Hurricanes and Climate Change: A Bayesian approach

By: Madeline Abbott, Aidan Teppema, Daisy Cho

December 18, 2017

Why Hurricanes?

In the past summer, two devastating category 4 hurricanes made landfall in North America. In August, people in Central America, Texas, Louisiana and other states lost their homes and loved ones to Hurricane Harvey. Soon after in early September, Hurricane Irma destroyed the coast of Florida.

Climate researchers suggest that although climate change may not directly increase the frequency of hurricanes, those hurricanes that do occur could be much stronger than those that have occurred in the past. They suggest that a combination of changes in the ocean and atmosphere may drive patterns in hurricane severity (Gray 2017).

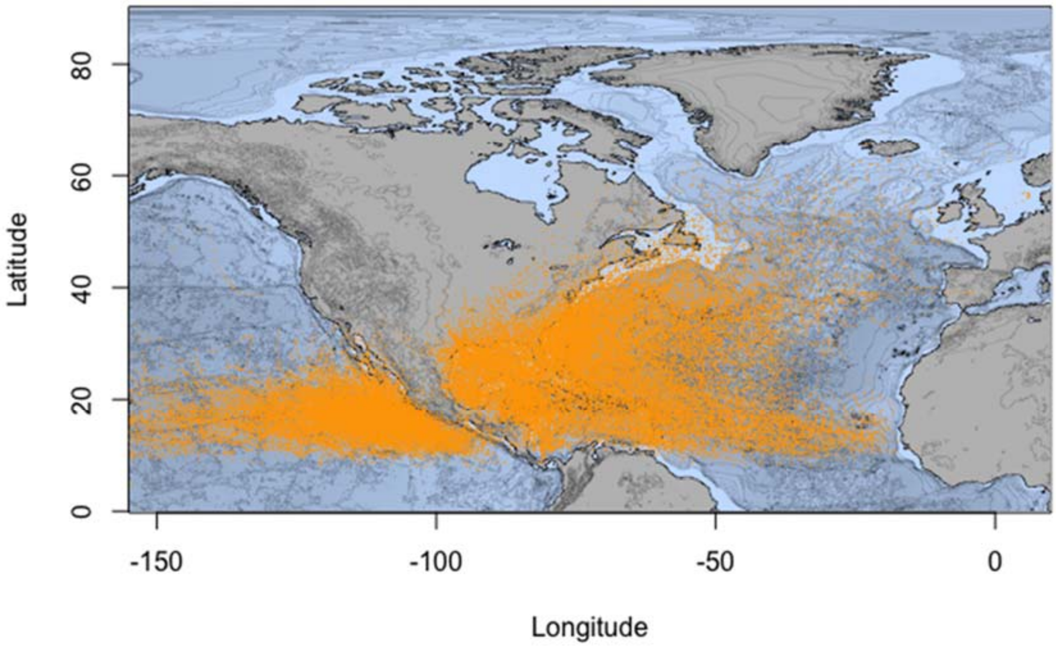
In the past, researchers have attempted to model annual hurricane counts using measures of ocean pressure oscillations (Elsner & Jagger 2006). Ocean oscillations are particularly useful at modeling hurricane frequency at a seasonal level because these measures fluctuate within each year.

Prompted by recent hurricanes, we attempt to model trends in hurricane characteristics using Bayesian regression. Because of the severity of the destruction from recent hurricanes, we model the average wind speed of hurricanes. Then, because of how quickly Hurricane Irma came after Hurricane Harvey, we investigate the frequency of annual hurricanes. Finally, we analyze the average durations of the hurricanes. For example, Hurricane Harvey lasted for 17 days and we wanted to see if the durations of hurricanes can be modeled. Looking at the wind speed, frequency, and durations of past hurricanes, we may be able to predict future hurricanes and be more prepared for them.

Figure 1:

Map of Hurricane Locations (1950-2008):

Graph of all hurricane tracks in Northern hemisphere since 1950



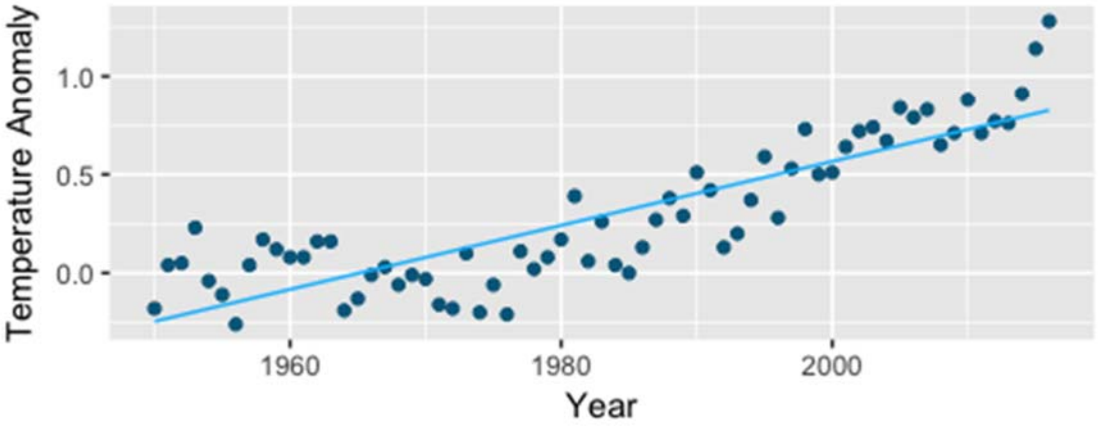
The Datasets

To measure climate changes, we looked at four different data sets. The first data set is our main data set on hurricanes, provided by the Homeland Infrastructure Foundation Level Data Committee (HIFLD, 2009). This dataset recorded the hurricane's wind speed, dates of occurrence, location, pressure, and category. This dataset includes all hurricanes and storms from 1850 to 2008 in both the North Atlantic and Pacific basins. We chose to exclude all hurricanes occurring before 1950 because the hurricane naming system was not standardized until 1950. Hurricanes are plotted in Figure 1.

Our predictor variables are temperature anomalies provided by NASA, CO_2 provided by the National Oceanic and Atmospheric Administration (Dlugokencky & Tans, 2017), and the North Atlantic Oscillation Index (NAOI) which is provided by the National Center for Atmospheric Research (NCAR, 2003).

Figure 2: Temperature Anomaly by Year

Average temperature anomaly in Northern hemisphere by year. Tempera anomaly is defined as the deviation of the absolute temperature from the reference temperature. The reference temperature is the average of 30 years of temperature data.

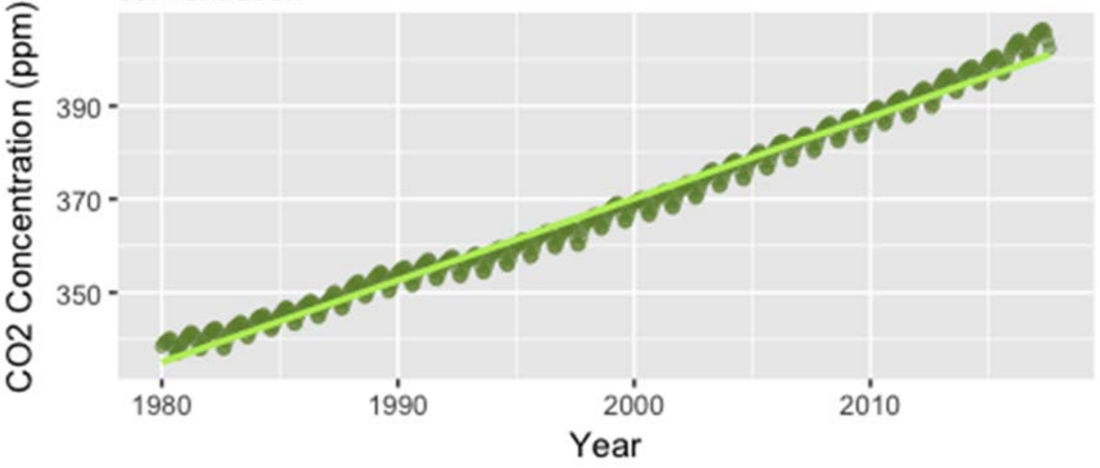


We chose to use temperature anomalies, rather than absolute temperature, as a predictor in our model due to the reliability of the measurements across space and time. Absolute temperatures, or actual temperatures measured in current time, are more prone to be affected by extraneous variables such as the height of the temperature measurement tower, whereas temperature anomaly are not as affected by these extraneous variables. Temperature anomalies over time are plotted in Figure 2. From the plot, we can see a general growth trend from a temperature anomaly measurement of -0.18 at 1950 (meaning a cooler than average year) to 1.28 at 2016 (indicating a warmer than average year).

CO_2 also shows an upward trend over the years. CO_2 concentration showed fluctuation within each year due to seasonal differences because during the summer months when plants are growing, photosynthesis outweighs respiration and as a result, CO_2 concentrations dip during the summer and increase slightly during the winter. More importantly, however, we observe an overall increase in atmospheric CO_2 concentration (see Figure 3).

Figure 3: CO2 Concentration by Year

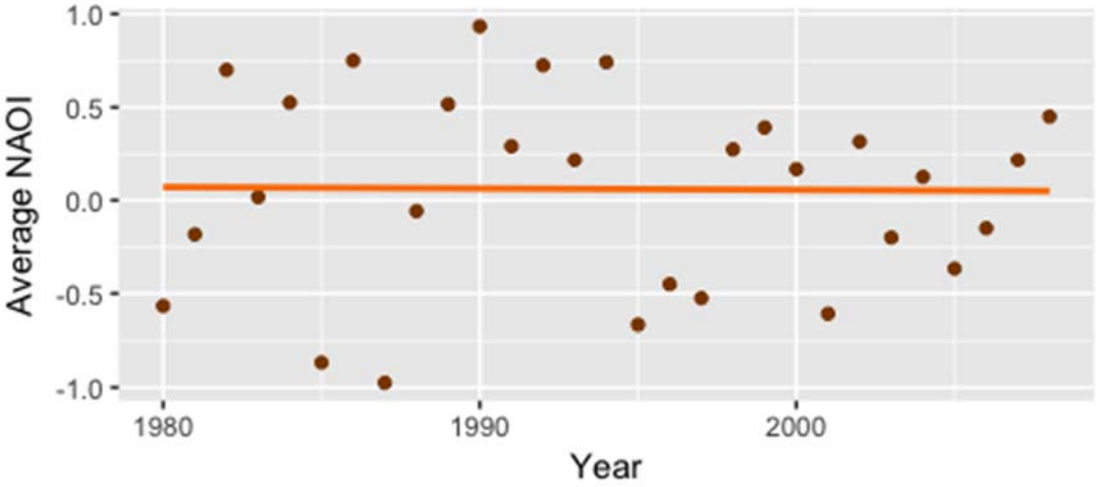
Atmospheric CO2 concentration by year. CO2 concentration is measured in parts per million (ppm). Oscillations are due to seasonal differences in CO2 concentrations. Notice overall upward trend in CO2 concentration



Our final predictor variable is North Atlantic Oscillation Index (NAOI). NAOI indicates fluctuations in atmospheric pressure at sea level in the Northern Atlantic ocean. High NAOI values have been correlated with increased intensity and frequency of storms in the North Atlantic Ocean (Hurricane Society n.d.).

Figure 4: North Atlantic Oscillation Index by Year

Shows NAOI by year, averaged across all months. NAOI indicates fluctuations in atmospheric pressure at sea level in the North Atlantic Ocean and effects the severity and direction of ocean storms.



While we see no strong trend in NAOI over the years (see Figure 4), we include NAOI as a predictor variable in our model with the hope that it will explain some of the variability in hurricane severity, frequency, and duration between years.

Before attempting to understand the relationship of these variables of CO$_2$ concentration, temperature anomalies, and NAOI with hurricane characteristics, we first looked at the relationship of these variables with each other. We tested for collinearity between these three variables and found that CO_2 and temperature anomalies were highly correlated (R = 0.8604). Thus, we excluded temperature from all models using CO$_2$ concentration as a predictor.

Creating Our Models

In order to understand possible trends in future hurricanes, we attempt to model patterns in hurricane severity, frequency, and duration. We use two different Bayesian models: a Normal-Normal model and Poisson regression model. A Normal-Normal model assumes a Normal distribution, or a "bell curve", with the density curve described by its mean and standard deviation, for both the prior and the posterior.

$$X \sim N(\mu, \sigma^2)$$

A Poisson regression model uses a Poisson distribution, which expresses the probability of a given number of events occurring in a fixed interval of time assuming these events occur with a known constant rate. It also assumes that the log of the expected value(equivalent to the mean) can be modeled by unknown parameters.

$$X \sim Pois(\lambda)$$

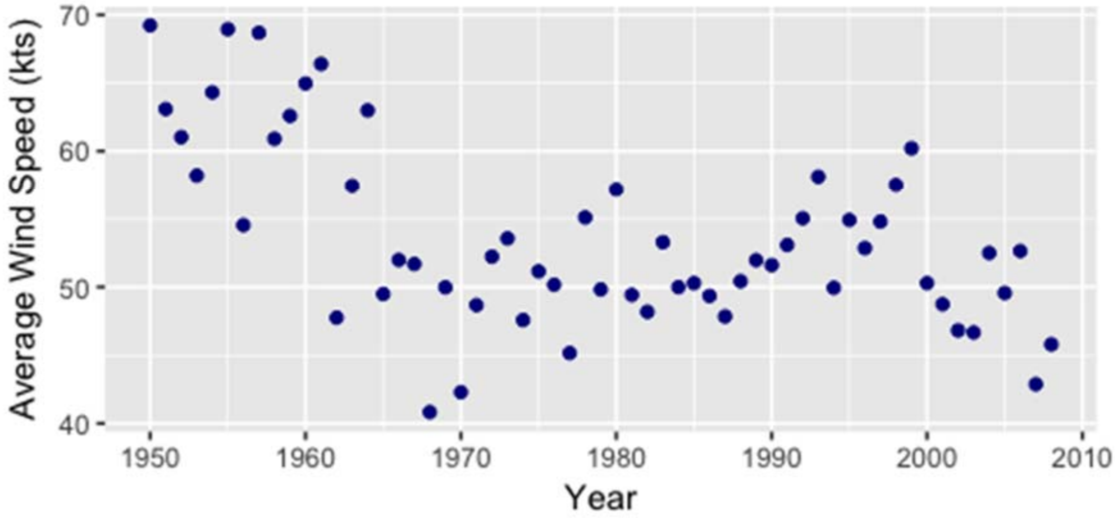
For all our models we ran Markov chains in order to simulate the posterior results after running our models in rjags (Plummer, 2016).

Hurricane Wind Speed

One notable characteristic of the hurricanes that occurred over the past summer was their severity and the extreme damage that they caused when they hit land. To further understand this possible pattern, we first modeled hurricane severity through time to see if there were any obvious trends throughout the years using NAOI and CO$_2$ concentration as predictors. First, in order to visually see if there was an obvious trend in hurricane severity, we graphed wind speed by year (Figure 5). We see that overall average wind speed of hurricanes is decreasing. Next, we attempt to model this trend.

Figure 5: Average Wind Speed by Year

Average wind speed (in knots) of all hurricanes occurring in the same year over time (in years).



To create our model, we used a Normal-Normal model to attempt to model wind speed with CO_2 and NAO over time (we excluded temperature anomaly from our model because it was highly collinear with CO_2). We used the highest wind speed for each hurricane as the variable for severity, since this is the measure that categorizes hurricanes.

All of the priors for this model are vague because, prior to creating and running our models, we didn't know much about the impact of NAO and CO_2on hurricane wind speed.

We define our model as follows:

Let:

Y_t = average highest wind speed per year t

X_{1t} = yearly CO_2 concentration in year t

X_{2t} = NAO index in year t

$$Y_t \sim N(\beta_0 + \beta_1 X_{1t} + \beta_2(X_{2t}), \tau^{-1})$$

$$\beta_0 \sim N(0, 100^2)$$

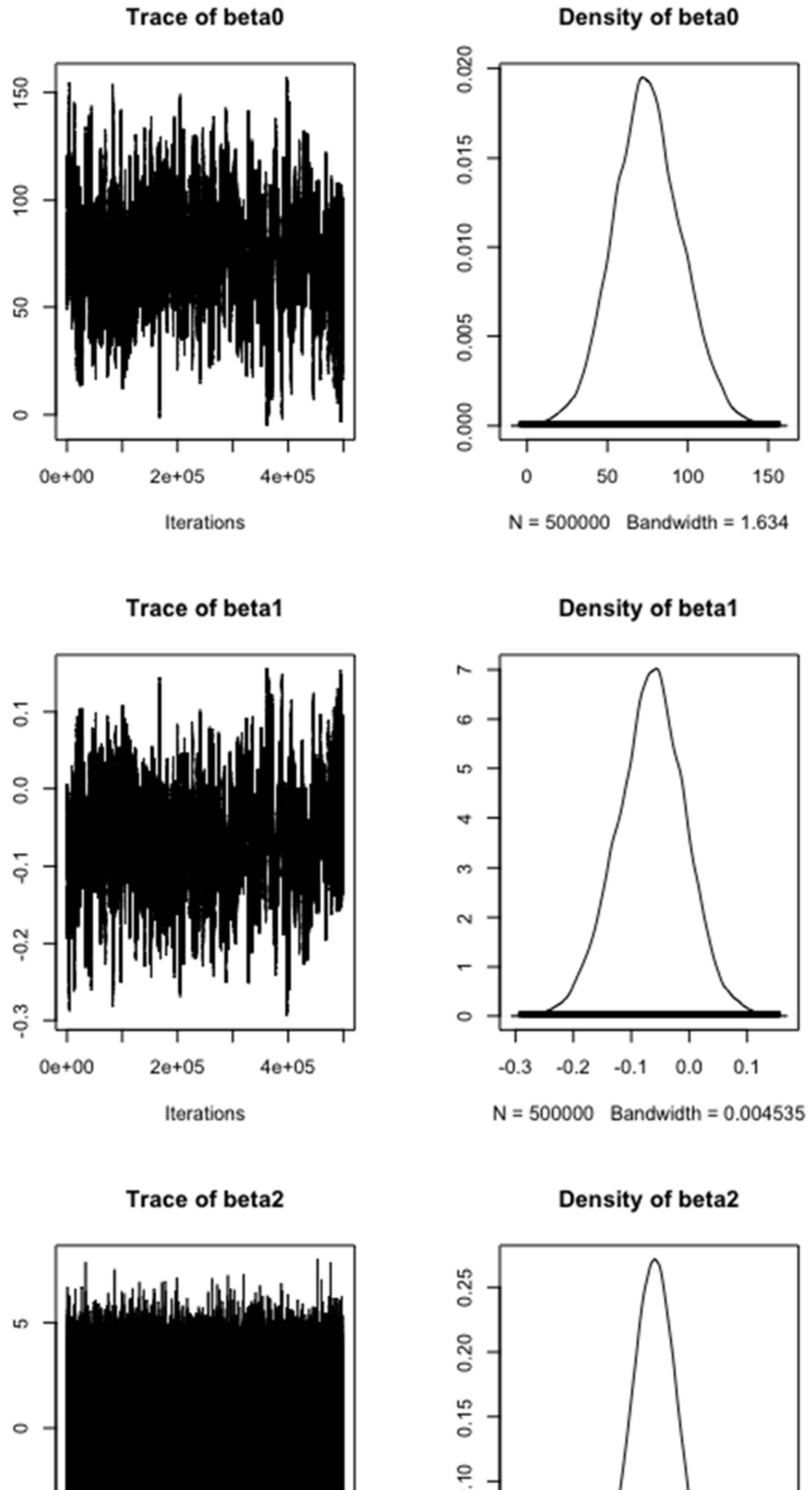
$$\beta_1 \sim N(0, 100^2)$$

$$\beta_2 \sim N(0, 100^2)$$

$$\tau \sim \Gamma(0.001{,}0.001)$$

Figure 6:

Trace plots and density plots of wind speed parameters.



Trace of beta0
150
100
50
0
0e+00
2e+05
4e+05
Iterations
Density of beta0
0.020
0.015
0.010
0.005
0.000
0
50
100
150
N = 500000   Bandwidth = 1.634
Trace of beta1
0.1
0.0
-0.1
-0.2
-0.3
0e+00
2e+05
4e+05
Iterations
Density of beta1
7
6
5
4
3
2
1
0
-0.3
-0.2
-0.1
0.0
0.1
N = 500000   Bandwidth = 0.004535
Trace of beta2
5
0
Density of beta2
0.25
0.20
0.15
.10

The approximated posterior parameter values and their density plots are summarized in Figure 6. The density plots show a smoothed histograms of all the posterior estimates of parameter values in the trace plot. After 500,000 iterations of the simulation, the running mean plots of the parameters converged. The summaries (means and 95% confidence intervals) of each of these is shown in the Table 1. Since they converged and the means are summarized in Table 1, the plots were not shown here, but if you would like to see how it converged, please refer to the appendix for the code.

Table 1:

Posterior summaries of each parameter.

Parameters	Mean	95% Confidence Interval
β_1	-0.0629	(-0.1742, 0.0565)
β_2	0.0004	(-2.9787, 2.9842)

β_1, the coefficient on the CO $_2$ variable, converges to approximately -0.0629 and β_2, the coefficient on the NAOI variable, converges to approximately 0.0004. We also decide to test the posterior probability of β_1 and β_2 to determine the chance that these variables are positively or negatively correlated with increased hurricane severity. These results are shown in Table 2.

Table 2:

Posterior probabilities of each parameter.

Parameter	Posterior Probability
β1	$\beta_1 < 0$ = 0.8292
β2	$\beta_2 > 0$ = 0.5002

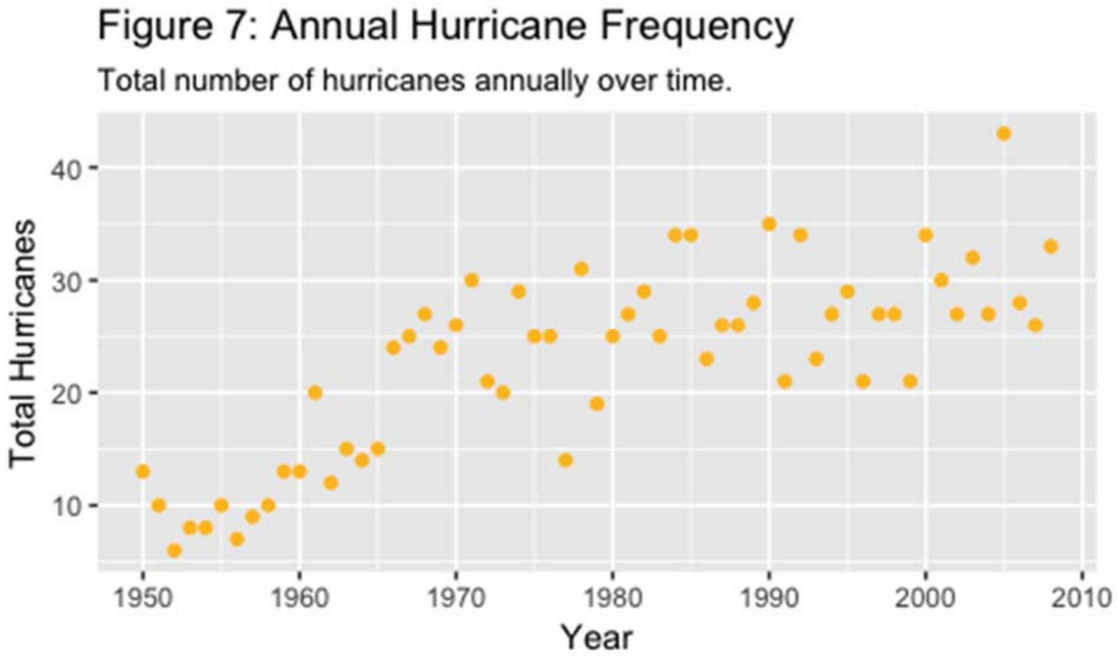
β_1 has a 0.8292 probability of being negatively correlated with average hurricane wind speed. β_2 has a 0.5002 probability of being positively correlated with average hurricane wind speed.

We determined that CO $_2$ and NAO are not useful predictors for hurricane severity (as measured by wind speed) based on the posterior estimations of the respective coefficients.

However, they may be good predictors for other aspects of hurricanes such as frequency, so we decided to model hurricane frequency using CO $_2$ concentration and the NAOI next.

Hurricane Frequency

Based on the past summer of hurricanes, it seems as if hurricanes are occurring more and more frequently. We plotted our data to see if it agreed, as shown in Figure 7.



We can see that from the beginning of our data in 1950 to the end of our data in 2008, there has been a substantial increase in total number of hurricanes per year. This prompted us to ask: Can we model annual hurricane frequency using CO $_2$ concentration and NAOI as predictors?

Because we are interested in modeling the number of hurricanes per year (a count), we use a Poisson distribution. Here, λ, the parameter describing the rate at which events occur, describes the expected number of hurricanes per year. We use Poisson regression to model the log of the expected rate of hurricanes per year. Our model is as follows:

Let:

Y_t = total hurricanes in year t

X_{1t} = yearly CO2 concentration in year t

X_{2t} = NAO index in year t

We propose the following model:

$$Y_t | \beta_0, \beta_1, \beta_2 \sim Pois(\lambda_i)$$

$$log(\lambda_i) = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t}$$

$$\beta_0 \sim N(0,1e6)$$

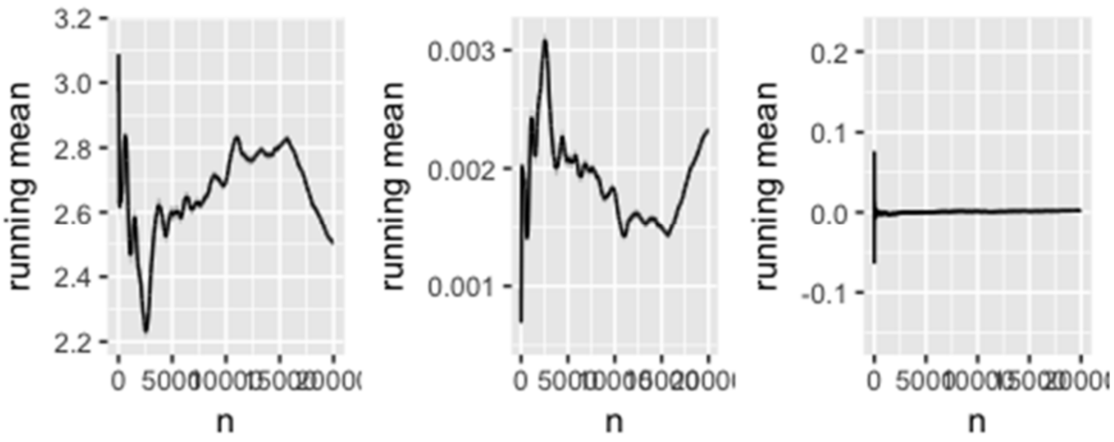
$$\beta_1 \sim N(0,1e6)$$

$$\beta_2 \sim N(0,1e6)$$

We run rjags for 40,000 iterations. To confirm that our parameter estimates have converged, we create running mean plots (see Figure 8).

Figure 8:

Running mean plots for β_0, β_1, and β_2. Plots shown below in that order.

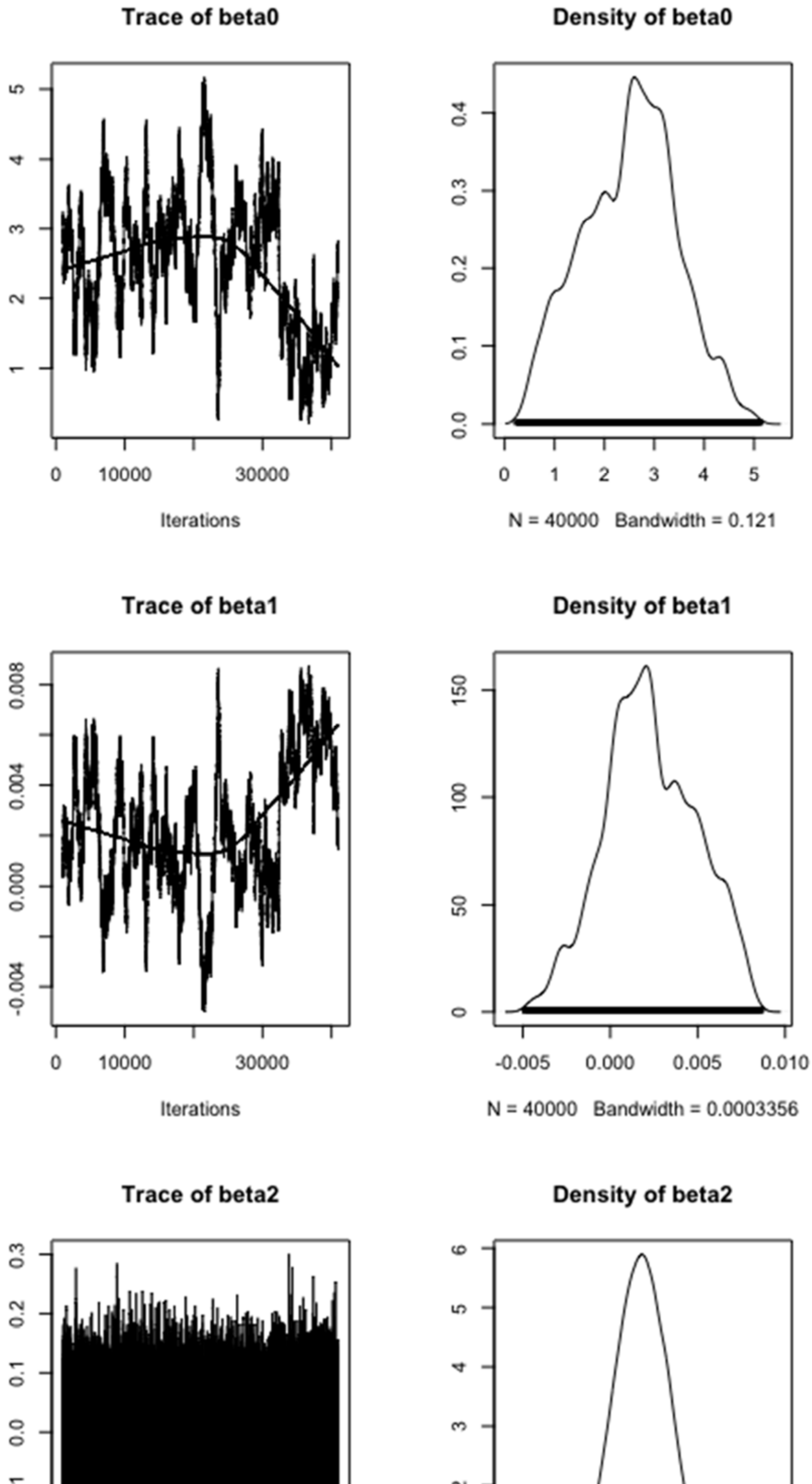


Based on the running mean plot, parameter estimates look like they have converged.

We can visualize the posterior distributions for these estimate parameters ($\beta_0, \beta_1, \beta_2$) in trace and density plots (see Figure 9).

Figure 9:

Posterior estimates of parameters summarized in trace and density plot.


Trace of beta0
Density of beta0
Iterations
N = 40000 Bandwidth = 0.121
0
10000
30000
1
2
3
4
5
0.0
0.1
0.2
0.3
0.4
Trace of beta1
Density of beta1
Iterations
N = 40000 Bandwidth = 0.0003356
-0.004
0.000
0.004
0.008
0
50
100
150
-0.005
0.000
0.005
0.010
Trace of beta2
Density of beta2
0.0
0.1
0.2
0.3
3
4
5
6

To gather further insight into any possible relationship between CO_2, NAOI, and hurricane frequency, we can look at the mean and 95% credible intervals the corresponding parameter estimations. Estimated parameter mean values and credible intervals are shown in Table 3.

Table 3:

Posterior summaries of each parameter.

Parameters	Mean	95% Confidence Interval
β1	0.00289	(-0.00123, 0.00706)
β2	0.00116	(-0.110, 0.135)

We notice that parameter estimates for β_1 and β_2 are centered around 0. This means that we expect there to be no relationship between CO $_2$ and NAO as predictors of hurricane frequency.

Using the density plots, we can also measure posterior probability that our parameters of interest, in this case β_1 and β_2, are negatively or positively associated with our outcome, here hurricane frequency. We calculate that the posterior probability of $\beta_1 < 0 = 0.6496$, meaning that CO2 is slightly negatively associated with increases in hurricane rate, and that the posterior probability of $\beta_2 > 0 = 0.670$, meaning that slightly positively associated with hurricane rate (results are summarised in Table 4).

Table 4:

Posterior probabilities of each parameter.

Parameter	Posterior Probability
β1	$\beta 1 < 0$ = 0.8612
β2	$\beta 2 > 0$ = 0.5031

From the model described above, we see that both NAO and CO $_2$ are weak predictors of hurricane frequency. But could this be due to differences in hurricane frequencies at different locations? To better understand this, we turn to our next question. Can we model annual hurricane frequency by zone using temperature anomalies as an explanatory variable?

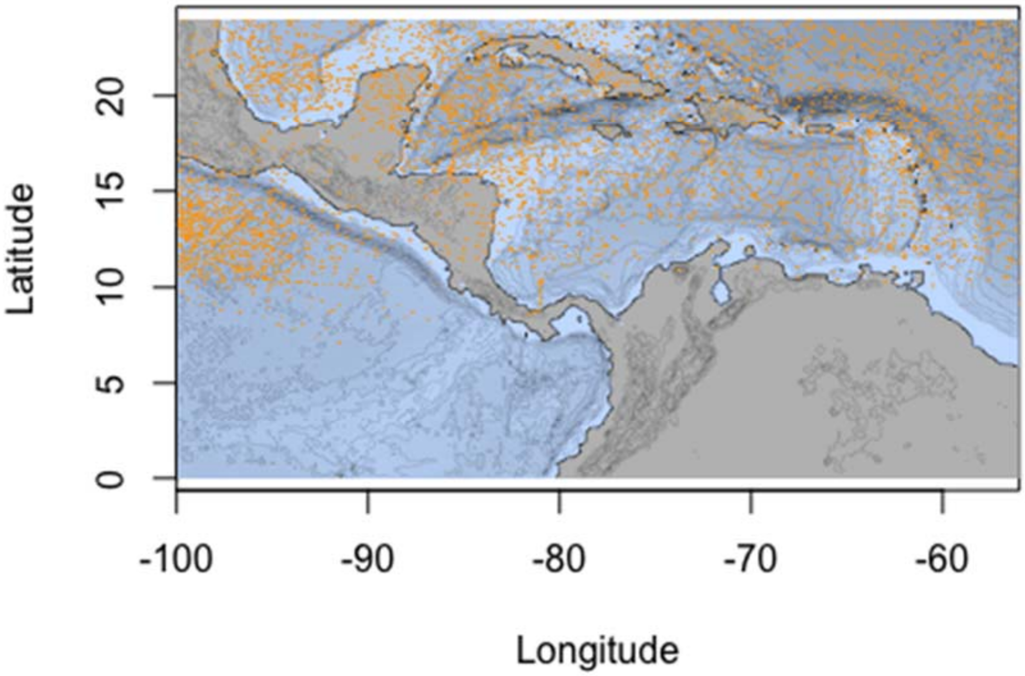
Hurricane Frequency by Latitudinal Zone

To start, we chose to restrict this model of hurricane frequency by zone to include only hurricanes occurring in the Atlantic basin. Due to limitations with temperature data, we further restrict our analysis to hurricanes occurring after 1980. We divide our data into four zones based on latitude: equator to 24N, 24N to 44N, 44N to 64N, and 64N to 90N. We selected these zones to match zones used in the available data on temperature anomalies. Hurricanes are plotted by zone in Figure 10:

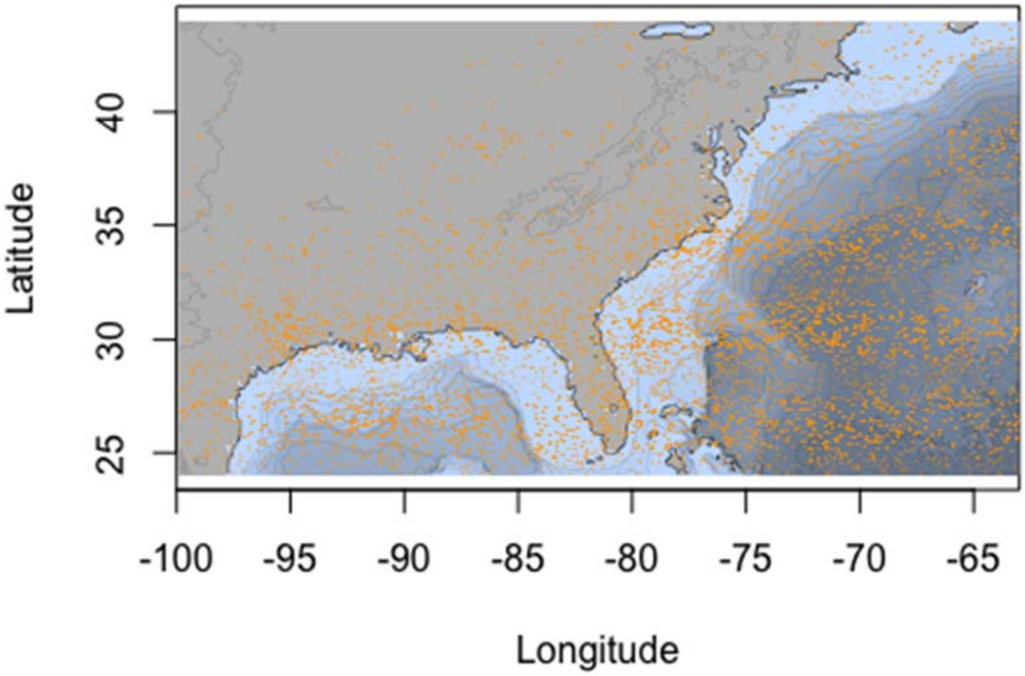
Figure 10:

Maps of Hurricanes by Zone

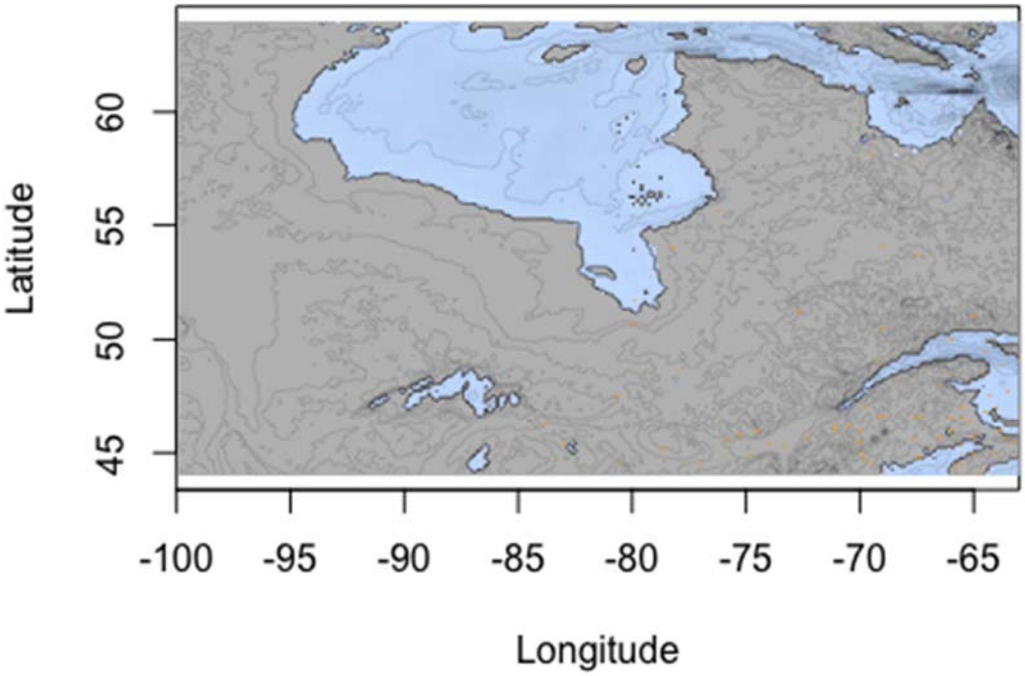
Graph of hurricane tracks from the equator to 24N



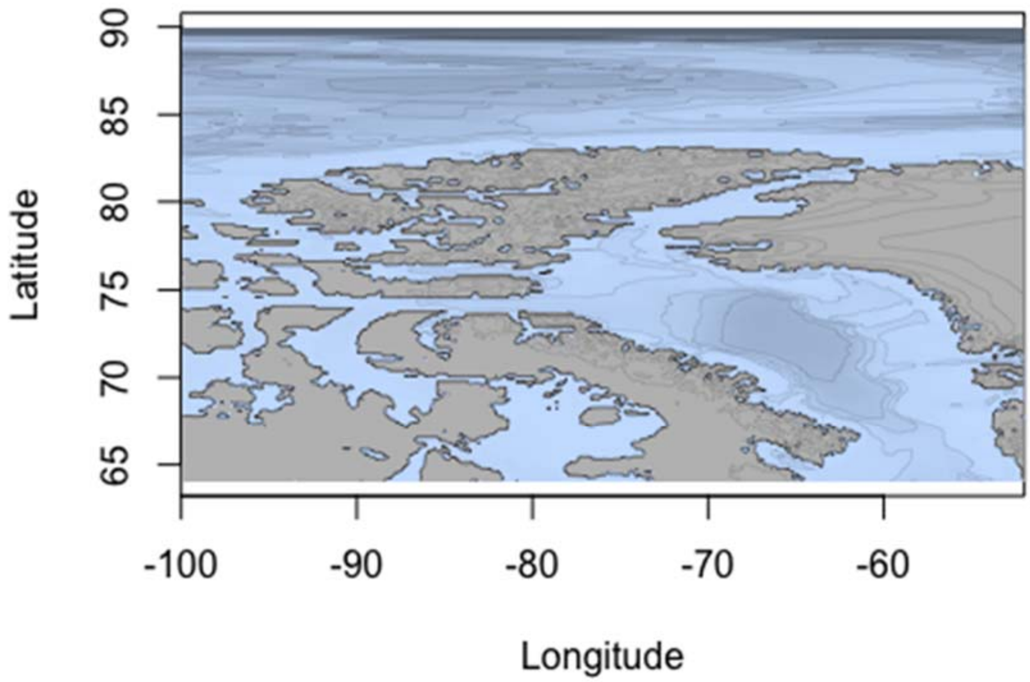
Hurricane tracks from 24N to 44N



Hurricane tracks from 44N to 64N



Hurricane tracks from 64N to 90N



Again, because we are modeling hurricane counts (although this time by zone), we use Poisson regression. Zone is included as a categorical variable of four levels. Because we are unsure of the relationship between zone and temperature with hurricane frequency, this model includes vague priors on all parameters. Our model is described below:

Let:

Y_t = total hurricanes in year t

X_{1t} = temperature anomaly by year t

X_{2t} = zone t

We propose the following model:

$$Y_t|\beta_0, \beta_1, \beta_2 \sim Pois(\lambda_i)$$

$$log(\lambda_i) = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t}$$

$$\beta_0 \sim N(0,1e6)$$

$$\beta_1 \sim N(0,1e6)$$

$$\beta_2 \sim N(0,1e6)$$

Again, we used rjags MCMC method to estimate parameter values. Although theoretically parameter value estimates should always converge given enough iterations, parameter estimates in our model failed to converge completely after 2.5 million iterations. Due to this limitation, it should be noted that our posterior estimation is not as exact as desired.

The mean of parameter estimations and 95% credible intervals are shown in the Table 5.

Table 5:

Posterior summaries of each parameter (all zones and temperature anomaly).

Parameters	Mean	95% Confidence Interval
$\beta_{64N-90N}$	-10.77	(-18.38, -4.46)
$\beta_{44N-64N}$	2.381	(2.292, 2.468)
$\beta_{24N-44N}$	4.997	(4.973, 5.021)
β_{0N-24N}	4.725	(4.698, 4.754)
$\beta_{temp-anom}$	0.140	(0.063, 0.230)

Because these credible intervals do not contain zero, we conclude that zone and temperature anomalies are useful predictors of hurricane frequency. Looking at the posterior estimates for the parameter values, we notice that overall the 24N to 44N zone has the highest baseline frequency of hurricanes, with hurricane rate decreasing as location diverges from this zone. The 64N to 90N zone has the lowest baseline rate of hurricanes, which makes sense considering that this zone includes the Arctic. For all zones, as temperature anomalies become more positive, hurricane rate increases.

But how useful is this model really? To further understand the usefulness of our model, we fit estimate the model parameters on training data and test its accuracy on a testing set. We begin by splitting our model into a training and testing set. The earliest 80% of data points are used as the training set and the remaining 20% are saved for testing the model.

Predictioned annual hurricane rates for each zone are shown below (see Figure 11).

Figure 11: Predictions vs Actual Hurricane Frequency by Zone

Figure 11: Predictions vs Actual Hurricane Frequency by Zone

Since zone categories and temperature anomalies had strong relationships with hurricane frequency according to our posterior approximations, we predicted that zone category and temperature anomalies may also have relationships with another aspect of hurricanes, such as hurricane duration. Therefore, we next attempted to modeled hurricane duration using temperature anomalies by zone.

Hurricane Duration

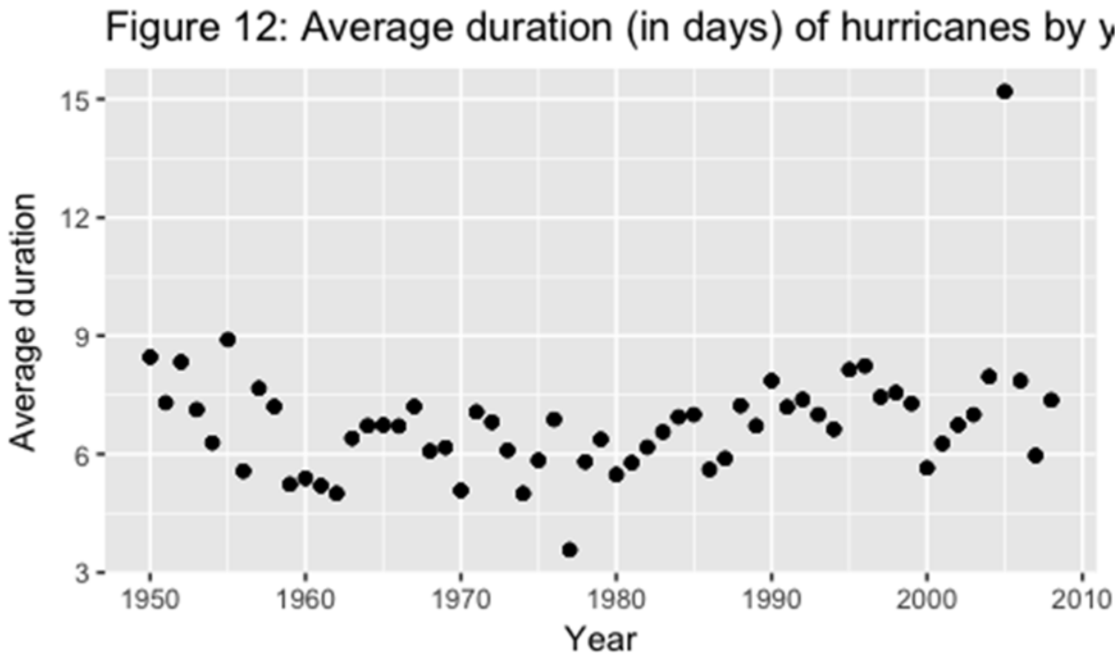


Figure 12: Average duration (in days) of hurricanes by y

We model average hurricane duration (in days) for all hurricanes occurring in the same year (Figure 12) to look for trends in the dataset. In order to model average duration of hurricanes, we decided to use Normal-Normal model, just like the hurricane wind speed model, but using zone categories and temperature anomalies as predictor variables.

Let:

Y_t = average hurricane duration in year t

X_{1t} = temperature anomaly by year t

X_{2t} = zone t

$$Y_t \sim N(\beta_0 + \beta_1 X_{1t} + \beta_2 (X_{2t}), \tau^{-1})$$

$$\beta_0 \sim N(0, 100^2)$$

$$\beta_1 \sim N(0, 100^2)$$

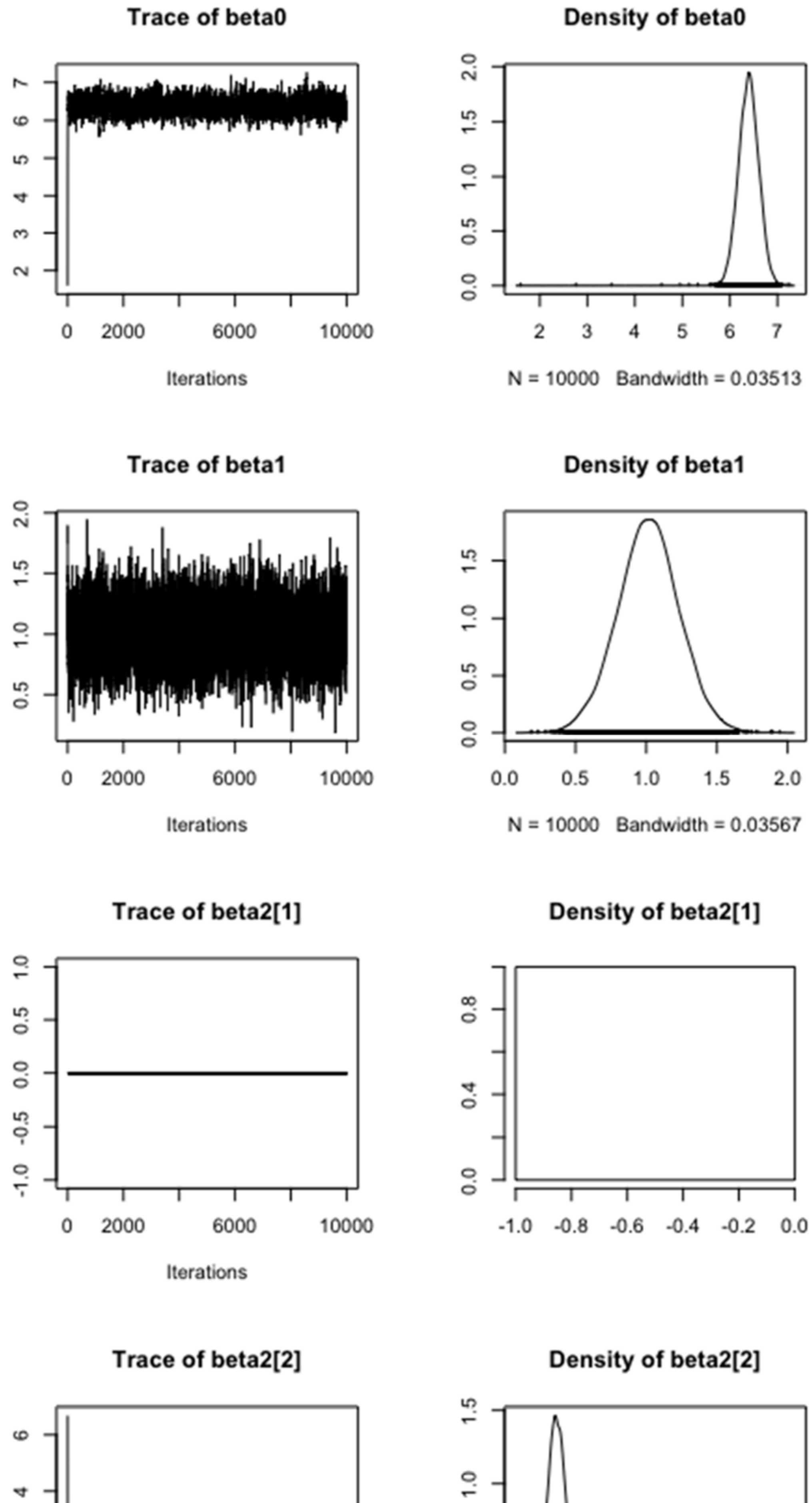
$$\beta_2 \sim N(0, 100^2)$$

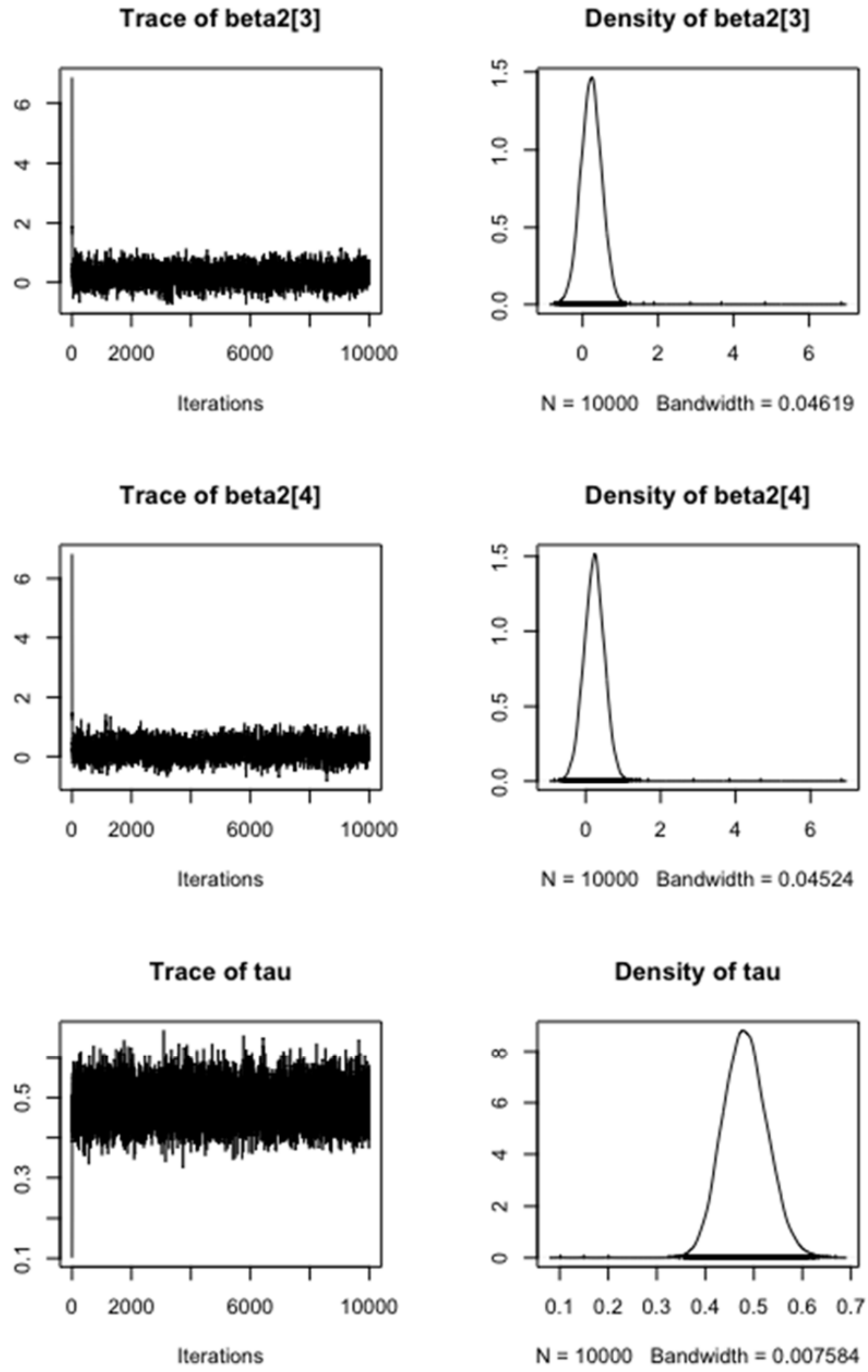
$$\tau \sim \Gamma(0.001, 0.001)$$

β_1 represented the coefficient for average temperature anomaly per year and β_2 represented the category zones.

Figure 13:

Trace plot and density plots for durations model parameters.


Trace of beta0
7
6
5
4
3
2
0
2000
6000
10000
Iterations
Density of beta0
2.0
1.5
1.0
0.5
0.0
2
3
4
5
6
7
N = 10000 Bandwidth = 0.03513
Trace of beta1
2.0
1.5
1.0
0.5
0
2000
6000
10000
Iterations
Density of beta1
1.5
1.0
0.5
0.0
0.0
0.5
1.0
1.5
2.0
N = 10000 Bandwidth = 0.03567
Trace of beta2[1]
1.0
0.5
0.0
-0.5
-1.0
0
2000
6000
10000
Iterations
Density of beta2[1]
0.8
0.4
0.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
Trace of beta2[2]
6
4
Density of beta2[2]
1.5
1.0


Trace of beta2[3]
6
4
2
0
0
2000
6000
10000
Iterations
Density of beta2[3]
1.5
1.0
0.5
0.0
0
2
4
6
N = 10000   Bandwidth = 0.04619
Trace of beta2[4]
6
4
2
0
0
2000
6000
10000
Iterations
Density of beta2[4]
1.5
1.0
0.5
0.0
0
2
4
6
N = 10000   Bandwidth = 0.04524
Trace of tau
0.5
0.3
0.1
0
2000
6000
10000
Iterations
Density of tau
8
6
4
2
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
N = 10000   Bandwidth = 0.007584

In Figure 13, the density plots show that the duration at the zone between latitudes 64N and 90N (shown as beta2.1.) is represented by the intercept, β_0. The other density plots show the differences from the sampled value from the intercept value. beta2.2. is the zone between 44N and 64N, beta2.3. Is the zone between 24N and 44N and the zone between the equator and 24N is represented by beta2.4.

Table 6:

Overall mean and 95% credible interval for all zones, temperature anomaly coefficient and precision.

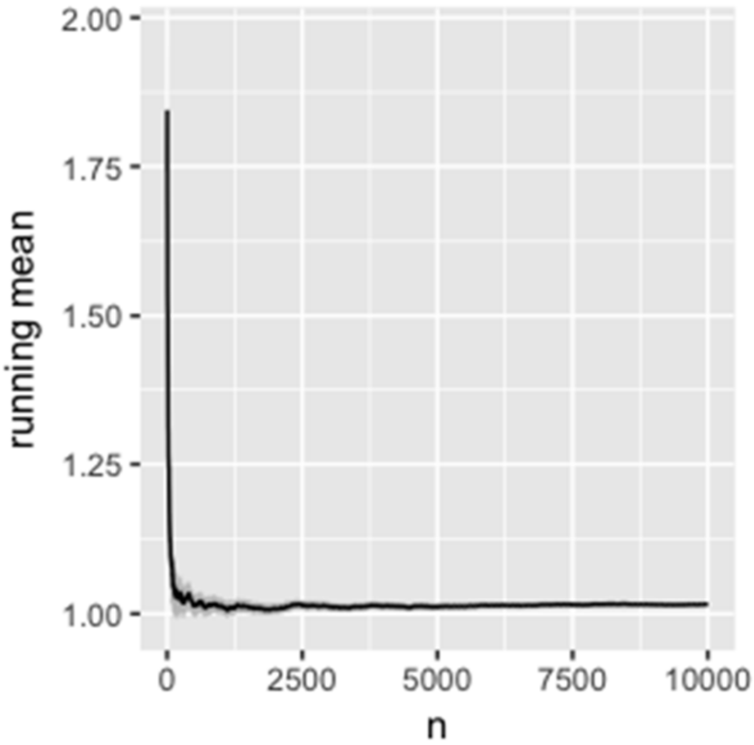
Parameters	Mean	95% Credible Interval
Equator to 24N	6.383291	(6.039368, 6.788171)
24N to 44N	6.475179	(6.149554, 6.869345)
44N to 64N	6.621217	(6.302772, 6.995594)
64N to 90N	6.621911	(6.310922, 6.982955)
β_1 Temperature Anomaly	1.02139	(0.676516, 1.438665)
τ Precision	0.4826497	(0.411310, 0.573710)

The means show the average posterior duration of hurricanes and the quantiles show the 95% confidence interval for each zone (Table 6).

In order to see if posterior parameter estimations converge, we looked at running mean plots (see Figure 14).

Figure 14:

Example running mean plot using temperature anomaly slope.



This is one example of a mean plot which looks at the convergence of the means of the temperature anomaly coefficients, or slope. To look at the convergence of the other parameters please refer to the appendix. Overall, all the parameters converged after 5000 iterations.

Table 7:

Posterior probabilities of each zone category.

Zone category	Posterior Probability
Equator to 24N	$\beta < 0$ = 0.8112
24N to 44N	$\beta < 0$ = 0.7085
44N to 64N	$\beta < 0$ = 0.5024

Just as we did for the first model for hurricane frequency using CO $_2$ concentration and NAOI, we calculated the posterior probabilities of our parameters (Table 7). We calculate that the posterior probability of $\beta_1 > 0 = 1$, meaning that, according to the approximated values, there is a positive association between temperature anomalies and hurricane duration. The posterior probability of the zone between the equator and 24N latitude, $\beta <$ 0 = 0.8112, had a negative relationship with hurricane duration just as zone 24N to 44N, $\beta < 0$ = 0.7085. However zone 44N and 64N had a weak relationship with hurricane duration with $\beta < 0$ = 0.5024, meaning that the slope when the hurricane is located in zone 44N and 64N is around 0.

Conclusion

From the models, we made some final conclusions. Overall, CO_2 and NAOI were not particularly useful predictors of hurricane characteristics. Models using temperature anomalies as a predictor of hurricane frequency and duration were slightly more successful.

CO_2 was negatively associated with severity and hurricane frequency. NAOI had little association with severity and hurricane frequency. Positive temperature anomalies was positively associated with hurricane frequency and average duration of the hurricane. Although the 24N to 44N latitudinal zone had the highest frequency of hurricanes, generally increased latitude was negatively associated with frequency. Additionally, latitude was weakly associated with hurricane duration.

Limitations and Future Research

The data used to explain patterns in hurricane characteristics summarized the explanatory variables at a general level, reducing the accuracy of our model to predict hurricanes at a more local level. For example, we used temporal averages of our predictor variables; therefore, even if a hurricane occurred in March, it would be modeled with an explanatory variable averaged across all months of that year.

In the future, we would like to allow for interactions between location and the climate variables. There variables of NAOI, CO_2, and temperature anomalies may affect hurricane characteristics differentially depending on location.

We would also suggest expanding upon these models to model the characteristics of individual hurricanes over time. In predicting natural disasters such as hurricanes, knowing the defining characteristics of the largest, most severe hurricane (rather than the average of all storms, both big and small) can provide more insight into the preparations needed to prevent damage.

We are also interesting in modeling hurricane locations at a finer scale. Additionally data providing information about temperature and wind and pressure conditions at a local level would allow for more precise predictions by our models. If climate data was available to describe conditions at a finer scale, we could also look at hurricanes at only specific points of their path, specifically landfall points.

References

- Achim Zeileis and Gabor Grothendieck (2005). zoo: S3 Infrastructure for Regular and Irregular Time Series. Journal of Statistical Software, 14(6), 1-27. doi:10.18637/jss.v014.i06
- Alicia Johnson (2017). MacBayes: Introductory Bayesian toolkit. R package version 0.0.0.9000.
- Ed Dlugokencky and Pieter Tans, NOAA/ESRL (www.esrl.noaa.gov/gmd/ccgg/trends/)

- Elsner, J. & Jagger, T. (2006). Prediction Models for Annual U.S. Hurricane Counts. Journal of Climate.
- Eric Pante, Benoit Simon-Bouhet (2013) marmap: A Package for Importing, Plotting and Analyzing Bathymetric and Topographic Data in R. PLoS ONE 8(9): e73051. doi:10.1371/journal.pone.0073051
- Garrett Grolemund, Hadley Wickham (2011). Dates and Times Made Easy with lubridate. Journal of Statistical Software, 40(3), 1-25. URL http://www.jstatsoft.org/v40/i03/.
- H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.
- Hadley Wickham (2007). Reshaping Data with the reshape Package. Journal of Statistical Software, 21(12), 1-20. URL http://www.jstatsoft.org/v21/i12/.
- Hadley Wickham, Jim Hester and Romain Francois (2017). readr: Read Rectangular Text Data. R package version 1.1.1. https://CRAN.R-project.org/package=readr
- Hadley Wickham and Lionel Henry (2017). tidyr: Easily Tidy Data with 'spread()' and 'gather()' Functions. R package version 0.7.2. https://CRAN.R-project.org/package=tidyr
- Hadley Wickham, Romain Francois, Lionel Henry and Kirill Müller (2017). dplyr: A Grammar of Data Manipulation. R package version 0.7.4. https://CRAN.R-project.org/package=dplyr
- Hansen, J., R. Ruedy, M. Sato, and K. Lo, 2010: Global surface temperature change, Rev. Geophys., 48, RG4004, doi:10.1029/2010RG000345.
- Historical Tropical Storm Tracks provided by the Homeland Infrastructure Foundation Level-Data, HIFLD, USA (2009). Accessed 13 Nov. 2017.
- Hurricane Society (n.d.) Variability of Hurricane Activity. Hurricanes: Science and Society. University of Rhode Island. http://www.hurricanescience.org/science/science/activity/.
- Martyn Plummer (2016). rjags: Bayesian Graphical Models using MCMC. R package version 4-6.https://CRAN.R-project.org/package=rjags
- Gray, R. (2017). Next-generation models revealing climate change effects on hurricanes. Horizon Magazine.
- NAO Index Data provided by the Climate Analysis Section, NCAR, Boulder, USA, Hurrell (2003). Updated regularly. Accessed 13 Nov. 2017.
- Venables, W. N. & Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0

Appendix

Packages used

```r
library(readr)
library(dplyr)
library(tidyr)
library(ggplot2)
library(zoo)
library(lubridate)
library(marmap)
library(reshape2)
library(MASS)
library(viridis)
library(rjags)
library(MacBayes)

# Loading the Data
hurricane_data <- read_csv("Historical_Tropical_Storm_Tracks.csv")
hurricane_data<-na.omit(hurricane_data)
hurricane_data<-subset(hurricane_data, select=-c(BTID, AD_TIME))

# Recent Hurricanes--select only hurricanes in 1950 and after
recent_hurricanes <- hurricane_data[hurricane_data$YEAR >= 1950, ]
recent_hurricanes <-
recent_hurricanes[!(recent_hurricanes$NAME=="NOTNAMED"),]

# Hurricane Frequency
# table of hurricanes per year
hurricanes_per_year <- recent_hurricanes %>%
  group_by(YEAR) %>%
  summarise(TOTAL_H = n_distinct(NAME))
hurricanes_per_basin <- recent_hurricanes %>%
  group_by(BASIN) %>%
  summarise(TOTAL_H = n_distinct(NAME, DAY))

#Format Coordinates
dat<-as.character(recent_hurricanes$LAT)
new<-substr(dat,1,nchar(dat))
lat<-as.numeric(new)
dat<-as.character(recent_hurricanes$LONG)
new<-substr(dat,1,nchar(dat))
long<-as.numeric(new)
long<- -abs(long)
coord<-as.data.frame(long)
#Add year and name
year<-as.character(recent_hurricanes$YEAR)
name<-as.character(recent_hurricanes$NAME)
hurricane_locations<-cbind(year, name, coord, lat)


# Temperature Data
temp_data_orig <- read.csv("ZonAnn.csv")
temp_data <- temp_data_orig
temp_data[!complete.cases(temp_data),]
temp_data<-na.omit(temp_data)
temp_data<-temp_data[,-(5:15),drop=FALSE]
temp_recent <- temp_data %>%
  filter(Year > 1949)
```

```r
# select only northern hemisphere temp data
tempN <- subset(temp_recent, select = c(Year, NHem))
# rename the columns
colnames(tempN) <- c("YEAR", "temp")
# join temperature to hurricanes per year table
hurricanes_per_year2 <- merge(x = hurricanes_per_year, y = tempN, by =
"YEAR", all.x = TRUE)

# CO2 Data
CO2 <- read_csv("CO2.csv")
# Alter CO2 to remove unnecessary variable trend
CO2 <- subset(CO2, select = -trend)
# get average by year
CO2_yr <- CO2 %>%
  group_by(year) %>%
  summarise(avg_CO2 = mean(average))
# rename columns
colnames(CO2_yr) <- c("YEAR", "avg_CO2")
# join with hurricane per year table
hurricanes_per_year3 <- merge(x = hurricanes_per_year2, y = CO2_yr, by =
"YEAR", all.x = TRUE)
# FIX THIS (bc only 1980 onward available)

# NAO Index Data
NAO_index1 <- read_csv("NAO_index_monthly.csv",
    col_types = cols(Apr = col_number(),
        Aug = col_number(), Dec = col_number(),
        Feb = col_number(), Jan = col_number(),
        Jul = col_number(), Jun = col_number(),
        Mar = col_number(), May = col_number(),
        Nov = col_number(), Oct = col_number(),
        Sep = col_number()))
NAO_index2 <- NAO_index1[, -c(14)]
NAO_index3 <- NAO_index2 %>%
  rename(year = X1) %>%
  mutate(Jul = replace(Jul, Jul < -50| Jul > 50, "")) %>%
  mutate(Aug = replace(Aug, Aug < -50| Aug > 50, "")) %>%
  mutate(Sep = replace(Sep, Sep < -50| Sep > 50, "")) %>%
  mutate(Oct = replace(Oct, Oct < -50| Oct > 50, "")) %>%
  mutate(Nov = replace(Nov, Nov < -50| Nov > 50, "")) %>%
  mutate(Dec = replace(Dec, Dec < -50| Dec > 50, ""))
NAO_index <- NAO_index3[-nrow(NAO_index3),]
# remove 2017 (data for year is incomplete)
NAO_index <- head(NAO_index, -1)
# reshape table
NAO_index <- melt(NAO_index, id=c("year"))
colnames(NAO_index)[2] <- "month"
colnames(NAO_index)[3] <- "index"
NAO_index$year <- as.factor(NAO_index$year)
```

```
NAO_index$index <- as.numeric(NAO_index$index)
# group by year
NAO_yr <- NAO_index %>%
  group_by(year) %>%
  summarise(avg_NAO = mean(index))
# rename columns
colnames(NAO_yr) <- c("YEAR", "avg_NAO")
# join with hurricane data
hurricanes_per_year4 <- merge(x = hurricanes_per_year3, y = NAO_yr, by =
"YEAR", all.x = TRUE)

# Remove all rows with missing data
hurricanes_per_year_clean<-na.omit(hurricanes_per_year4)

# Wind speed data
# Create new data set with wind speed and year
windspeed_yr <- hurricane_data %>%
  group_by(YEAR) %>%
  summarise(avg_windspeed = mean(WIND_KTS))

# Merge this with the other data set
HURRICANES <- merge(x = hurricanes_per_year_clean, y = windspeed_yr, by =
"YEAR", all.x = TRUE)
```

Checking for collinearity

```
# First, we check for collinearity between the explanatory variables of
temperature, CO2, and NAO index.

cor(x = as.matrix(hurricanes_per_year_clean$temp), y =
as.matrix(hurricanes_per_year_clean$avg_CO2))

cor(x = as.matrix(hurricanes_per_year_clean$avg_NAO), y =
as.matrix(hurricanes_per_year_clean$temp))

cor(x = as.matrix(hurricanes_per_year_clean$avg_NAO), y =
as.matrix(hurricanes_per_year_clean$avg_CO2))
```

Markov Chain for Wind speed

```
# specify the model
windSpeed_model <- "model{
    #Data
    for(i in 1:length(y)) {
        y[i] ~ dnorm(beta0 + beta1*x1[i] + beta2*x2[i], tau)

    }
    #Priors
    beta0 ~ dnorm(0, 1/(1000)^2) #PRECISION
    beta1 ~ dnorm(0, 1/(1000)^2) #PRECISION
    beta2 ~ dnorm(0, 1/(1000)^2) #PRECISION
    tau ~ dgamma(0.001, 0.001)

}"

#*set up an algorithm to simulate the posterior by
#*combining the model (games_model) and data (x)
#*set the random number seed
windSpeed_jags <-
jags.model(textConnection(windSpeed_model),data=list(y=HURRICANES$avg_windspe
ed,x1=HURRICANES$avg_CO2,x2=HURRICANES$avg_NAO), inits = list(beta0 = 5,
beta1 = 0, beta2 = 0))

#*simulate a sample from the posterior
#*note that we specify both mu and tau variables
windSpeed_sim <- coda.samples(windSpeed_jags,
variable.names=c("beta0","beta1","beta2"), n.iter=500000)


#*store the samples in a data frame:
windSpeed_sample <- data.frame(step=1:500000, windSpeed_sim[[1]])

running_mean_plot(x=windSpeed_sample$beta1, se=TRUE)
running_mean_plot(x=windSpeed_sample$beta2, se=TRUE)
```

Markov Chains for Frequency

Creating dataset for frequency, temperature, and zones

```
# Adding zones to the temperature data and limiting time interval to between
1950-2008.
temp_data_orig<-na.omit(temp_data_orig) %>%
  filter(1950 <= Year) %>%
  filter(Year <= 2008)
my_temp <- subset(temp_data_orig, select = c("Year", "X64N.90N", "X44N.64N",
"X24N.44N", "EQU.24N"))
colnames(my_temp) <- c("Year", "N64to90", "N44to64", "N24to44", "EQUtoN24")
my_temp <- melt(my_temp, id=c("Year"))
colnames(my_temp) <- c("year", "zone", "temp")

hurricane_locations_Atlantic <- merge(hurricane_locations, recent_hurricanes,
by.x = c("name", "year", "lat", "long"), by.y = c("NAME", "YEAR", "LAT",
"LONG"), all.x = TRUE) %>%
  filter(BASIN == "North Atlantic") %>%
  dplyr::select(c(year, name, long, lat))

hurricanes_zones <- hurricane_locations_Atlantic %>%
  mutate(zone=cut(lat, breaks=c(0, 24, 44, 64, 90), labels=c("EQUtoN24",
"N24to44", "N44to64", "N64to90"))) %>%
  group_by(year, zone)  %>%
  summarise(total = n())
```

```
# Combine the  hurricane and temperature dataset together
hurricanes_per_zone <- merge(my_temp, hurricanes_zones, by = c("year",
"zone"), all.x = TRUE, all.y = TRUE)
hurricanes_per_zone[is.na(hurricanes_per_zone)] <- 0
```

Using CO_2 and NAOI

```
# specify the model
hur_mod <- " model {
  for (i in 1:length(TOTAL_H)) {
      TOTAL_H[i] ~ dpois(lam[i])
      log(lam[i]) = beta0 + beta1*X1[i] + beta2*X2[i]
  }

  beta0 ~ dnorm(0.0, 1.0/1e6)
  beta1 ~ dnorm(0.0, 1.0/1e4)
  beta2 ~ dnorm(0.0, 1.0/1e4)
} "

# set up an algorithm to simulate the posterior by combining the model and
data (x)
# set the random number seed

data_jags = as.list(hurricanes_per_year_clean[,2:5])
str(data_jags)

#freq_jags <- jags.model(textConnection(hur_mod),data=list(TOTAL_H =
hurricanes_per_year_clean$TOTAL_H, X1 = hurricanes_per_year_clean$temp, X2 =
hurricanes_per_year_clean$avg_CO2,X3 = hurricanes_per_year_clean$avg_NAO),
inits=list(.RNG.name="base::Wichmann-Hill", .RNG.seed=2000))

# manually initialize parameters
freq_jags <- jags.model(textConnection(hur_mod),data=list(TOTAL_H =
hurricanes_per_year_clean$TOTAL_H, X1 = hurricanes_per_year_clean$avg_CO2,X2
= hurricanes_per_year_clean$avg_NAO), inits=list(beta0=5,beta1=0, beta2=0))

# simulate a sample from the posterior
# note that we specify both mu and tau variables
freq_sim <- coda.samples(freq_jags, variable.names = c("beta0", "beta1",
"beta2"), n.iter=500000)

# store the samples in a data frame:
freq_sample <- data.frame(step = 1:500000, freq_sim[[1]])
head(freq_sample, 10)


# make a dataframe of parameter values every other step to make plotting
faster
Nth.delete<-function(dataframe, n)dataframe[-
(seq(n,to=nrow(dataframe),by=n)),]
freq_sample_small <- Nth.delete(freq_sample, 2)
```

Using Temperature and Zone categories

Create testing and training data

```
train_zone <- hurricanes_per_zone %>%
  filter(year < 1996)
test_zone <- hurricanes_per_zone %>%
  filter(year >= 1996)
```

Poisson regression--with temperature and location zone

```
#specify the model
hur_mod3 <- " model {
  for (i in 1:length(total)) {
      total[i] ~ dpois(lam[i])
      log(lam[i]) = beta0 + beta1*X1[i] + beta2[X2[i]]
  }

  beta0 ~ dnorm(0.0, 1.0/1e4)
  beta1 ~ dnorm(0.0, 1.0/1e4)
  beta2[1] <- 0
  for (i in 2:4) {
    beta2[i] ~ dnorm(0.0, 1.0/1e4)
  }
} "

# set up an algorithm to simulate the posterior by combining the model and
data (x)
# set the random number seed
#freq_jags3 <- jags.model(textConnection(hur_mod3),data=list(total =
train_zone$total, X1 = train_zone$temp, X2 = train_zone$zone),
inits=list(.RNG.name="base::Wichmann-Hill", .RNG.seed=2000))

# try to manually initialize parameters
freq_jags3 <- jags.model(textConnection(hur_mod3),data=list(total =
train_zone$total, X1 = train_zone$temp, X2 = train_zone$zone),
inits=list(beta0=0, beta1=0, beta2 = c(NA, 8, 10, 10)))


# simulate a sample from the posterior
freq_sim3 <- coda.samples(freq_jags3, variable.names = c("beta0", "beta1",
"beta2"), n.iter=2500000)

# store the samples in a data frame:
freq_sample3 <- data.frame(step = 1:2500000, freq_sim3[[1]])
head(freq_sample3, 10)
plot(freq_sim3)
```

```
# make step size smaller to plot faster
freq_sample3_small <- Nth.delete(freq_sample3, 2)

plot_freq_sample3 <- freq_sample3 %>%
  filter(step > (2500000-100000))

# Check for convergence for each parameter
running_mean_plot(x=plot_freq_sample3$beta2.1., se=TRUE)
running_mean_plot(x=plot_freq_sample3$beta2.2., se=TRUE)
running_mean_plot(x=plot_freq_sample3$beta2.3., se=TRUE)
running_mean_plot(x=plot_freq_sample3$beta2.4., se=TRUE)
```

Testing the model

```
# Testing Zone N64to90--begin with the northern-most zone
# select data in this zone
test_zoneN64to90 <- test_zone %>%
  dplyr::filter(c(zone == "N64to90")) %>%
  dplyr::select(-c(zone))
# create prediction function--takes 1000 draws from Poisson distribution
hurri_prediction_zoneN64to90 <- function(X1){
  pred <- rpois(1000, exp(freq_sample3$beta0 + freq_sample3$beta1*X1))
  return(pred)
}
# create empty vector to store predictions
hurri_preds_zoneN64to90 <- rep(0, nrow(test_zoneN64to90))
# get predictions--mean predicted value for each year in this zone
for (i in 1:nrow(test_zoneN64to90)){
  hurri_preds_zoneN64to90[i] <- mean(hurri_prediction_zoneN64to90(X1 =
test_zoneN64to90$temp[i]))
}


# Testing Zone N44to64--next northern-most zone
# select data in this zone
test_zoneN44to64 <- test_zone %>%
  dplyr::filter(c(zone == "N44to64")) %>%
  dplyr::select(-c(zone))
# create prediction function--takes 1000 draws from poisson distribution
hurri_prediction_zoneN44to64 <- function(X1){
  pred <- rpois(1000, exp(freq_sample3$beta0 + freq_sample3$beta2.2. +
freq_sample3$beta1*X1))
  return(pred)
}
# create empty vector to store predictions
hurri_preds_zoneN44to64 <- rep(0, nrow(test_zoneN44to64))
# get predictions--mean predicted value for each year in this zone
for (i in 1:nrow(test_zoneN44to64)){
  hurri_preds_zoneN44to64[i] <- mean(hurri_prediction_zoneN44to64(X1 =
test_zoneN44to64$temp[i]))
}


# Testing Zone N24to44--the next northern-most zone
# select data in this zone
test_zoneN24to44 <- test_zone %>%
  dplyr::filter(c(zone == "N24to44")) %>%
  dplyr::select(-c(zone))
# create prediction function--takes 1000 draws from poisson distribution
hurri_prediction_zoneN24to44 <- function(X1){
  pred <- rpois(1000, exp(freq_sample3$beta0 + freq_sample3$beta2.3. +
freq_sample3$beta1*X1))
  return(pred)
}
# create empty vector to store predictions
hurri_preds_zoneN24to44 <- rep(0, nrow(test_zoneN24to44))
# get predictions--mean predicted value for each year in this zone
for (i in 1:nrow(test_zoneN24to44)){
  hurri_preds_zoneN24to44[i] <- mean(hurri_prediction_zoneN24to44(X1 =
test_zoneN24to44$temp[i]))
}


# Testing Zone EQUtoN24--southern-most zone
# select data in this zone
test_zoneEQUtoN24 <- test_zone %>%
  dplyr::filter(c(zone == "EQUtoN24")) %>%
  dplyr::select(-c(zone))
# create prediction function--takes 1000 draws from poisson distribution
hurri_prediction_zoneEQUtoN24 <- function(X1){
  pred <- rpois(1000, exp(freq_sample3$beta0 + freq_sample3$beta1*X1 +
freq_sample3$beta2.4.))
  return(pred)
}
# create empty vector to store predictions
hurri_preds_zoneEQUtoN24 <- rep(0, nrow(test_zoneEQUtoN24))
# get predictions--mean predicted value for each year in this zone
for (i in 1:nrow(test_zoneEQUtoN24)){
  hurri_preds_zoneEQUtoN24[i] <- mean(hurri_prediction_zoneEQUtoN24(X1 =
test_zoneEQUtoN24$temp[i]))
}

# Compare results to actual data
zoneEQUtoN24 <- cbind(preds = as.matrix(hurri_preds_zoneEQUtoN24),
test_zoneEQUtoN24)
zoneEQUtoN24
ggplot(zoneEQUtoN24, aes(x = year)) + geom_point(aes(y = preds, color =
"predicted")) + geom_point(aes(y = total, color = "actual")) + xlim(1996,
2008) + ylim(0, 400) + ylab("total hurricanes") + ggtitle("0N to 24N")

zoneN24to44 <- cbind(preds = as.matrix(hurri_preds_zoneN24to44),
test_zoneN24to44)
ggplot(zoneN24to44, aes(x = year)) + geom_point(aes(y = preds, color =
"predicted")) + geom_point(aes(y = total, color = "actual")) + xlim(1996,
2008) + ylim(0, 400)+ ylab("total hurricanes") + ggtitle("24N to 44N")

zoneN44to64 <- cbind(preds = as.matrix(hurri_preds_zoneN44to64),
test_zoneN44to64)
ggplot(zoneN44to64, aes(x = year)) + geom_point(aes(y = preds, color =
"predicted")) + geom_point(aes(y = total, color = "actual")) + xlim(1996,
2008) + ylim(0, 400)+ ylab("total hurricanes") + ggtitle("44N to 64N")

zoneN64to90 <- cbind(preds = as.matrix(hurri_preds_zoneN64to90),
test_zoneN64to90)
ggplot(zoneN64to90, aes(x = year)) + geom_point(aes(y = preds, color =
"predicted"), position = "jitter") + geom_point(aes(y = total, color =
"actual")) + xlim(1996, 2008) + ylim(0, 400)+ ylab("total hurricanes") +
ggtitle("64N to 90N")


# find mean squared error
all_predictions <- rbind(zoneEQUtoN24, zoneN24to44, zoneN44to64, zoneN64to90)


sum_diff <- 0
for (i in 1:nrow(all_predictions)) {
  sq_diff <- (all_predictions$total[i] - all_predictions$preds[i])^2
  sum_diff <- sum_diff + sq_diff
  print(sq_diff)
}

sum_diff/nrow(all_predictions)
```

Markov Chain for Average Duration

Creating a new dataset with duration, zones, and temperature

```
# Hurricane data only looking at Date and name
hurricane_time<-subset(recent_hurricanes, select = c(NAME,YEAR,MONTH,DAY))

#Group data by year and name and subtracting starting time from end time for
each hurricane (duration)
foo <- hurricane_time %>%
  group_by(YEAR, NAME) %>%
  mutate(date = as.Date(paste(YEAR, MONTH, DAY, sep='-')), "%Y-%m-%d") %>%
  summarise(duration = max(date) - min(date))

#removed outlier
```

```
foo<-foo[!(foo$YEAR==1954 & foo$NAME=="ALICE"),]

# Averaging duration per year
foo<-group_by(foo, YEAR)%>%
  summarise(mean(duration))
# Name second column of data as duration and first column as year
colnames(foo)[2] <- "duration"
colnames(foo)[1]<-"year"
#create durations with temperature data, CO2 data, NAOI data and frequency
data
durations <- merge(foo, hurricanes_per_zone, by = c("year"), all.x = TRUE,
all.y = TRUE)
```

Running the model

```
library(rjags)

#specify the model
duration_model <- "model{
    #Data
    for(i in 1:length(y)) {
        y[i] ~ dnorm(beta0 + beta1*X1[i] + beta2[X2[i]],tau)
    }

    #Priors
    beta0 ~ dnorm(0.0, 1.0/1e4)
    beta1 ~ dnorm(0.0, 1.0/1e4)
    beta2[1] <- 0
    for (i in 2:4) {
      beta2[i] ~ dnorm(0.0, 1.0/1e4)
  }
    tau ~ dgamma(.001, .001)
}"

# initialize parameters
duration_jags <- jags.model(textConnection(duration_model), data=list(y =
durations$duration, X1 = durations$temp, X2 = durations$zone),
inits=list(.RNG.name="base::Wichmann-Hill", .RNG.seed=1989))
# simulate sample from posterior
duration_sim <- coda.samples(duration_jags, variable.names=c("beta0","beta1",
"beta2","tau"), n.iter=10000)

#store samples in data frame
duration_samples <- data.frame(duration_sim[[1]])
summary(duration_sim)
plot(duration_sim)

# running mean plot for each parameter and category
running_mean_plot(x=duration_samples$beta0, se=TRUE)
running_mean_plot(x=duration_samples$beta1, se=TRUE)
```

```r
running_mean_plot(x=duration_samples$beta2.2., se=TRUE)
running_mean_plot(x=duration_samples$beta2.3., se=TRUE)
running_mean_plot(x=duration_samples$beta2.4., se=TRUE)
running_mean_plot(x=duration_samples$tau, se=TRUE)

set.seed(19)

# 95% conidence interval
quantile(duration_samples$beta0, c(0.05, 0.975))
quantile(duration_samples$beta1, c(0.05, 0.975))
quantile(duration_samples$beta2.1.+duration_samples$beta0, c(0.05, 0.975))
quantile(duration_samples$beta2.2.+duration_samples$beta0, c(0.05, 0.975))
quantile(duration_samples$beta2.3.+duration_samples$beta0, c(0.05, 0.975))
quantile(duration_samples$beta2.4.+duration_samples$beta0, c(0.05, 0.975))

# Mean of all parameters
mean(duration_samples$beta0)
mean(duration_samples$beta1)
mean(duration_samples$beta2.1.+duration_samples$beta0)
mean(duration_samples$beta2.2.+duration_samples$beta0)
mean(duration_samples$beta2.3.+duration_samples$beta0)
mean(duration_samples$beta2.4.+duration_samples$beta0)
mean(duration_samples$tau)

# Posterior probability
mean(duration_samples$beta2.2.<0)
mean(duration_samples$beta2.3.<0)
mean(duration_samples$beta2.4.<0)
```