# P8160 - Hurricane Project Report

Yimiao Pang, Xiao Ma, Wen Cheng, Tucker Morgan, Jie Liu

5/5/2022

# 1. Introduction

## 1.1. Background

Hurricanes are dangerous and can cause major damage from storm surge, wind damage, rip currents and flooding. They can happen along any U.S. coast or in any territory in the Atlantic or Pacific oceans. The amount of damage depends on the strength of a storm and what it hits. High winds are one of the primary causes of hurricane-inflicted loss of life and property damage. For better planning and prevention ahead to secure people from destructive hurricanes, it is extremely important and necessary to explore trajectories of hurricanes and predict each hurricane's wind speed.

## 1.2. Objectives

In this study, two data sets were explored.In the first part, we attempted to use the track data of 703 hurricanes in the North Atlantic area since 1950 to explore the seasonal differences and if there is any evidence showing that the hurricane wind speed has been increasing over years. First, we calculated the posterior distribution of four parameters $(B, \beta, \sigma^2, \Sigma^{-1})$ in proposed Bayesian model. Next, we designed an MCMC algorithm to generate the posterior distribution. Then, we used the MCMC chain we developed to estimate the parameters, and checked to see how well the model fits the data.

Furthermore, in order to forecast hurricane damage and deaths, another data set containing the damages and deaths caused by 46 hurricanes in the United States were used. We constructed a model and to determine which traits of hurricanes are more associated to damage and deaths.

# 2. Methods

## 2.1. Data Cleaning and Exploratory Analysis

In this study, there are two data sets. First one contains 703 hurricanes in the North Atlantic since 1950. It recorded the location (longitude and latitude) and maximum wind speed every 6 hours for every hurricanes. There are 8 variables and 22038 observations.

The second data set contains the damages and deaths caused by 46 hurricanes in the United States. There are 14 variables and 43 observations.

## 2.2. Posterior Distributions

The following Bayesian model was suggested.

$$Y_i(t+6) = \beta_{0,i} + \beta_{1,i}Y_i(t) + \beta_{2,i}\Delta_{i,1}(t) + \beta_{3,i}\Delta_{i,2}(t) + \beta_{4,i}\Delta_{i,3}(t) + \epsilon_i(t)$$

where $Y_i(t)$ the wind speed at time $t$ (i.e. 6 hours earlier), $\Delta_{i,1}(t)$, $\Delta_{i,2}(t)$ and $\Delta_{i,3}(t)$ are the changes of latitude, longitude and wind speed between $t$ and $t-6$, and $\epsilon_{i,t}$ follows a normal distributions with mean zero and variance $\sigma^2$, independent across $t$.

In the model, $\boldsymbol{\beta}_i = (\beta_{0,i}, \beta_{1,i}, ..., \beta_{7,i})$ are the random coefficients associated the $i$th hurricane, we assume that

$$\boldsymbol{\beta}_i \sim \mathcal{N}(\boldsymbol{\beta}, \boldsymbol{\Sigma}),$$

and we assume the following non-informative or weak prior distributions for $\sigma^2$, $\boldsymbol{\beta}$ and $\Sigma$.

$$P(\sigma^2) \propto \frac{1}{\sigma^2}; \quad P(\boldsymbol{\beta}) \propto 1; \quad P(\Sigma^{-1}) \propto |\Sigma|^{-(d+1)} \exp(-\frac{1}{2}\Sigma^{-1})$$

$d$ is dimension of $\beta$.

Note from given Bayesian model:

$$\epsilon_i(t) = Y_i(t+6) - \Big(\beta_{0,i} + \beta_{1,i}Y_i(t) + \beta_{2,i}\Delta_{i,1}(t) + \beta_{3,i}\Delta_{i,2}(t) + \beta_{4,i}\Delta_{i,3}(t)\Big) \overset{i.i.d}{\sim} N(0, \sigma^2)$$

or

$$Y_i(t+6) \sim N(\boldsymbol{X}_i(t)\boldsymbol{\beta}_i^T, \sigma^2)$$

where $\boldsymbol{X}_i(t) = (1, Y_i(t), \Delta_{i,1}(t), \Delta_{i,2}(t), \Delta_{i,3}(t))$, and $\boldsymbol{\beta}_i = (\beta_{0,i}, \beta_{1,i}, \beta_{2,i}, \beta_{3,i}, \beta_{4,i})$. Therefore,

$$f_{Y_i(t+6)}(y_i(t+6) \mid \boldsymbol{X}_i(t), \boldsymbol{\beta}_i, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\Big\{ -\frac{1}{2\sigma^2}\Big(y_i(t+6) - \boldsymbol{X}_i(t)\boldsymbol{\beta}_i^T\Big)^2\Big\}$$

for hurricane $i$ at time $t$. To show the likelihood function for hurricane $i$ across all time points, $t$, we can write the multivariate normal distribution

$$(\boldsymbol{Y}_i \mid \boldsymbol{X}_i, \boldsymbol{\beta}_i, \sigma^2) \sim \mathcal{N}(\boldsymbol{X}_i\boldsymbol{\beta}_i^T, \sigma^2 I)$$

where $Y_i$ is an $m_i$-dimensional vector and $\boldsymbol{X}_i$ is a $m_i \times d$ matrix. Finally, the joint likelihood function of all hurricanes can be expressed as

$$L_Y(\mathbf{B}, \sigma^2 I) = \prod_{i=1}^{n} \Big\{ \det(2\pi\sigma^2 I)^{-\frac{1}{2}} \exp\Big( -\frac{1}{2}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_i^T)^T(\sigma^2 I)^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_i^T)\Big)\Big\},$$

where $I$ is an identity matrix with dimension consistent with $Y_i$. We can find the posterior distribution for $\Theta$ by

$$\pi(\mathbf{B}, \boldsymbol{\beta}, \sigma^2, \Sigma^{-1} \mid Y) \propto L_Y(\mathbf{B}, \sigma^2 I) \times \pi(\mathbf{B} \mid \boldsymbol{\beta}, \Sigma^{-1}) \times \pi(\boldsymbol{\beta}) \times \pi(\sigma^2) \times \pi(\Sigma^{-1}),$$

where $\pi(\mathbf{B} \mid \boldsymbol{\beta}, \Sigma)$ is the joint multivariate normal density of $\beta$,

$$\pi(\mathbf{B} \mid \boldsymbol{\beta}, \Sigma^{-1}) = \prod_{i=1}^{n} \Big\{ \det(2\pi\Sigma)^{-\frac{1}{2}} \exp(-\frac{1}{2}(\boldsymbol{\beta}_i - \boldsymbol{\beta})\Sigma^{-1}(\boldsymbol{\beta}_i - \boldsymbol{\beta})^T)\Big\}.$$

So we have the following joint posterior distribution of $\Theta$:

$$\pi(\mathbf{B}, \boldsymbol{\beta}, \sigma^2, \Sigma^{-1} \mid Y) \propto \prod_{i=1}^{n} \Big\{ (2\pi\sigma^2)^{-m_i/2} \exp\big\{ -\frac{1}{2}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_i^T)^T(\sigma^2 I)^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_i^T)\big\}\Big\}$$
$$\times \prod_{i=1}^{n} \Big\{ \det(2\pi\Sigma)^{-\frac{1}{2}} \exp\big\{ -\frac{1}{2}(\boldsymbol{\beta}_i - \boldsymbol{\beta})\Sigma^{-1}(\boldsymbol{\beta}_i - \boldsymbol{\beta})^T\big\}\Big\} \times \frac{1}{\sigma^2} \times |\Sigma|^{-(d+1)} \exp\big\{ -\frac{1}{2}\Sigma^{-1}\big\}.$$

We can use the joint posterior distribution to derive conditional posterior distributions for each of our parameters.

Let $\tau = \sigma^2$, then

$$(\tau|\beta, B, \Sigma^{-1}, Y) \propto \tau^{1+\frac{\sum_{i=1}^{n} m_i}{2}} exp(-\tau \times \frac{1}{2}\sum_{i=1}^{n}(Y_i - X_i\beta_i^T)^T I^{-1}(Y_i - X_i\beta_i^T)$$

Thus, $\sigma^2$ is from inverse-gamma distribution

$$\sigma^2 \sim \text{Inv-Gamma}(\frac{\sum_{i=1}^{n} m_i}{2}, \frac{1}{2}\sum_{i=1}^{n}(Y_i - X_i\beta_i^T)(Y_i - X_i\beta_i^T).$$

Parameter $\mathbf{B}$ has the following conditional posterior:

$$\pi(B|\beta, \sigma^2, \Sigma^{-1}, Y) \propto exp(-\frac{1}{2}\sum_{i=1}^{n}\big[(Y_i - X_i\beta_i^T)^T(\sigma^2 I)^{-1}(Y_i - X_i\beta_i^T) + (\beta_i - \beta)\Sigma^{-1}(\beta_1 - \beta)^T\big]) \tag{1}$$
$$\propto exp(-\frac{1}{2}\sum_{i=1}^{n}\big[\beta_i(X_i^T(\sigma^2 I)^{-1}X_i + \Sigma^{-1})\beta_i^T - 2\beta_i(X_i(\sigma^2 I)^{-1})Y_i + \Sigma^{-1}\beta^T\big]) \tag{2}$$

Let $V_i = X_i^T(\sigma^2 I)^{-1})X_i + \Sigma^{-1}$, and $U_i = X_i(\sigma^2 I)^{-1}Y_i + \Sigma^{-1}\beta^T$, then

$$\boldsymbol{\beta}_i \sim \mathcal{M}VN(V_i^{-1}U_i, V_i^{-1}).$$

Similarly, parameter $\boldsymbol{\beta}$ has a conditional posterior of:

$$\pi(\beta|B, \sigma^2, \Sigma^{-1}, Y) \propto exp(-\frac{1}{2}\sum_{i=1}^{n}(\beta_i - \beta)\Sigma^{-1}(\beta_1 - \beta)^T) \tag{3}$$
$$\propto exp(-\frac{1}{2}\sum_{i=1}^{n}\big[\beta\Sigma^{-1}\beta^T - 2\beta\Sigma^{-1}\beta_i^T\big]) \tag{4}$$

Let $V = n\Sigma^{-1}, U = \sum_{i=1}^{n}\Sigma^{-1}\beta_i^T$, then

$$(\boldsymbol{\beta}|B, \sigma^2, \Sigma^{-1}, Y) \sim \mathcal{M}VN(V^{-1}U, V^{-1}).$$

Finally, parameter $\Sigma$ has the conditional posterior:

$$\pi(\Sigma^{-1}|\beta, B, \sigma^2, Y) \propto |\Sigma|^{-(d+1)}exp(-\frac{1}{2}tr(\Sigma^{-1})|\Sigma|^{-\frac{n}{2}}exp(-\frac{1}{2}\sum_{i=1}^{n}(\beta_i - \beta)\Sigma^{-1}(\beta_1 - \beta)^T) \tag{5}$$
$$\propto |\Sigma|^{-(d+1+\frac{n}{2})}exp(-\frac{1}{2}\big[tr(\Sigma^{-1}) + tr((\beta_i - \beta)\Sigma^{-1}(\beta_1 - \beta)^T)\big] \tag{6}$$
$$\propto |\Sigma|^{\frac{-(d+1+n+d+1)}{2}}exp(-\frac{1}{2}tr(\left[I + \sum_{i=1}^{n}(\beta_i - \beta)^T(\beta_i - \beta)\right]\Sigma^{-1})) \tag{7}$$

Thus,

$$\Sigma \sim \mathcal{W}_d^{-1}(\Psi, v),$$

where $v = d + 1 + n$, and $\Psi = I + \sum_{i=1}^{n}(\beta_i - \beta)^T(\beta_i - \beta)$.

## 2.3. Gibbs Sampling Algorithm

Now that we have conditional posterior distributions for each of our parameters, we can utilize the Gibbs Sampling MCMC algorithm to estimate model parameters. In Gibbs sampling, we use starting values $(\boldsymbol{\beta}_0, \Sigma_0, \sigma_0^2, \mathbf{B}_0)$ and for each $j = 1, 2, ..., n$:

1. Generate $\boldsymbol{\beta}_j$ from $\pi(\boldsymbol{\beta} \mid \Sigma = \Sigma_{j-1}, \sigma^2 = \sigma_{j-1}^2, \mathbf{B} = \mathbf{B}_{j-1})$;
2. Generate $\Sigma_j$ from $\pi(\Sigma \mid \boldsymbol{\beta} = \boldsymbol{\beta}_j, \sigma^2 = \sigma_{j-1}^2, \mathbf{B} = \mathbf{B}_{j-1})$;
3. Generate $\sigma^2$ from $\pi(\sigma^2 \mid \boldsymbol{\beta} = \boldsymbol{\beta}_j, \Sigma = \Sigma_j, \mathbf{B} = \mathbf{B}_{j-1})$;
4. Generate $\mathbf{B}$ from $\pi(\mathbf{B} \mid \boldsymbol{\beta} = \boldsymbol{\beta}_j, \Sigma = \Sigma_j \sigma^2 = \sigma_j^2)$

to yield $\Theta_j$. As $j$ increases and the Markov Chain continues, estimates stabilize, and we can obtain our results by taking the mean of the Gibbs-generated parameters.

# 3. Results

## 3.1. Bayesian Parameter Estimates

To obtain estimates of $\Theta = (\mathbf{B}^T, \boldsymbol{\beta}^T, \sigma^2, \Sigma)$, we first generated 1000 iterations in the Markov Chain. As an illustration, **Figure 1** shows the Markov Chain for $\boldsymbol{\beta}$ estimates. Although there is some noise in the chart, we can see that $\beta_0$ converges to a value approximately equal to 4, $\beta_1$ fluctuates about 1, $\beta_2$ and $\beta_4$ are between 0 and 1, and $\beta_3$ is less than zero.

Figure 1: Markov Chain of Beta Estimates

We find our estimates by taking the mean of $\Theta_j$, $j = 501, ..., 1000$. **Table 1** shows estimates for $\boldsymbol{\beta}$, and **Table 2** shows a selection of $\boldsymbol{\beta}_i$, $i = 1, ..., 6$ estimates from **B**. **Figure 2** shows the distribution of $\boldsymbol{\beta}_i$ coefficients across the population of hurricanes.

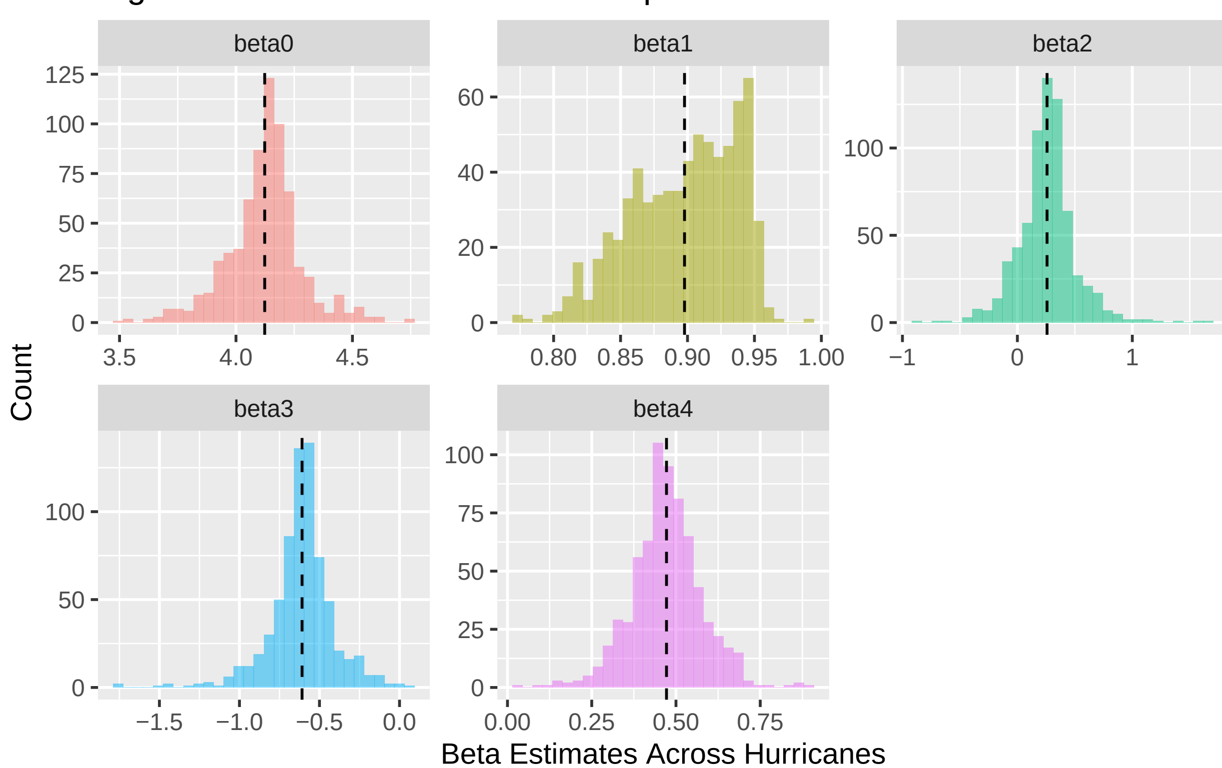
Table 1: Beta Parameter Estimates

| beta | avg_est |
|---|---|
| beta0 | 4.1173396 |
| beta1 | 0.9319335 |
| beta2 | 0.2587513 |
| beta3 | -0.6090160 |
| beta4 | 0.4765512 |

Table 2: Sample of Beta Estimates for i-th Hurricanes

| i | beta0 | beta1 | beta2 | beta3 | beta4 |
|---|---|---|---|---|---|
| 1 | 4.020239 | 0.9473525 | -0.0965604 | -0.7170444 | 0.4825699 |
| 2 | 4.059386 | 0.9129088 | 0.2621460 | -0.4725510 | 0.5880370 |
| 3 | 3.825276 | 0.9395315 | 0.2853879 | -0.4219473 | 0.4145077 |
| 4 | 4.061244 | 0.9516740 | 0.1761047 | -0.5171807 | 0.5559185 |
| 5 | 3.791682 | 0.9364665 | 0.0294229 | -0.3995459 | 0.5311705 |
| 6 | 4.023829 | 0.9524201 | -0.0741399 | -0.8908773 | 0.5419813 |

Figure 2: Beta Estimates Across Population of Hurricanes



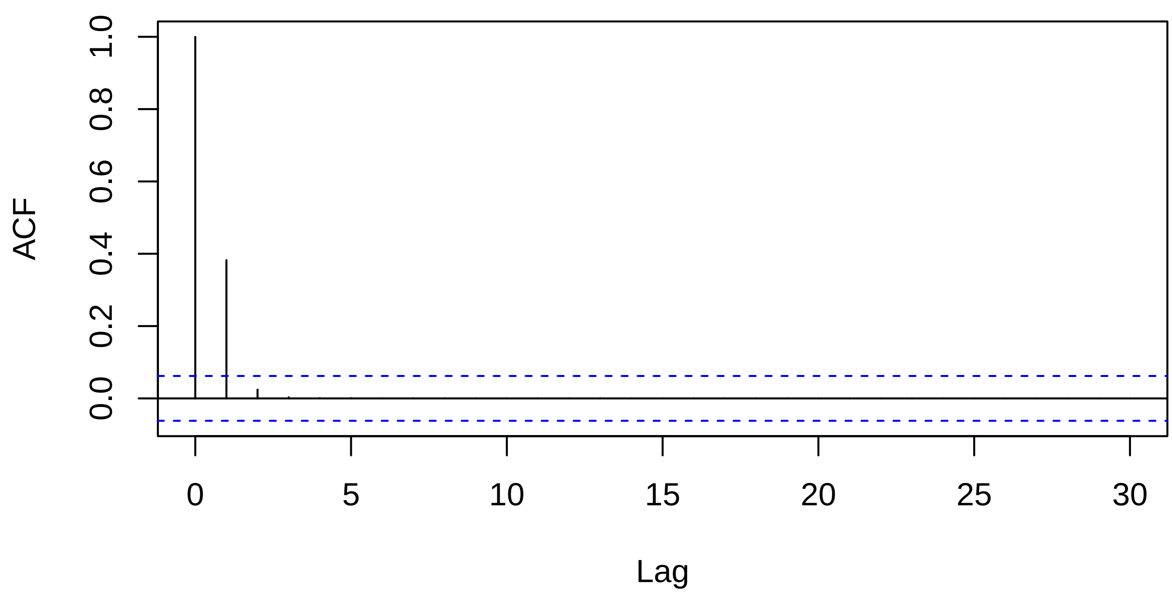

Note: Vertical lines indicate average estimates of beta values across hurricanes.

In **Table 3**, we see the estimated $\Sigma$ matrix. The estimated value of $\sigma^2$ is 31.53.

Table 3: Estimated Sigma Matrix

| | | | | |
|---|---|---|---|---|
| 0.4327031 | -0.0070313 | 0.0850938 | -0.0751975 | -0.0025897 |
| -0.0070313 | 0.1251760 | -0.0080831 | 0.0028970 | 0.0007579 |
| 0.0850938 | -0.0080831 | 0.5208588 | -0.1448723 | -0.0030389 |
| -0.0751975 | 0.0028970 | -0.1448723 | 0.2533229 | 0.0050132 |
| -0.0025897 | 0.0007579 | -0.0030389 | 0.0050132 | 0.0719811 |

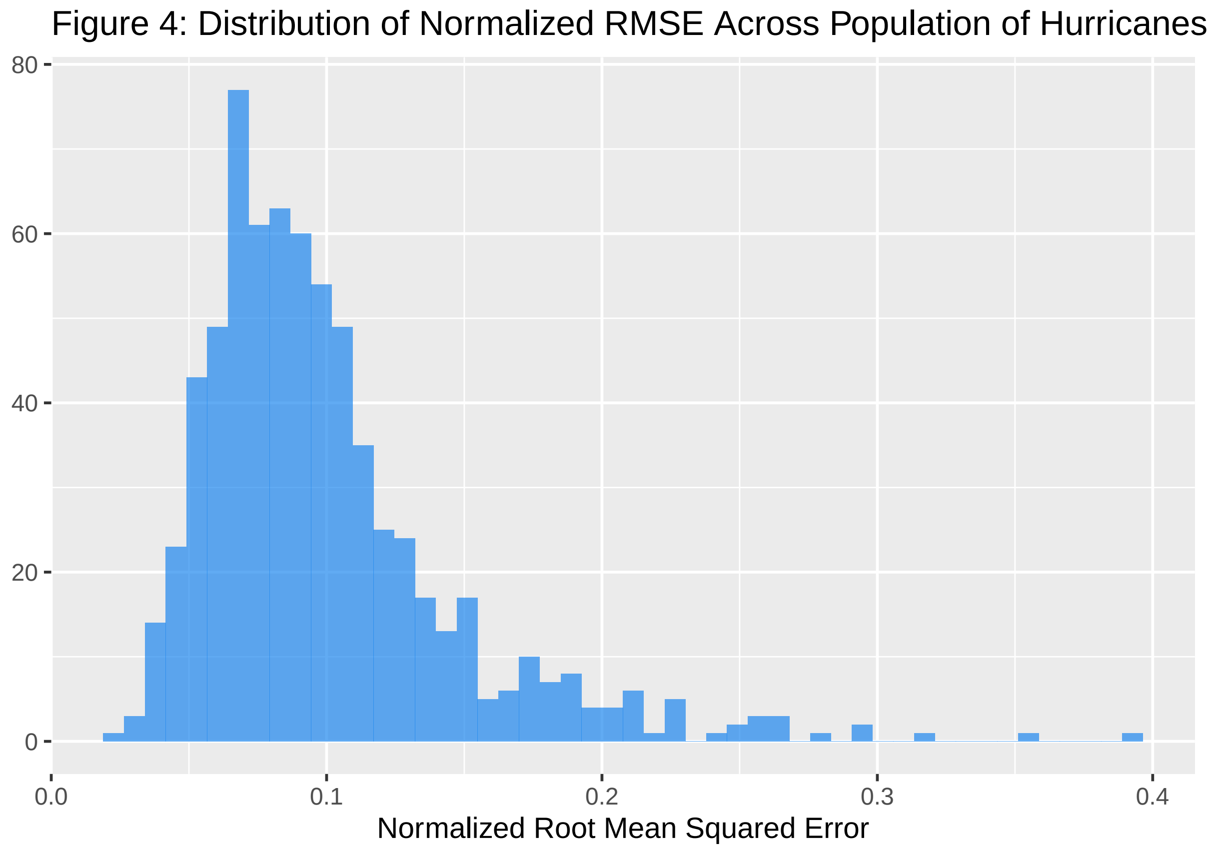
**Figure 3: Autocorrelation of Sigma^2 Markov Chain**




### 3.2. Bayesian Model Predictions

We can use the above estimates with the Bayesian model below and our predictor variables to estimate $\widehat{Y}_i(t+6)$ for each hurricane.

$$Y_i(t+6) = \beta_{0,i} + \beta_{1,i}Y_i(t) + \beta_{2,i}\Delta_{i,1}(t) + \beta_{3,i}\Delta_{i,2}(t) + \beta_{4,i}\Delta_{i,3}(t) + \epsilon_i(t)$$




Our model performed somewhat well predicting $Y_i(t+6)$ with most of our predictions yielding a normalized root-mean-squared-error (RMSE) of less than 0.1. However, there are a few instances of very poor predictions with a normalized RMSE of greater than 0.3.

### 3.3. Changes in Hurricanes over Time

We can use time variables to examine seasonal differences between hurricanes as well as changes over years. Let $x_{i,1}$ be the month of year when the i-th hurricane began, $x_{i,2}$ be the calendar year in which hurricane i began, and $x_{i,3}$ be the type of the i-th hurricane.

### 3.4. Analyzing Damage Caused by Hurricanes

## 4. Discussion

In **Figure 2** we see the distribution of $\beta_i$ estimates across the population of hurricanes. Some of these coefficients, like $\beta_1$ and $\beta_4$, do not differ much across the population, having a narrow range of values. This indicates the effect of wind speed at time $t$ and the effect of change in wind speed from time $t-6$ to $t$ on future wind speed do not vary much between hurricanes. The larger variance of $\beta_2$ and $\beta_3$ indicates the effect of change in position (latitude and longitude) on future wind speed varies more among the population of hurricanes. This makes sense because change in position can have a different effect on wind speed depending on the location of the hurricane. Hurricanes typically weaken closer to land, therefore change in location may have a different effect on wind speed depending on how close the hurricane is to land **[1]**.

Lastly, we see that $\beta_0$ has a larger magnitude of effect on future wind speed and a similar variance to $\beta_2$ and $\beta_3$. This indicates there are hurricane characteristics not captured in our model that have an influence on future wind speed, and these underlying characteristics differ among the population of hurricanes.

### 4.1. Limitations

There are a few limitations in the model estimation technique here. First, Bayesian models can be sensitive to the selection of prior distributions. The assumption of prior distributions in this scenario were non-informative or weak, however different prior distributions may change results. Additionally, the Gibbs sampling technique may fail under certain conditions. For instance, if the conditional posterior distributions result in extreme probability states, Gibbs sampling may become "trapped" in one of the high-probability conditions. Additionally, Gibbs sampling requires knowledge of conditional posterior distributions, however these distributions can be difficult to derive or intractable in some cases.

### 4.2. Strengths

The Gibbs sampling algorithm is relatively simple and easy to understand in concept. Additionally, we found the resulting Markov Chain stabilized relatively quickly, which can reduce the computational overhead required to perform this kind of estimation.

## References


1. Interaction between a hurricane and the land. Hurricanes. (n.d.). Retrieved May 7, 2022, from http://www.hurricanescience.org/science/science/hurricaneandland/