{1}{\sigma^2}$: Gamma$(0.01, 0.01)$

-Adaptation length: 100

-Markov chain length after adaptation: 10000

Appendix C

Supplementary Material for Chapter 4

C.1 Proof of Separable Spatio-temporal Series

Given matrices A, B, C, and D, the following Kronecker product and vec properties hold:

1. $(A \otimes B)(C \otimes D) = AC \otimes BD$
2. $(A \otimes B)^T = A^T \otimes B^T$
3. $\text{vec}(ABC) = \left(C^T \otimes A\right) \text{vec}(B)$

Thus, by 3,

$$\text{vec}(\mathbf{Y}_{S\times T}) = \text{vec}\left(\Sigma_S^{1/2}\mathbf{Z}_{S\times T}\Sigma_T^{1/2}\right) \tag{C.1}$$

$$= \left(\left(\Sigma_T^{1/2}\right)^T \otimes \Sigma_S^{1/2}\right) \text{vec}(\mathbf{Z}_{S\times T}) \tag{C.2}$$

Because Σ_T is symmetric, $\Sigma_T^T = \Sigma_T$. Thus,

$$\text{vec}(\mathbf{Y}_{S\times T}) = \left(\Sigma_T^{1/2} \otimes \Sigma_S^{1/2}\right) \text{vec}(\mathbf{Z}_{S\times T}) \tag{C.3}$$

$$= \left(\Sigma_T^{1/2} \otimes \Sigma_S^{1/2}\right) \mathbf{Z}_{ST}. \tag{C.4}$$

$$\mathbb{V}\left(\mathrm{vec}\left(\mathbf{Y}_{S\times T}\right)\right) = \mathbb{V}\left(\left(\Sigma_T^{1/2}\otimes\Sigma_S^{1/2}\right)\mathbf{Z}_{ST}\right). \tag{C.5}$$

$$= \left(\Sigma_T^{1/2}\otimes\Sigma_S^{1/2}\right)\mathbb{V}\left(\mathbf{Z}_{ST}\right)\left(\Sigma_T^{1/2}\otimes\Sigma_S^{1/2}\right) \tag{C.6}$$

$$= \left(\Sigma_T^{1/2}\otimes\Sigma_S^{1/2}\right) I_{ST} \left(\Sigma_T^{1/2}\otimes\Sigma_S^{1/2}\right)^T \tag{C.7}$$

$$= \left(\Sigma_T^{1/2}\otimes\Sigma_S^{1/2}\right)\left(\Sigma_T^{1/2}\otimes\Sigma_S^{1/2}\right)^T \tag{C.8}$$

$$= \left(\Sigma_T^{1/2}\otimes\Sigma_S^{1/2}\right)\left(\left(\Sigma_T^{1/2}\right)^T\otimes\left(\Sigma_S^{1/2}\right)^T\right) \text{ (Property 2)} \tag{C.9}$$

$$= \left(\Sigma_T^{1/2}\left(\Sigma_T^{1/2}\right)^T\right)\otimes\left(\Sigma_S^{1/2}\left(\Sigma_S^{1/2}\right)^T\right) \text{ (Property 1)} \tag{C.10}$$

$$= \left(\Sigma_T^{1/2}\Sigma_T^{1/2}\right)\otimes\left(\Sigma_S^{1/2}\Sigma_S^{1/2}\right) \text{ (symmetry)} \tag{C.11}$$

$$= \Sigma_T\otimes\Sigma_S. \tag{C.12}$$

$$\Sigma = \left(\begin{array}{c|c|c} \phi_{11}\Sigma_S & \dots & \phi_{1T}\Sigma_S \\ \hline \vdots & \ddots & \vdots \\ \hline \phi_{T1}\Sigma_S & \dots & \phi_{TT}\Sigma_S \end{array}\right) \tag{C.13}$$

$$= \left(\begin{array}{ccc|c|ccc} \phi_{11}\cdot\nu_{11} & \dots & \phi_{11}\cdot\nu_{1S} & & \phi_{1T}\cdot\nu_{11} & \dots & \phi_{1T}\cdot\nu_{1S} \\ \vdots & \ddots & \vdots & \dots & \vdots & \ddots & \vdots \\ \phi_{11}\cdot\nu_{S1} & \dots & \phi_{11}\cdot\nu_{SS} & & \phi_{1T}\cdot\nu_{S1} & \dots & \phi_{1T}\cdot\nu_{SS} \\ \hline \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \hline \phi_{T1}\cdot\nu_{11} & \dots & \phi_{T1}\cdot\nu_{1S} & & \phi_{TT}\cdot\nu_{11} & \dots & \phi_{TT}\cdot\nu_{1S} \\ \vdots & \ddots & \vdots & \dots & \vdots & \ddots & \vdots \\ \phi_{T1}\cdot\nu_{S1} & \dots & \phi_{T1}\cdot\nu_{SS} & & \phi_{TT}\cdot\nu_{S1} & \dots & \phi_{TT}\cdot\nu_{SS} \end{array}\right). \tag{C.14}$$

C.2 SAS Procedure for Pseudo-linearized Data

Using matrix rather than case notation, we define the same generalized linear mixed model (GLMM) using the natural exponential family (NEF) as in Chapter 3. We now incorporate the canonical link function g which converts from canonical parameter $\boldsymbol{\eta}$ to a mean parameter $\boldsymbol{\mu} = \mathbb{E}[\mathbf{Y}|\mathbf{U} = \mathbf{u}]$. The model is now written as:

$$f(\mathbf{Y}|\mathbf{U} = \mathbf{u}) = \text{NEF}(\boldsymbol{\eta}); \tag{C.15}$$

$$\boldsymbol{\mu} = g^{-1}(\boldsymbol{\eta}) = g^{-1}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}), \tag{C.16}$$

$$\mathbf{U} \sim \mathcal{N}(\mathbf{0}, \Sigma_U). \tag{C.17}$$

Following (Wolfinger and O'Connell, 1993), a first order Taylor expansion about $\tilde{\boldsymbol{\beta}}$ and $\tilde{\mathbf{u}}$ gives:

$$g^{-1}(\mathbf{X}\tilde{\boldsymbol{\beta}} + \mathbf{Z}\tilde{\mathbf{u}}) \doteq g^{-1}(\tilde{\boldsymbol{\eta}}) + \tilde{\Delta}\mathbf{X}(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}) + \tilde{\Delta}\mathbf{Z}(\mathbf{u} - \tilde{\mathbf{u}}), \tag{C.18}$$

where $\tilde{\Delta} = \left(\frac{\partial g^{-1}(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\right)_{\tilde{\boldsymbol{\beta}}, \tilde{\mathbf{u}}}$.

Thus, rearranging (C.18) gives:

$$\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} \doteq \tilde{\Delta}^{-1}\left[\mathbf{Y} - g^{-1}(\tilde{\boldsymbol{\eta}})\right] + \mathbf{X}\tilde{\boldsymbol{\beta}} + \mathbf{Z}\tilde{\mathbf{u}}. \tag{C.19}$$

so the estimated pseudo-linearized data, $\tilde{\mathbf{p}}$, at each optimization step is set to:

$$\tilde{\mathbf{p}} \overset{set}{=} \tilde{\Delta}^{-1}\left[\mathbf{Y} - g^{-1}(\tilde{\boldsymbol{\eta}})\right] + \mathbf{X}\tilde{\boldsymbol{\beta}} + \mathbf{Z}\tilde{\mathbf{u}}. \tag{C.20}$$

The final update of this $\tilde{\mathbf{p}}$ is used as residuals for the hypothesis testing.