P8160 - Project 3
Baysian modeling of hurricane

Renjie Wei, Hao Zheng, Xinran Sun
Wentong Liu, Shengzhi Luo

2022-05-09

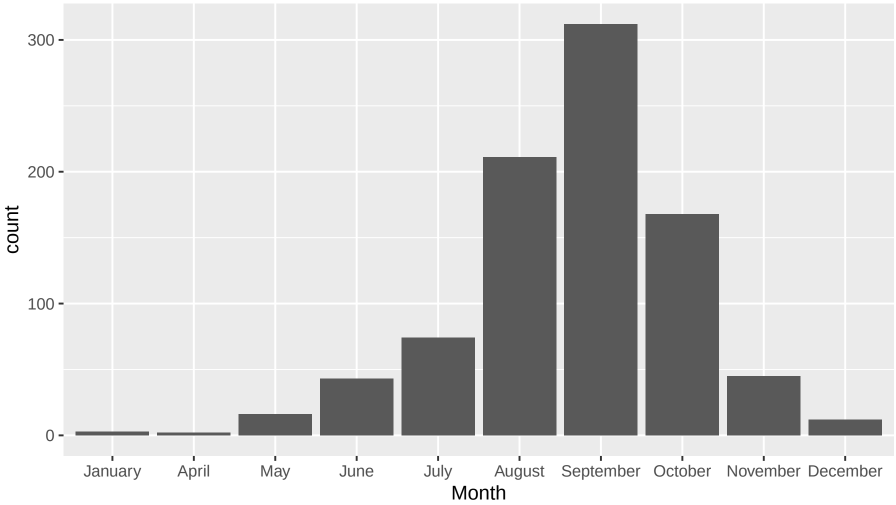
Introduction

- Hurricanes can result in death and economical damage
- There is an increasing desire to predict the speed and damage of the hurricanes
- Use Bayesian Model and Markov Chain Monte Carlo algorithm
  - Predict the wind speed of hurricanes
  - Study how hurricanes is related to death and financial loss

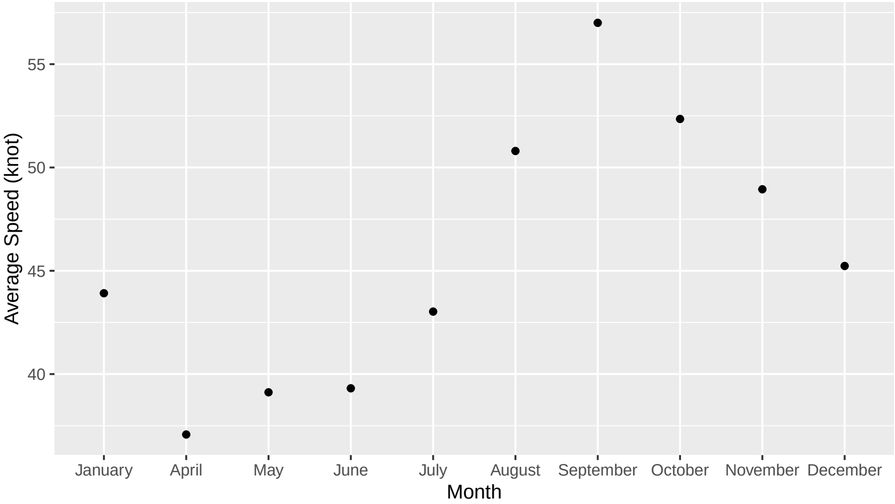
Dataset

- Hurrican703 dataset: 22038 observations $\times$ 8 variables
  - 702 hurricanes in the North Atlantic area in year 1950-2013
- Processed dataset: add 5 more variables into hurrican703
- Hurricanoutcome2 dataset: 43 observations $\times$ 14 variables

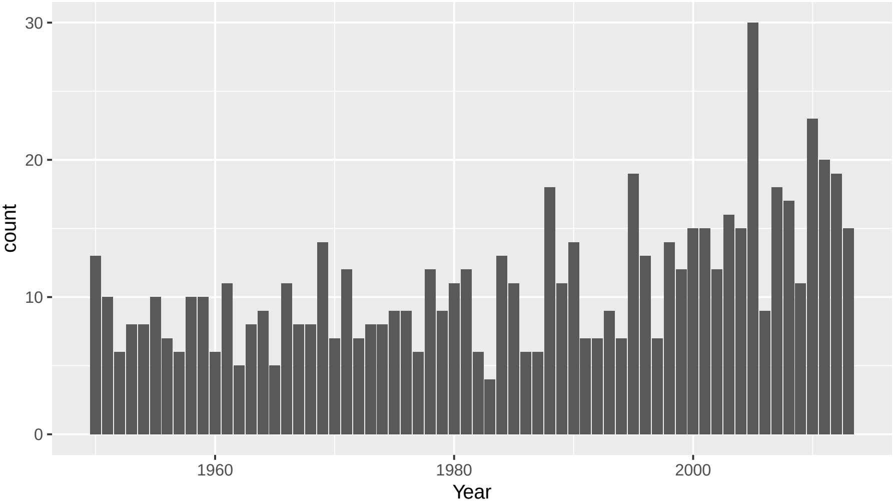
EDA - Count of Hurricanes in Each Month



EDA - Average Speed (knot) of Hurricanes in Each Month

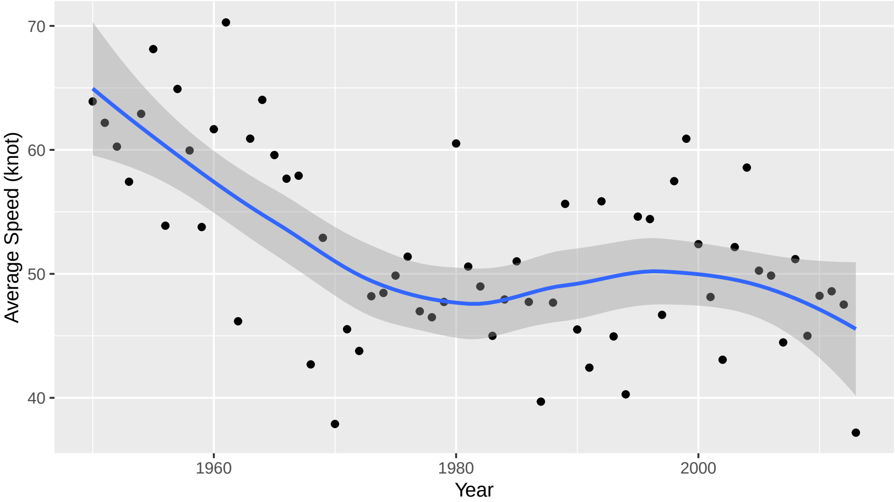


EDA - Count of Hurricanes in Each Year

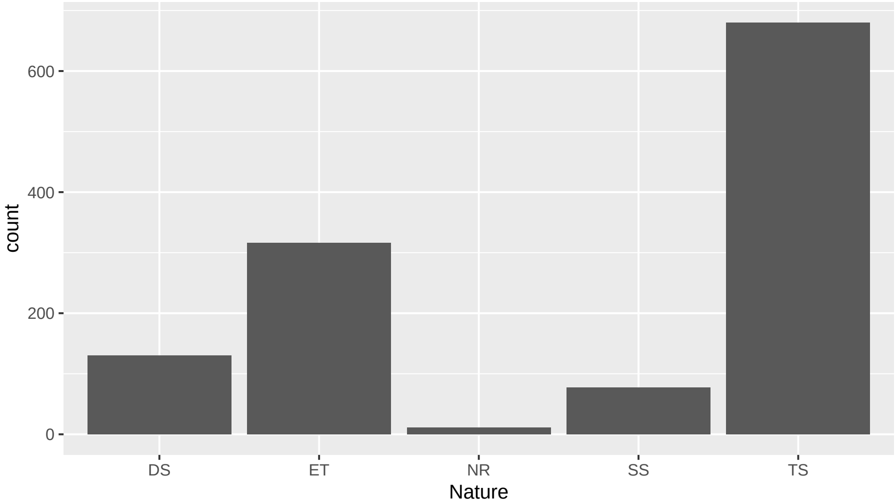


EDA - Average Speed (knot) of Hurricanes in Each Year

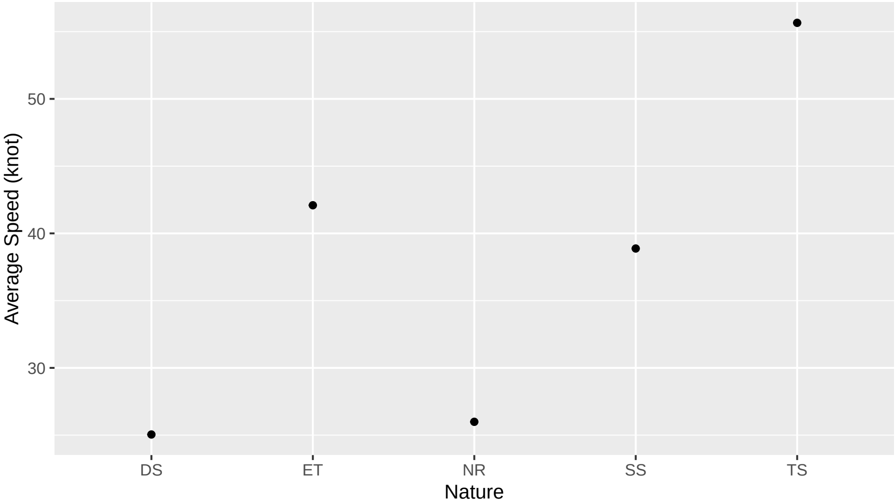
```
## `geom_smooth()` using formula 'y ~ x'
```



EDA - Count of Hurricanes in Each Nature



EDA - Average Speed (knot) of Hurricanes in Each Nature



Bayesian Model Setting

Model

The suggested Bayesian model is

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

- where $Y_i(t)$ 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 $t-6$, and $\epsilon_{i,t}$ follows a normal distributions with mean zero and variance σ^2, independent across t.
- $\beta_i = (\beta_{0,i}, \beta_{1,i}, ..., \beta_{5,i})$, we assume that $\beta_i \sim N(\mu, \Sigma_{d\times d})$, where d is dimension of β_i.

Priors

$$P(\sigma^2) \propto \frac{1}{\sigma^2}; \quad P(\mu) \propto 1; \quad P(\Sigma^{-1}) \propto |\Sigma|^{-(d+1)} \exp(-\frac{1}{2}\Sigma^{-1})$$

Posterior

- Derive $\pi(\Theta|Y)$, where $\Theta = (\mathbf{B}^\top, \mu^\top, \sigma^2, \Sigma)$, $\mathbf{B} = (\beta_1^\top, ..., \beta_n^\top)^\top$

Joint posterior

Notations

- $X_i(t)\beta_i^\top = \beta_{0,i} + \beta_{1,i}Y_i(t) + \beta_{2,i}\Delta_{i,1}(t) + \beta_{3,i}\Delta_{i,2}(t) + \beta_{4,i}\Delta_{i,3}(t)$
- For i^{th} hurricane, there may be m_i times of record (excluding the first and second observation), let

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

- Hence, $Y_i \mid X_i, \beta_i, \sigma^2 \sim N(X_i\beta_i^\top, \sigma^2 I)$
- Where, X_i is a $m_i \times d$ dimensional matrix

$$X_i = \begin{pmatrix} 1 & Y_i(t_0) & \Delta_{i,1}(t_0) & \Delta_{i,2}(t_0) & \Delta_{i,3}(t_0) \\ 1 & Y_i(t_1) & \Delta_{i,1}(t_1) & \Delta_{i,2}(t_1) & \Delta_{i,3}(t_1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & Y_i(t_{m_i-1}) & \Delta_{i,1}(t_{m_i-1}) & \Delta_{i,2}(t_{m_i-1}) & \Delta_{i,3}(t_{m_i-1}) \end{pmatrix}$$

Joint posterior

Posterior

$$
\begin{aligned}
\pi(\Theta|Y) &= \pi(\mathbf{B}^\top, \mu^\top, \sigma^2, \Sigma \mid Y) \\
&\propto \underbrace{\prod_{i=1}^{n} f(Y_i \mid \beta_i, \sigma^2)}_{\text{likelihood of } Y} \underbrace{\prod_{i=1}^{n} \pi(\beta_i \mid \mu, \Sigma)}_{\text{distribution of } \mathbf{B}} \underbrace{P(\sigma^2) P(\mu) P(\Sigma^{-1})}_{\text{priors}} \\
&\propto \prod_{i=1}^{n} \Big\{ (2\pi\sigma^2)^{-m_i/2} \exp\Big\{ -\frac{1}{2}(Y_i - X_i\beta_i^\top)^\top (\sigma^2 I)^{-1} (Y_i - X_i\beta_i^\top) \\
&\times \prod_{i=1}^{n} \Big\{ \det(2\pi\Sigma)^{-\frac{1}{2}} \exp\Big\{ -\frac{1}{2}(\beta_i - \mu)\Sigma^{-1}(\beta_i - \mu)^\top \Big\} \Big\} \\
&\times \frac{1}{\sigma^2} \times \det(\Sigma)^{-(d+1)} \exp\Big\{ -\frac{1}{2}\Sigma^{-1} \Big\}
\end{aligned}
$$

MCMC Algorithm

- Monte Carlo Method
  - Random sampling method to estimate quantity
- Markov Chain
  - Generates a sequence of random variables where the current state only depends on the nearest past
- Example: Gibbs Sampler
  - MCMC approaches with known conditional distributions
  - Samples from each random variables in turn given the value of all the others in the distribution

MCMC Algorithm

Conditional Posterior

- 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.
  - $\pi(\mathbf{B}|Y, \mu^\top, \sigma^2, \Sigma)$
  - $\pi(\sigma^2|Y, \mathbf{B}^\top, \mu^\top, \Sigma)$
  - $\pi(\Sigma|Y, \mathbf{B}^\top, \mu^\top, \sigma^2)$
  - $\pi(\mu|Y, \mathbf{B}^\top, \sigma^2, \Sigma)$

MCMC Algorithm - Conditional Posterior

- β_i: $\pi(\beta_i|Y, \mu^\top, \sigma^2, \Sigma) \sim \mathcal{N}(\hat{\beta}_i, \hat{\Sigma}_{\beta_i})$
  - where $\hat{\beta}_i = (\Sigma^{-1} + X_i^\top(\sigma^2 I)^{-1}X_i)^{-1}Y_i^\top(\sigma^2 I)^{-1}X_i + \mu\Sigma^{-1}, \hat{\Sigma}_{\beta_i} = (\Sigma^{-1} + X_i^\top(\sigma^2 I)^{-1}X_i)^{-1}$
- σ^2: $\pi(\sigma^2|Y, \mathbf{B}^\top, \mu^\top, \Sigma) \sim IG(\frac{1}{2}\sum_{i=1}^{n} m_i, \frac{1}{2}\sum_{i=1}^{n}(Y_i - X_i\beta_i^\top)^\top(Y_i - X_i\beta_i^\top))$
- Σ: $\pi(\Sigma|Y, \mathbf{B}^\top, \mu^\top, \sigma^2) \sim IW(n + d + 1,\ I + \sum_{i=1}^{n}(\beta_i - \mu)(\beta_i - \mu)^\top)$
- μ: $\pi(\mu|Y, \mathbf{B}^\top, \sigma^2, \Sigma) \sim \mathcal{N}(\frac{1}{n}\sum_{i=1}^{n}\beta_i, \frac{1}{n}\Sigma)$

MCMC Algorithm - Parameter Updates

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

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

MCMC Algorithm - Inital Values

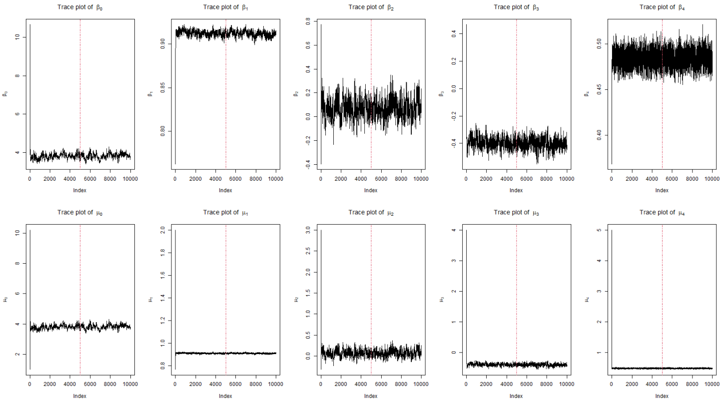
1. For initial value of **B**, we run multivariate linear regressions for each hurricane and use the regression coefficients β_i^{MLR} as the initial value for β_i. Then, the initial value of **B** can be represented as $\mathbf{B}_{init} = (\beta_1^{MLR^\top}, ..., \beta_n^{MLR^\top})^\top$.
2. For initial value of μ, we take the average of β_i^{MLR}, that is $\mu_{init} = \frac{1}{n}\sum_{i=1}^{n} \beta_n^{MLR}$
3. For initial value of σ^2, we take the average of the MSE for i hurricanes.
4. For initial value of Σ, we just set it to a simple diagonal matrix, i.e. $\Sigma_{init} = diag(1, 2, 3, 4, 5)$

MCMC Results

Details

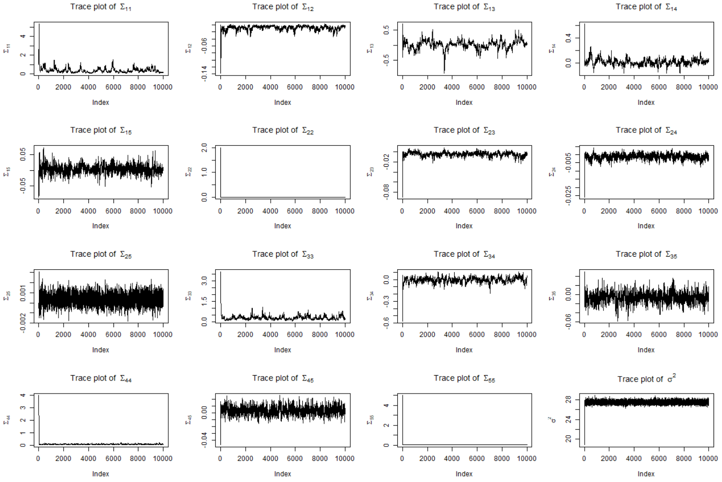
- 10000 iterations
- First 5000 iterations as burn-in period
- Estimates and inferences based on last 5000 MCMC samples

MCMC Results - Trace Plots 1



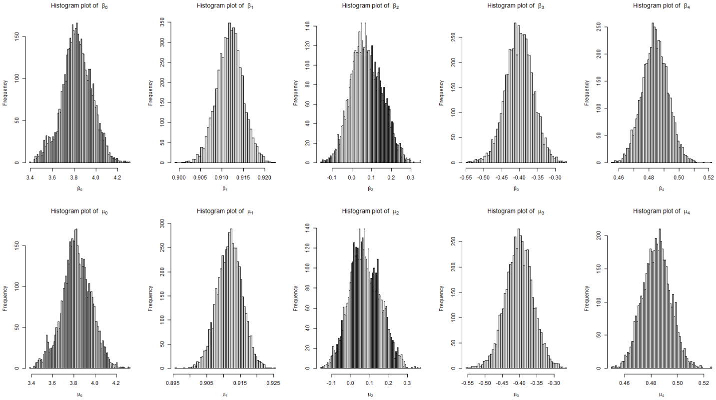
Trace plots of model parameters, based on 10000 MCMC sample

MCMC Results - Trace Plots 2



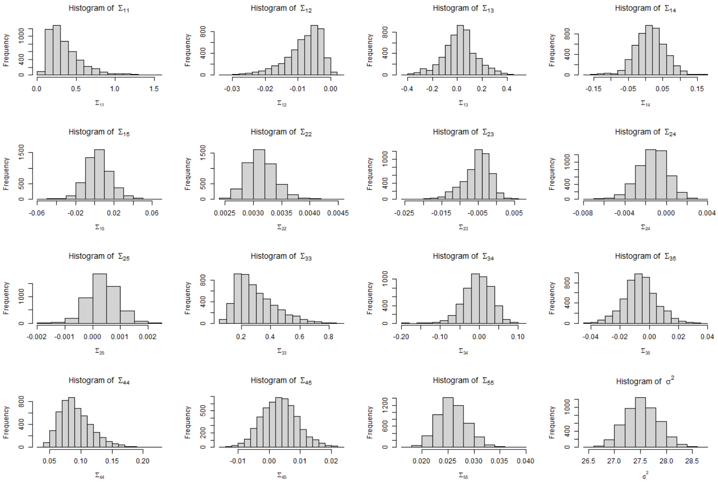
Trace plots of variance parameters, based on 10000 MCMC sample

MCMC Results - Histograms 1



Histograms of model parameters, based on last 5000 MCMC sample

MCMC Results - Histograms 2



Histograms of variance parameters, based on last 5000 MCMC sample

MCMC Results - Model Parameter Estimations and Inferences

Variables	$\bar{\beta}_i$	Var($\bar{\beta}_i$)	95% CI of $\bar{\beta}_i$	$\bar{\mu}$	Var($\bar{\mu}$)	95% CI of $\bar{\mu}$
intercept	3.8252	0.0185	(3.5588,4.0916)	3.8166	0.0190	(3.5468,4.0865)
Wind_prev	0.9118	0.0000	(0.9059,0.9177)	0.9121	0.0000	(0.9049,0.9194)
Lat_change	0.0744	0.0060	(-0.0776,0.2264)	0.0720	0.0065	(-0.0857,0.2298)
Long_change	-0.4014	0.0015	(-0.4771,-0.3257)	-0.3968	0.0016	(-0.4759,-0.3177)
Wind_change	0.4841	0.0001	(0.4674,0.5009)	0.4847	0.0001	(0.464,0.5053)

Bayesian posterior estimates for model parameters

MCMC Results - Variance Parameter Estimations and Inferences

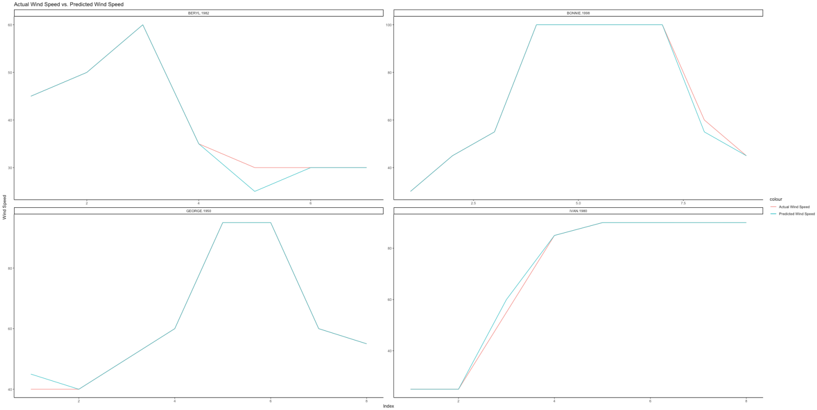
Parameters	Estimates	Variance	95% CI
Σ_{11}	0.3493	0.0435	(-0.0595,0.7581)
Σ_{12}	-0.0081	0.0000	(-0.0189,0.0027)
Σ_{13}	0.0201	0.0176	(-0.2399,0.2801)
Σ_{14}	0.0131	0.0019	(-0.0725,0.0987)
Σ_{15}	0.0035	0.0002	(-0.0215,0.0285)
Σ_{22}	0.0031	0.0000	(0.0026,0.0036)
Σ_{23}	-0.0053	0.0000	(-0.0125,0.0019)
Σ_{24}	-0.0013	0.0000	(-0.0041,0.0014)
Σ_{25}	0.0004	0.0000	(-7e-04,0.0014)
Σ_{33}	0.2960	0.0176	(0.0362,0.5558)
Σ_{34}	-0.0031	0.0012	(-0.0716,0.0653)
Σ_{35}	-0.0060	0.0001	(-0.0276,0.0156)
Σ_{44}	0.0918	0.0007	(0.0412,0.1424)
Σ_{45}	0.0034	0.0000	(-0.008,0.0148)
Σ_{55}	0.0258	0.0000	(0.0203,0.0313)

Bayesian Model Performance

	ID	r_square	rmse	n_obs
1	BONNIE.1998	0.996	1.667	9
2	IVAN.1980	0.996	1.767	8
3	GEORGE.1950	0.993	1.768	8
4	MARIA.2011	0.964	1.768	8
5	BERYL.1982	0.971	1.889	7
6	FLORENCE.1960	0.927	1.890	7
7	LOIS.1966	0.990	1.890	7
8	ERIN.1989	0.991	1.890	7
9	GRETA.1970	0.893	2.041	6
10	HILDA.1964	0.995	2.236	5

R Squared and RMSE

Bayesian Model Performance



Actual Wind Speed vs. Predicted Wind Speed

Seasonal Difference Exploration

	Beta 0		Beta 1		Beta 2	
	Estimate	Pr(>|t|)	Estimate	Pr(>|t|)	Estimate	
(Intercept)	4.4810021	0.0000000	1.3431063	0.0000000	0.0413063	0.9
monthApril	0.0232609	0.8346449	0.0147943	0.6696787	0.0165579	0.9
monthMay	0.0259813	0.7827813	-0.0001180	0.9967888	0.0708822	0.6
monthJune	0.0275693	0.7650618	0.0053935	0.8509869	-0.0070875	0.9
monthJuly	0.0125400	0.8914489	0.0154032	0.5901741	-0.0090910	0.9
monthAugust	-0.0198034	0.8284715	0.0233206	0.4124181	-0.0522548	0.7
monthSeptember	-0.0070528	0.9384385	0.0261005	0.3585599	-0.0361073	0.8
monthOctober	0.0093435	0.9185853	0.0210829	0.4587183	-0.0286163	0.8
monthNovember	0.0145692	0.8748155	0.0246144	0.3925264	0.0239972	0.8
monthDecember	0.0057977	0.9526542	0.0088244	0.7715305	-0.0543131	0.7
year	-0.0003419	0.0717253	-0.0002252	0.0001471	0.0000365	0.9
natureET	0.0008449	0.9774141	0.0037334	0.6877086	-0.0702038	0.1
natureNR	0.0008122	0.9866387	-0.0146142	0.3331114	0.0058967	0.9
natureSS	0.0141564	0.4904257	-0.0033299	0.6021721	-0.0013517	0.9

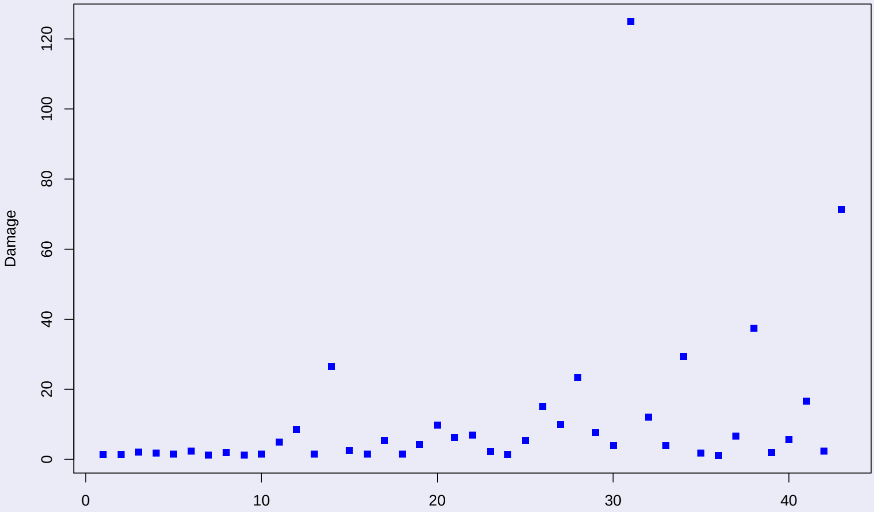
Seasonal Difference Exploration

	Beta 0		
	Estimate	Pr(>|t|)	Estim
(Intercept)	3.8365500	0.0000000	0.8942
seasonSummer	-0.0305003	0.2048954	0.0152
seasonAutumn	-0.0235346	0.3248438	0.0209
seasonWinter	-0.0186542	0.6535827	0.0034

	Beta 0		Bet
	Estimate	Pr(>|t|)	Estimate
(Intercept)	4.5142875	0.0000000	1.3448481

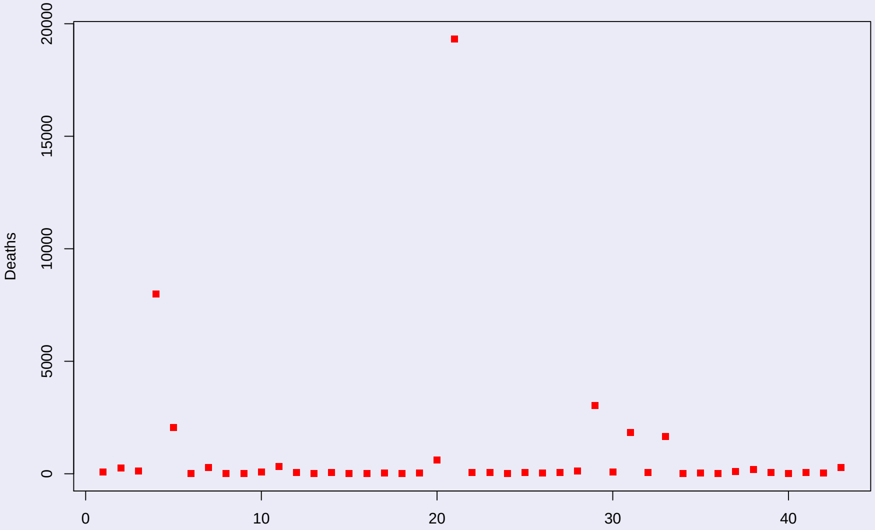
Predictions of Damage and Deaths

Basic plot of Damage and Deaths



Predictions of Damage and Deaths

Basic plot of Damage and Deaths



Coefficient Table

```
 [1] "|id              | intercept|     beta1|      beta2|      beta3|     beta4|"
 [2] "|:---------------|---------:|---------:|----------:|----------:|---------:|"
 [3] "|agnes.1972      |  3.950974| 0.9224097|  0.0059532| -0.3103372| 0.5453543|"
 [4] "|alex.2010       |  3.798737| 0.9370333|  0.0698849| -0.3937358| 0.5400187|"
 [5] "|alicia.1983     |  3.897408| 0.9036878| -0.0748341| -0.3994486| 0.5477718|"
 [6] "|allen.1980      |  3.687070| 0.9655304|  0.1306393| -0.5460144| 0.5466129|"
 [7] "|andrew.1992     |  3.676279| 0.9375384| -0.2843257| -0.5782973| 0.5370158|"
 [8] "|betsy.1965      |  3.808396| 0.9513766| -0.4500720| -0.3890718| 0.4244575|"
 [9] "|bob.1991        |  3.629466| 0.9232143|  0.0279527| -0.5751636| 0.4382048|"
[10] "|camille.1969    |  3.994355| 0.9355674|  0.0729188| -0.5734830| 0.6703910|"
```

Fitted results of beta models

Predict Damage

	term	estimate	std.error	statistic	p.value
1	(Intercept)	-2.179428e+02	63.786161983	-3.416772	6.336828e-04
2	intercept	5.044916e+00	0.872632934	5.781258	7.414400e-09
3	beta1	6.283543e+01	14.027126920	4.479565	7.479523e-06
4	beta2	-1.095810e+00	0.424325439	-2.582476	9.809426e-03
5	beta3	3.378223e+00	0.816050104	4.139725	3.477231e-05
6	nobs	4.921117e-02	0.008036275	6.123630	9.146733e-10
7	Season	7.497698e-02	0.012627373	5.937655	2.891284e-09
8	MonthJune	-3.416174e+00	0.762110791	-4.482516	7.376795e-06
9	MonthNovember	-1.902107e+00	0.789148853	-2.410327	1.593822e-02
10	MonthOctober	-1.290673e+00	0.298201079	-4.328198	1.503344e-05
11	MonthSeptember	-1.764116e+00	0.243173467	-7.254558	4.029764e-13
12	NatureNR	-4.317468e+00	1.126675716	-3.832042	1.270843e-04
13	NatureTS	-2.038481e+00	0.452900892	-4.500942	6.765302e-06
14	Maxspeed	5.044572e-02	0.006764325	7.457613	8.810369e-14
15	Meanspeed	-6.565465e-02	0.015403789	-4.262240	2.023877e-05
16	Percent.Poor	-3.819578e-02	0.005858677	-6.519522	7.053169e-11
17	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Predict Deaths

	term	estimate	std.error	statistic	p.value
1	(Intercept)	1.164978e+02	1.257956e+01	9.260883	2.027487e-20
2	intercept	1.167475e+01	2.564192e-01	45.529931	0.000000e+00
3	beta1	1.141195e+02	2.200144e+00	51.869091	0.000000e+00
4	beta2	5.528798e+00	1.226329e-01	45.084128	0.000000e+00
5	beta3	8.561691e+00	2.853214e-01	30.007184	7.908823e-198
6	beta4	-1.049211e+01	3.058279e-01	-34.307225	6.123346e-258
7	nobs	3.430943e-03	1.116605e-03	3.072657	2.121619e-03
8	Season	6.102077e-03	2.093747e-03	2.914429	3.563401e-03
9	MonthJuly	-1.183782e+00	1.448847e-01	-8.170505	3.071002e-16
10	MonthJune	-1.291597e+00	8.968191e-02	-14.401980	5.028215e-47
11	MonthNovember	-2.533192e+00	1.551869e-01	-16.323490	6.718278e-60
12	MonthOctober	-1.546676e+00	6.466487e-02	-23.918335	1.974205e-126
13	MonthSeptember	-2.751167e-01	4.588850e-02	-5.995331	2.030720e-09
14	NatureNR	2.348783e+00	1.290216e-01	18.204563	4.748263e-74
15	NatureTS	3.563406e+00	1.209962e-01	29.450564	1.238185e-190
16	Meanspeed	-3.676417e-02	3.143216e-03	-11.696356	1.330451e-31
17	Maxpressure	-2.686076e-01	9.670821e-03	-27.775052	8.684053e-170
18	Meanpressure	5.377225e-03	2.009523e-04	26.758717	9.775966e-158
19	Total.Pop	9.410461e-07	2.587520e-08	36.368659	1.332659e-289
20	Percent.Poor	3.599824e-02	8.024514e-04	44.860342	0.000000e+00
21	Percent.USA	-7.214139e-03	5.570867e-04	-12.949761	2.356879e-38