Baysian modeling of hurricane trajectories

P8160 Group Project 3 Baysian modeling of hurricane trajectories

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

Columbia University

2023-04-30

Content

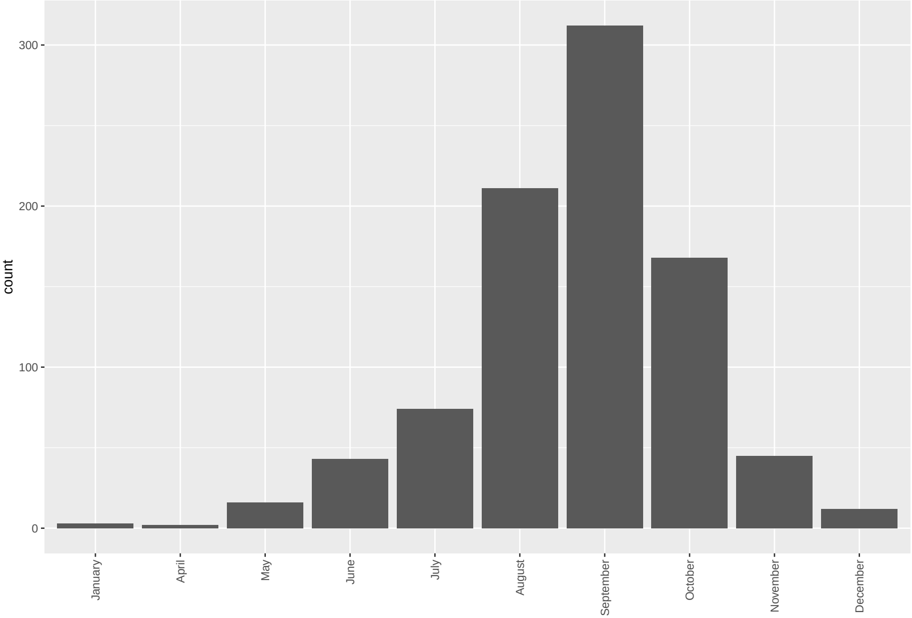
Introduction

- 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.

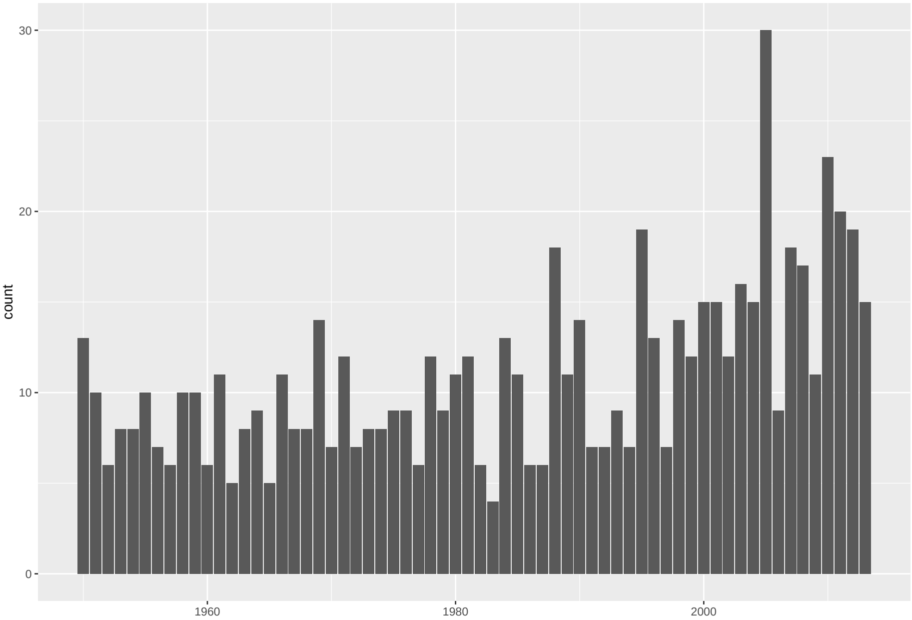
** Dataset **

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

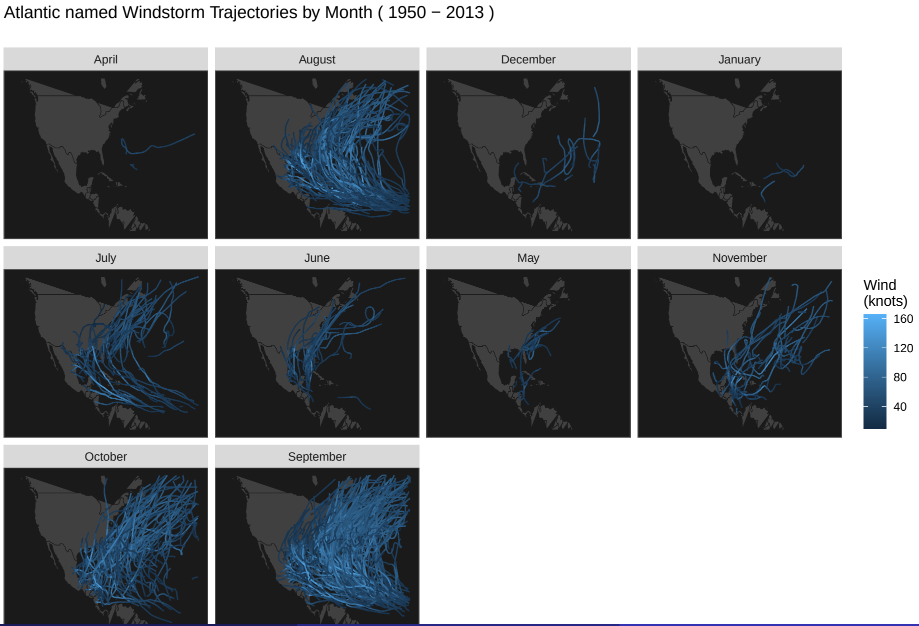
EDA-Count of Hurricanes in each Month



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

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

$$P(\Sigma^{-1}) \propto |\Sigma|^{-\frac{(\nu+d+1)}{2}} \exp(-\frac{1}{2}tr(S\Sigma^{-1})) \quad (3)$$

$$P(\gamma) \propto exp(-\frac{\gamma^2}{2 * (0.05)^2}) = e^{-200\gamma^2} \quad (4)$$

$$P(\sigma) = \frac{2\alpha}{\pi + \alpha^2} \propto \frac{1}{\sigma^2 + \alpha^2} \quad (5)$$

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

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

We can find that

$$Y_i \sim MVN(Z_i\beta_i, \sigma^2 I) \tag{6}$$

The likelihood for our data is

$$L(Y_i \mid B_i, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma}} \exp\{-\frac{1}{2}(y_i - Z_i\beta_i - X_i\gamma_i)^\top(\sigma^2 I)^{-1}(y_i - Z_i\beta_{i}$$

$$\propto(\sigma^{-\frac{N}{2}}) \prod_{i=1}^{n} \exp\{ -\frac{1}{2}(Y_i - Z_i\beta_i - X_i\gamma_i)^\top(\sigma^2 I)^{-1}(Y_i - Z \tag{7}$$

Joint Posrerior

$$
\begin{aligned}
P(\Theta|Y) =& P(B, \mu, \sigma^2, A, \gamma|Y) \\
& \propto L(Y|B, \sigma^2) L(B|\mu, \Sigma) p(\mu) p(\sigma) p(A) p(\gamma) \\
& \propto \frac{1}{\sigma^N(\sigma^2 + 10^2)} \prod_{i=1}^{n} \exp\{-\frac{1}{2}(Y_i - Z_i\beta_i - X_i\gamma_i)^\top(\sigma^2 I)^{-1}(Y_i \\
& \times \exp\{-\frac{1}{2}\sum_{i}^{n}(\beta_i - \mu)^\top A(\beta_i - \mu)\}|A|^{\frac{N+d+6}{2}} \exp\{-\frac{1}{2}tr(SA)\}|V
\end{aligned}
\tag{9}
$$

Conditional Distributions

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

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

$$\propto \prod_{i=1}^{n} \exp(-\frac{1}{2}(\beta_j^\top(\frac{Z_j^\top Z_j}{\sigma^2} + A)\beta_j - 2(\frac{Y_j^\top Z_j - X_j^\top}{\sigma^2}$$

$$\propto \prod_{i=1}^{n} \exp\{-\frac{1}{2}(\beta_j - (\frac{Z_j^\top Z_j}{\sigma^2} + A)(\beta_j - (\frac{Z_j^\top Z_j}{\sigma^2} + A)^{-1}(\frac{Y_j^\top Z_j - X_j^\top}{\sigma^2}$$

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

MCMC Algorithm

$$f(\mathbf{B}|\mu^\top, \sigma^2, A, \gamma, Y^\top) \sim MVN_d((\frac{Z_j^\top Z_j}{\sigma^2} + A)^{-1}(\frac{Y_j^\top Z_j^\top - X_j^\top Z_j\gamma}{\sigma^2} + \mu^\top A)^\top,$$

$$f(\mu_t|B^\top, \sigma^2, A, \gamma, Y^\top) \sim MVN_d((NA + \frac{1}{v})^{-1}(\sum_i^n \beta_j A)^\top, (NA + \frac{1}{v})^{-1})$$

$$f(\sigma|B^\top, \mu^\top, A, \gamma, Y^\top) \sim MVN_d((X_i^\top \sigma^{-2} I X_i + 400I)^{-1}(X_i^\top \sigma^{-2} I Y_i - X_i^\top \sigma$$

$$f(A|B^\top \sigma^2, \gamma, Y^\top) \sim w^{-1}(S + \sum_i^n (\beta_i - \mu)(\beta_i - \mu)^\top, n + v)$$

$$f(\gamma|B^\top, \mu^t op, \sigma^2, A, Y^\top) \sim MVN((\frac{X_j^\top X_j}{\sigma^2} + 400I)^{-1}(\sum^n Y_j^\top X_j - \sum^n X_j Z$$

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

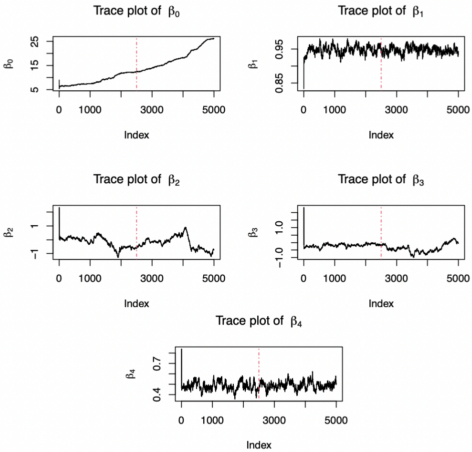
We first fit a Generalize Linear Mixed Models(GLMM)

- $\beta_i^{(0)}$: 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

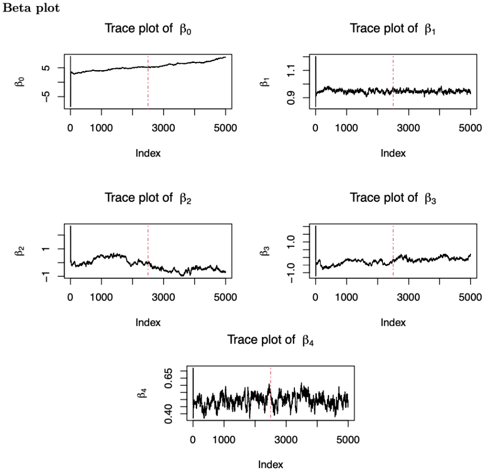
MCMC Results - Beta Plots 1

- 5000 iterations
- Estimates and inferences based on 5000 MCMC samples

Beta plot



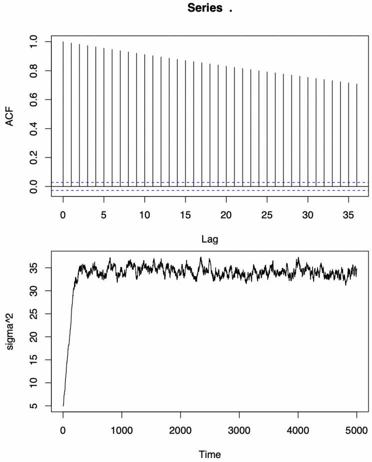
MCMC Results - Beta Plots 2

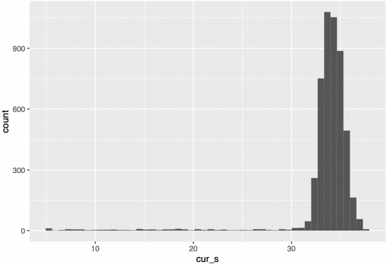


-Trace plots of variance parameters, based on 5000 MCMC sample.

MCMC Results - sigma^2 Plots 1

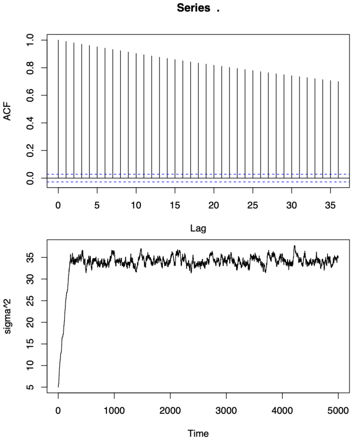
sigma^2 plot

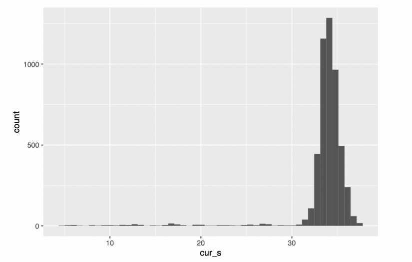




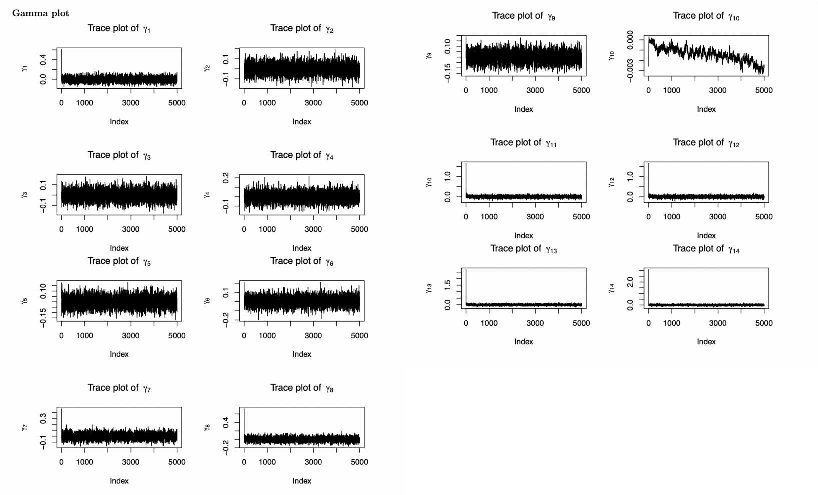
MCMC Results - sigma^2 Plots 2

sigma^2 plot

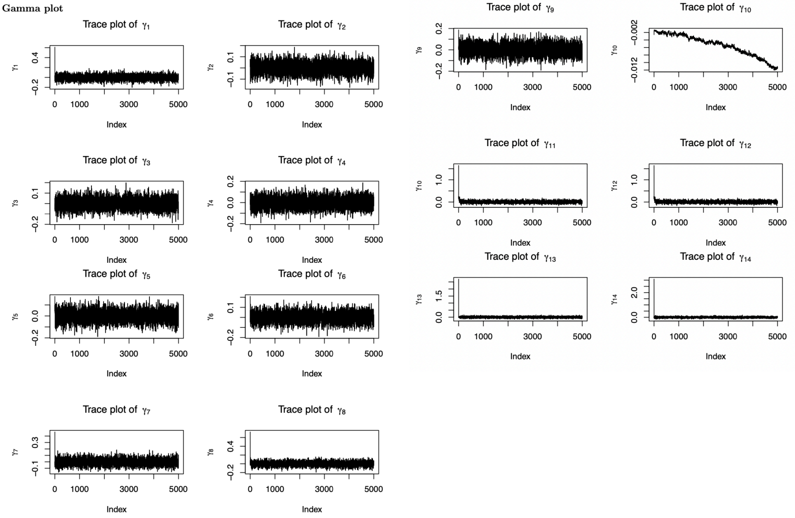




MCMC Results - Gamma Plots 1

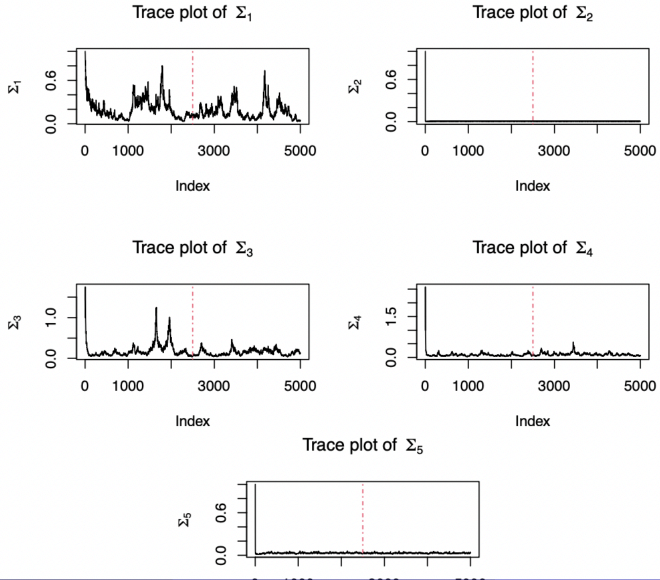


MCMC Results - Gamma Plots 2



MCMC Results - Sigma Inverse Plots 1

Sigma_inverse plot



MCMC Results - Sigma Inverse Plots 2

Sigma_inverse plot