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 to determine which traits of hurricanes are more associated to damage and deaths.

2. Methods

2.1. Data

In this study, there are two data sets. The first one contains 702 hurricanes in the North Atlantic since 1950 including the location (longitude and latitude) and maximum wind speed every 6 hours for each hurricane. There are 8 variables and 22038 observations in the original data set. Due to the lagged nature of the data, a hurricane needs at least 3 observations for the model; we removed three hurricanes with fewer than three observations. We created four new predictor variables: wind lag, latitude change, longitude change, and wind change for the further steps.

The second data set contains the damages and deaths caused by 43 hurricanes along with 14 variables. In order to predict the damage and deaths caused by hurricane, we combine information from our estimated Bayesian model with this data set by the ID of hurricanes.

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 σ^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 ith 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 σ^2, $\boldsymbol{\beta}$ and Σ.

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

d is dimension of β.

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

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

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 Θ 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 β,

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

So we have the following joint posterior distribution of Θ:

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

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

Let $\tau = 1/\sigma^2$, then

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

Thus, σ^2 is from inverse-gamma distribution

$$(\sigma^2 \mid \boldsymbol{\beta}, \mathbf{B}, \Sigma^{-1}, Y) \sim \text{Inv-Gamma}(\frac{\sum_{i=1}^{n} m_i}{2}, \frac{1}{2} \sum_{i=1}^{n} (Y_i - X_i \beta_i^T)(Y_i - X_i \beta_i^T).$$

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

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

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

$$(\boldsymbol{\beta}_i \mid \boldsymbol{\beta}, \Sigma^{-1}, \sigma^2, Y) \sim \mathcal{M}VN(V_i^{-1} U_i, V_i^{-1}).$$

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

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

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

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

Finally, parameter Σ has the conditional posterior:

$$\pi(\Sigma^{-1}|\boldsymbol{\beta},\mathbf{B},\sigma^2,Y) \propto |\Sigma|^{-(d+1)} exp(-\frac{1}{2}tr(\Sigma^{-1})|\Sigma|^{-\frac{n}{2}} exp(-\frac{1}{2}\sum_{i=1}^{n}(\beta_i-\beta)\Sigma^{-1}(\beta_i-\beta)^T) \tag{5}$$

$$\propto |\Sigma^{-1}|^{d+1+\frac{n}{2}} exp(-\frac{1}{2}\left[tr(\Sigma^{-1}) + tr(\sum_{i=1}^{n}(\beta_i-\beta)\Sigma^{-1}(\beta_i-\beta)^T)\right] \tag{6}$$

$$\propto |\Sigma^{-1}|^{3d+3+n-d-1} exp(-\frac{1}{2}tr(\left[I + \sum_{i=1}^{n}(\beta_i-\beta)^T(\beta_i-\beta)\right]\Sigma^{-1})) \tag{7}$$

Thus,

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

where $v = 3d + 3 + n$, 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 $j = 1, 2, ..., n$:

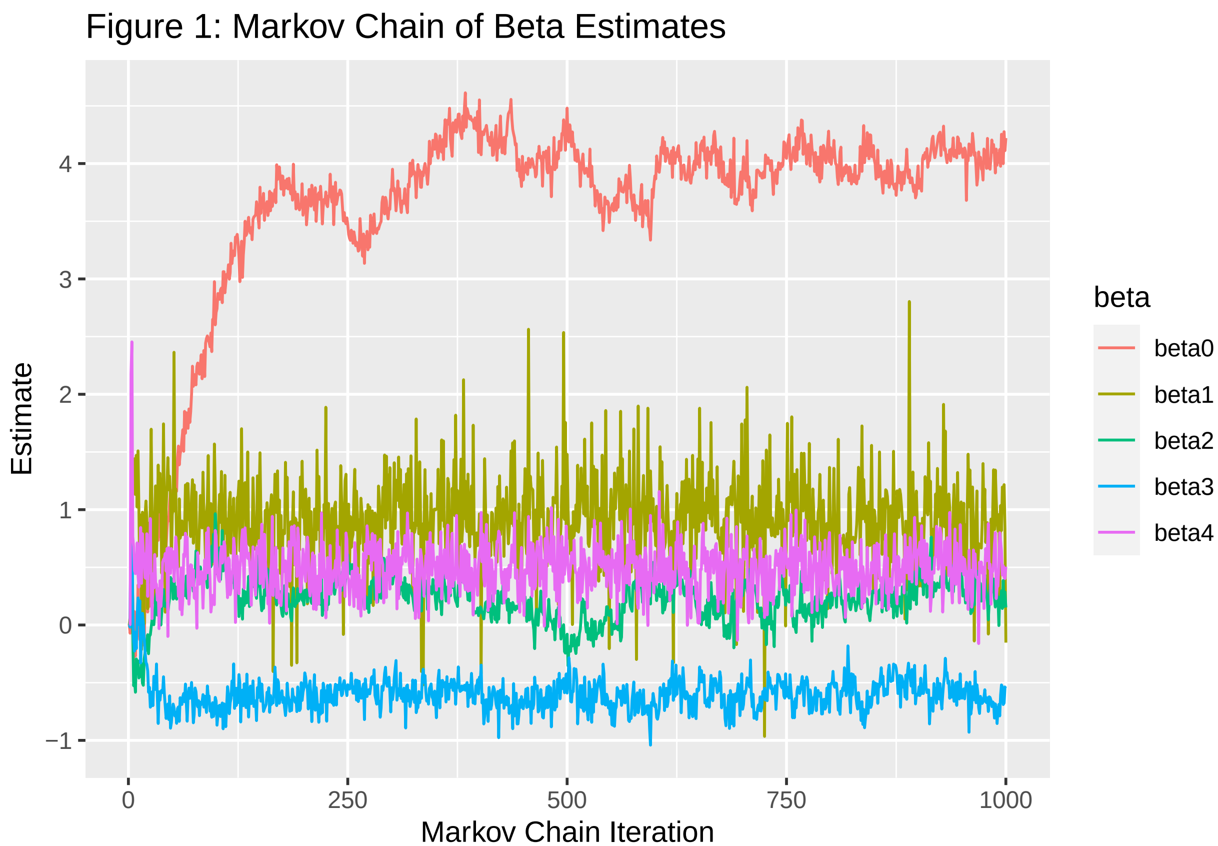
1. Generate $\boldsymbol{\beta}_j$ from $\pi(\boldsymbol{\beta} \mid \Sigma = \Sigma_{j-1}, \sigma^2 = \sigma^2_{j-1}, \mathbf{B} = \mathbf{B}_{j-1})$;
2. Generate Σ_j from $\pi(\Sigma \mid \boldsymbol{\beta} = \boldsymbol{\beta}_j, \sigma^2 = \sigma^2_{j-1}, \mathbf{B} = \mathbf{B}_{j-1})$;
3. Generate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

to yield Θ_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. Example code for this algorithm cane be seen in **Appendix A**.

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 β_0 converges to a value approximately equal to 4, β_1 fluctuates about 1, β_2 and β_4 are between 0 and 1, and β_3 is less than zero.



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.

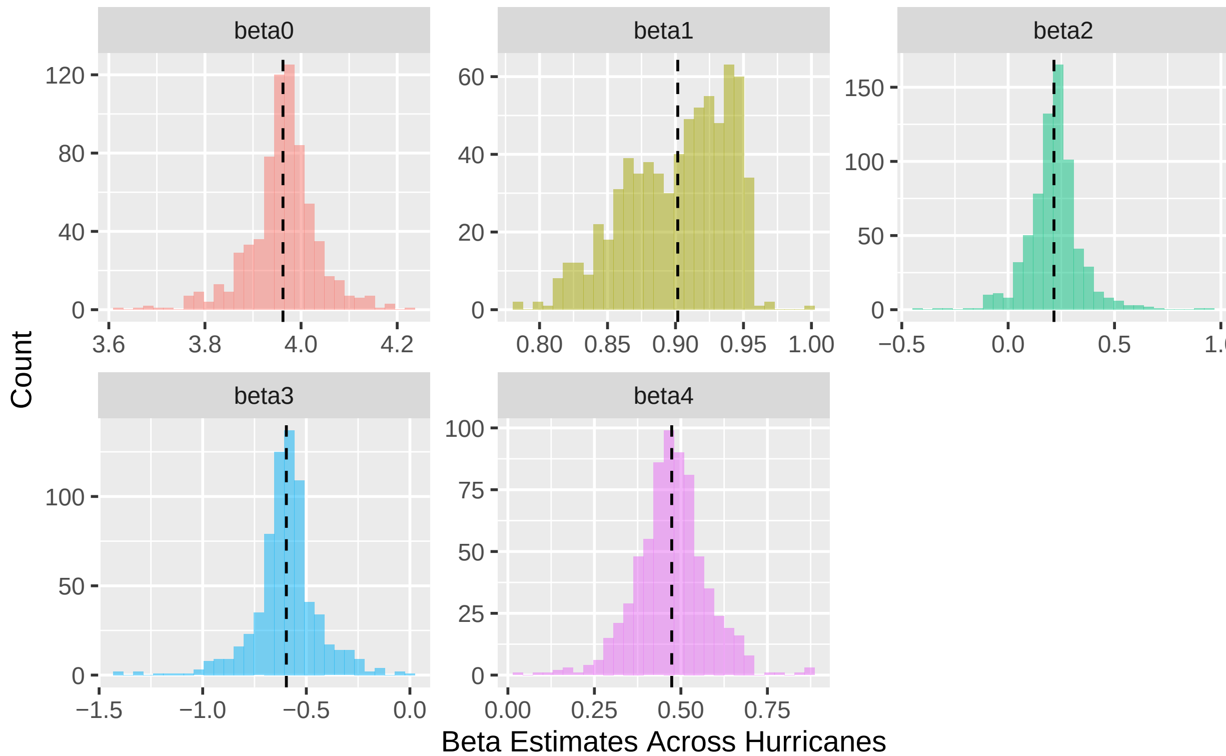
Table 1: Beta Parameter Estimates

beta	avg_est
beta0	3.9560971
beta1	0.9289637
beta2	0.2217102
beta3	-0.6042475
beta4	0.4780620

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

i	beta0	beta1	beta2	beta3	beta4
1	3.897373	0.9473008	0.0378833	-0.7136875	0.4879133
2	3.949851	0.9146437	0.2291781	-0.5094849	0.5806599
3	3.819281	0.9411540	0.2326227	-0.4532977	0.4158447
4	3.932027	0.9526669	0.1557606	-0.5055851	0.5674213
5	3.796049	0.9382914	0.0935038	-0.4641068	0.5271523
6	3.909442	0.9518325	0.0966168	-0.9006977	0.5453941

Figure 2: Beta Estimates Across Population of Hurricanes



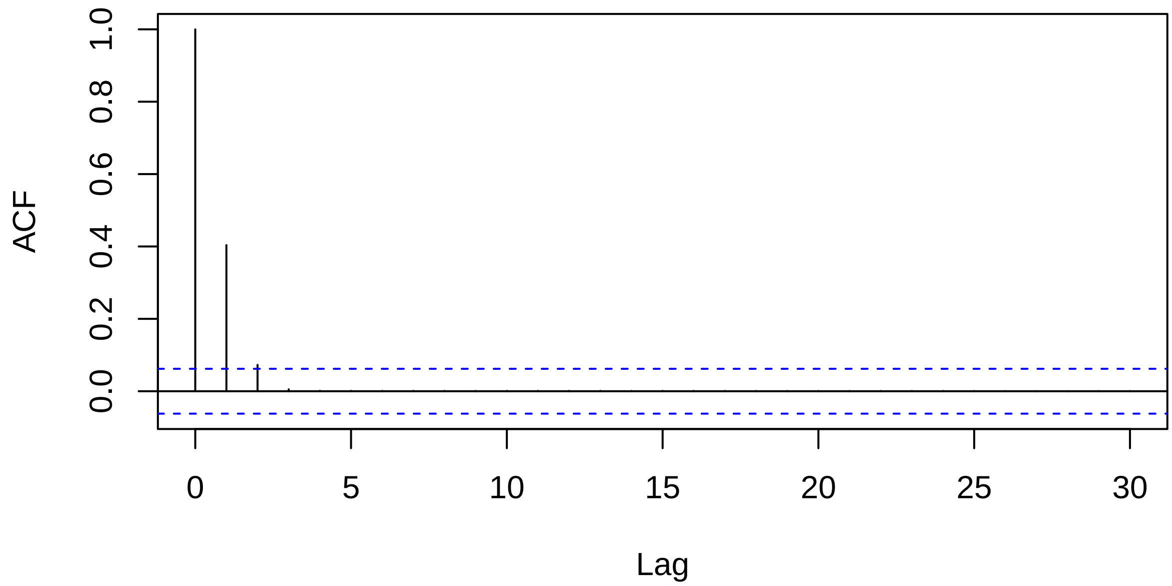
Note: Vertical lines indicate means.

In **Table 3**, we see the estimated Σ matrix. The estimated value of σ^2 is 31.92.

Table 3: Estimated Sigma Matrix

0.1956703	-0.0024467	0.0194313	-0.0155578	-0.0005066
-0.0024467	0.1390597	-0.0007349	-0.0003993	-0.0009256
0.0194313	-0.0007349	0.2306833	-0.0677137	-0.0017668
-0.0155578	-0.0003993	-0.0677137	0.1913711	0.0029405
-0.0005066	-0.0009256	-0.0017668	0.0029405	0.0712005

Figure 3: Autocorrelation of Sigma^2 Markov Chain

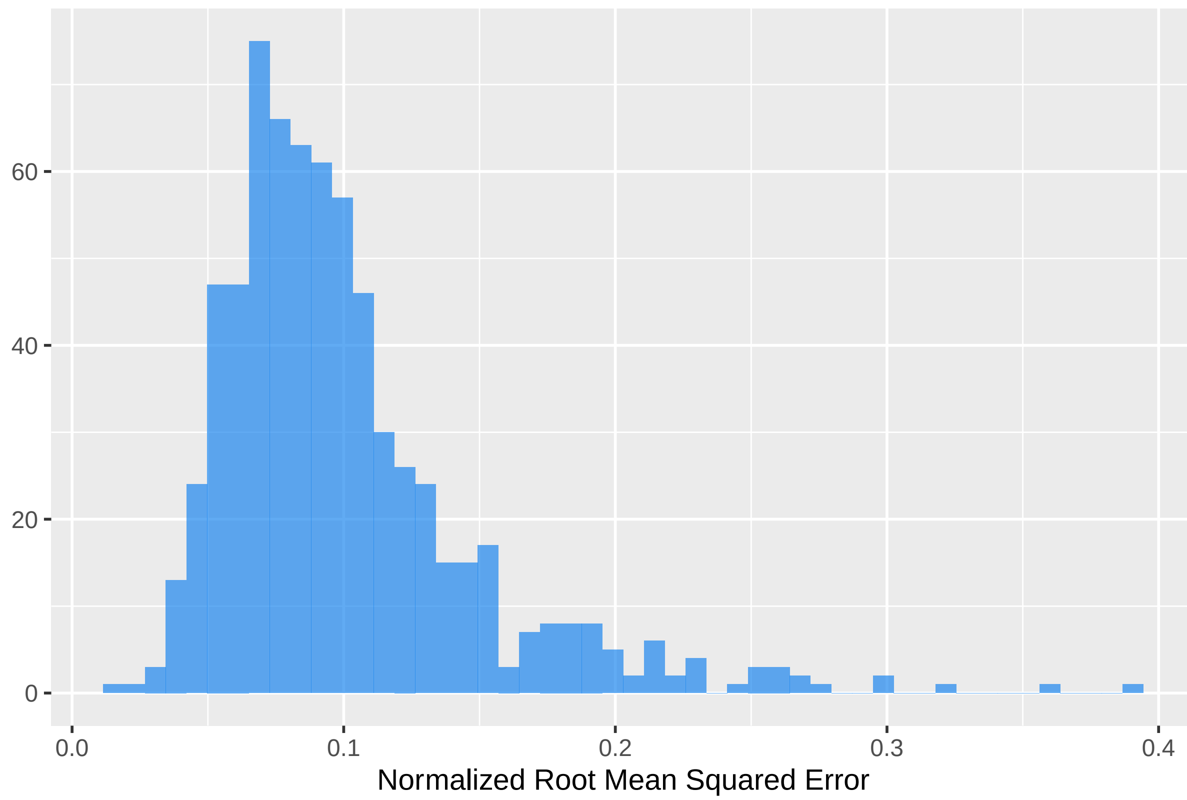


3.2. Bayesian Model Predictions

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.

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

Figure 4: Distribution of Normalized RMSE Across Population of Hurricanes



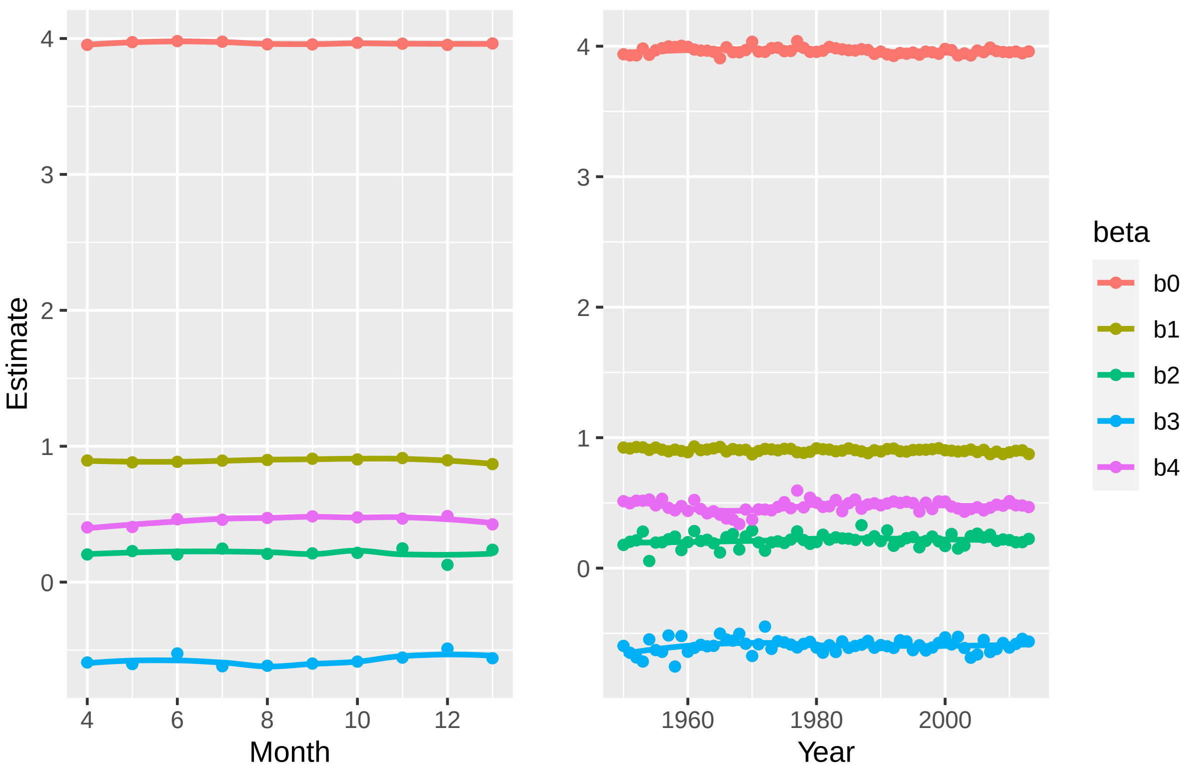
Our model performed somewhat well predicting $Y_i(t+6)$ with most of our predictions yielding a normalized root-mean-squared-error (RMSE) of less than 0.1. However, there are a few instances of very poor predictions with a normalized RMSE of greater than 0.3.

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. Using code similar to that in **Appendix B**, we performed linear regression with $x_{i,1}, x_{i,2}$, and $x_{i,3}$ as predictors of each $\boldsymbol{\beta}_i$ coefficient. For these regressions, our data begins in April and progresses through January to align with typical "hurricane seasons". Note there were no hurricanes observed in February and March.

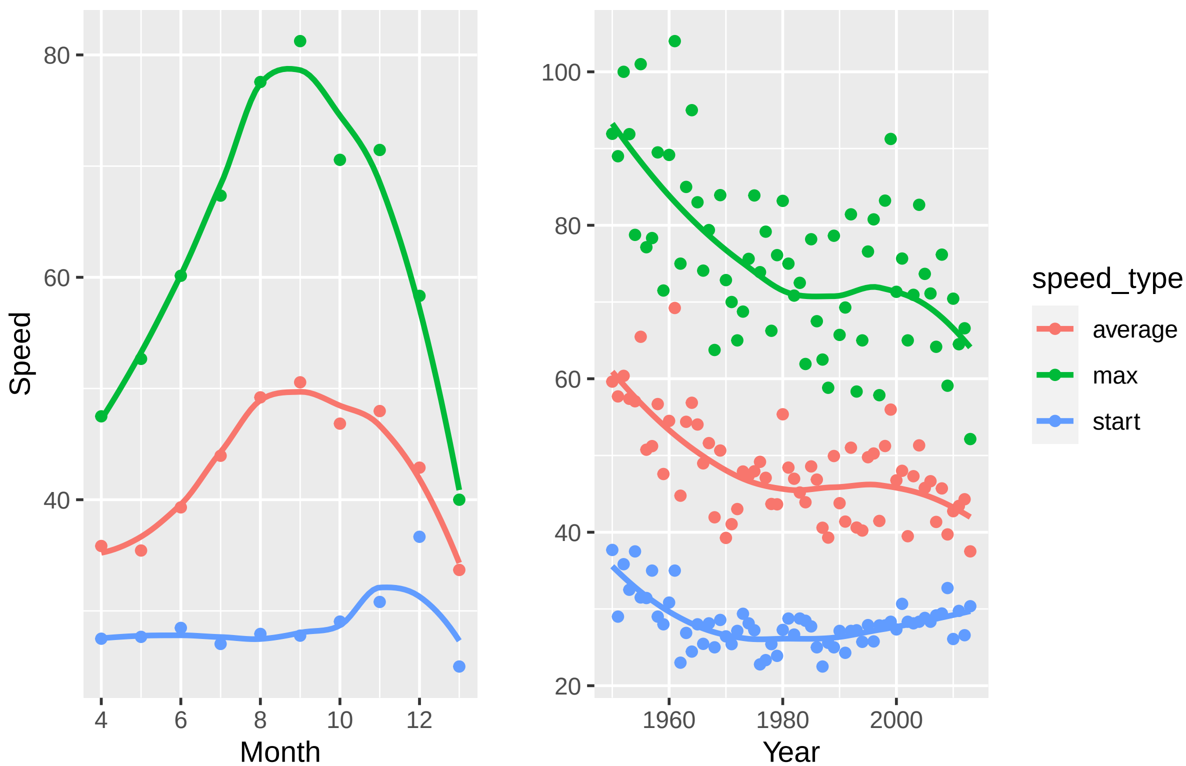
Before performing regressions, we explored the overall trends of β coefficients over time. In **Figure 5**, we can see there are no clear changes in β values across months or across years.

Figure 5: Beta Estimates Across Months and Years



Accordingly, we saw no significant relationship between time predictors and $\beta_0, \beta_2, \beta_3$ or β_4. There was a significant relationship between year and β_1, the effect of wind speed at time t, however the estimated coefficient is approximately -3.7×10^{-4}, which is quite small. We did see a significant relationship between β_2 and β_3, the effect of change in latitude and change in longitude, and subtropical hurricanes compared to tropical storms. This could mean that the wind speed of subtropical hurricanes is more sensitive to changes in location than for tropical storms. In conclusion, we did not find any significant trends between our estimated β_i coefficients and time - years and months. We could interpret this to mean that the effects of change in position, change in wind speed, and previous wind speed do not significantly change for hurricanes at different times.

Figure 6: Hurricane Speed Across Months and Years



We did see changes in wind speed across months and years (see **Figure 6**). For starting wind speed of hurricanes, we saw significant relationships with the month of December compared to April. All months besides January showed increased starting wind speed compared to April. Extra-tropical hurricanes showed a statistically significantly higher starting speed compared to tropical storms, whereas disturbances and not rated hurricanes showed statistically significantly lower starting speed.

For max speed, the later months (July through December) tend to have higher max speeds compared to April, however these are not statistically significant. Despite what some may claim - that wind speed has increased over the years - we find a statistically significant relationship between year and max wind speed indicating that wind speed decreases as year increases. We also find max wind speed is statistically significantly lower for sub-tropical hurricanes compared to tropical storms.

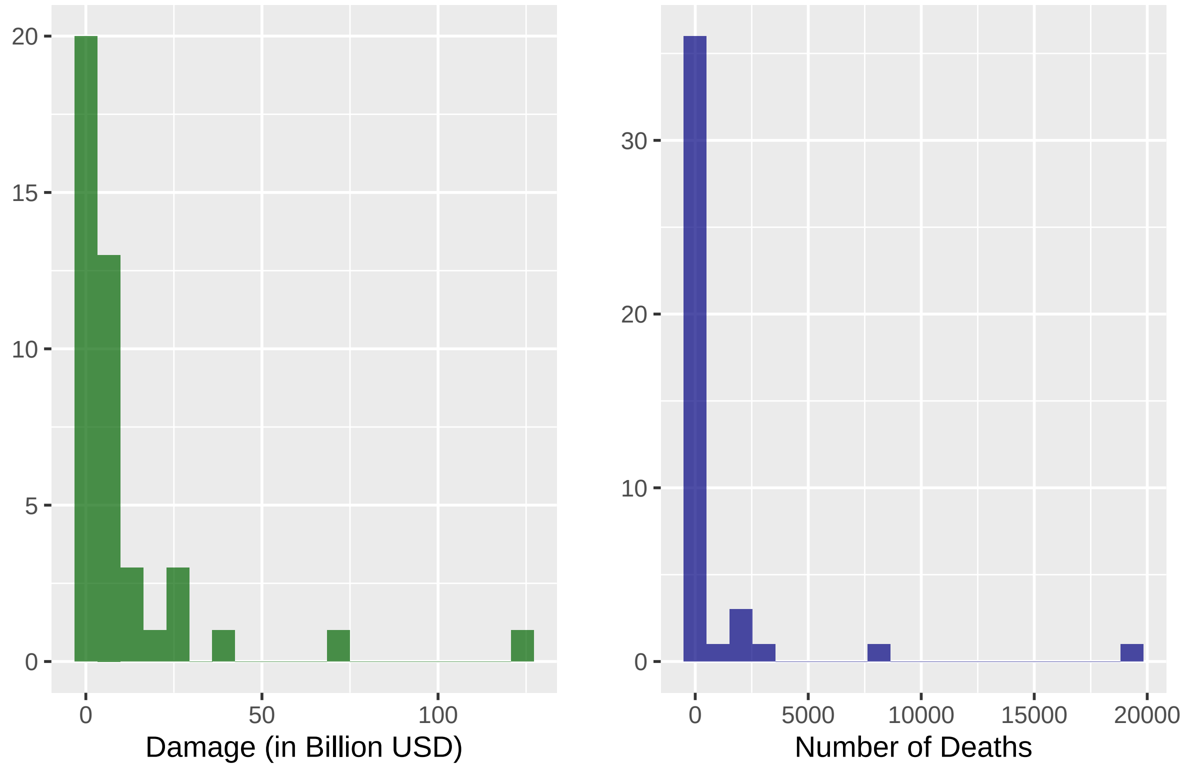
Lastly, we find similar results for average wind speed over time. The months of June through December tend to have higher wind speed compared to April, however this is not statistically significant. Again, we see a statistically significant result that average wind speed tends to decrease as year increases. These trends can be seen in **Figure 6**.

3.4. Analyzing Damage and Deaths Caused by Hurricanes

For analyzing damage and deaths caused by 43 hurricanes, we fit the Elastic Net model to perform variable selection.

According to **Figure 7**, we find that damage is not normally distributed; therefore, we took the log of damage to be our response variable, which makes the distribution closer to normal.

Figure 7: Damage and Deaths Caused by Hurricanes



According to the result, season (or year), maximum recorded wind speed of the hurricane, and percentage affected population that resides in the United States are three statistically significant characteristics related to damage caused by hurricanes. When season, max speed, and percent of USA increase, the average of damage is expected to increase. The average damage increases by 1.0467, 1.0226, 1.0078 with one unit increase in season, max speed and percent of USA, respectively. We see mean wind speed is positively associated with average damage while β_4, or the effect of previous wind speed change on future wind speed, is negatively associated. We also see that not rated hurricanes and tropical storms tend to cause less damage than other types of hurricanes. However, these coefficients are not statistically significant. The mean square error of the model is 0.66. Regressions were performed using code similar to that in **Appendix C**.

For deaths, although it is count data and could only be integers, we also log the response to ensure the distribution is closer to normal. For deaths, we centralized all the parameters to keep consistency. According to model results, max pressure and percent poor are two important parameters. When max pressure decreases, the average number of death increases, which makes sense because lower air pressure tends to be associated with more intense hurricanes [**1**]. The average number of deaths increases 1.0386 with one unit increase in percent poor. The mean square error is 2.13. Regressions were performed using code similar to that in **Appendix C**.

4. Discussion

In **Figure 2** we see the distributions of β_i estimates across the population of hurricanes. Some of these coefficients, like β_1 and β_4, have a relatively 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 β_2 and β_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 [**2**].

Lastly, we see that β_0 has a larger magnitude of effect on future wind speed. 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. As an example of something not captured in this model, wind temperature tends to have influence on the intensity of hurricanes and might play a role in this model.

When examining trends across time, we did not see changes in β_i coefficients in different months or years. This indicates the effects of change in position, change in wind speed, and previous wind speed do not significantly change for hurricanes at different times of year or in different years.

We found that season (or year), max wind speed, mean wind speed, and percent of residents in USA to be positively associated with damage caused by hurricanes. These results align with our understanding of the nature of hurricanes - higher wind speeds can lead to more damage. Season (or year) is also positively associated with damage, which could mean that hurricane severity is increasing or it could be a reflection of an increase in value of property that can be damaged in affected areas. For instance, if a beach community receives investment and builds larger, more expensive buildings, a powerful hurricane could cause more damage simply because more property is there to be damaged. We also saw that β_4, the effect of change in wind speed on future wind speed, and type not rated and tropical storm tended to be negatively associated with damages.

We found that max pressure of a hurricane was negatively associated with deaths. When air pressure decreases, the hurricane tends to intensify with higher winds; therefore a decrease in pressure would indicate a stronger hurricane and a higher potential for death. We saw the percentage of the affected population residing in low GDP countries (i.e., GDP per capita $\leq$ 10,000) was positively associated with deaths. This could be caused by poorer communities having fewer safety measures in place in the event of a hurricane or residents of these lower income areas may have less access to evacuation resources in the event of a hurricane. These results are important to consider when creating evacuation and storm-response policies in order to limit loss of property, or more importantly, loss of life.

4.1. Limitations

There are a few limitations in the model estimation technique. First, Bayesian models are 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.

When evaluating prediction accuracy, we trained our model on the full data set and then compared predictions of that data set to the actual values. This may more closely approximate training error rather than test error. In order to build a more robust model, we would need to split our data into training and testing data sets before analysis. One way to do this would be to partition the data for each hurricane, using the majority of observations to train parameters and the remaining data to evaluate predictions.

4.2. Strengths

One important decision in Bayesian analysis is the assumption of prior distributions. In this application, we assume uninformative or weak prior distributions, so we do not assume too much prior knowledge. Additionally, we found the resulting Markov Chain stabilized relatively quickly, which can reduce the computational overhead required to perform this kind of estimation. Our results from sections 3.3 and 3.4 are easily interpreted due to the simplicity of models used, and our results align with what we know about hurricanes from outside sources (see References).

4.3. Future Work

In the future, it would be interesting to include more variables that may have effects on hurricanes such as air temperature and time over water. Additionally, model evaluation could be improved by splitting data into training and testing sets. When examining damages caused by hurricanes, it would be interesting to examine damage per capita of the affected regions or damage per total asset value in affected areas. Similarly, it would potentially be more informative to examine death rates of hurricanes rather than the raw number of deaths.

Contributions

T.M. drafted posterior distribution of parameters and conducted evaluation of MCMC model performance. Y.P. revised posterior distribution, derived conditional distributions, and analyzed trends of coefficients and hurricane speeds across time. T.M. with Y.P. designed MCMC algorithm in R using Gibbs sampling method to estimate parameters. J.L., W.C., and X.M. built models to analyze damages and deaths caused by hurricanes. T.M., J.L., W.C., X.M., Y.P. drafted the report with T.M. and Y.P. editing it.

References

1. Hurricane Life Cycle. Hurricanes. (n.d.). Retrieved May 7, 2022, from http://www.hurricanescience.org/science/science/hurricanelifecycle/

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

Appendix

Appendix A

```
gibbs_function <- function(iter, start, data){
  # lists to store results
  beta_list <- list()
  sigma_list <- list()
  sigma2_list <- list()
  B_list <- list()
  # starting values
  beta_list[[1]] <- start$beta
  sigma_list[[1]] <- start$sigma_m
  sigma2_list[[1]] <- start$sigma2
  B_list[[1]] <- theta$B
  # for loop to iteratively generate Markov Chain
  for (i in 2:iter){
    beta_list[[i]] <- beta_dist(sigma = sigma_list[[i-1]], B = B_list[[i-1]][[1]])
    sigma_list[[i]] <- sigma_m_dist(beta = beta_list[[i]][[1]], B = B_list[[i-1]][[1]])
    sigma2_list[[i]] <- sigma2_dist(B = B_list[[i-1]][[1]], x = data$x, y = data$y)
    B_list[[i]] <- B_dist(x = data$x, y = data$y, sigma2 = sigma2_list[[i]],
                          sigma_m = sigma_list[[i]], beta = beta_list[[i]][[1]])
  }

  return(list(beta = beta_list, sigma_m = sigma_list,
             sigma2 = sigma2_list, B = B_list))

}
```

Appendix B

```
# beta0 regression
reg(y = beta0, month, year, type)
# beta1 regression
reg(y = beta1, month, year, type)
# beta2 regression
reg(y = beta2, month, year, type)
# beta3 regression
reg(y = beta3, month, year, type)
# beta4 regression
reg(y = beta4, month, year, type)
```

```r
# start_speed regression
reg(y = start_speed, month, year, type)
# max_speed regression
reg(y = max_speed, month, year, type)
# avg_speed regression
reg(y = avg_speed, month, year, type)
```

Appendix C

```r
# model training using caret package
train(log(damage)~.,
      data = hurricane,
      method = "glmnet",
      tuneGrid=expand.grid(alpha = seq(0,1,length=21),
                           lambda = exp(seq(3,-5,length=100))),
      trControl = trainControl(method = "repeatedcv", number = 10, repeats = 5))
# fitting model for damage with selected predictors
lm(log(damage) ~ season + nature + maxspeed + meanspeed + percent_usa + beta4,
   data = hurricane)
```