P8160 - Bayesian Modeling of Hurricane Trajectories

Paulina Han, Waveley Qiu, Yida Wang, Lin Yang, Jibei Zheng, Haolin Zhong

2022-05-09

Hurricane Ida

CAPITAL WEATHER GANG

Ida's impact from the Gulf Coast to Northeast — by the numbers

The storm caused more than 40 deaths in the Northeast, brought tornadoes in six states and unleashed 172 mph winds in Louisiana





From: Livingston, I., The Washington Post, 2021

Saffir-Simpson Wind Scale

protective action, including evacuating from areas vulnerable to storm surge.



From: NHC NOAA

Proposed Hierarchical Bayesian Model

The following hierarchical Bayesian model was proposed to predict the wind speed of the i^{th} hurricane at time t + 6:

$$Y_{i}(t+6) = \beta_{0,i} + \beta_{1,i} Y_{i}(t) + \beta_{2,i} \Delta_{i,1}(t) + \beta_{3,i} \Delta_{i,2}(t) + \beta_{4,i} \Delta_{i,3}(t) + \varepsilon_{i}(t),$$

where $Y_i(t)$ is the wind speed at time t, $\Delta_{i,1}(t)$, $\Delta_{i,2}(t)$, $\Delta_{i,3}(t)$ are the changes in latitude, longitude, and wind speed between times t and t-6, $\varepsilon_i(t)$ is the random error associated with each $Y_i(t+6)$

We want to estimate the random coefficients, $\beta_i = (\beta_{0,i}, \beta_{1,i}, \beta_{2,i}, \beta_{3,i}.\beta_{4,i})$, for each hurricane.

Assumed Prior Distributions

The prior distributions for each of these parameters are assumed to be as follows:

$$eta_i \sim \textit{MVN}(\mu, \Sigma)$$
 $P(\mu) \propto 1$
$$P\left(\Sigma^{-1}\right) \propto |\Sigma|^{-(d+1)} \exp\left(-\frac{1}{2} \textit{tr}(\Sigma^{-1})\right),$$
 where d is the dimension of β_i

$$\epsilon_i(t) \sim \textit{N}(0,\sigma^2),$$
 which are independent across t $P\left(\sigma^2\right) \propto \frac{1}{\sigma^2}$

Goals

- 1. Construct an MCMC algorithm from which we can sample from a posterior distribution to estimate $\Theta = (\boldsymbol{B}^T, \boldsymbol{\mu}^T, \sigma^2, \boldsymbol{\Sigma})$.
- Conduct analysis using estimated parameters to understand their properties.
 - a. Seasonal changes in any of the coefficients
 - b. Predictive influence of these coefficients on forecasting hurricane impact.



Data

ID: ID of hurricanes

Year: In which the hurricane occurred

Month: In which the hurricane occurred

Nature: Nature of the hurricane

ET: Extra TropicalDS: Disturbance

► NR: Not Rated

SS: Sub Tropical

TS: Tropical Storm

Time: dates and time of the record

Latitude and Longitude: The location of a hurricane check point

Wind.kt: Maximum wind speed (in Knot) at each check point

MCMC for Hierarchical Bayesian Model:Data Partition

Data Partition

- ► For each hurricane, 80% of records were randomly assigned to the training set and the remaining 20% were assigned to testing set.
- Hurricanes with less than 7 records were removed: at least 5 observations are included in the training set and 1 observation is included in the testing set.

Distribution Derivation:

- Gibbs sampler is used to generate random variables from given distribution:
- Let $\Theta = (\mathbf{B}^T, \boldsymbol{\mu}^T, \boldsymbol{\sigma}^2, \boldsymbol{\Sigma})$, the posterior distribution can be written as:

$$P(\Theta \mid Y) \propto f(Y|\Theta)P(\Theta)$$

= $f(Y \mid \mathbf{B}, \boldsymbol{\mu}, \sigma^2, \boldsymbol{\Sigma})f(\mathbf{B} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})P(\boldsymbol{\mu})P(\sigma^2)P(\boldsymbol{\Sigma}^{-1})$

Posterior Distribution of Θ :

▶ $Y_i \sim MVN(X_i\beta_i^T, \sigma^2 I_{n_i})$, where n_i is the number of observation of the i^{th} hurricane.

$$f(Y \mid \mathbf{B}, \boldsymbol{\mu}, \sigma^{2}, \boldsymbol{\Sigma}) = \prod_{i=1}^{N} f(Y_{i} \mid B, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \sigma^{2})$$

$$= \prod_{i=1}^{N} (2\pi)^{-\frac{n_{i}}{2}} \left| \sigma^{2} I_{n_{i}} \right|^{-\frac{1}{2}} \exp(-\frac{1}{2} (Y_{i} - X_{i} \beta_{i}^{T})^{T} (\sigma^{2} I_{n_{i}})^{-1}$$

▶
$$β_i$$
 ~ MVN($μ$, $Σ$):

$$f(B|\mu, \Sigma) = \prod_{i=1}^{N} (2\pi)^{-\frac{5}{2}} |\sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\beta_i - \mu)^T \Sigma^{-1}(\beta_i - \mu)\right)$$

Posterior Distribution of Θ :

• the posterior distribution of Θ , where $A = \Sigma^{-1}$:

$$P(\Theta \mid Y) \propto \sigma^{-\sum_{i=1}^{N} n_i - 2} |A|^{d + \frac{N}{2} + 1}$$

$$\exp \left[-\frac{1}{2} \sum_{i=1}^{N} \left[(Y_i - X_i \beta_i^T)^T (\sigma^2 I_{n_i})^{-1} (Y_i - X_i \beta_i^T) + (\beta_i - \mu)^T (\gamma_i - X_i \beta_i^T) \right] \right]$$

$$-\frac{1}{2} tr(A)$$

Conditional Distribution of each parameter:

- $\beta_i \sim MVN(V^{-1}M, V^{-1}), \text{ where } V = \Sigma^{-1} + X_i^T \sigma^{-2} I_{ni} X_i,$ $M = Y_i^T \sigma^{-2} I_{ni} X_i + \mu^T \Sigma^{-1}$
- $\blacktriangleright \mu \sim MVN(V^{-1}M, V^{-1})$, where $V = N\Sigma^{-1}$, $M = \sum_{i=1}^{N} \Sigma^{-1}\beta_i$

Gibbs Sampling

Initialize
$$\Theta_0 = (\mathbf{B}_0, \boldsymbol{\mu}_0, \sigma_0^2, \boldsymbol{\Sigma}_0)$$

for iteration i = 1,2,... do

Sample
$$\mathbf{B}_i \sim \pi(\mathbf{B}|oldsymbol{\mu}_{i-1}, \sigma^2_{i-1}, \mathbf{\Sigma}_{i-1}, \mathbf{Y})$$

Sample
$$\mu_i \sim \pi(\mu|\mathbf{B}_i, \sigma_{i-1}^2, \mathbf{\Sigma}_{i-1}, \mathbf{Y})$$

Sample
$$\sigma_i^2 \sim \pi(\sigma^2 | \mathbf{B}_i, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_{i-1}, \mathbf{Y})$$

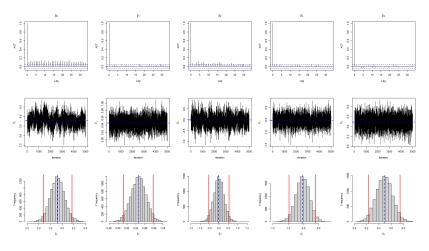
Sample
$$\mathbf{\Sigma}_i^{-1} \sim \pi(\mathbf{\Sigma}^{-1}|\mathbf{B}_i, \boldsymbol{\mu}_i, \sigma_i^2, \mathbf{Y})$$

take inverse

end for

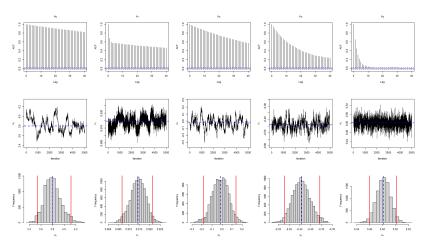
Convergence Plots and Distributions of B

chain length: 10000; burn: 5000



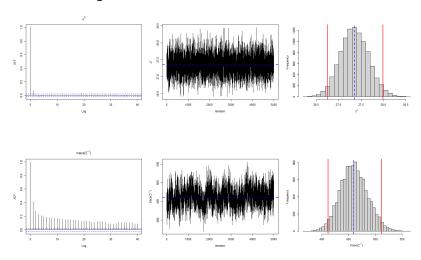
Convergence Plots and Distributions of μ

chain length: 10000; burn: 5000



Convergence Plots and Distributions of σ^2 and Σ^{-1}

chain length: 10000; burn: 5000



Estimates for **B** and μ

Hurricane	β_0	β_1	β_2	β_3	β_4
ABLE.1950	3.79[3.19,4.42]	0.95[0.92,0.97]	-0.12[-0.65,0.4]	-0.53[-0.98,-0.1]	0.54[0.32,0.77]
BAKER.1950	3.79[3.2,4.4]	0.92[0.89, 0.95]	-0.1[-0.65,0.44]	-0.39[-0.84,0.07]	0.68[0.49, 0.87]
CHARLIE.1950	3.78[3.17,4.38]	0.94[0.92, 0.97]	-0.01[-0.53,0.51]	-0.42[-0.85,0.04]	0.45[0.18, 0.71]
DOG.1950	3.81[3.22,4.43]	0.96[0.94, 0.97]	-0.06[-0.58,0.45]	-0.39[-0.79,-0.01]	0.53[0.31, 0.76]
EASY.1950	3.8[3.19, 4.43]	0.92[0.88, 0.95]	-0.01[-0.53, 0.53]	-0.43[-0.89, 0.02]	0.54[0.33, 0.74]
FOX.1950	3.79[3.16,4.4]	0.95[0.93, 0.98]	-0.1[-0.66,0.42]	-0.56[-1.02,-0.11]	0.56[0.31, 0.81]
GEORGE.1950	3.81[3.21,4.4]	0.95[0.93, 0.98]	-0.03[-0.56,0.49]	-0.38[-0.78,0.02]	0.46[0.19, 0.73]
HOW.1950	3.82[3.22,4.43]	0.89[0.82, 0.96]	-0.02[-0.55,0.52]	-0.43[-0.89,0.03]	0.47[0.17, 0.79]
ITEM.1950	3.83[3.23,4.45]	0.92[0.88, 0.97]	-0.05[-0.57,0.49]	-0.45[-0.91,0.01]	0.5[0.3, 0.71]
JIG.1950	3.83[3.23, 4.45]	0.95[0.92, 0.98]	-0.02[-0.55,0.5]	-0.47[-0.92,-0.03]	0.48[0.22, 0.75]

μ_0	μ_1	μ_2	μ_3	μ_4
3.82[3.59, 4.12]	0.91[0.91, 0.92]	-0.03[-0.2,0.12]	-0.44[-0.52, -0.36]	0.48[0.46, 0.5]

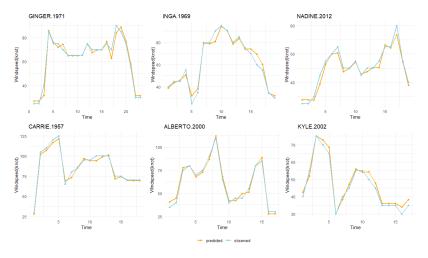
Estimates for σ^2 and Σ^{-1}

$$\sigma^2 = 27.36[26.76, 27.98]$$

Σ^{-1} :

17.0148877	7.740045	-0.1368787	0.8033023	1.8493396
7.7400455	360.894502	5.5883769	3.7666486	-9.9431197
-0.1368787	5.588377	19.0227527	1.0456746	0.5528947
0.8033023	3.766649	1.0456746	22.7116754	-3.4877498
1.8493396	-9.943120	0.5528947	-3.4877498	40.9917065

Prediction Performance



Predicted vs. observed values of windspeed

Prediction Performance

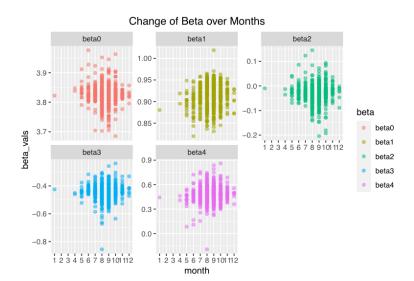
Overall RMSE = 5.78

Hurricane	RMSE	0.20	Λ				
ABLE.1950	3.142189		11				
BAKER.1950	6.665961		131				
CHARLIE.1950	2.420988	0.15					
DOG.1950	3.352041	Queusity 0.10	1 1				
EASY.1950	7.954826	B 0.10	1 1				
FOX.1950	3.360563						
GEORGE.1950	3.966812	0.05					
HOW.1950	3.212678						
ITEM.1950	15.515327	0.00	\sim	~~~			
JIG.1950	2.198730	0.00	5.78 10.00	20.00	30.00 RMSE	40.00	50.0

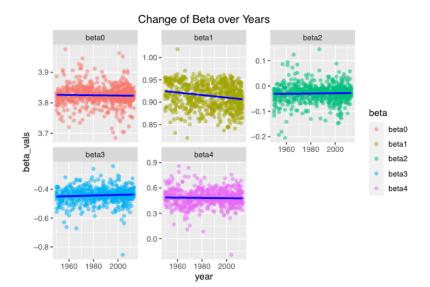
RMSE table and density plot

Seasonal Analysis

$\hat{\beta}_i$ by Starting Month



$\hat{\beta}_i$ by Year



Seasonal Analysis

Model 1:

$$\beta_i = \alpha_0 + \alpha_{1m}I(Month = M) + \alpha_2 \times Year + \alpha_{3n}I(Type = N)$$

M: April-December (reference: January)

N: ET, NR, SS, TS (reference: DS)

Coefficients of Model 1 for β_i

	Beta0		Beta	a1	Bet	a2	Beta	a3	Beta4		
	Estimate	Pr(> t)									
(Intercept)	3.8722305	0.0000000	1.4200232	0.0000000	-0.0807485	0.6123701	-0.7395516	0.0012896	0.8052773	0.0590557	
factor(month)4	0.0211436	0.5897652	0.0316515	0.3472330	-0.0115682	0.7896176	-0.0234552	0.7064830	-0.0053003	0.9635270	
factor(month)5	0.0202412	0.5405684	0.0281544	0.3215509	-0.0113447	0.7563959	-0.0120033	0.8192595	-0.0390352	0.6897011	
factor(month)6	0.0159306	0.6239449	0.0246013	0.3779221	-0.0148705	0.6789563	0.0098662	0.8483886	0.0112067	0.9071129	
factor(month)7	0.0078141	0.8090619	0.0404591	0.1453332	-0.0165819	0.6428643	-0.0027461	0.9573635	0.0199011	0.8350861	
factor(month)8	0.0019068	0.9527290	0.0425000	0.1241205	-0.0260312	0.4643250	-0.0112584	0.8255944	0.0229217	0.8095080	
factor(month)9	0.0009337	0.9768273	0.0472980	0.0868820	-0.0233893	0.5105760	-0.0075900	0.8818256	0.0389302	0.6820459	
factor(month)10	0.0074737	0.8163761	0.0411045	0.1371659	-0.0168883	0.6351370	-0.0007253	0.9886799	0.0268667	0.7776519	
factor(month)11	0.0057884	0.8588527	0.0448708	0.1086605	-0.0079430	0.8253327	0.0043686	0.9326594	0.0387334	0.6872898	
factor(month)12	0.0048248	0.8874129	0.0308019	0.2926123	-0.0208686	0.5797518	0.0072339	0.8936869	0.0283150	0.7786590	
year	-0.0000290	0.6794636	-0.0002769	0.0000050	0.0000378	0.6260392	0.0001587	0.1544308	-0.0001713	0.4088111	
factor(type)ET	0.0075408	0.4379949	0.0086401	0.3006911	-0.0108462	0.3131136	-0.0192894	0.2118151	-0.0222770	0.4382889	
factor(type)NR	0.0005575	0.9705947	0.0072156	0.5784605	-0.0132239	0.4292334	-0.0418405	0.0819304	0.0070952	0.8739144	
factor(type)SS	0.0074733	0.2505823	0.0082071	0.1417607	-0.0038667	0.5906999	0.0003254	0.9748646	-0.0225589	0.2407159	
factor(type)TS	0.0057948	0.2474418	0.0009877	0.8182988	-0.0024415	0.6592917	-0.0141969	0.0746373	-0.0108813	0.4623952	

Seasonal Analysis

Model 2:

 $Beta_i = \alpha_0 + \alpha_{1s}I(Season = S) + \alpha_2 \times Year + \alpha_{3n}I(Type = N)$

S: Spring, Summer, Winter (reference: Fall)

N: ET, NR, SS, TS (reference: DS)

Coefficients of Model 2 for β_i

	Beta 0		Beta 1		Beta 2		Beta 3		Beta 4	
	Estimate	Pr(> t)								
(Intercept)	3.8777185	0.0000000	1.4515958	0.0000000	-0.1108749	0.4777627	-0.7432739	0.0009368	0.8246599	0.0469629
factor(season)spring	0.0165743	0.0386414	-0.0161701	0.0195181	0.0080903	0.3613959	-0.0095303	0.4529259	-0.0700505	0.0029939
factor(season)summer	0.0017442	0.4921921	-0.0054774	0.0126824	-0.0021923	0.4356701	-0.0020772	0.6061277	-0.0145208	0.0520294
factor(season)winter	0.0004419	0.9695014	-0.0175939	0.0782636	0.0012626	0.9214169	0.0101974	0.5781775	-0.0099913	0.7687012
year	-0.0000297	0.6721280	-0.0002706	0.0000093	0.0000429	0.5814228	0.0001589	0.1539417	-0.0001639	0.4270083
factor(type)ET	0.0090086	0.3383681	0.0086688	0.2860439	-0.0058182	0.5765098	-0.0167604	0.2616162	-0.0206719	0.4547698
factor(type)NR	0.0017339	0.9077767	0.0079185	0.5401551	-0.0073224	0.6586863	-0.0384250	0.1059825	0.0085343	0.8462012
factor(type)SS	0.0077248	0.2318002	0.0080589	0.1486481	-0.0023967	0.7374924	0.0006525	0.9492157	-0.0222179	0.2419930
factor(type)TS	0.0047623	0.3404235	0.0024950	0.5629051	-0.0029696	0.5913088	-0.0157040	0.0477883	-0.0093683	0.5233984

Forecasting Hurricane Impact

Forecasting Hurricane Impact

ID: ID of the hurricanes

Season: In which year the hurricane occurred **Month**: In which month the hurricane occurred

Nature: Nature of the hurricane

Damage: Financial loss (in Billion U.S. dollars) caused by

hurricanes

Deaths: Number of death caused by hurricanes

Maxspeed: Maximum recorded wind speed of the hurricane

Meanspeed: Average wind speed of the hurricane

Maxpressure: Maximum recorded central pressure of the hurricane

Meanpressure: Average central pressure of the hurricane

Hours: Duration of the hurricane in hours

Total.Pop: Total affected population

Percent.Poor: % affected population that reside in low GDP

counties

Percent.USA: % affected population that reside in the United

States

LASSO Model for Damage

	Coefficients
(Intercept)	-533.5099174
season	3.3837824
deaths	0.0000000
monthJuly	0.0000000
monthJune	0.0000000
monthNovember	0.0000000
monthOctober	0.0000000
monthSeptember	0.0000000
natureNR	0.0000000
natureTS	0.0000000
maxspeed	1.2117851
meanspeed	0.0000000
maxpressure	0.0000000
meanpressure	0.0000000
hours	0.0000000
total pop	0.3187361
percent_poor	0.0000000
percent_usa	0.7073409
beta0	0.0000000
beta1	0.0000000
beta2	0.0000000
beta3	0.0000000
beta4	0.0000000

Refitted Linear Regression Model

Model: Y = γ_0 + γ_1 x season + γ_2 x maxspeed + γ_3 x total_pop + γ_4 x percent_usa

	Coefficients
(Intercept)	-1316.7386136
season	0.6485139
maxspeed	0.1968674
$total_pop$	0.0000033
percent_usa	0.1356486

Poisson Model for Deaths

 $Y_i \sim \mathsf{Poisson}(\mu_i)$, where $\mu_i = \mathit{hours}_i * \lambda_i$, λ_i is the number of deaths per hour

 $\mathsf{Model:} \ \mathit{log}(\lambda_i) = \boldsymbol{X_i^T} \boldsymbol{\gamma} + \mathit{log}(\mathit{hours}_i)$

	Estimate	Std. Error	z value	$\Pr(> z)$
(Intercept)	-199.5331639	11.8792784	-16.7967411	0.0000000
season	-0.0404185	0.0028048	-14.4104991	0.0000000
damage	0.0220163	0.0005679	38.7649762	0.0000000
monthJuly	-10.2286750	0.1604645	-63.7441688	0.0000000
monthJune	0.3928062	0.0989170	3.9710698	0.0000716
monthNovember	1.8733767	0.1664682	11.2536625	0.0000000
monthOctober	-1.6041896	0.0787720	-20.3649754	0.0000000
monthSeptember	1.2490015	0.0575033	21.7205350	0.0000000
natureNR	2.0903864	0.1371766	15.2386495	0.0000000
natureTS	-1.1903051	0.1118619	-10.6408484	0.0000000
maxspeed	0.0035207	0.0013988	2.5168778	0.0118400
meanspeed	-0.1978651	0.0039977	-49.4953412	0.0000000
maxpressure	0.0048106	0.0075485	0.6372945	0.5239331
meanpressure	0.0021204	0.0001759	12.0515409	0.0000000
total_pop	0.0000009	0.0000000	31.4237737	0.0000000
percent_poor	0.0873434	0.0010058	86.8433730	0.0000000
percent_usa	-0.0080185	0.0004884	-16.4173000	0.0000000
beta0	41.3531048	0.5634443	73.3934206	0.0000000
beta1	132.8784572	1.9305164	68.8305252	0.0000000
beta2	-10.7339527	0.5001340	-21.4621524	0.0000000
beta3	-0.4736994	0.5091748	-0.9303277	0.3522014
beta4	4.4919244	0.1971025	22.7897893	0.00000000

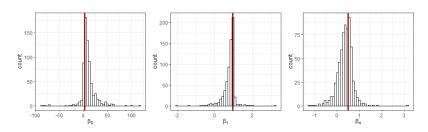
Discussion

Strength & Limitation of MCMC methods

- Bypass coefficient optimization process and directly sample coefficients from their assumed distributions
- ▶ Often computationally expensive and can be inefficient
- Convergence is not guaranteed

Why Non-convergence?

 $lackbox{egin{aligned} }eta_i\sim \textit{MVN}(\mu,\Sigma) \text{ may be a too strong of an assumption } \end{aligned}$



Distribution of β_i s obtained by performing OLS for each hurricane (red line: μ obtained by performing OLS on the whole training dataset)

Future work: use a more adequate distributional assumption of β_i which can account for skewness



References

- Livingston, I. (2021, September 3). Ida's impact from the Gulf Coast to northeast - by the numbers. The Washington Post. https://www.washingtonpost.com/weather/2021/09/03/hurr icane-ida-numbers-surge-wind-pressure-damage/
- Saffir-Simpson Hurricane Wind Scale. (n.d.). https://www.nhc.noaa.gov/aboutsshws.php