P8160 - Project 3

P8160 Group Project 3 Baysian modeling of hurricane trajectories

Jingchen Chai, Yi Huang, Zining Qi, Ziyi Wang, Ruihan Zhang

Columbia University

2023-05-01

Content

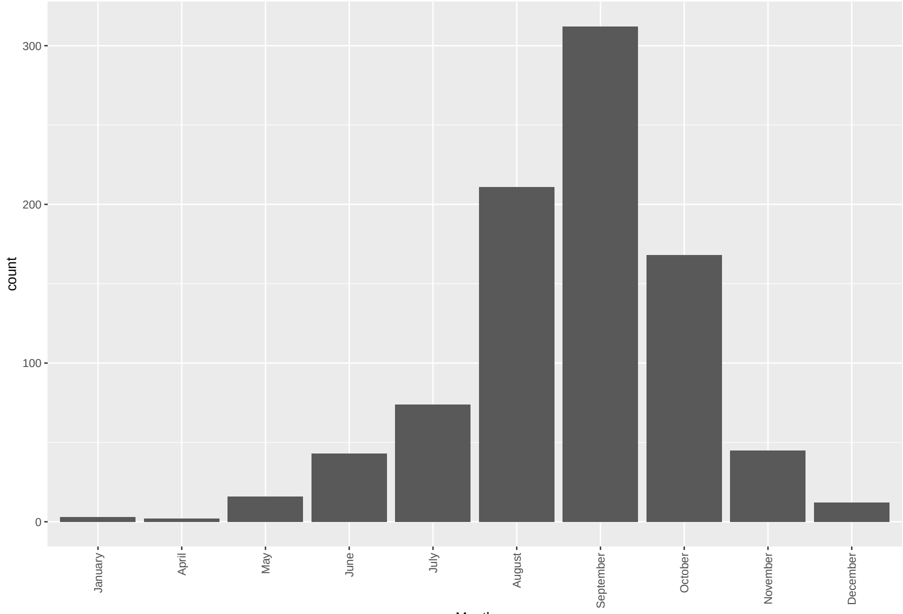
Introduction

- Hurricanes cause fatalities and property damage
- There is a growing need to accurately predict hurricane behavior, including location and speed
- This project aims to forecast wind speeds by modeling hurricane trajectories using a Hierarchical Bayesian Model.

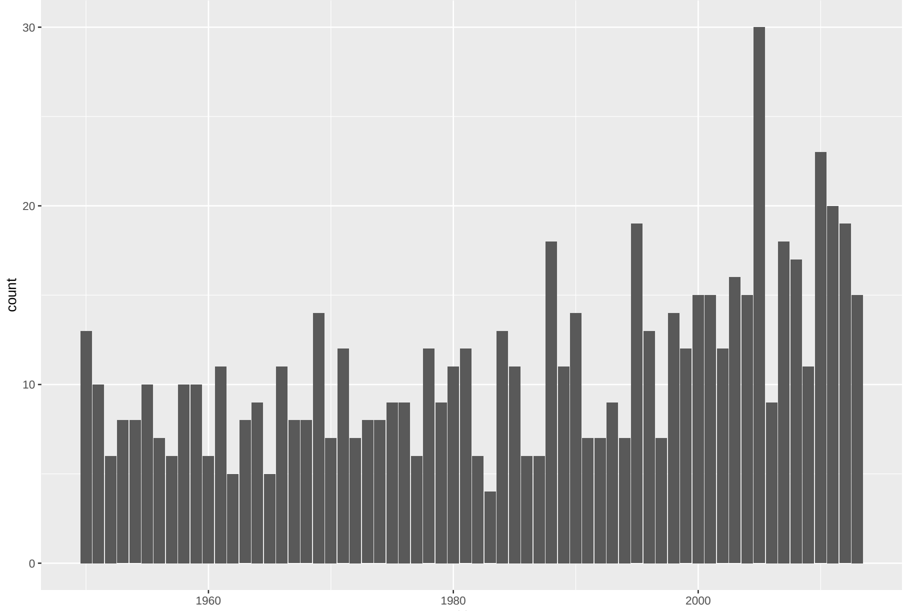
Dataset

- Hurrican703 dataset: 22038 observations $\times$ 8 variables
  - 702 hurricanes in the North Atlantic area since 1950

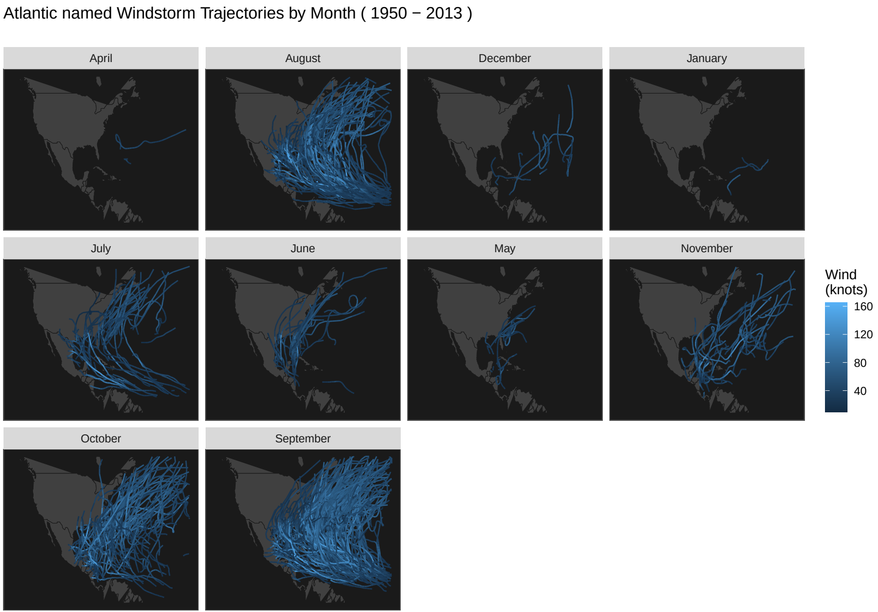
EDA-Count of Hurricanes in each Month



EDA-Count of Hurricanes in each Year



Show hurricance tracks by month



Bayesian Model

The suggested Bayesian model is $Y_i(t+6) = \beta_{0,i} + \beta_{1,i}Y_i(t) + \beta_{2,i}\Delta_{i,1}(t) + \beta_{3,i}\Delta_{i,2}(t) + \beta_{4,i}\Delta_{i,3}(t) + X_i\gamma + \epsilon_i(t)$

- 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.
- $\beta_i = (\beta_{0,i}, \beta_{1,i}, ..., \beta_{5,i})$, we assume that $\beta_i \sim N(\mu, \Sigma)$, where d is dimension of β_i.

Priors

$$P(\mu) = \frac{1}{\sqrt{2\pi}|V|^{\frac{1}{2}}} \exp\{-\frac{1}{2}\mu^\top V^{-1}\mu\} \propto |V|^{-\frac{1}{2}} \exp\{-\frac{1}{2}\mu^\top V^{-1}\mu\}$$

where V is a variance-covariance matrix

$$P(\Sigma) \propto |\Sigma|^{-\frac{(\nu+d+1)}{2}} \exp(-\frac{1}{2}tr(S\Sigma^{-1}))$$

$$P(\gamma) \propto exp(-\frac{\gamma^2}{2*(0.05)^2}) = e^{-200\gamma^2}$$

$$P(\sigma) = \frac{2\alpha}{\pi + \alpha^2} \propto \frac{1}{\sigma^2 + \alpha^2}$$

Posterior

Let $\mathbf{B} = (\beta_1^\top, ..., \beta_n^\top)^\top$, derive the posterior distribution of the parameters $\Theta = (\mathbf{B}^\top, \mu^\top, \sigma^2, \Sigma, \gamma)$.

Let

$$Z_i(t)\beta_i^\top = \beta_{0,i} + \beta_{1,i}Y_i(t) + \beta_{2,i}\Delta_{i,1}(t) + \beta_{3,i}\Delta_{i,2}(t) + \beta_{4,i}\Delta_{i,3}(t) + X_i\gamma + \epsilon_i(t)$$

We can find that

$$Y_i \sim MVN(Z_i\beta_i, \sigma^2 I)$$

The likelihood for our data is

$$f(Y \mid B, \mu, \sigma^2, \Sigma, \gamma) =$$

$$\prod_{i=1}^{N} f(Y_i|B, \mu, \Sigma, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma}} \exp\{-\frac{1}{2}(y_i - Z_i\beta_i - X_i\gamma_i)^\top(\sigma^2 I)^{-1}(y_i - Z_i\beta_i - X_i\gamma_i)\}$$

$$\propto (2\pi\sigma^2)^{-\frac{N}{2}} \prod_{i=1}^{n} \exp\{-\frac{1}{2}(Y_i - Z_i\beta_i - X_i\gamma_i)^\top(\sigma^2 I)^{-1}(Y_i - Z_i\beta_i - X_i\gamma$$

where N is the total number of hurricanes. ## Joint Posterior

Conditional Distributions

Generate B_t from $f(B_t, \sigma_{t-1}, \mu_{t-1}, \Sigma_{t-1}^{-1})$

$$
\begin{aligned}
f(B^\top | \sigma^2, \mu, \Sigma) &\propto \prod_{i=1}^{n} \exp\{-\frac{(Y_j - Z_j\beta_j - X_j\gamma)^\top (Y_j Z_j \beta_j - X_j\gamma)}{2\sigma^2} - \frac{(\beta_j - \mu)^\top \Sigma^{-1} (\beta_j - \mu)}{2}\} \\
&\propto \prod_{i=1}^{n} \exp(-\frac{1}{2}(\beta_j^\top (\frac{Z_j^\top Z_j}{\sigma^2} + \Sigma^{-1})\beta_j - 2(\frac{Y_j^\top Z_j - X_j^\top Z_j \gamma}{\sigma^2} + \mu^\top \Sigma^{-1})\beta_j)) \\
&\propto \prod_{i=1}^{n} \exp\{-\frac{1}{2}(\beta_j - (\frac{Z_j^\top Z_j}{\sigma^2} + \Sigma^{-1})(\beta_j - (\frac{Z_j^\top Z_j}{\sigma^2} + \Sigma^{-1})^{-1}(\frac{Y_j^\top Z_j - X_j^\top Z_j \gamma}{\sigma^2}
\end{aligned}
$$

$$
MVN_d((\frac{Z_j^\top Z_j}{\sigma^2} + \Sigma^{-1})^{-1}(\frac{Y_j^\top Z_j^\top - X_j^\top Z_j \gamma}{\sigma^2} + \mu^\top \Sigma^{-1})^\top, \frac{Z_j^\top Z_j}{\sigma^2} + \Sigma^{-1})^{-1})
$$

MCMC Algorithm

$$\beta_i \sim MVN_d(N^{-1}M^\top, N^{-1})$$

where $N = \frac{Z_i^\top Z_i}{\sigma^2} + \Sigma^{-1}$ and $M = \frac{Y_i^\top Z_i^\top - X_i^\top Z_i \gamma}{\sigma^2} + \mu^\top \Sigma^{-1}$

$$\mu_t \sim MVN_d(N^{-1}M^\top, N^{-1})$$

where $N = NA + \frac{1}{v}$ and $M = \sum_i^n \beta_i A$, and v is the degree of freedom.

$$\Sigma \sim w^{-1}(S + \sum_i^n (\beta_i - \mu)(\beta_i - \mu)^\top, n + v)$$

$$\gamma \sim MVN(N^{-1}M^\top, N^{-1})$$

where $N = \frac{X_i^\top X_i}{\sigma^2} + 400I$ and $M = \sum_i^n Y_i^\top X_i - \sum_i^n X_i Z_i \beta_i$

MCMC Algorithm - Metropolis-Hastings

*Target distribution is

$$\pi(\sigma|Y, \mathbf{B}^\top, \mu^\top, \Sigma, \gamma) \propto \frac{1}{\sigma^N(\sigma^2 + 10^2)} \times \prod_{i=1}^{n} \exp\{-\frac{1}{2(\sigma^2 I)}(Y_i - Z_i\beta_i - X_i\gamma_i)^\top (Y_i - Z_i\beta_i - X_i\gamma_i)\}$$

- Choose a random walk with step size distributed as a uniform random variable
- The conditional density is $q(x|y) = \frac{1}{2a} 1_{[y-a,y+a]}(x)$
- Proposed q is symmetric, thus the acceptance rate is only depend on $P(\sigma|B, \mu, A, \gamma, Y)$

MCMC Algorithm - Metropolis-Hastings

- The acceptance rate $\alpha_{XY} = \min(1, \frac{P(X|B,\mu,A,\gamma,Y)}{P(Y|B,\mu,A,\gamma,Y)})$
- Accept X if $U < \alpha_{XY}$
- Iterate over 1000 times
- New σ is the mean of last 200 values in the chain

MCMC Algorithm - Gibbs Sampling

We apply a MCMC algorithm consisting of Gibb Samping and Metropolis-Hastings steps.

Parameters are updated component-wise for each $k = 1, ..., N, N = 5000$:

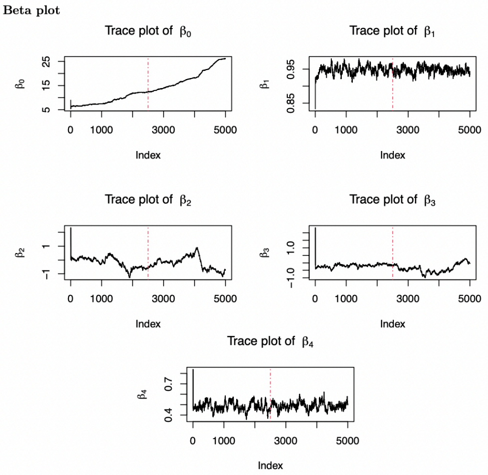
- Generate $\beta_{ij}, j = 0, 1, 2, 3, 4$ for i^{th} hurricane from $\pi(\mathbf{B}|Y, \mu_{k-1}^{\top}, \sigma_{k-1}, \Sigma_{k-1}, \gamma_{k-1})$
- Generate $\mu_j, j = 0, 1, 2, 3, 4$ from $\pi(\mu|Y, \mathbf{B}_k, \sigma_{k-1}, \Sigma_{k-1}, \gamma_{k-1})$
- Generate σ_k from the Metropolis-Hastings steps
- Generate Σ_k from $\pi(\Sigma|Y, \mathbf{B}_k, \mu_k, \sigma_k, \gamma_{k-1})$
- Generate γ_k from $\pi(\gamma|Y, \mathbf{B}_k, \mu_k, \sigma_k, \Sigma_k)$

MCMC Algorithm - Initial Values

We first fit a Generalize Linear Mixed Models(GLMM)

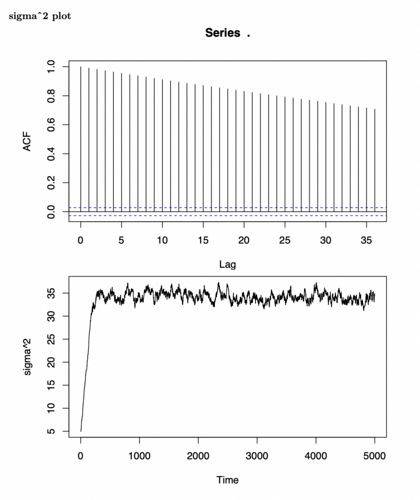
- $\beta_i^{(0)}$: The random effect for i^{th} hurricane from GLMM as start values
- $\mu^{(0)}$: Average over $\beta_i^{(0)}$
- $\sigma^{(0)}$: Residuals from the GLMM
- $\Sigma^{(0)}$: Variance-Covariance matrix of $\beta_i^{(0)}$
- $\gamma^{(0)}$: Fixed effects from the GLMM

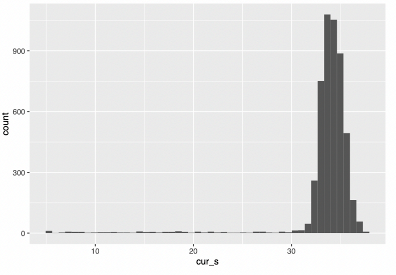
MCMC Results - Beta Plots



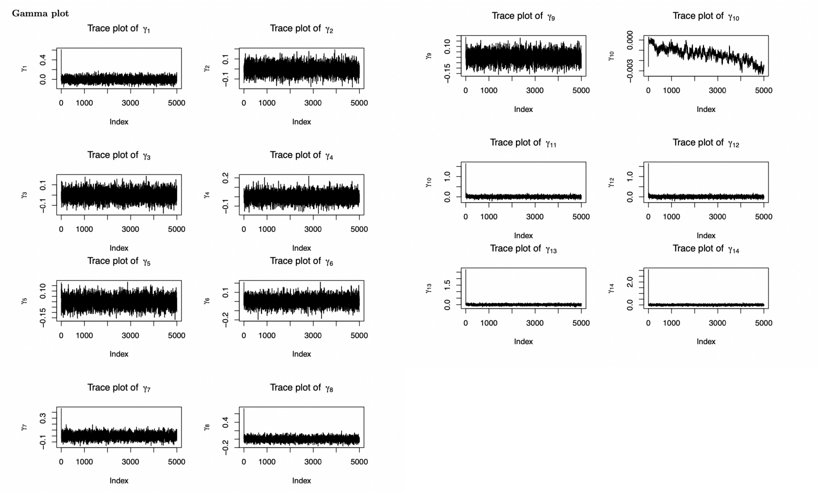
-Trace plots of variance parameters, based on 5000 MCMC sample.

MCMC Results - sigma^2 Plots



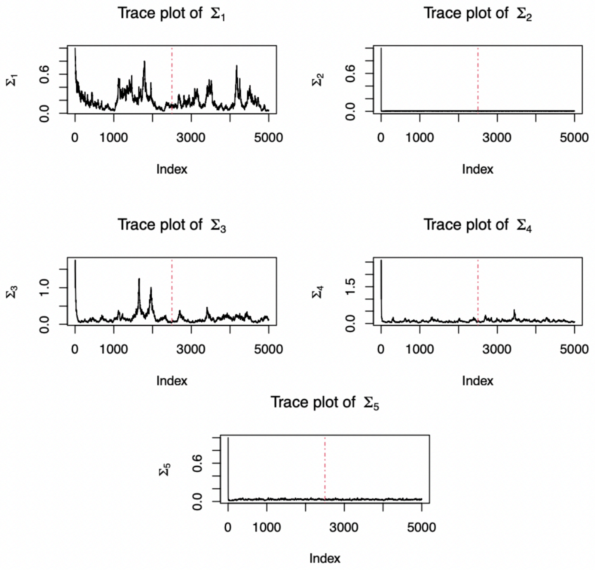


MCMC Results - Gamma Plots



MCMC Results - Sigma Plots

Sigma_inverse plot

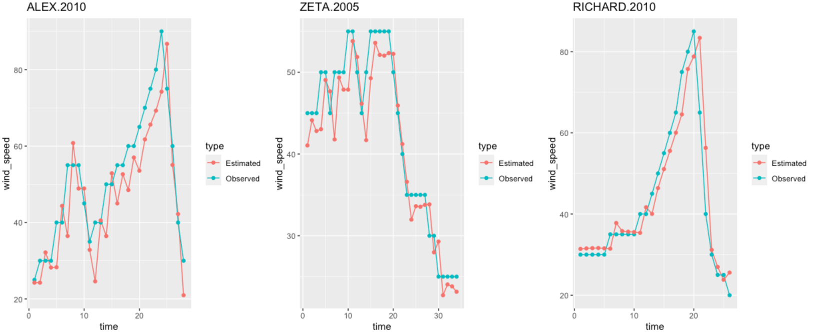


Bayesian Model Performance

-The overall mean RMSE is 6.467.

	ID	r_square	rmse
1	SUBTROP:UNNAMED.1974	0.655	4.867
2	JEANNE.1980	0.921	5.437
3	FRANCES.2004	0.978	5.628
4	CHANTAL.1995	0.947	2.388
5	ETHEL.1960	0.473	27.218
6	PHILIPPE.2011	0.843	5.598
7	JOSEPHINE.1984	0.956	4.095
8	FRANCES.1976	0.895	6.114
9	BEULAH.1963	0.930	3.873
10	HOLLY.1969	0.873	5.670
11	ISAAC.2000	0.957	5.631
12	DAVID.1979	0.949	7.899
13	ALMA.1966	0.913	6.557
14	ERIN.1995	0.883	8.036
15	ANA.1997	0.880	2.156
16	DEBBIE.1969	0.851	8.869
17	HARVEY.2005	0.941	2.836
18	ALLISON.1995	0.768	4.339
19	LAURA.1971	0.967	2.112
20	EDNA.1968	0.957	2.006

Bayesian Model Performance



Estimated Wind Speed vs. Predicted Wind Speed