

# derive posterior distribution

Yi Huang

2023-04-26

The suggested Bayesian model

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

$Y_i(t)$  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  $t+6$ ,

$\mathbf{X}_i = (x_{i,1}, x_{i,2}, x_{i,3})$  are covariates with fixed effect  $\gamma$ , where  $x_{i,1}$  be the month of year when the  $i$ -th hurricane started,  $x_{i,2}$  be the calendar year of the  $i$  hurricane, and  $x_{i,3}$  be the type of the  $i$ -th hurricane.

and  $\epsilon_{i,t}$  follows a normal distributions with mean zero and variance  $\sigma^2$ , independent across  $t$ .

In the model,  $\beta_i = (\beta_{0,i}, \beta_{1,i}, \dots, \beta_{4,i})$  are the random coefficients associated the  $i$ th hurricane, we assume that

$$\beta_i \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

follows a multivariate normal distributions with mean  $\boldsymbol{\beta}$  and covariance matrix  $\boldsymbol{\Sigma}$ .

## Prior distributions

1.  $\boldsymbol{\mu}$ : a normal distribution with mean vector  $\mathbf{0}$  and variance-covariance matrix  $\mathbf{V}$ , reflecting the prior knowledge that the mean coefficients should be centered around zero but allowing for some variability across hurricanes. The variance-covariance matrix  $\mathbf{V}$  can be set to a diagonal matrix with large variances on the diagonal and small covariances off-diagonal, reflecting the prior knowledge that the coefficients may have some correlation but are largely independent across hurricanes.

Assume  $\text{diag}(\mathbf{V}) = \tau^2$

$$P(\boldsymbol{\mu}) = \frac{1}{\sqrt{2\pi\tau}} \exp\left(-\frac{\mu^2}{2\tau^2}\right) \propto 1$$

2.  $\boldsymbol{\Sigma}$ : an inverse-Wishart distribution with degrees of freedom  $\nu$  and scale matrix  $\mathbf{S}$ , reflecting the prior knowledge that the covariance matrix of the coefficients should be positive definite and have some structure. The degrees of freedom  $\nu$  can be set to a small value (e.g., 5) to reflect a relatively weak prior, while the scale matrix  $\mathbf{S}$  can be set to a diagonal matrix with large variances on the diagonal and small covariances off-diagonal, reflecting the prior knowledge that the covariance matrix should be diagonal or nearly diagonal.

$$P(\boldsymbol{\Sigma}^{-1}) \propto |\boldsymbol{\Sigma}|^{-\frac{(v+d+1)}{2}} \exp\left(-\frac{1}{2}\text{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1})\right)$$

where  $d$  is dimension of scale matrix  $\mathbf{S}$

3. All the fixed effects  $\gamma \sim N(0, 0.05^2)$

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

4.  $\sigma$ : a half-Cauchy distribution with scale parameter 10, reflecting the prior knowledge that the residual variance should be positive and large enough to account for any unexplained variability in the wind speed data.

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

where  $\alpha = 10$

## Posterior Distributions

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

Let

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

where  $\mathbf{X}_i(t) = (1, Y_i(t), \Delta_{i,1}(t), \Delta_{i,2}(t), \Delta_{i,3}(t))$ ,  $\beta_i = (\beta_{0,i}, \beta_{1,i}, \beta_{2,i}, \beta_{3,i}, \beta_{4,i})$

then, we can find that

$$Y_i(t+6) \sim N(\mathbf{X}_i(t)\beta_i^\top, \sigma^2)$$

For  $i^{th}$  hurricane, there may be  $m_i$  times of record (excluding the first and second observation), let

$$\mathbf{Y}_i = \begin{pmatrix} Y_i(t_0+6) \\ Y_i(t_1+6) \\ \vdots \\ Y_i(t_{m_i-1}+6) \end{pmatrix}_{m_i \times 1}$$

denotes the  $m_i$ -dimensional result vector for the  $i^{th}$  hurricane. Therefore, since  $Y_i(t)$ 's are independent across  $t$ , we can show that the conditional distribution of  $\mathbf{Y}_i$  is

$$\mathbf{Y}_i \mid \mathbf{X}_i, \beta_i, \sigma^2 \sim N(\mathbf{X}_i\beta_i^\top, \sigma^2 I)$$

where

$$\mathbf{X}_i = \begin{pmatrix} 1 & Y_i(t_0) & \Delta_{i,1}(t_0) & \Delta_{i,2}(t_0) & \Delta_{i,3}(t_0) \\ 1 & Y_i(t_1) & \Delta_{i,1}(t_1) & \Delta_{i,2}(t_1) & \Delta_{i,3}(t_1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & Y_i(t_{m_i-1}) & \Delta_{i,1}(t_{m_i-1}) & \Delta_{i,2}(t_{m_i-1}) & \Delta_{i,3}(t_{m_i-1}) \end{pmatrix}_{m_i \times d}$$

and the pdf of  $\mathbf{Y}_i$  is

$$\begin{aligned} f(\mathbf{Y}_i \mid \beta_i, \sigma^2) &= \det(2\pi\sigma^2 I_{(m_i \times m_i)})^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{Y}_i - \mathbf{X}_i\beta_i^\top)^\top (\sigma^2 I_{(m_i \times m_i)})^{-1} (\mathbf{Y}_i - \mathbf{X}_i\beta_i^\top)\right\} \\ &= (2\pi\sigma^2)^{-m_i/2} \exp\left\{-\frac{1}{2}(\mathbf{Y}_i - \mathbf{X}_i\beta_i^\top)^\top (\sigma^2 I_{(m_i \times m_i)})^{-1} (\mathbf{Y}_i - \mathbf{X}_i\beta_i^\top)\right\} \end{aligned}$$

Since

$$\beta_i \sim N(\boldsymbol{\mu}, \Sigma)$$

Therefore the pdf of  $\beta_i$  is

$$\pi(\beta_i \mid \boldsymbol{\mu}, \Sigma) = \det(2\pi\Sigma)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\beta_i - \boldsymbol{\mu})\Sigma^{-1}(\beta_i - \boldsymbol{\mu})^\top\right)$$

Notice that  $\mathbf{B} = (\beta_1^\top, \dots, \beta_n^\top)^\top$ , i.e.

$$\mathbf{B} = \begin{pmatrix} \beta_{0,1} & \beta_{1,1} & \beta_{2,1} & \beta_{3,1} & \beta_{4,1} \\ \beta_{0,2} & \beta_{1,2} & \beta_{2,2} & \beta_{3,2} & \beta_{4,2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \beta_{0,n} & \beta_{1,n} & \beta_{2,n} & \beta_{3,n} & \beta_{4,n} \end{pmatrix}_{n \times d}$$

So, by using Bayesian rule, we can show the posterior distribution of  $\Theta$  is,

$$\begin{aligned}
\pi(\Theta|Y) &= \pi(\mathbf{B}^\top, \boldsymbol{\mu}^\top, \sigma^2, \boldsymbol{\Sigma} | Y) \propto \prod_{i=1}^n f(\mathbf{Y}_i | \beta_i, \sigma^2) \prod_{i=1}^n \pi(\beta_i | \boldsymbol{\mu}, \boldsymbol{\Sigma}) P(\sigma^2) P(\boldsymbol{\mu}) P(\boldsymbol{\Sigma}^{-1}) \\
&\propto \prod_{i=1}^n \left\{ (2\pi\sigma^2)^{-m_i/2} \exp \left\{ -\frac{1}{2} (\mathbf{Y}_i - \mathbf{X}_i \beta_i^\top)^\top (\sigma^2 I)^{-1} (\mathbf{Y}_i - \mathbf{X}_i \beta_i^\top) \right\} \right\} \\
&\times \prod_{i=1}^n \left\{ \det(2\pi\boldsymbol{\Sigma})^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\beta_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\beta_i - \boldsymbol{\mu}) \right\} \right\} \\
&\times \frac{1}{\sigma^2} \times \det(\boldsymbol{\Sigma})^{-(d+1)} \exp \left\{ -\frac{1}{2} \boldsymbol{\Sigma}^{-1} \right\}
\end{aligned}$$

## Markov chain Monte Carlo Algorithm to Generate the posterior distribution

Due to the high-dimensional problem of the full joint posterior distribution, considering the computational complexity plus we actually well know the form of the joint posterior, we suggest to use Gibbs sampling algorithm instead of Metropolis-Hastings algorithm.

To apply MCMC using Gibbs sampling, we need to find conditional posterior distribution of each parameter, then we can implement Gibbs sampling on these conditional posterior distributions.

Since our suggested model mainly focus on the parameter  $\mathbf{B}$ , we decided to derive its conditional posterior distribution.

### 1. The posterior distribution of $\mathbf{B}$

Since finding the posterior distribution of  $\mathbf{B}$  is the same to find the posterior distribution of  $\beta_i$ , we try to derive the conditional distribution  $\pi(\beta_i | Y, \boldsymbol{\mu}^\top, \sigma^2, \boldsymbol{\Sigma})$

$$\begin{aligned}
\pi(\mathbf{B} | Y, \boldsymbol{\mu}^\top, \sigma^2, \boldsymbol{\Sigma}) &\propto L_Y(\mathbf{B}^\top, \sigma^2) \times \pi(\mathbf{B} | \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\
&\propto \prod_{i=1}^n f(\mathbf{Y}_i | \beta_i, \sigma^2) \prod_{i=1}^n \pi(\beta_i | \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\
&\propto \prod_{i=1}^n \left\{ (2\pi\sigma^2)^{-m_i/2} \exp \left( -\frac{1}{2} (\mathbf{Y}_i - \mathbf{X}_i \beta_i^\top)^\top (\sigma^2 I)^{-1} (\mathbf{Y}_i - \mathbf{X}_i \beta_i^\top) \right) \right\} \\
&\times \prod_{i=1}^n \left\{ \det(2\pi\boldsymbol{\Sigma})^{-\frac{1}{2}} \exp \left( -\frac{1}{2} (\beta_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\beta_i - \boldsymbol{\mu}) \right) \right\} \\
&\propto \prod_{i=1}^n \exp \left\{ -\frac{1}{2} \left( (\mathbf{Y}_i - \mathbf{X}_i \beta_i^\top)^\top (\sigma^2 I)^{-1} (\mathbf{Y}_i - \mathbf{X}_i \beta_i^\top) + (\beta_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\beta_i - \boldsymbol{\mu}) \right) \right\} \\
&= \exp \left\{ -\frac{1}{2} \left( \mathbf{Y}_i^\top (\sigma^2 I)^{-1} \mathbf{Y}_i + \beta_i \mathbf{X}_i^\top (\sigma^2 I)^{-1} \mathbf{X}_i \beta_i^\top - \mathbf{Y}_i^\top (\sigma^2 I)^{-1} \mathbf{X}_i \beta_i^\top \right. \right. \\
&\quad \left. \left. - \beta_i \mathbf{X}_i^\top (\sigma^2 I)^{-1} \mathbf{Y}_i + \beta_i \boldsymbol{\Sigma}^{-1} \beta_i^\top + \boldsymbol{\mu} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}^\top - \boldsymbol{\mu} \boldsymbol{\Sigma}^{-1} \beta_i^\top - \beta_i \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}^\top \right) \right\} \\
&= \exp \left\{ -\frac{1}{2} \left( \beta_i (\boldsymbol{\Sigma}^{-1} + \mathbf{X}_i^\top (\sigma^2 I)^{-1} \mathbf{X}_i) \beta_i^\top - 2(\mathbf{Y}_i^\top (\sigma^2 I)^{-1} \mathbf{X}_i + \boldsymbol{\mu} \boldsymbol{\Sigma}^{-1}) \beta_i^\top + \mathbf{C} \right) \right\}
\end{aligned}$$

where,

$$\mathbf{C} = \mathbf{Y}_i^\top (\sigma^2 I)^{-1} \mathbf{Y}_i + \boldsymbol{\mu} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}^\top$$

By re-writing the conditional posterior distribution, and ignoring some constant terms, we can show that

$$\pi(\mathbf{B} | Y, \boldsymbol{\mu}^\top, \sigma^2, \boldsymbol{\Sigma}) \propto \prod_{i=1}^n \exp \{ (\beta_i^\top - \hat{\beta}_i)^\top (\hat{\boldsymbol{\Sigma}}_{\beta_i})^{-1} (\beta_i^\top - \hat{\beta}_i) \}$$

Hence, each  $\beta_i$  has a conditional posterior multivariate normal distribution

$$\pi(\beta_i | \mathbf{Y}, \boldsymbol{\mu}^\top, \sigma^2, \boldsymbol{\Sigma}) \sim \mathcal{N}(\hat{\beta}_i, \hat{\boldsymbol{\Sigma}}_{\beta_i})$$

where

$$\begin{aligned}\hat{\beta}_i &= (\boldsymbol{\Sigma}^{-1} + \mathbf{X}_i^\top (\sigma^2 I)^{-1} \mathbf{X}_i)^{-1} \mathbf{Y}_i^\top (\sigma^2 I)^{-1} \mathbf{X}_i + \boldsymbol{\mu} \boldsymbol{\Sigma}^{-1} \\ \hat{\boldsymbol{\Sigma}}_{\beta_i} &= (\boldsymbol{\Sigma}^{-1} + \mathbf{X}_i^\top (\sigma^2 I)^{-1} \mathbf{X}_i)^{-1}\end{aligned}$$

2. The posterior distribution of  $\pi(\sigma^2 | \mathbf{Y}, \mathbf{B}^\top, \boldsymbol{\mu}^\top, \boldsymbol{\Sigma})$

$$\begin{aligned}\pi(\sigma^2 | \mathbf{Y}, \mathbf{B}^\top, \boldsymbol{\mu}^\top, \boldsymbol{\Sigma}) &\propto L_Y(\mathbf{B}^\top, \sigma^2) \times \pi(\sigma^2) \\ &\propto \frac{1}{\sigma^2} \prod_{i=1}^n \left\{ (2\pi\sigma^2)^{-m_i/2} \exp\left(-\frac{1}{2}(\mathbf{Y}_i - \mathbf{X}_i \beta_i^\top)^\top (\sigma^2 I)^{-1} (\mathbf{Y}_i - \mathbf{X}_i \beta_i^\top)\right) \right\} \\ &\propto \frac{1}{\sigma^2} \left( \frac{\sum_{i=1}^n m_i}{2} + 1 \right) \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (\mathbf{Y}_i - \mathbf{X}_i \beta_i^\top)^\top (\mathbf{Y}_i - \mathbf{X}_i \beta_i^\top)\right\}\end{aligned}$$

which follows the form of pdf of inverse gamma distribution

$$f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{1}{x} \exp\left\{-\frac{\beta}{x}\right\}$$

in this case,  $x$  is replaced by  $\sigma^2$ ,  $\alpha$  is replaced by  $\frac{1}{2} \sum_{i=1}^n m_i$ ,  $\beta$  is replaced by  $\frac{1}{2} \sum_{i=1}^n (\mathbf{Y}_i - \mathbf{X}_i \beta_i^\top)^\top (\mathbf{Y}_i - \mathbf{X}_i \beta_i^\top)$  i.e.

$$\pi(\sigma^2 | \mathbf{Y}, \mathbf{B}^\top, \boldsymbol{\mu}^\top, \boldsymbol{\Sigma}) \sim IG\left(\frac{1}{2} \sum_{i=1}^n m_i, \frac{1}{2} \sum_{i=1}^n (\mathbf{Y}_i - \mathbf{X}_i \beta_i^\top)^\top (\mathbf{Y}_i - \mathbf{X}_i \beta_i^\top)\right)$$

3. The posterior distribution of  $\pi(\boldsymbol{\Sigma} | \mathbf{Y}, \mathbf{B}^\top, \boldsymbol{\mu}^\top, \sigma^2)$

$$\begin{aligned}\pi(\boldsymbol{\Sigma} | \mathbf{Y}, \mathbf{B}^\top, \boldsymbol{\mu}^\top, \sigma^2) &\propto \pi(\mathbf{B} | \boldsymbol{\mu}, \boldsymbol{\Sigma}) \pi(\boldsymbol{\Sigma}^{-1}) \\ &\propto \prod_{i=1}^n \left\{ \det(2\pi\boldsymbol{\Sigma})^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\beta_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\beta_i - \boldsymbol{\mu})\right) \right\} |\boldsymbol{\Sigma}|^{-(d+1)} \exp\left(-\frac{1}{2}\text{tr}(\boldsymbol{\Sigma}^{-1})\right) \\ &\propto |\boldsymbol{\Sigma}|^{-(n+d+1+d+1)/2} \exp\left\{-\frac{1}{2}(\beta_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\beta_i - \boldsymbol{\mu}) - \frac{1}{2}\text{tr}(\boldsymbol{\Sigma}^{-1})\right\} \\ &\propto |\boldsymbol{\Sigma}|^{-(n+d+1+d+1)/2} \exp\left\{-\frac{1}{2}\text{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1})\right\}\end{aligned}$$

where

$$\mathbf{S} = \mathbf{I} + \sum_{i=1}^n (\beta_i - \boldsymbol{\mu})(\beta_i - \boldsymbol{\mu})^\top$$

which is the form of pdf of the inverse wishart distribution Inverse Wishart( $\mathbf{V}, \mathbf{S}$ ), where  $\mathbf{V} = n + d + 1$ , i.e.

$$\pi(\boldsymbol{\Sigma} | \mathbf{Y}, \mathbf{B}^\top, \boldsymbol{\mu}^\top, \sigma^2) \sim IW(n + d + 1, \mathbf{I} + \sum_{i=1}^n (\beta_i - \boldsymbol{\mu})(\beta_i - \boldsymbol{\mu})^\top)$$

4. The posterior distribution of  $\pi(\boldsymbol{\mu} | \mathbf{Y}, \mathbf{B}^\top, \sigma^2, \boldsymbol{\Sigma})$

$$\begin{aligned}
\pi(\boldsymbol{\mu} | \mathbf{Y}, \mathbf{B}^\top, \boldsymbol{\sigma}^2, \boldsymbol{\Sigma}) &\propto \pi(\mathbf{B} | \boldsymbol{\mu}, \boldsymbol{\Sigma}) \pi(\boldsymbol{\mu}) \\
&= \prod_{i=1}^n \left\{ \det(2\pi\boldsymbol{\Sigma})^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{\beta}_i - \boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}_i - \boldsymbol{\mu})^\top\right) \right\} \\
&\propto \exp\left\{-\frac{1}{2} \sum_{i=1}^n (\boldsymbol{\beta}_i - \boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}_i - \boldsymbol{\mu})^\top\right\} \\
&\propto \exp\left\{-\frac{1}{2} \left( \sum_{i=1}^n \boldsymbol{\beta}_i \boldsymbol{\Sigma}^{-1} \boldsymbol{\beta}_i^\top + n\boldsymbol{\mu}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^\top - 2 \sum_{i=1}^n \boldsymbol{\beta}_i \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}^\top \right)\right\} \\
&= \exp\left\{-\frac{1}{2} \left( \boldsymbol{\mu}(n\boldsymbol{\Sigma}^{-1})\boldsymbol{\mu}^\top - 2 \left( \sum_{i=1}^n \boldsymbol{\beta}_i \boldsymbol{\Sigma}^{-1} \right) \boldsymbol{\mu}^\top + \mathbf{C}' \right)\right\} \\
&\propto \exp\left\{-\frac{1}{2} \left( \boldsymbol{\mu} - \frac{1}{n} \sum_{i=1}^n \boldsymbol{\beta}_i \right) (n\boldsymbol{\Sigma}^{-1}) \left( \boldsymbol{\mu} - \frac{1}{n} \sum_{i=1}^n \boldsymbol{\beta}_i \right)^\top \right\}
\end{aligned}$$

where

$$\mathbf{C}' = \sum_{i=1}^n \boldsymbol{\beta}_i \boldsymbol{\Sigma}^{-1} \boldsymbol{\beta}_i^\top$$

Hence

$$\pi(\boldsymbol{\mu} | \mathbf{Y}, \mathbf{B}^\top, \boldsymbol{\sigma}^2, \boldsymbol{\Sigma}) \sim \mathcal{N}\left(\frac{1}{n} \sum_{i=1}^n \boldsymbol{\beta}_i, \frac{1}{n} \boldsymbol{\Sigma}\right)$$

## Markov Chain Monte Carlo

Because our hierarchical Bayesian Model exploited non-informative priors for four parameters, the Gibbs Sampling method would be implemented, updating parameters in the following order from their conditional posteriors distributions,  $\mathbf{B}$ ,  $\sigma^2$ ,  $\boldsymbol{\Sigma}$  and  $\boldsymbol{\mu}$ .

## Algorithm Implemtation and Estimation

Since we have derived the conditional posterior of the four parameters, we implemented Gibbs sampling algorithm, updating parameters by randomly generating samples from their conditional posterior distributions.

The update of parameters is component wise, at  $(t+1)^{\text{th}}$  step, updating parameters in the following the order:

- Sample  $\mathbf{B}^{(t+1)}$ , i.e., sample each  $\boldsymbol{\beta}_i^{(t+1)}$  from  $\mathcal{N}(\hat{\boldsymbol{\beta}}_i^{(t)}, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}_i}^{(t)})$
- Then, sample  $\sigma^2$  from  $IG(\frac{1}{2} \sum_{i=1}^n m_i, \frac{1}{2} \sum_{i=1}^n (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}_i^{(t+1)})^\top (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}_i^{(t+1)})^\top)$
- Next, sample  $\boldsymbol{\Sigma}^{(t+1)}$  from  $IW(n + d + 1, \mathbf{I} + \sum_{i=1}^n (\boldsymbol{\beta}_i^{(t+1)} - \boldsymbol{\mu}^{(t)})(\boldsymbol{\beta}_i^{(t+1)} - \boldsymbol{\mu}^{(t)})^\top)$
- Finally, sample  $\boldsymbol{\mu}^{(t+1)}$  from  $\mathcal{N}(\frac{1}{n} \sum_{i=1}^n \boldsymbol{\beta}_i^{(t+1)}, \frac{1}{n} \boldsymbol{\Sigma}^{(t+1)})$

## Train Test Splits

For model training as well as the performance evaluation of our Bayesian model, we split the dataset into train and test set. We first drop the hurricanes with less than 3 observations, and removed observations

without transformed predictors values. Then we got a dataset with 697 different hurricanes. Our train-test split is within each hurricane's observations, that is, for each hurricane, we randomly set 80% of observations to the training set and left 20% to the test set. We then trained Bayesian model based on training data and evaluate the model performance on test dataset.

### ### Initial Values

For a good performance and to speed up the convergence of our algorithm, also keep some uncertainty in

For initial value of  $\textbf{B}$ , we run multivariate linear regressions for each hurricane and use the

For initial value of  $\boldsymbol{\mu}$ , we take the average of  $\boldsymbol{\beta}_i^{\text{MLR}}$ , that is

For initial value of  $\sigma^2$ , we take the average of the MSE for  $i$  hurricanes.

For initial value of  $\boldsymbol{\Sigma}$ , we just set it to a simple diagonal matrix, i.e.  $\boldsymbol{\Sigma}$