Data Science Research – Hurricane Rapid Intensification
Project Proposal

Cheuk Hang Ng (a1821087)

University of Adelaide, a1821087@student.adelaide.edu.au

1 INTRODUCTION

This project aims to explore and investigate a dataset which consists of Atlantic storms from 1979 to 2015 with modern data science methods. The project is composed by two parts. The first part will focus on performing descriptive data analysis on the dataset, such as providing a statistical summarization of the dataset, and producing plots of storm speed over time, the moving track of hurricanes and the frequency of hurricanes in different months and years, and will attempt to predict hurricane strength and the presence or absence of hurricanes by the existing data as predictors. In the second part of this project, the rapid intensification of hurricanes will be studied, and machine learning algorithms may be used to predict and model the rapid intensification process.

2 BACKGROUND

Hurricane behaviors and processes of formation have always been popular topics in the scope of atmospheric science. Some had done prediction on the frequency of Atlantic hurricane with decadal data [1], some studied structural change over time and modelled the rapid intensification of single hurricane [2], while some looked at the relationship between rapid intensification and environmental features [3]. Yet, it seems that there is a lack of predicting and modelling the rapid intensification and decadal hurricane data with modern data science methods.

In this project, the dataset given consists of 10 unorganized files named as <feature abbreviation + “store”>, with each file contains numbers of rows and columns that represent either the data at a certain time frame or hurricanes. The files are said to be unorganized by means of all columns are without column names, and the data contained in each file could only be interpreted by the file names. To be precise, it was discovered that the 10 files contain data of “day”, “hour”, “month”, “year”, “latitude”, “longitude”, “pressure”, “wind speed”, “wind shear strength” and “potential wind strength” for some number of hurricanes in a certain time span. However, the time span of the data could only be revealed by investigating the dimensions in “day”, “hour”, “month” and year.

3 METHODS

The project will most be done on R, a powerful tool for data science and statistical modelling. It starts by cleaning the dataset provided, mainly using the data handling libraries in R such as Tidyverse. This steps include but not limited to: Exploring the files, renaming all the columns in each file, zero values and missing data handling, and producing summarization data frame which consists of statistical summarization of each file for the hurricanes. For example, mean, maximum and minimum will be used to reduce the dimensions of the numerical data. The next is exploratory data analysis. In this step, the built-in plot function from R will be used to plot histograms

and scatter plots to visualize the distributions of varied data, such as the time data, pressure and wind speed data. The movement track of a few hurricanes will also be plotted using the latitude and longitude data. In addition, the relations between different features will be visualized by bivariate plots.

If the above procedures are carried out successfully, these features will be used to attempt to predict storm strength by some statistical models, for instance regression model is one of the choices. The presence or absence of hurricanes will also be predicted if the storm strength prediction results are successful and reasonable, with the use of some other features. Even if the above prediction is not successful, we will look for important features by analyzing the modelling and prediction results, and extract these features for fine-tuning the models. Either the successful models or the important features would be concluded as the results of the first part and used as the input for the second part of this project.

By combining the above results, together with the in-depth understanding of rapid intensification of hurricanes, the occurrence of rapid intensification will be predicted with machine learning model and if the results are good enough, visualization of fully predicted hurricane from formation to rapid intensification and weakening, and the corresponding movement track will also be presented as the research result. Similar to the first part, if the results are not satisfying, find-tuning will be carried out and important features will be extracted for analysis.

REFERENCES

[1] Doug M. Smith, Rosie Eade, Nick J. Dunstone, David Fereday, James M. Murphy, Holger Pohlmann and Adam A. Scaife. 2010. Skilful multi-year predictions of Atlantic hurricane frequency. *Nature Geoscience* 3, 12 (Nov. 2010), 846–849, DOI: https://doi.org/10.1038/ngeo1004

[2] Chen Hua, Zhang Da-Lin, Carton James, Atlas Robert. 2011. On the Rapid Intensification of Hurricane Wilma (2005). Part I: Model Prediction and Structural Changes. *Weather and Forecasting* 26, 6, (Dec. 2011), 885–901, DOI: https://doi.org/10.1175/WAF-D-11-00001.1

[3] Wadler, J. B., Zhang, J. A., Rogers, R. F., Jaimes, B., and Shay, L. K. 2021. The Rapid Intensification of Hurricane Michael (2018): Storm Structure and the Relationship to Environmental and Air–Sea Interactions. *Monthly Weather Review* 149, 1 (Jan. 2021), 245-267, DOI: https://doi.org/10.1175/MWR-D-20-0145.1