

# Part\_1-4

Tianchuan Gao, Ruiqi Yan, Jiacheng Wu

5/4/2022

## Hurrican Data

hurricane703.csv collected the track data of 703 hurricanes in the North Atlantic area since 1950. For all the storms, their location (longitude & latitude) and maximum wind speed were recorded every 6 hours. The data includes the following variables

1. **ID**: ID of the hurricanes
2. **Season**: In which **year** the hurricane occurred
3. **Month**: In which **month** the hurricane occurred
4. **Nature**: Nature of the hurricane
  - ET: Extra Tropical
  - DS: Disturbance
  - NR: Not Rated
  - SS: Sub Tropical
  - TS: Tropical Storm
5. **time**: dates and time of the record
6. **Latitude** and **Longitude**: The location of a hurricane check point
7. **Wind.kt** Maximum wind speed (in Knot) at each check point

## load and clean data

### Part I: Likelihood, Prior, Posterior and Conditional Posterior

#### Likelihood

$$L(\mathbf{Y} | \boldsymbol{\theta}) \propto \prod_{i=1}^m (\sigma^2)^{-\frac{n_i}{2}} \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}_i)^T (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}_i) \right\}$$

where  $m$  is the number of hurricane and  $n_i$  is the number of observations for  $i^{th}$  hurricane

#### Prior

$$\pi(\boldsymbol{\theta} = (\mathbf{B}, \boldsymbol{\mu}, \Sigma^{-1}, \sigma^2)) \propto (\sigma^2)^{-1} |\Sigma^{-1}|^{d+1} \exp \left\{ -\frac{1}{2} \text{tr}(\Sigma^{-1}) \right\} \prod_{i=1}^m |\Sigma^{-1}|^{\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\boldsymbol{\beta}_i - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{\beta}_i - \boldsymbol{\mu}) \right\}$$

where  $d$  is the dimension of  $\boldsymbol{\mu}$

## Posterior

$$\begin{aligned}
\pi(\boldsymbol{\theta} \mid \mathbf{Y}) &\propto (\sigma^2)^{-1} |\Sigma^{-1}|^{d+1} \exp \left\{ -\frac{1}{2} + \text{tr}(\Sigma^{-1}) \right\} \\
&\times \prod_{i=1}^m (\sigma^2)^{-\frac{n_i}{2}} |\Sigma^{-1}|^{\frac{1}{2}} \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}_i)^T (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}_i) \right\} \exp \left\{ -\frac{1}{2} (\boldsymbol{\beta}_i - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{\beta}_i - \boldsymbol{\mu}) \right\} \\
&= (\sigma^2)^{-\left(1 + \frac{\sum_{i=1}^m n_i}{2}\right)} |\Sigma^{-1}|^{d+1+\frac{m}{2}} \exp \left\{ -\frac{1}{2} \text{tr}(\Sigma^{-1}) \right\} \exp \left\{ -\frac{1}{2} \sum_{i=1}^m (\boldsymbol{\beta}_i - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{\beta}_i - \boldsymbol{\mu}) \right\} \\
&\times \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^m (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}_i)^T (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}_i) \right\}
\end{aligned}$$

## Conditional Posterior

$$\boldsymbol{\beta}_i : \pi(\boldsymbol{\beta}_i \mid \boldsymbol{\theta}_{(-\boldsymbol{\beta}_i)} \mathbf{Y}) \propto \exp \left\{ -\frac{1}{2} (\boldsymbol{\beta}_i - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{\beta}_i - \boldsymbol{\mu}) - \frac{1}{2\sigma^2} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}_i)^T (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}_i) \right\}$$

$$\boldsymbol{\mu} : \pi(\boldsymbol{\mu} \mid \boldsymbol{\theta}_{(-\boldsymbol{\mu})}, \mathbf{Y}) \propto \exp \left\{ -\frac{1}{2} \sum_{i=1}^m (\boldsymbol{\beta}_i - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{\beta}_i - \boldsymbol{\mu}) \right\}$$

$$\begin{aligned}
(\boldsymbol{\beta}_i - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{\beta}_i - \boldsymbol{\mu}) &= \text{tr} \left( (\boldsymbol{\beta}_i - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{\beta}_i - \boldsymbol{\mu}) \right) \\
&= \text{tr} \left( \Sigma^{-1} (\boldsymbol{\beta}_i - \boldsymbol{\mu}) (\boldsymbol{\beta}_i - \boldsymbol{\mu})^T \right)
\end{aligned}$$

$$\begin{aligned}
\pi(\boldsymbol{\mu} \mid \boldsymbol{\theta}_{(-\boldsymbol{\mu})}, \mathbf{Y}) &\propto \exp \left\{ -\frac{1}{2} \text{tr} \left( \Sigma^{-1} \sum_{i=1}^m (\boldsymbol{\beta}_i - \boldsymbol{\mu}) (\boldsymbol{\beta}_i - \boldsymbol{\mu})^T \right) \right\} \\
&= \exp \left\{ -\frac{1}{2} \text{tr} \left( \Sigma^{-1} m (\boldsymbol{\mu} - \bar{\boldsymbol{\beta}}) (\boldsymbol{\mu} - \bar{\boldsymbol{\beta}})^T \right) \right\} \\
&= \exp \left\{ -\frac{1}{2} (\boldsymbol{\mu} - \bar{\boldsymbol{\beta}})^T \Sigma^{-1} m (\boldsymbol{\mu} - \bar{\boldsymbol{\beta}}) \right\} \\
&\Rightarrow \boldsymbol{\mu} \mid \boldsymbol{\theta}_{(-\boldsymbol{\mu})}, \mathbf{Y} \sim N(\bar{\boldsymbol{\beta}}, \Sigma/m)
\end{aligned}$$

$$\sigma^2 : \pi(\sigma^2 \mid \boldsymbol{\theta}_{(-\sigma^2)}, \mathbf{Y}) \propto (\sigma^2)^{-\left(1 + \frac{\sum_{i=1}^m n_i}{2}\right)} \times \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^m (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}_i)^T (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}_i) \right\}$$

$$\begin{aligned}
\Sigma^{-1} : \pi(\Sigma^{-1} \mid \boldsymbol{\theta}_{(-\Sigma^{-1})}, \mathbf{Y}) &\propto |\Sigma^{-1}|^{d+1+\frac{m}{2}} \exp \left\{ -\frac{1}{2} \text{tr}(\Sigma^{-1}) \right\} \exp \left\{ -\frac{1}{2} \sum_{i=1}^m (\boldsymbol{\beta}_i - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{\beta}_i - \boldsymbol{\mu}) \right\} \\
&= |\Sigma^{-1}|^{d+1+\frac{m}{2}} \exp \left\{ -\frac{1}{2} \text{tr}(\Sigma^{-1}) - \frac{1}{2} \text{tr} \left( \Sigma^{-1} \sum_{i=1}^m (\boldsymbol{\beta}_i - \boldsymbol{\mu}) (\boldsymbol{\beta}_i - \boldsymbol{\mu})^T \right) \right\} \\
&= |\Sigma^{-1}|^{d+1+\frac{m}{2}} \exp \left\{ -\frac{1}{2} \text{tr} \left( \Sigma^{-1} \left( \mathbf{I} + \sum_{i=1}^m (\boldsymbol{\beta}_i - \boldsymbol{\mu}) (\boldsymbol{\beta}_i - \boldsymbol{\mu})^T \right) \right) \right\} \\
&\Rightarrow \Sigma^{-1} \mid \boldsymbol{\theta}_{(-\Sigma^{-1})}, \mathbf{Y} \sim \text{Wishart} \left( 3d + 3 + m, \left( \mathbf{I} + \sum_{i=1}^m (\boldsymbol{\beta}_i - \boldsymbol{\mu}) (\boldsymbol{\beta}_i - \boldsymbol{\mu})^T \right)^{-1} \right)
\end{aligned}$$

## Part II: MCMC Algorithm

We could generate full conditional posterior distribution for some parameters but not all, so we apply hybrid algorithm consisting with Metropolis-Hastings steps and Gibbs steps.

MH steps for  $\beta_{ij}$ : Sampling proposed  $\beta'_{ij}$ ,  $j = 0, 1, \dots, 4$  for  $i^{th}$  hurricane from proposal distribution  $U\left(\beta_{ij}^{(t)} - a_{ij}, \beta_{ij}^{(t)} + a_{ij}\right)$ , where  $a_{ij}$  is the search window for  $\beta_{ij}$ . Since the proposals are symmetry, the accepting or rejecting the proposed  $\beta'_{ij}$  depends on the ratio of posterior distribution. Some of the parameters in  $\theta$  could be cancelled out, so the ratio simplified to be:

$$\frac{\pi\left(\beta'_i, \theta_{(-\beta_i)}^{(t)} \mid \mathbf{Y}\right)}{\pi\left(\beta_i^{(t)}, \theta_{(-\beta_i)}^{(t)} \mid \mathbf{Y}\right)} = \frac{\exp\left\{-\frac{1}{2\sigma^{2(t)}}\left(\mathbf{Y}_i - \mathbf{X}_i\beta'_i\right)^T\left(\mathbf{Y}_i - \mathbf{X}_i\beta'_i\right) - \frac{1}{2}\left(\beta'_i - \mu^{(t)}\right)^T \Sigma^{-1(t)}\left(\beta'_i - \mu^{(t)}\right)\right\}}{\exp\left\{-\frac{1}{2\sigma^{2(t)}}\left(\mathbf{Y}_i - \mathbf{X}_i\beta_i^{(t)}\right)^T\left(\mathbf{Y}_i - \mathbf{X}_i\beta_i^{(t)}\right) - \frac{1}{2}\left(\beta_i^{(t)} - \mu^{(t)}\right)^T \Sigma^{-1(t)}\left(\beta_i^{(t)} - \mu^{(t)}\right)\right\}}$$

where  $\beta'_i$  consisting with  $\beta'_{ij}$ ,  $\beta'_{ik}$  for  $k > j$  and  $\beta_{ik}^{(t+1)}$  for  $k < j$ .

The log of the ratio:

$$\begin{aligned} \log \frac{\pi\left(\beta'_i, \theta_{(-\beta_i)}^{(t)} \mid \mathbf{Y}\right)}{\pi\left(\beta_i^{(t)}, \theta_{(-\beta_i)}^{(t)} \mid \mathbf{Y}\right)} &= -\frac{1}{2} \left( \frac{\left(\mathbf{Y}_i - \mathbf{X}_i\beta'_i\right)^T\left(\mathbf{Y}_i - \mathbf{X}_i\beta'_i\right)}{\sigma^{2(t)}} + \left(\beta'_i - \mu^{(t)}\right)^T \Sigma^{-1(t)}\left(\beta'_i - \mu^{(t)}\right) \right) \\ &\quad + \frac{1}{2} \left( \frac{\left(\mathbf{Y}_i - \mathbf{X}_i\beta_i^{(t)}\right)^T\left(\mathbf{Y}_i - \mathbf{X}_i\beta_i^{(t)}\right)}{\sigma^{2(t)}} + \left(\beta_i^{(t)} - \mu^{(t)}\right)^T \Sigma^{-1(t)}\left(\beta_i^{(t)} - \mu^{(t)}\right) \right) \end{aligned}$$

Then we randomly sample  $u$  from  $U(0, 1)$  and compare  $\log(u)$  with the log ratio. If the  $\log(u)$  is smaller, we accept  $\beta'_{ij} = \beta_{ij}^{(t+1)}$ , otherwise we reject  $\beta'_{ij}$  and  $\beta_{ij}^{(t)} = \beta_{ij}^{(t+1)}$ .

Then, Gibb step for  $\mu$ : Sample  $\mu^{(t+1)}$  from  $N\left(\bar{\beta}^{(t+1)}, \Sigma^{(t)}/m\right)$ , where  $\bar{\beta}^{(t+1)}$  is the average over  $\beta_i^{(t+1)}$ .

Next, MH step to generate  $\sigma^{2'}$  from  $U\left(\sigma^{2(t)} - a_{\sigma^2}, \sigma^{2(t)} + a_{\sigma^2}\right)$ . Firstly, check whether  $\sigma^{2'}$  is positive, if not, we reject  $\sigma^{2'}$ . Then, we randomly sample  $u$  from  $U(0, 1)$  and compare  $\log(u)$  with the log posterior ratio. The log posterior ratio

$$\begin{aligned} \log \frac{\pi\left(\sigma^{2'}, \theta_{(-\sigma^2)}^{(t)} \mid \mathbf{Y}\right)}{\pi\left(\sigma^{2(t)}, \theta_{(-\sigma^2)}^{(t)} \mid \mathbf{Y}\right)} &= -\left(1 + \frac{M}{2}\right) \log(\sigma^{2'}) - \frac{1}{2\sigma^{2'}} \sum_{i=1}^m \left(\mathbf{Y}_i - \mathbf{X}_i\beta_i^{(t+1)}\right)^T \left(\mathbf{Y}_i - \mathbf{X}_i\beta_i^{(t+1)}\right) \\ &\quad + \left(1 + \frac{M}{2}\right) \log(\sigma^{2(t)}) + \frac{1}{2\sigma^{2(t)}} \sum_{i=1}^m \left(\mathbf{Y}_i - \mathbf{X}_i\beta_i^{(t+1)}\right)^T \left(\mathbf{Y}_i - \mathbf{X}_i\beta_i^{(t+1)}\right) \end{aligned}$$

where  $M$  is total number of observation for all hurricanes. If the  $\log(u)$  is smaller, we accept  $\sigma^{2'} = \sigma^{2(t+1)}$ .

Finally, we sample  $\Sigma^{-1(t+1)}$  from Wishart  $\left(3d + 3 + m, \left(\mathbf{I} + \sum_{i=1}^m \left(\beta_i^{(t+1)} - \mu^{(t+1)}\right) \left(\beta_i^{(t+1)} - \mu^{(t+1)}\right)^T\right)^{-1}\right)$

### Part III: Implementation and Estimations

#### Initial Value

Initial Values:

- $\beta_i$  - Fit multivariate linear regression (MLR) for  $i^{th}$  hurricane and use the coefficients as  $\beta_{i0}$
- $\mu$  - Average over all  $\beta_{i0}$  as  $\mu_0$
- $\sigma^2$  -  $\hat{\sigma}_i^2$  is the mean square residuals of the MLR model on  $i^{th}$  hurricane. Take the mean over all  $\hat{\sigma}_i^2$  as  $\sigma_0^2$
- $\Sigma^{-1}$  - Generate the covaraiance matrix of  $\beta_{i0}$  and take the inverse of the matrix as  $\Sigma_0^{-1}$

## Implement 10000 MCMC iterations

```
set.seed(1971)
theta_chain = componentwisemixedMH(a, a_sigma = 2, beta_0 = starting_beta,
                                   sigma_0, cov_0, nrep=10000)
```

## Search Window and Acceptance Rate

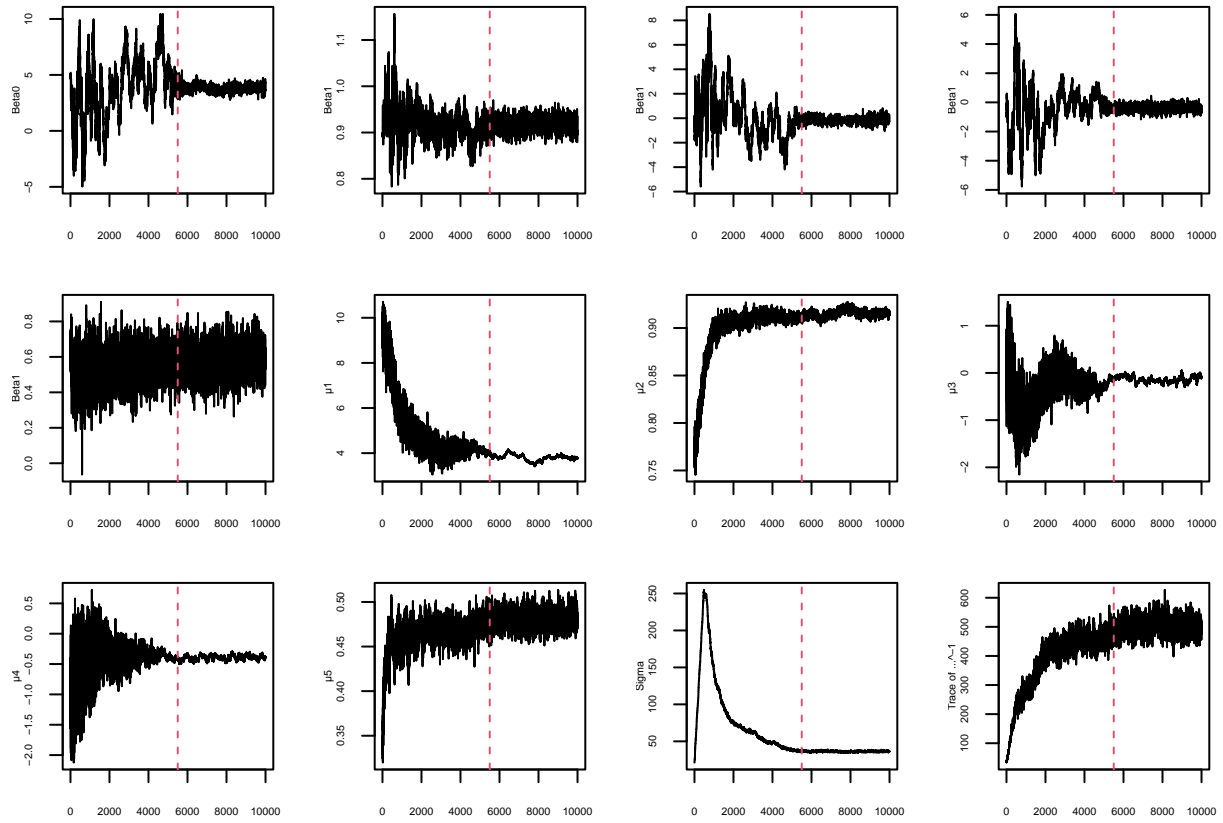
There is MH step in the MCMC algorithm so the choice of search window is important. The acceptance rate of the MH step is around 0.3167 to 0.6368, so the search window of our algorithm is appropriate. We tune the search windows multiple times to achieve this result. Table 1 demonstrates the range of search windows  $a$  and associated acceptance rates for  $\beta$  and  $\sigma^2$ .

Table 1: Range of Search Window and Acceptance Rate for parameters used MH step

	search window	acceptance rate (%)
$\beta_0$	1.1	45.87-51.36
$\beta_1$	0.04-0.1	31.67-63.68
$\beta_2$	0.8-1	38.6-45.6
$\beta_3$	0.5-0.6	33.2-61.32
$\beta_4$	0.4-0.5	34.95-60.45
$\sigma^2$	2	44.93

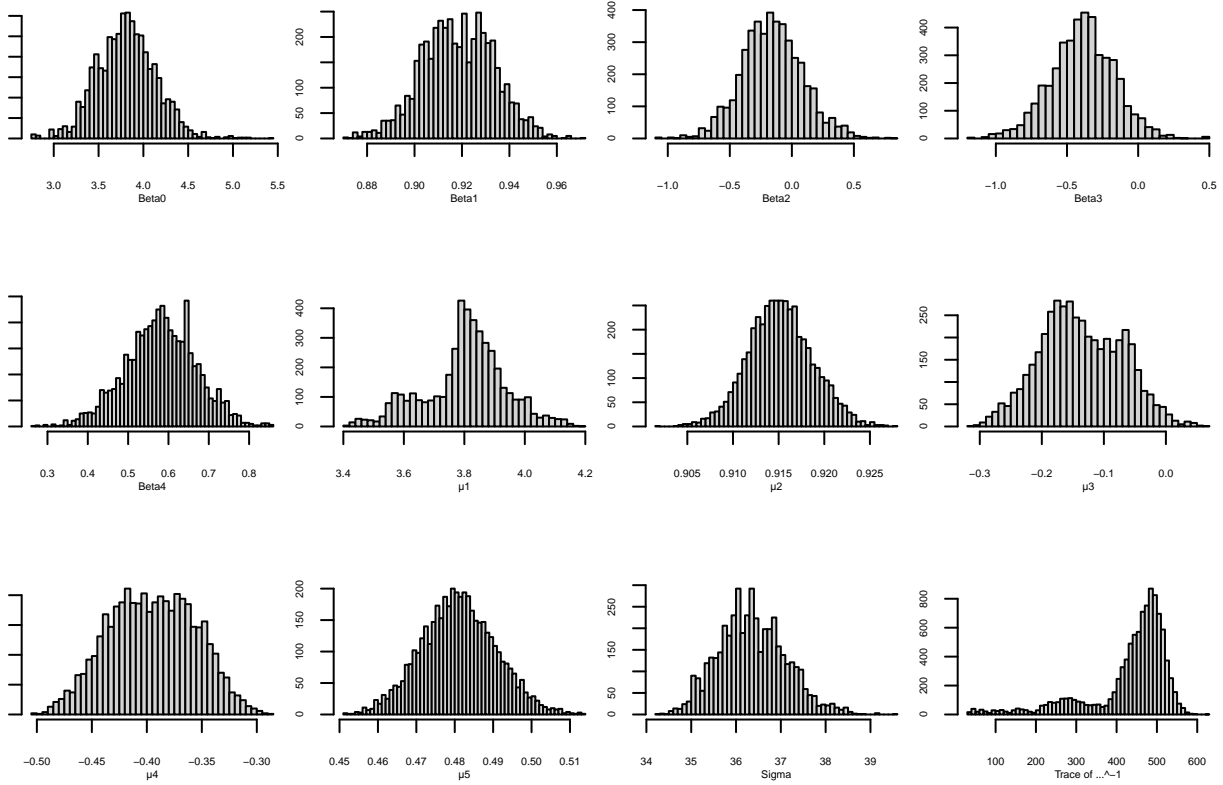
## Convergence trace plots

Covergence Plots of Selected Parameters



## Histograms

Histograms of Selected Parameters



Estimates (burn-in first 5500 runs)

Histograms of estimated Betas of All Hurricanes

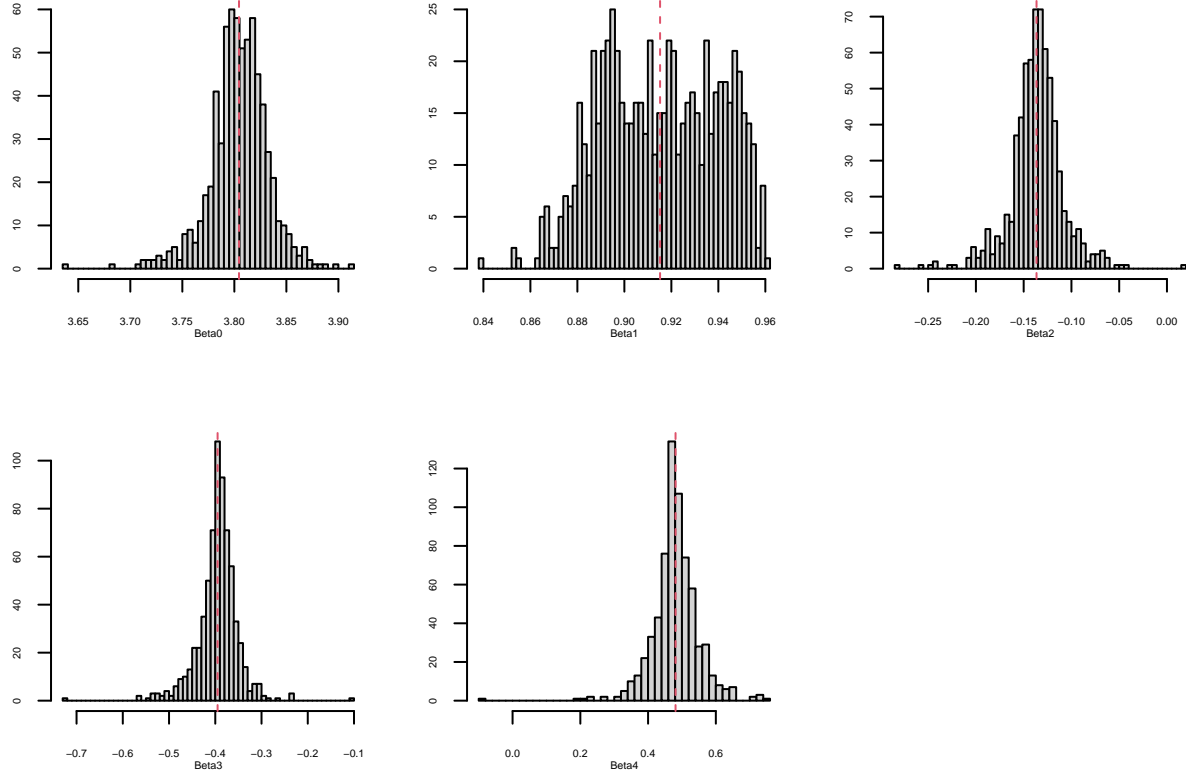


Chart of the estimates of mu, sigma and inverseCov

Table 2: Bayesian Estimates for  $\mu$  and  $\sigma^2$

	$\mu_0$	$\mu_1$	$\mu_2$	$\mu_3$	$\mu_4$	$\sigma^2$
Estimates	3.8	0.92	-0.14	-0.39	0.48	36.36

Table 3: Bayesian Estimate for  $\Sigma^{-1}$

	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$
$\beta_0$	16.07	7.63	-1.78	5.39	1.48
$\beta_1$	7.63	0.34	0.47	3.96	1.14
$\beta_2$	-1.78	0.47	381.32	-6.32	24.35
$\beta_3$	5.39	3.96	-6.32	21.43	-3.30
$\beta_4$	1.48	1.14	24.35	-3.30	55.12

## Part IV: Model Performance

The overall adjusted  $R^2$  of the estimated Bayesian model is 0.9524156 and P-value for goodness-of-fit test of all observations is 1.

Furthermore, we check the model performance of individual hurricane. Even though most of the estimated models track hurricanes well with great adjusted  $R^2$ , some estimated models track the hurricanes extremely

Table 4:  $R^2_{adj}$  for each hurricane

$R^2_{adj}$	Count of Hurricanes	Percentage(%)
0.6-1	522	76.7
0.2-0.6	79	11.6
< 0.2	80	11.7

bad. 23.3% of the estimated models do not track the hurricane well and have adjusted  $R^2$  less than 0.6. Few models have negative adjusted  $R^2$ . Table 2 show the distribution of the adjusted  $R^2$  of the estimated Bayesian models for individual hurricane.

We also perform the goodness-of-fit test for individual estimated model. 119 of the estimated models have p-value less than 0.05, implying that those models do not fit well.

### Observed Wind Speed vs. Bayesian Model Estimates

