P8160 Report - Bayesian Modeling of Hurricane Trajectories

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

5/1/2023

Contents

1 Introduction

1.1 Background

A hurricane is a powerful tropical storm characterized by high winds, heavy rain, storm surges, and flooding. Hurricanes are also known as cyclones or typhoons, depending on the region where they occur.

Hurricanes typically form over warm ocean waters and can travel for thousands of miles, causing widespread destruction and disruption to communities in their path. They are categorized on a scale of 1 to 5 based on their wind speed and potential for damage, with Category 5 being the most severe.

Hurricanes can cause significant damage to infrastructure, homes, and businesses, and can also result in loss of life. As such, it is essential to take precautions and follow instructions from emergency management officials in the event of a hurricane.

1.2 Motivation

Climate researcher are interested in modeling hurricane trajectories for early warning and preparedness, resource allocation, planning and response, and scientific research. Overall, accurate modeling of hurricane trajectories is essential for mitigating the impact of hurricanes on communities, infrastructure, and the environment, as well as for advancing our scientific understanding of these powerful storms.

In this project, we are particularly interested in forecasting the wind speed of hurricanes.

1.3 Dataset

The dataset, `hurrican703.csv`, collected the track data of 702 hurricanes in the North Atlantic area from 1950 to 2013. For all the storms, their location (longitude & latitude) and maximum wind speed were recorded every 6 hours. The data includes the following variables:

- ID: ID of the hurricanes
- Season: In which year the hurricane occurred
- Month: In which month the hurricane occurred
- Nature: Nature of the hurricane ET: Extra Tropical

  DS: 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

1.4 Data pre-processing

First we need to pre-process the data. We only kept observations that occurred on 6 consecutive hour intervals. Through this step, we found that some hurricanes had the same ID but were actually different ones (eg. ALICE). Hurricanes that had fewer than 3 observations were excluded. For the purpose of seasonal comparison, we defined August, September, and October as hurricane-active season, and the rest as hurricane-inactive season. After data cleaning, there are 21691 observations across 704 unique hurricanes.

2 Bayesian Model

2.1 Bayesian hierarchical model for wind speed

3 MCMC

3.1 Posterior distribution of the parameters

3.2 MCMC algorithm implementation

3.3 Parameter convergence diagnostic

3.4 Posterior summaries of gamma

Posterior summaries and 95% credible intervals of γ is shown in table 1 and table 2. We are mainly interested in answering two questions. Are there seasonal differences in hurricane wind speeds? Is there evidence to support the claim that hurricane wind speeds have been increasing over the years? The convergence of parameter γ_{year} and γ_{active} is shown in figure 1 and figure 2. The two parameters converged after MCMC, and γ_{active} had a better performance on convergence than γ_{year}. Thus we could answer the two questions using the 95% credible intervals. Since the credible intervals for both parameters contain zero, the conclusion is that there is no seasonal or yearly difference in hurricane wind speeds.

3.5 Hurricane wind speed prediction

Given parameters $\boldsymbol{\gamma}$ and $\boldsymbol{\beta_i}$ which is the coefficients associated the ith hurricane, the predicted wind speed of the i-th hurricane at $t+6$ time stamp, denoted as $\hat{Y}_i(t+6)$ could be computed as:

$$\hat{Y}_i(t+6) = \boldsymbol{\beta_i}^T \mathbf{Z_i}(t) + \mathbf{X_i}\boldsymbol{\gamma}$$

where $\boldsymbol{\beta_i} = (\beta_{0,i}, \beta_{1,i}, ..., \beta_{4,i})$, and $\mathbf{Z_i}(t) = (1,\ Y_i(t),\ \Delta_{i,1}(t),\ \Delta_{i,2}(t),\ \Delta_{i,3}(t))$.

Based on the burn-in number 8000 as concluded in MCMC parameter convergence diagnostic section, we calculated the parameters to be used to make prediction, and predict the wind speeds of each hurricane at each time stamp, except for the first two time stamps. We computed the RMSE and R-squared of each hurricane, as partly shown in the table 3. It can be observed that different hurricanes perform variously on RMSE and R-squared. Generally the RMSE for selected hurricanes is around 3 to 8, and the R-squared is within 0.7 to 1.0. However, there is also extremely low R-squared(0.30 as shown) and high RMSE(9.64 as shown).

The prediction performance on several chosen example hurricanes is shown in figure 3 and figure 4. As shown in figure3, the predicted wind speed overall nicely fit the observed wind speed at each time index of those hurricanes, though performance is better on example hurricane 'ABLE.1951' than other examples. According to figure4, the prediction and observation data points fit well as a linear line with slope = 1. Digging deeper, the Bayesian model performs much better when the wind speed is stable, but worse when there exits some fluctuations of wind speed. More specifically, the Bayesian model's prediction has a latency in responding to the fluctuation of observed wind speed, and the changing of predicted speed is generally one unit time index behind observed speed.

The overall RMSE performance of difference hurricanes is shown in figure 5. And the overall R-squared performance of difference hurricanes is shown in figure 6. The RMSE follows an very right-skewed distribution, with most RMSEs distributed around 2 to 6, and distribution peak is around 3.5. The distribution of R-squared then follows a very left-skewed distribution with most of the values distributed from 0.875 to 1, and the peak is around 0.94. There also exists some extremely large RMSEs and small R-squareds, which may need further investigation.

Furthermore, we examined the distribution of RMSE across various properties of hurricanes is shown in figure 7 and figure 8. We observed that hurricanes belong to nature of "Not Rated" and "Tropical Storm" have an average higher RMSE than other hurricanes. Besides, hurricanes happening in summer from June to September, also have higher RMSE than other months.

4 Discussion

4.1 MCMC parameters convergence problem

4.2 Prediction latency of wind speed change

Group Contributions

Appendices

Figures and Tables

Table 1: Posterior Summaries of γ

parameter	Mean	Standard Deviation	Median
γ_{year}	-0.010	0.009	-0.011
γ_{active}	0.002	0.053	0.002
γ_{ET}	0.003	0.050	0.003
γ_{NR}	0.0002	0.049	-0.0003
γ_{SS}	0.005	0.050	0.005
γ_{TS}	0.01	0.057	0.01

Table 2: 95% Credible Intervals of γ

parameter	2.5% Quantile	97.5% Quantile
γ_{year}	-0.029	0.008
γ_{active}	-0.102	0.104
γ_{ET}	-0.096	0.098
γ_{NR}	-0.098	0.095
γ_{SS}	-0.096	0.104
γ_{TS}	-0.105	0.119

Table 3: 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

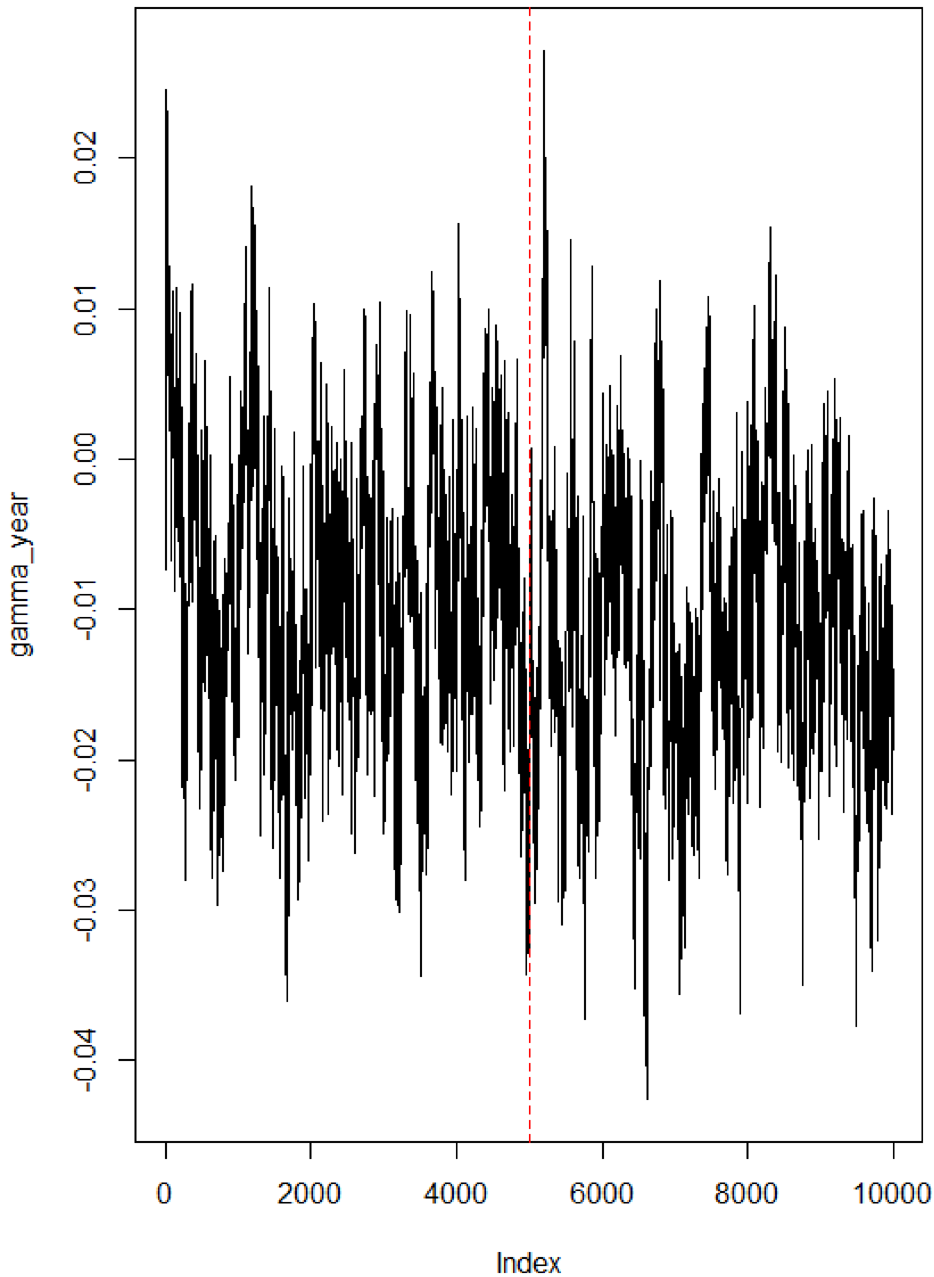

Figure 1: Time series plot of γ_{year}

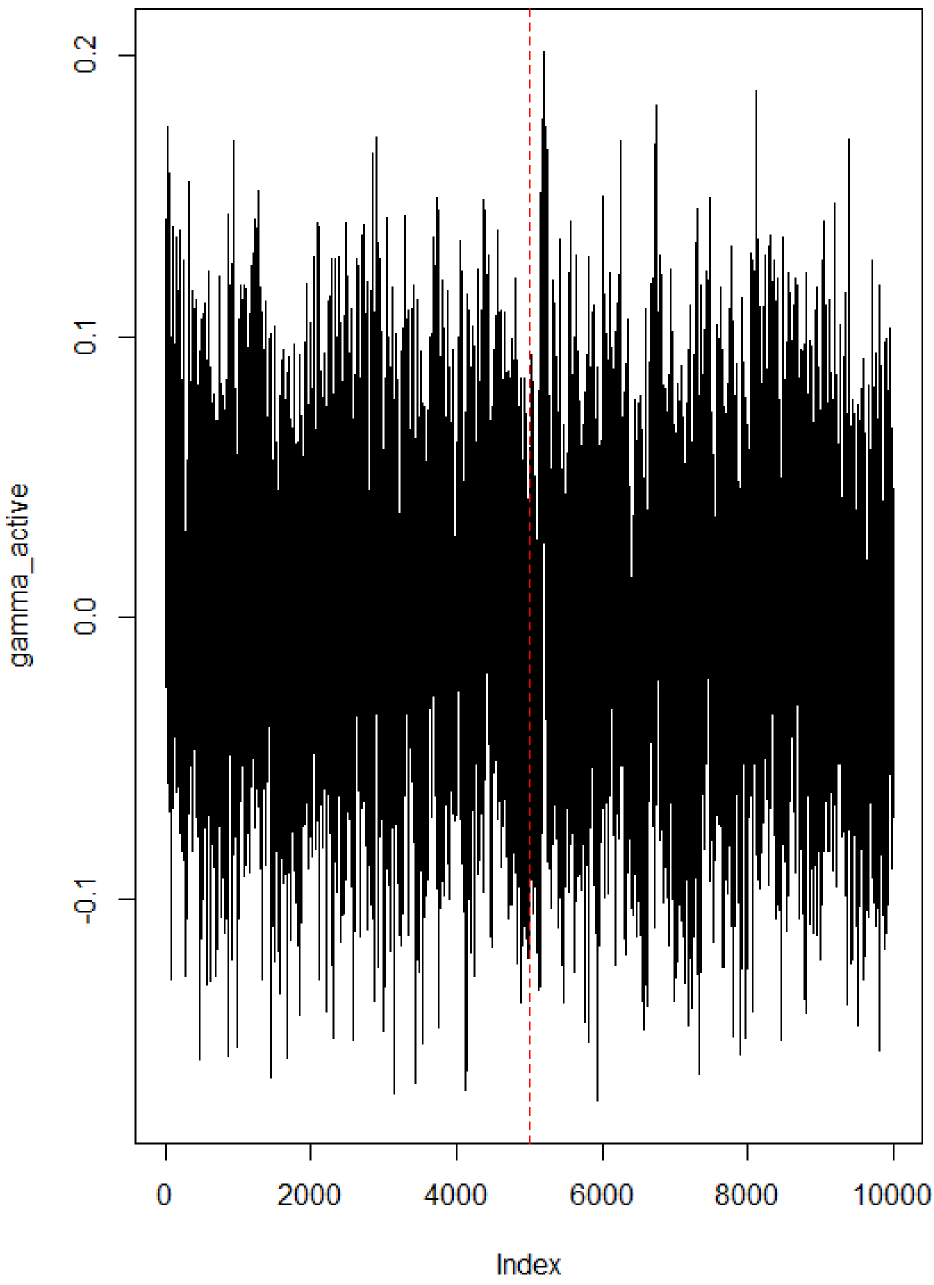

Figure 2: Time series plot of γ_{active}

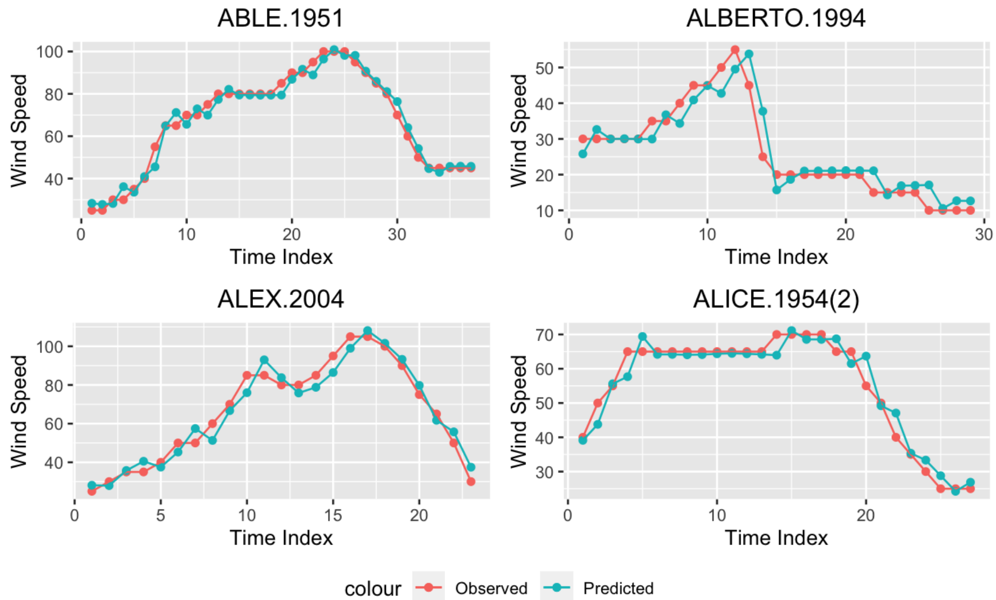

Figure 3: Time series prediction plot

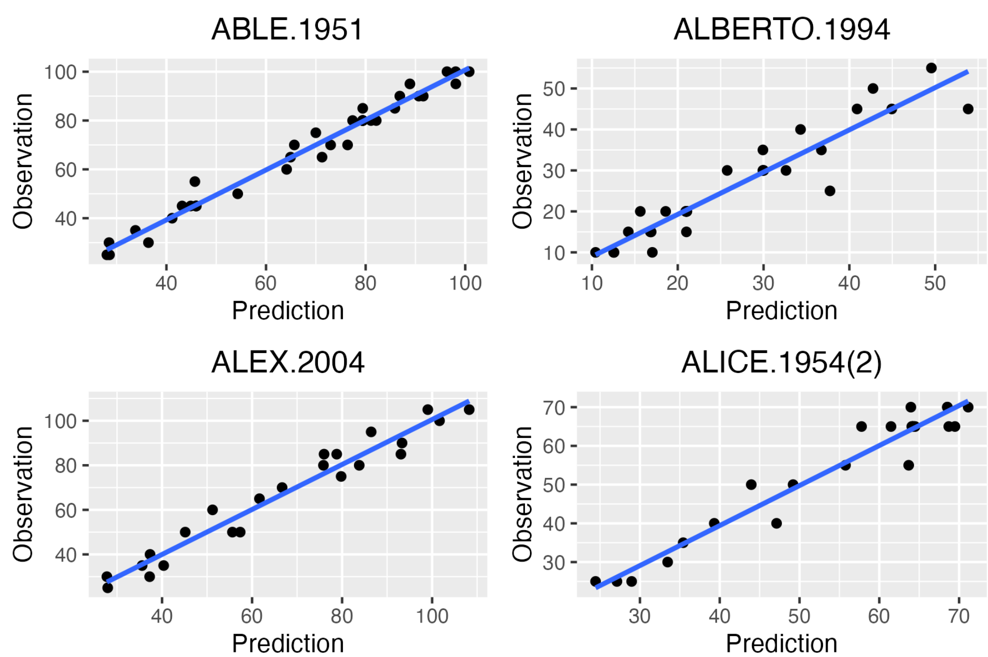

Figure 4: Prediction vs. observation

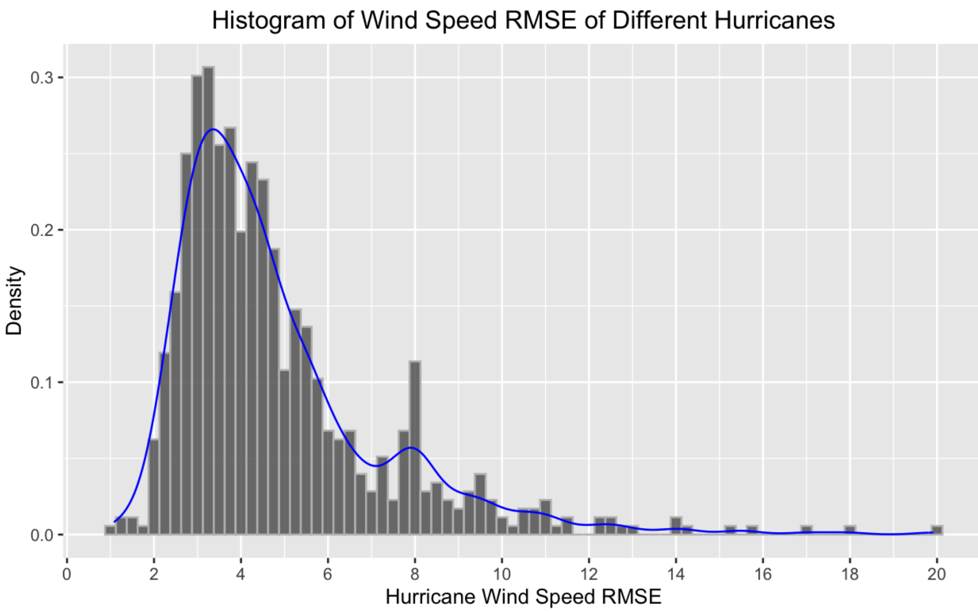

Figure 5: RMSE distribution

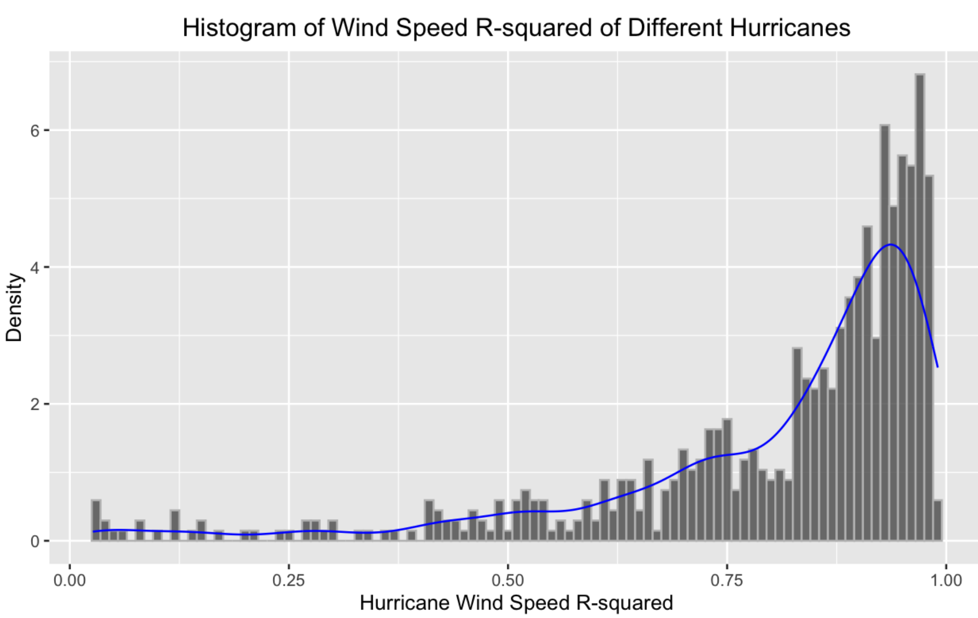

Figure 6: R^2 distribution

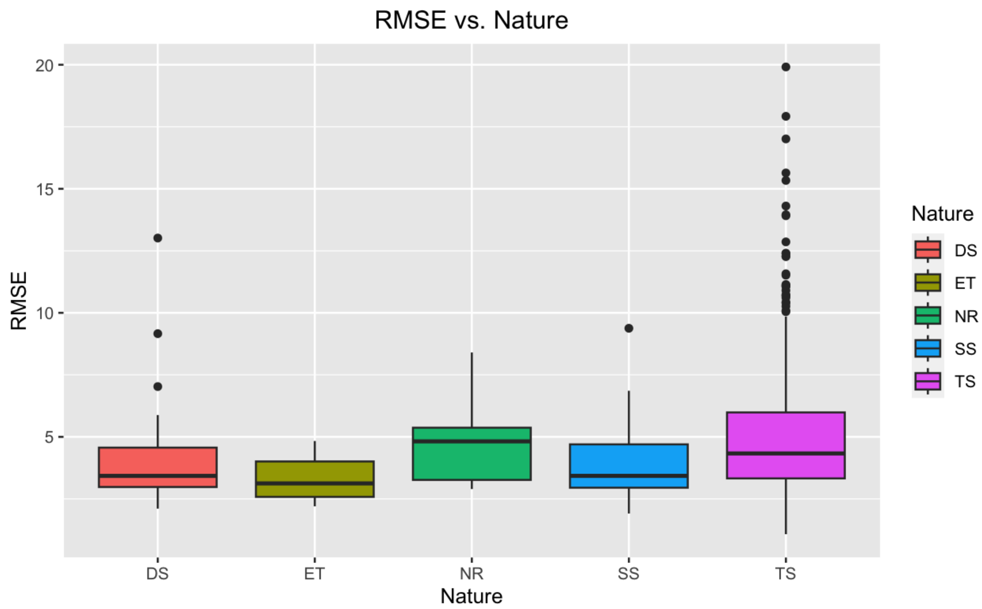

Figure 7: RMSE under different natures

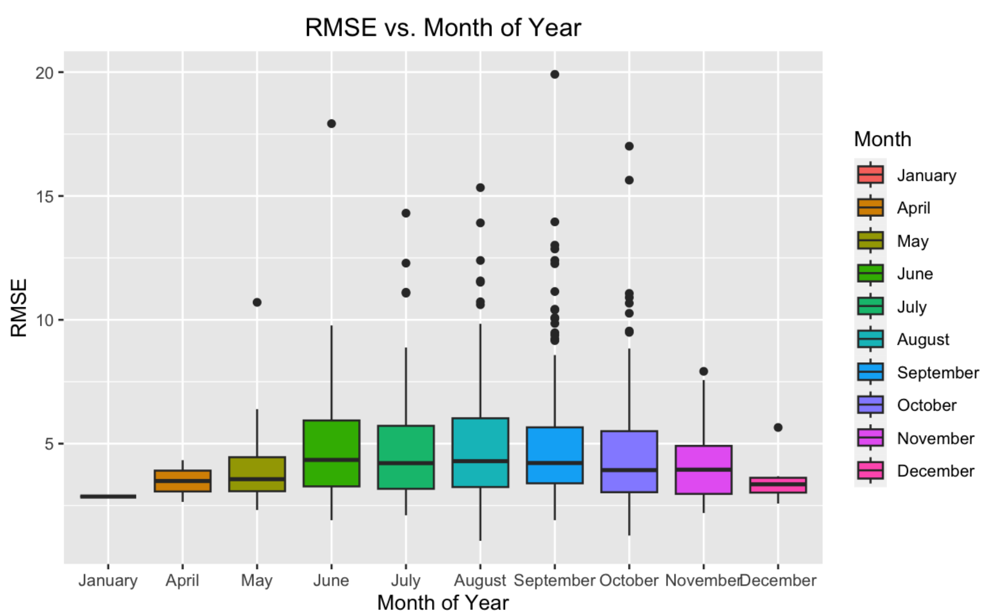

Figure 8: RMSE under different months