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

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

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

Prior

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

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

Posterior

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

Conditional Posterior

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

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

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

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

$$\begin{aligned}
\Sigma^{-1} : \pi\left(\Sigma^{-1} \mid \boldsymbol{\theta}_{(-\Sigma^{-1})}, \boldsymbol{Y}\right) &\propto \left|\Sigma^{-1}\right|^{d+1+\frac{m}{2}} \exp\left\{-\frac{1}{2}\operatorname{tr}\left(\Sigma^{-1}\right)\right\} \exp\left\{-\frac{1}{2}\sum_{i=1}^{m}\left(\boldsymbol{\beta}_i - \boldsymbol{\mu}\right)^T \Sigma^{-1}\left(\boldsymbol{\beta}_i - \boldsymbol{\mu}\right)\right\} \\
&= \left|\Sigma^{-1}\right|^{d+1+\frac{m}{2}} \exp\left\{-\frac{1}{2}\operatorname{tr}\left(\Sigma^{-1}\right) - \frac{1}{2}\operatorname{tr}\left(\Sigma^{-1}\sum_{i=1}^{m}\left(\boldsymbol{\beta}_i - \boldsymbol{\mu}\right)\left(\boldsymbol{\beta}_i - \boldsymbol{\mu}\right)^T\right)\right\} \\
&= \left|\Sigma^{-1}\right|^{d+1+\frac{m}{2}} \exp\left\{-\frac{1}{2}\operatorname{tr}\left(\Sigma^{-1}\left(I + \sum_{i=1}^{m}\left(\beta_i - \mu\right)\left(\beta_1 - \mu\right)^T\right)\right)\right\} \\
&\Rightarrow \Sigma^{-1} \mid \boldsymbol{\theta}_{(-\Sigma^{-1})}, \boldsymbol{Y} \sim \text{Wishart}\left(3d + 3 + m, \left(\boldsymbol{I} + \sum_{i=1}^{m}\left(\boldsymbol{\beta}_i - \boldsymbol{\mu}\right)\left(\boldsymbol{\beta}_i - \boldsymbol{\mu}\right)^T\right)^{-1}\right)
\end{aligned}$$

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 β_{ij}: Sampling proposed β'_{ij}, $j = 0, 1...4$ 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 β_{ij}. Since the proposals are symmetry, the accepting or rejecting the proposed β'_{ij} depends on the ratio of posterior distribution. Some of the parameters in $\boldsymbol{\theta}$ could be cancelled out, so the ratio simplified to be:

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

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

The log of the ratio:

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

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 $\beta'_{ij} = \beta^{(t+1)}_{ij}$, otherwise we reject β'_{ij} and $\beta^{(t)}_{ij} = \beta^{(t+1)}_{ij}$.

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

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

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

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 $\left(3d + 3 + m, \left(\boldsymbol{I} + \sum_{i=1}^{m}\left(\boldsymbol{\beta}^{(t+1)}_i - \boldsymbol{\mu}^{(t+1)}\right)\left(\boldsymbol{\beta}^{(t+1)}_i - \boldsymbol{\mu}^{(t+1)}\right)^T\right)^{-1}\right)$

Part III: Implementation and Estimations

Iinitial Value

Initial Values:

- $\boldsymbol{\beta}_i$ - Fit multivariate linear regression (MLR) for i^{th} hurricane and use the coefficients as $\boldsymbol{\beta}_{i0}$
- $\boldsymbol{\mu}$ - Average over all $\boldsymbol{\beta}_{i0}$ as $\boldsymbol{\mu}_0$
- σ^2 - $\hat{\sigma}^2_i$ is the mean square residuals of the MLR model on i^{th} hurricane. Take the mean over all $\hat{\sigma}^2_i$ as σ^2_0
- Σ^{-1} - Generate the covaraiance matrix of $\boldsymbol{\beta}_{i0}$ and take the inverse of the matrix as Σ^{-1}_0

Implement 10000 MCMC iterations

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

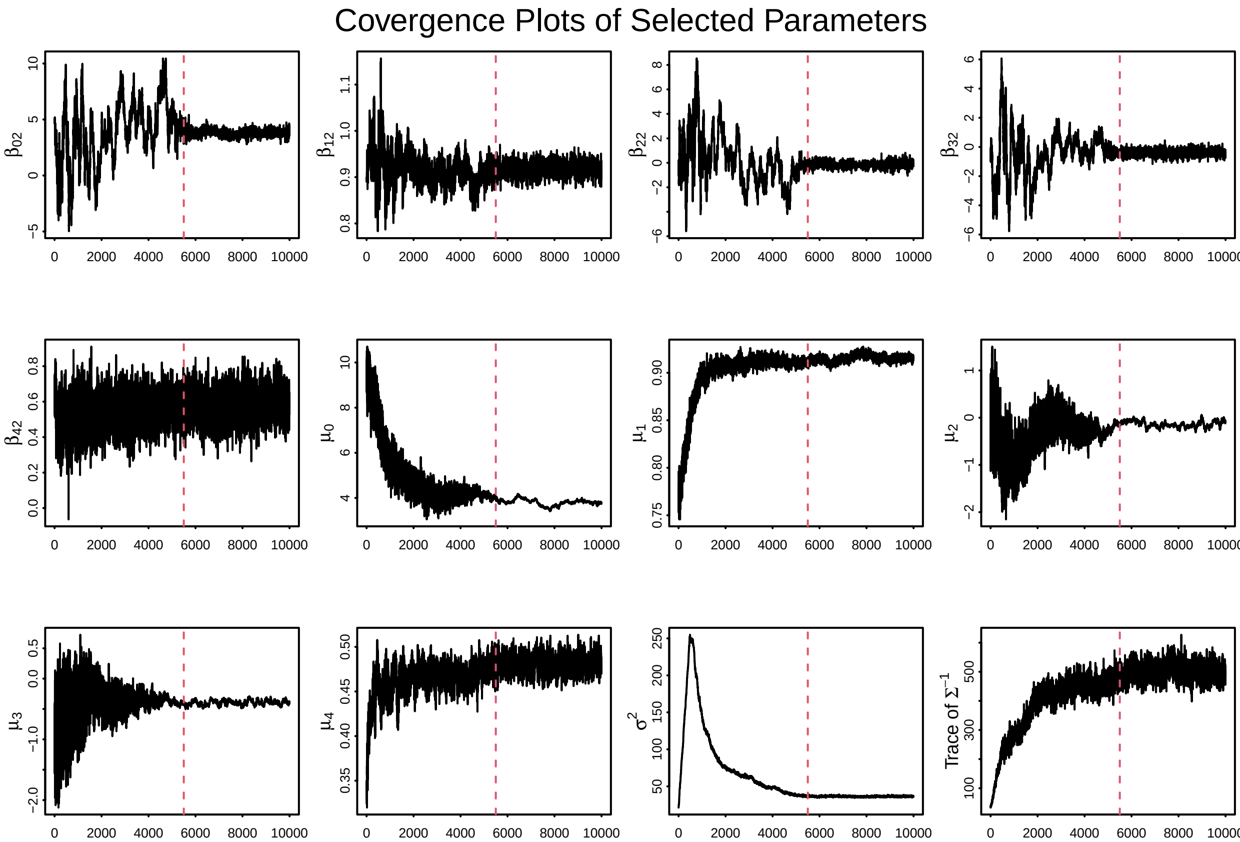
Search Window and Acceptance Rate

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 $\boldsymbol{\beta}$ and σ^2.

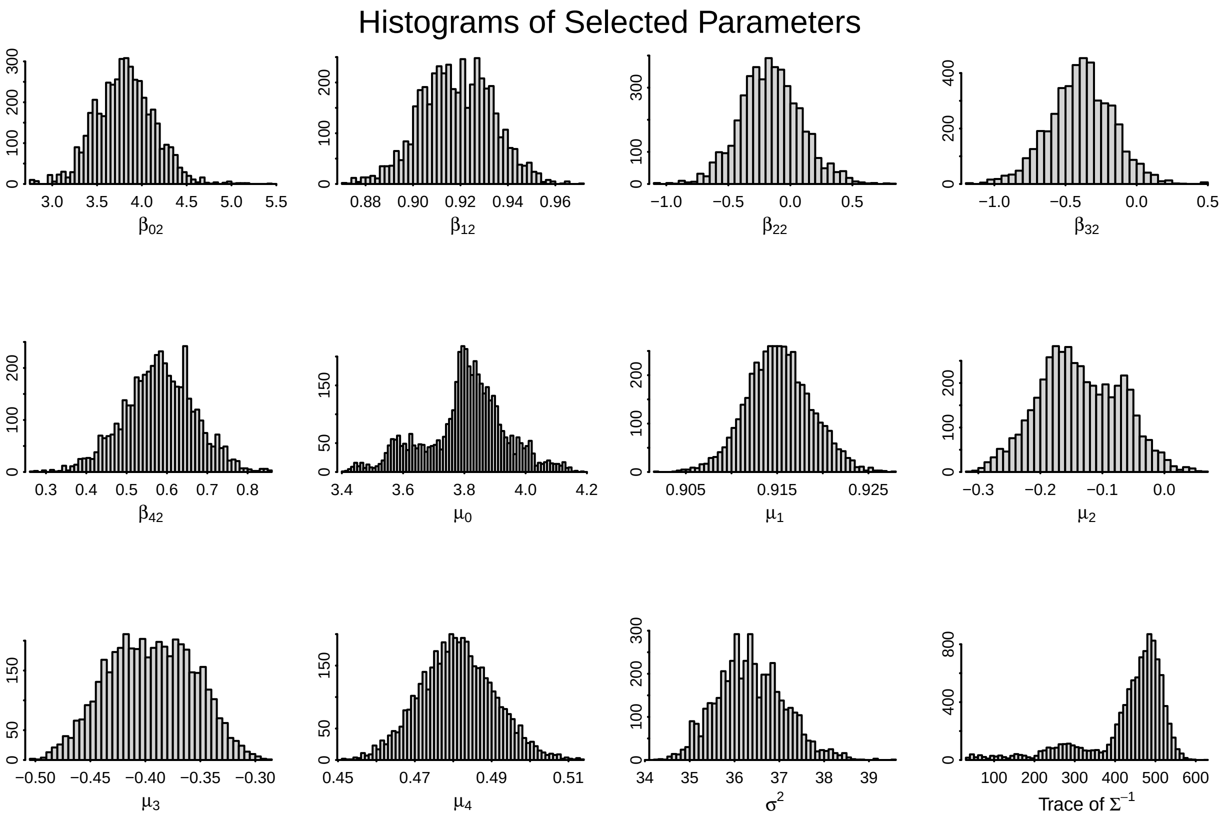
Table 1: Range of Search Window and Acceptance Rate for paraemters used MH step

	search window	acceptance rate (%)
$\boldsymbol{\beta}_0$	1.1	45.87-51.36
$\boldsymbol{\beta}_1$	0.04-0.1	31.67-63.68
$\boldsymbol{\beta}_2$	0.8-1	38.6-45.6
$\boldsymbol{\beta}_3$	0.5-0.6	33.2-61.32
$\boldsymbol{\beta}_4$	0.4-0.5	34.95-60.45
$\boldsymbol{\sigma}^2$	2	44.93

Convergence trace plots



Histograms


Histograms of Selected Parameters
β02
β12
β22
β32
β42
μ0
μ1
μ2
μ3
μ4
σ2
Trace of Σ−1

Estimates (burn-in first 5500 runs)

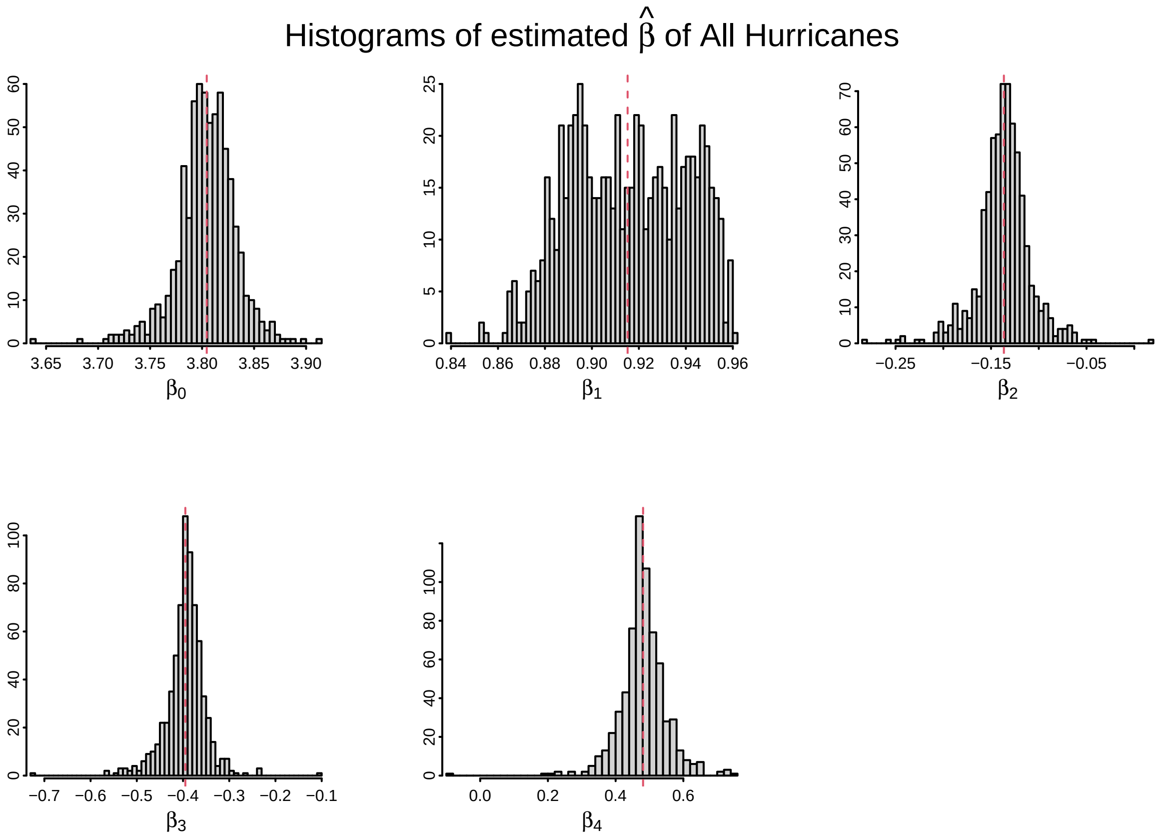


Chart of the estimates of μ, σ^2 and covaraince matrix

Table 2: Bayesian Estimates for $\boldsymbol{\mu}$ and σ^2

	μ_0	μ_1	μ_2	μ_3	μ_4	σ^2	σ_{00}^2	σ_{11}^2	σ_{22}^2	σ_{33}^2	σ_{44}^2
Estimates	3.8	0.92	-0.14	-0.39	0.48	36.36	0.063	0.003	0.047	0.042	0.018

$$\hat{\Sigma}^{-1} = \begin{bmatrix} 16.07 & 7.63 & 0.34 & -1.78 & 0.47 \\ 7.63 & 381.32 & 5.39 & 3.96 & -6.32 \\ 0.34 & 5.39 & 21.43 & 1.48 & 1.14 \\ -1.78 & 3.96 & 1.48 & 24.35 & -3.3 \\ 0.47 & -6.32 & 1.14 & -3.3 & 55.12 \end{bmatrix} \quad \hat{\rho} = \begin{bmatrix} 1 & -0.101 & -0.018 & 0.094 & -0.011 \\ -0.101 & 1 & -0.057 & -0.043 & 0.043 \\ -0.018 & -0.057 & 1 & -0.067 & -0.042 \\ 0.094 & -0.043 & -0.067 & 1 & 0.089 \\ -0.011 & 0.043 & -0.042 & 0.089 & 1 \end{bmatrix}$$

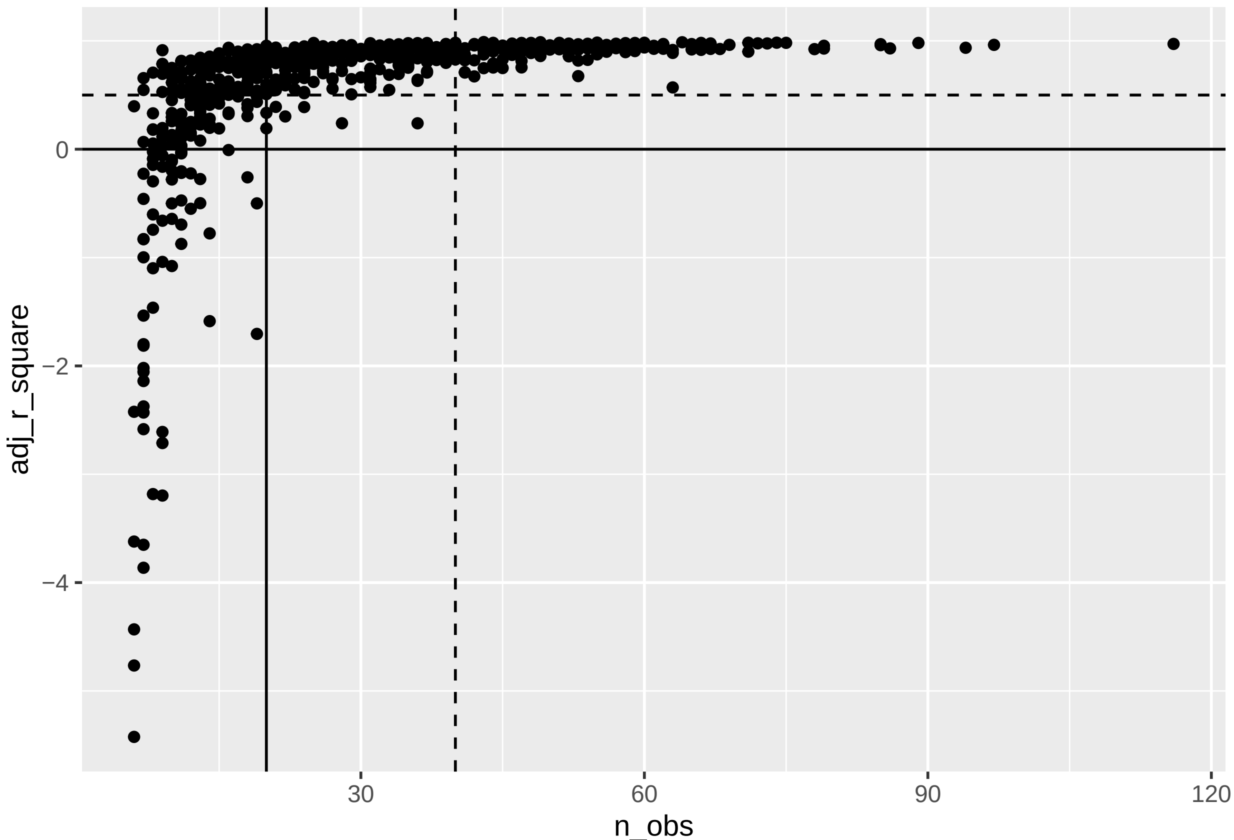
Part IV: Model Performance

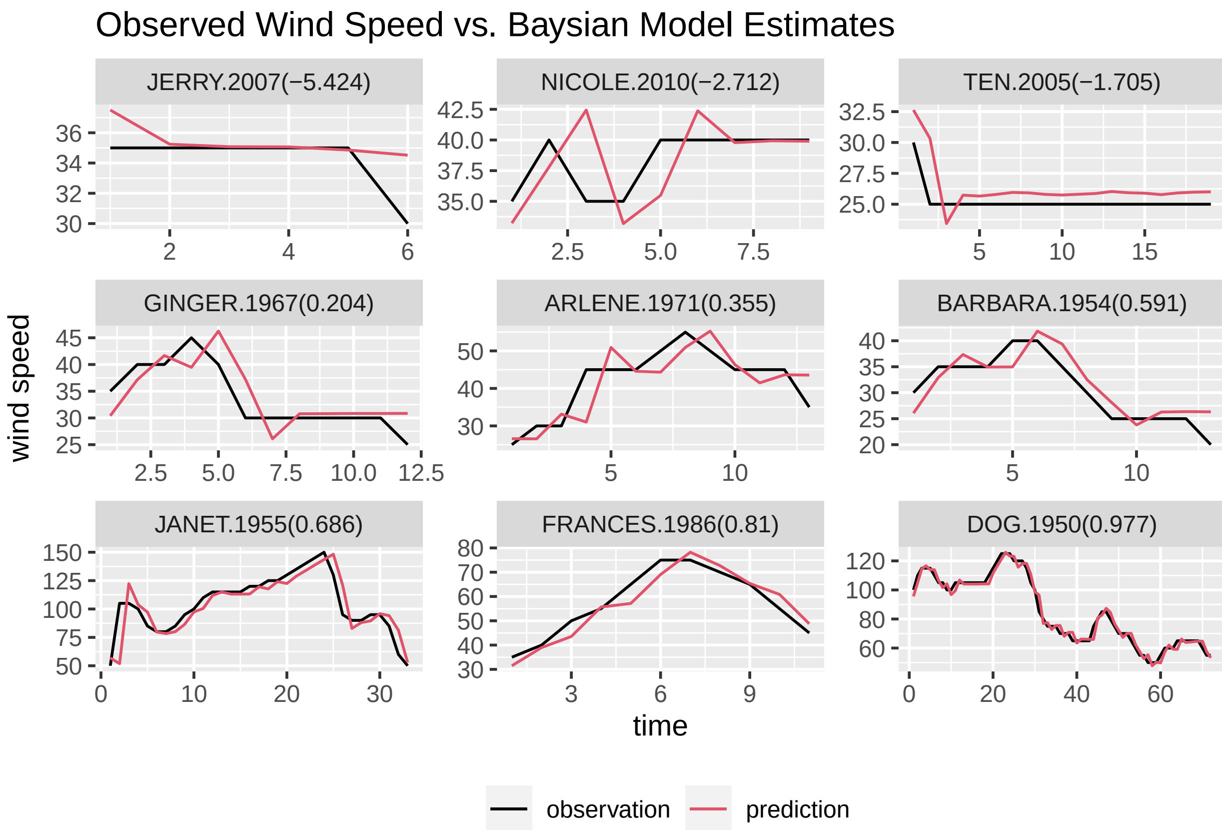
The overall adjusted R^2 of the estimated Bayesian model is 0.9524156.

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 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 Baysian models for individual hurricane.

Table 3: 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





We also perform the goodness-of-fit test for individual estimated model. We used residuals of Bayesian estimates, r_{ij}, of the j^{th} observation in i^{th} hurricane, to calculate the test statistics. $\chi^2_{stat} = \frac{\sum_{j=1}^{n_i} r_{ij}^2}{\sigma^2}$, where σ^2 is the estimate σ^2 from MCMC. Based on the normal assumption in intro, $\chi^2_{stat} \sim \chi^2_{n_i}$, where n_i is the number of observation for i^{th} hurricane. The estimated models of 88 hurricanes have p-value less than 0.05, implying that those models do not fit well.