Bayesian Modeling of Hurricane Trajectories

Hongjie Liu, Xicheng Xie, Jiajun Tao, Zijian Xu, Shaohan Chen

May 1st, 2023

Outline

- Background
- Task 1
- Task 2
- Task 3
- Task 4
- Discussions
- Reference
- Q&A

Background

Bayesian Modeling of Hurricane Trajectories:

- In this project we are interested in modeling the hurricane trajectories to forecast the wind speed.

Background

Data Source:

- “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.

Bayesian Model

Let $Y_i(t)$ denote the wind speed of the ith hurricane at time t (in hours) since the hurricane began. The following Bayesian model was suggested to model the wind speed of the ith hurricane 6 hours later:

$$Y_i(t+6) = \beta_{0,i} + \beta_{1,i} Y_i(t) + \beta_{2,i}\Delta_{i,1}(t) + \beta_{3,i}\Delta_{i,2}(t) + \beta_{4,i}\Delta_{i,3}(t) \\ + \mathbf{X}_i^\top \boldsymbol{\gamma} + \epsilon_i(t),$$

where

- $\Delta_{i,1}(t)$, $\Delta_{i,2}(t)$ and $\Delta_{i,3}(t)$: changes of latitude, longitude and wind speed between $t-6$ and t, with random coefficients $\boldsymbol{\beta}_i = (\beta_{0,i}, \beta_{1,i}, ..., \beta_{4,i})^\top$
- $\mathbf{X}_i = (x_{i,1}, \ldots, x_{i,6})^\top$: covariates with fixed effects $\boldsymbol{\gamma}$, where
  - $x_{i,1}$: the calendar year of the i-th hurricane
  - $x_{i,2}$: indicator variable of the month in active season (August-October) when the i-th hurricane started
  - $x_{i,3}, \ldots, x_{i,6}$: indicator variables of the type (ES, NR, SS, TS) of the i-th hurricane
- $\epsilon_{i,t} \sim N(0, \sigma^2)$, independent across t

Task 1 - Prior Distributions

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

We assume that

- $\beta_i \overset{i.i.d.}{\sim} N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
- $\boldsymbol{\mu} \sim N(\mathbf{0}, \boldsymbol{V})$
- $\boldsymbol{\Sigma}$: an inverse-Wishart distribution with d.f. ν and scale matrix $\boldsymbol{S}$
- $\gamma \sim N(\mathbf{0}, 0.05^2 \boldsymbol{I}_6)$
- σ: a half-Cauchy distribution with scale parameter 10

We set $\boldsymbol{V} = \boldsymbol{S} = \boldsymbol{I}_5$, and $\nu = 5$.

Task 1 - Joint Prior Distribution of Parameters

Let n denote the number of hurricanes in the dataset. The prior distribution of $\Theta = (\mathbf{B}^\top, \boldsymbol{\mu}^\top, \boldsymbol{\Sigma}, \boldsymbol{\gamma}^\top, \sigma)$ is given by

$$\begin{aligned}
\pi(\Theta) &= \pi(\mathbf{B}^\top, \boldsymbol{\mu}^\top, \boldsymbol{\Sigma}, \boldsymbol{\gamma}^\top, \sigma) \\
&= \pi(\mathbf{B}^\top \mid \boldsymbol{\mu}^\top, \boldsymbol{\Sigma})\pi(\boldsymbol{\mu}^\top, \boldsymbol{\Sigma})\pi(\boldsymbol{\gamma})\pi(\sigma) \\
&= \left(\prod_{i=1}^{n} \pi(\boldsymbol{\beta}_i^\top \mid \boldsymbol{\mu}^\top, \boldsymbol{\Sigma})\right) \pi(\boldsymbol{\mu})\pi(\boldsymbol{\Sigma})\pi(\boldsymbol{\gamma})\pi(\sigma) \\
&\propto |\boldsymbol{\Sigma}|^{-n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{\beta}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}_i - \boldsymbol{\mu})\right) \\
&\quad \times \exp\left(-\frac{1}{2}\boldsymbol{\mu}^\top \boldsymbol{V}^{-1}\boldsymbol{\mu}\right) \\
&\quad \times |\boldsymbol{\Sigma}|^{-(\nu+6)/2} \exp\left(-\frac{1}{2}\mathrm{tr}(\boldsymbol{S}\boldsymbol{\Sigma}^{-1})\right) \\
&\quad \times \exp\left(-\frac{1}{2}\cdot 400\boldsymbol{\gamma}^\top\boldsymbol{\gamma}\right) \times \frac{I(\sigma > 0)}{1 + (\sigma/10)^2}.
\end{aligned}$$

Task 1 - Likelihood

Let m_i denote the number of observations and $\boldsymbol{Y}_i = (Y_{i,1}, \ldots Y_{i,m_i})^\top$ denote the wind speed data of the i-th hurricane (excluding the first and second observations), where $Y_{i,k} = Y_i(6k+6)$. Denote $\boldsymbol{Y} = (\boldsymbol{Y}_1^\top, \boldsymbol{Y}_2^\top, \ldots, \boldsymbol{Y}_n^\top)^\top$, and $\mathbf{Z}_{i,k} = (1, Y_{i,k}, \Delta_{i,1}(6k+6), \Delta_{i,2}(6k+6), \Delta_{i,3}(6k+6))^\top$.

Given that

$$Y_{i,j} \mid (\boldsymbol{\beta}_i^\top, \boldsymbol{\gamma}^\top, \sigma) \sim N(\mathbf{Z}_{i,j-1}^\top \boldsymbol{\beta}_i + \mathbf{X}_i^\top \boldsymbol{\gamma}, \sigma^2),$$

we have

$$\begin{aligned} L(\Theta \mid \boldsymbol{Y}^\top) &= \prod_{i=1}^{n} \prod_{j=1}^{m_i} L(\Theta \mid Y_{i,j}) \\ &= \prod_{i=1}^{n} \prod_{j=1}^{m_i} \left[\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(Y_{i,j} - \mathbf{Z}_{i,j-1}^\top \boldsymbol{\beta}_i - \mathbf{X}_i^\top \boldsymbol{\gamma})^2}{2\sigma^2} \right) \right]. \end{aligned}$$

Task 1 - Joint Posterior Distribution of Parameters

$$
\begin{aligned}
\pi(\Theta \mid \boldsymbol{Y}^\top) &\propto L(\Theta \mid \boldsymbol{Y}^\top)\pi(\Theta) \\
&\propto \frac{I(\sigma > 0)\sigma^{-\sum_{i=1}^{n} m_i}}{1 + (\sigma/10)^2} |\boldsymbol{\Sigma}|^{-(n+\nu+6)/2} \\
&\quad \times \exp\Bigg[-\frac{1}{2}\Bigg(\sum_{i=1}^{n} (\boldsymbol{\beta}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{\beta}_i - \boldsymbol{\mu}) + \boldsymbol{\mu}^\top \boldsymbol{V}^{-1} \boldsymbol{\mu} \\
&\qquad\qquad + \operatorname{tr}(\boldsymbol{S}\boldsymbol{\Sigma}^{-1}) + 400\|\gamma\|_2^2 \Bigg)\Bigg] \\
&\quad \times \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} \sum_{j=1}^{m_i} (Y_{i,j} - \mathbf{Z}_{i,j-1}^\top \beta_i - \mathbf{X}_i^\top \gamma)^2 \right).
\end{aligned}
$$

Task 2

Objective:

Design and implement a custom MCMC algorithm for the outlined Bayesian hierarchical model. Monitor the convergence of the MCMC chains, using diagnostic plots and summary statistics to check for any issues.

Task 2 - MCMC Algorithm

Algorithm 1 MCMC Algorithm (Part 1)

Require: $\mathbf{Y}$; $\boldsymbol{\beta}_0, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, \sigma_0, \gamma_0$

Ensure: $\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}, \widehat{\sigma}, \widehat{\gamma} \approx \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \sigma, \gamma$

$i \leftarrow 0$, where i is the current number of iterations

while iteration times is not met **do**

$i \leftarrow i + 1$

Gibble sampling for $\boldsymbol{\beta}$

$$\boldsymbol{\beta}^{(k)} \sim \mathcal{N}(\boldsymbol{\beta}|\mathbf{Y}, \boldsymbol{\mu}^{(k-1)}, \boldsymbol{\Sigma}^{(k-1)}, \sigma^{(k-1)}, \boldsymbol{\gamma}^{(k-1)})$$

Gibble sampling for $\boldsymbol{\mu}$

$$\boldsymbol{\mu}^{(k)} \sim \mathcal{N}(\boldsymbol{\mu}|\mathbf{Y}, \boldsymbol{\beta}^{(k)}, \boldsymbol{\Sigma}^{(k-1)}, \sigma^{(k-1)}, \boldsymbol{\gamma}^{(k-1)})$$

Gibble sampling for Σ

$$\boldsymbol{\Sigma}^{(k)} \sim \mathcal{W}\rangle\int\langle\dashv\nabla\sqcup(\boldsymbol{\Sigma}|\mathbf{Y}, \boldsymbol{\beta}^{(k)}, \boldsymbol{\mu}^{(k)}, \sigma^{(k-1)}, \gamma^{(k-1)})$$

Gibble sampling for $\boldsymbol{\gamma}$ $\boldsymbol{\gamma}^{(k)} \sim p(\boldsymbol{\Sigma}|\mathbf{Y}, \boldsymbol{\beta}^{(k)}, \boldsymbol{\mu}^{(k)}, , \Sigma^{(k)}, \sigma^{(k-1)})$

Task 2 - MCMC Algorithm

Algorithm 2 MCMC Algorithm (Part 2)

Metropolis-Hastings for σ

Propose a new value σ^* from a normal distribution with mean $\sigma^{(k-1)}$ and a small variance.

Compute the acceptance ratio

$$\gamma = \frac{p(\gamma^{(k)}|\mathbf{Y}, \boldsymbol{\beta}^{(k)}, \boldsymbol{\mu}^{(k)}, \boldsymbol{\Sigma}^{(k)}, \sigma^{(k)})q(\gamma^{(k-1)}|\gamma^{(k)})}{p(\gamma^{(k-1)}|\mathbf{Y}, \boldsymbol{\beta}^{(k)}, \boldsymbol{\mu}^{(k)}, \boldsymbol{\Sigma}^{(k)}, \sigma^{(k)})q(\gamma^{(k)}|\gamma^{(k-1)})}$$

Generate a random number u from a uniform distribution between 0 and 1.

If $u \leq r$, set $\sigma^{(k)} = \sigma^*$, otherwise set $\sigma^{(k)} = \sigma^{(k-1)}$.

end while

Task 2 - MCMC Algorithm

Initial Values:

- $\boldsymbol{\beta}_i$: This can be obtained through the random effects term in the lmm model. The random effects term can be added to the fixed effects term to obtain $\boldsymbol{\beta}_i^{(0)}$.
- $\boldsymbol{\mu}$: This can be obtained through the fixed effects term lmm$coefficients[1]$.
- σ^2: This can be obtained through the model residual sigma0.
- Σ^{-1}: This can be obtained through the VarCorr(lmm) function which returns the covariance matrix of the random effects in the model. The inverse of this matrix can be taken to obtain ${\Sigma^{-1}}^{(0)}$.

Table 1: Initial Values

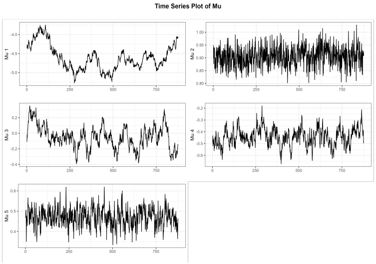
Parameter	Value
$\boldsymbol{\mu}$	$(24.25, 0.94, -0.02, -0.24, 0.47)$
$\boldsymbol{\gamma}$	$(-0.01, 0.35, 0.28, 0.37, 0.12, 0.08)$
Σ	$\begin{pmatrix} 0.36 & -0.01 & 0.04 & 0.12 & 0.03 \\ -0.01 & 0.00 & -0.00 & -0.00 & 0.00 \\ 0.04 & -0.00 & 0.04 & 0.03 & -0.02 \\ 0.12 & -0.00 & 0.03 & 0.07 & 0.00 \\ 0.03 & 0.00 & -0.02 & 0.00 & 0.02 \end{pmatrix}$
σ^2	5.27

Task 2 - MCMC Algorithm R code

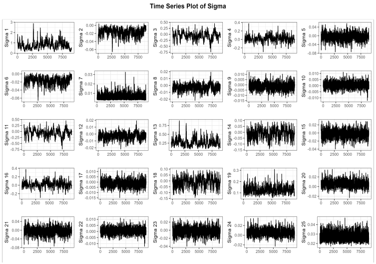
```r
# Gimbble sampling algorithm
B_sample <- function(mu, Sigma, gamma, sigma) {
  Sigma.inv <- solve(Sigma)
  B_mean_cov <- function(i) {
    cov <- solve(Sigma.inv + 1/sigma^2 * t(Z[[i]]) %*% Z[[i]])
    mean <- cov %*% (Sigma.inv %*% mu + 1/sigma^2 * colSums((Y[[i]] - (X[i,] %*% gamma)[,]) * Z[[i]]))
    list(mean = mean, cov = cov)
  }
  mean_cov_list <- lapply(1:n, B_mean_cov)
  B <- sapply(mean_cov_list, function(x) {mvrnorm(mu = x$mean, Sigma = x$cov)})
  return(B)
}

# MH algorithm (random walk)
sigma_sample <- function(sigma, B, gamma, a) {
  sigma_new <- sigma + (runif(1) - 0.5) * 2 * a # candidate sigma
  if (sigma_new <= 0) {
    return(sigma)
  }
  RSS <- sum(sapply(1:n, function(i) sum((Y[[i]] - Z[[i]] %*% B[,i] - (X[i,] %*% gamma)[,])^2)))
  log_kernal_ratio <- -sum(m) * log(sigma_new/sigma) +
    log(1 + (sigma/10)^2) - log(1 + (sigma_new/10)^2) -
    0.5 * (1/sigma_new^2 - 1/sigma^2) * RSS
  log_prob <- min(0, log_kernal_ratio)
  sigma <- ifelse(log_prob > log(runif(1)), sigma_new, sigma)
  return(sigma)
}
```

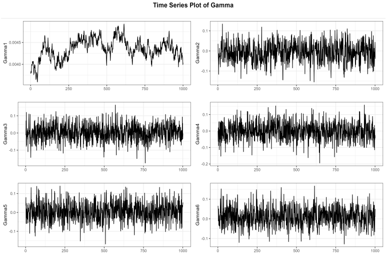
Task 3 - Results Presentation



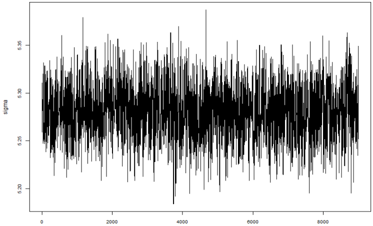
(a) Mu



(b) Sigma



(c) Gamma



(d) sigma

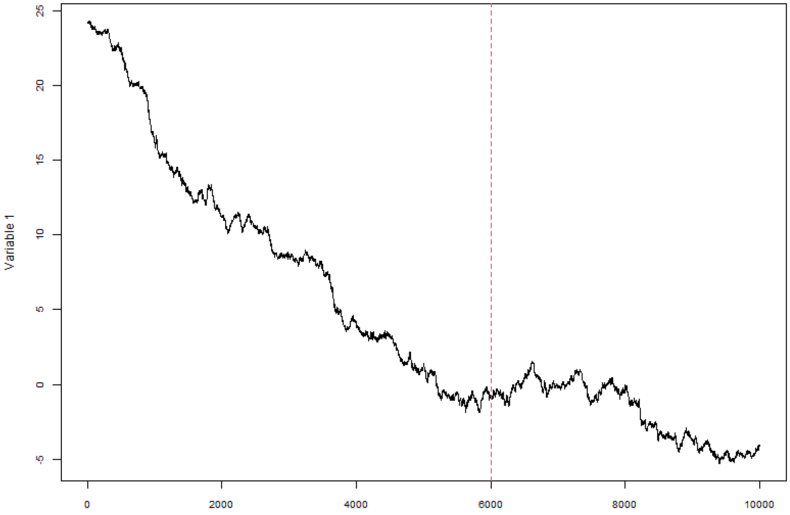
Task 3

Objective:

Compute posterior summaries and 95% credible intervals of γ, the fixed effects associated with the covariates in the model. Using the estimated Bayesian model, answer the following questions:

(1) Are there seasonal differences in hurricane wind speeds?
(2) Is there evidence to support the claim that hurricane wind speeds have been increasing over the years?

Task 3 - Parameters Convergence



Task 4 - Objective

Objective:

With the estimated model parameters and covariate values, you can calculate the predicted wind speed for each time point using the model equation. This way, you can track the hurricane and compare the predicted wind speeds with the actual wind speeds recorded during the hurricane. Please evaluate how well the estimated Bayesian model can track individual hurricanes.

Task 4

Prediction: Using the parameters after burn-in, we can obtain the predicted value for each hurricane.

$$Y_i(t+6) = \beta_{0,i} + \beta_{1,i} Y_i(t) + \beta_{2,i}\Delta_{i,1}(t) + \beta_{3,i}\Delta_{i,2}(t) + \beta_{4,i}\Delta_{i,3}(t) + \mathbf{X}_i{}^{\top}\boldsymbol{\gamma}$$

Performance evaluation: For each hurricane, we can evaluate the estimated Bayesian model performance by calculating

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Task 4

Table 2: Summary of RMSE and R-squared for selected hurricanes

ID	Year	RMSE	R-squared
ABBY.1960	1960	8.8804	0.7700
ABBY.1964	1964	9.6430	0.3033
ABBY.1968	1968	3.5043	0.9360
ABLE.1950	1950	3.6755	0.9813
ABLE.1951	1951	3.4802	0.9767
ABLE.1952	1952	4.5183	0.9583
AGNES.1972	1972	5.2483	0.8881
ALBERTO.1982	1982	8.0473	0.7499
ALBERTO.1988	1988	2.6121	0.7420
ALBERTO.1994	1994	4.3941	0.8807
ALBERTO.2000	2000	3.7896	0.9625
ALBERTO.2006	2006	4.3591	0.7882
ALBERTO.2012	2012	3.2193	0.8036
ALEX.1998	1998	2.9351	0.7289
ALEX.2004	2004	5.4552	0.9539

Task 4

Prediction performance on random chosen example hurricanes.

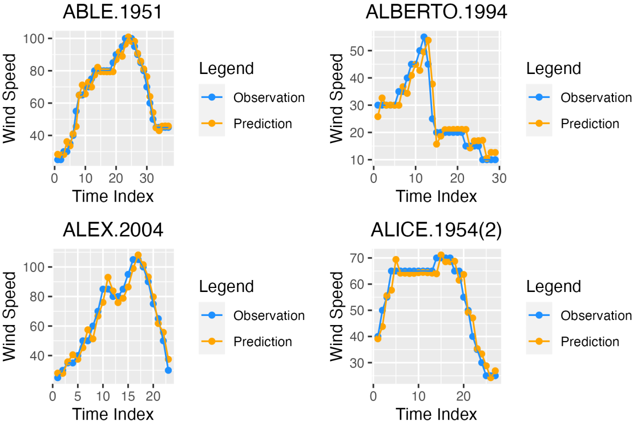


Figure 1: Time series prediction plot

Task 4

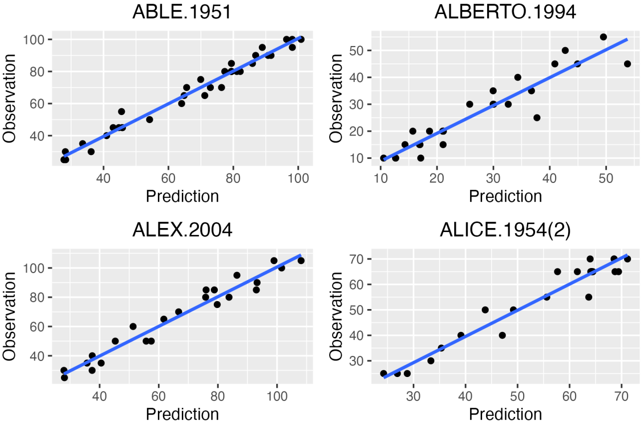


Figure 2: Prediction vs. observation

Discussions

- Parameters Convergence Problem

Reference

Reference

Q&A

- Thanks for listening!