Hurricane Prediction

Kolton Baldwin, Sam Cochran, Thomas Fackrell, and Chelsey Noorda

March 2021

1 Research Question and Overview of the Data

One to seven hurricanes make landfall every year in the United States causing millions of dollars worth of damage. The ability to accurately predict a hurricane's route could save lives, money, and property by allowing states and cities to take more targeted precautionary actions. Using data from the National Oceanic and Atmospheric Administration (NOAA), we hope to accurately predict when and where (if at all) a hurricane will make landfall. The more accurate the predicted trajectories, the more time residents will have to prepare for and evacuate before a storm makes landfall. The implications of early and accurate warning of hurricane trajectories are profound. More narrowly applied evacuation orders could decongest state evacuation routes, allowing those who are actually in the storm path a quick and safe escape while avoiding unnecessary evacuations. Similarly, airports and ports could better coordinate departures and arrivals, allowing visiting tourists and business people to leave the impacted area. Advanced warning could allow businesses and homes to be better protected. We believe that property damage and loss of live can be reduced with accurate trajectory projections.

2 Ethical Implications

Because our data is composed of scientific readings of various hurricanes, relatively few ethical issues can arise from our project. The data we use does not directly impact humans, and comes from a publicly available dataset created by NOAA. While we hope that our work can contribute to bettering current forecasting practices, we note that our work should not yet be construed as complete. Those facing potential hurricane landfall should rely on the the more robust work done by the NOAA and the National Weather service.

3 Data Cleaning

We have data for hurricanes in both the Atlantic and Pacific oceans. Because the underlying mechanisms that determine the behavior of hurricanes are different for these two oceans, we treat the oceans separately throughout the paper.

The raw data has distinct row types for data and headers. We first parsed the data to identify row types. We next discarded data that was too old to have good measurements. Specifically, the data for Wind Radii Max Extent has only been available since 2003, so we dropped all hurricane data from before these measurements existed for both the Atlantic and Pacific Datasets.

Next, we one-hot encoded the categorical Status data. This data reflects the intensity of the storm at each time stamp (for example, the status might be tropical storm at one point in time, and later if the storm intensifies it might be categorized as a full-on hurricane at another point in time).

Finally, we drop the record identifier column. This column contains information about key changes in the behavior of each storm (for example, when it makes landfall, when it reaches peak intensity, when it achieves minimum pressure, etc.). We choose not to train our models on this column because it is very sporadically recorded (most entries in this column are empty) and the information it contains can be inferred in terms of the other data columns. Furthermore, the data recorded herein is somewhat backward-looking. Because we hope to supply robust methods which could be applied to future storms, it makes sense to only use the data that could be theoretically available in real time. Measurements such as peak intensity and minimum pressure require a comprehensive knowledge of the full hurricane path, which would render our attempts at prediction moot.

Our end result is a list of time series for each ocean, where each element contains a time series dataframe for one hurricane. We have 271 time series for Atlantic hurricanes and 323 time series for Pacific hurricanes.

4 Robustness

While our code for processing the data should be quite robust, as long as the data fed in has the same format as the NOAA datasets for which it was designed. Two data cleaning considerations are paramount to functionality with our work.

First, we would need to identify when the Wind Radii Max Extent data becomes available for the new data, find the unique identifier for the first hurricane in the new dataset which contains all the requisite data, and then drop all hurricanes which precede it.

Second, we would need to identify and fix any errors in the data. Our Pacific Dataset required this type of correction. There were some miscoded entries in the Status column. For example, TS (tropical storm) was input as ST a small number of times, which is meaningless given the data descriptions. We had to identify errors like this by inspection, replace the faulty instances with NaN values, and then impute the now missing data from surrounding values. New data, it may be assumed, would need similar inspection.

5 Data Visualization and Basic Analysis

Given the preliminary data before experimentation on latitude and longitude transformed to polar coordinates, we were able to plot the trajectories of the five longest-lasting hurricanes found in our data. With further data exploration and model development, we will have more visualizations to show.

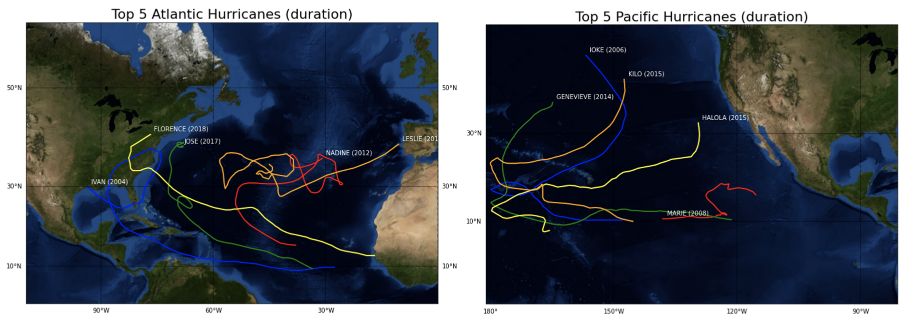

Figure 1: The trajectories of the longest 5 hurricanes in both the Pacific and Atlantic regions

6 Learning Algorithms and In-depth Analysis

Since our data is a time series and a stochastic process, we began by running our data through an ARMA model. Currently our ARMA model can only handle one-dimensional time series, so we predicted the latitude and longitude coordinates of one of the hurricanes for the following five days. These initial results can be seen in the following graphs.

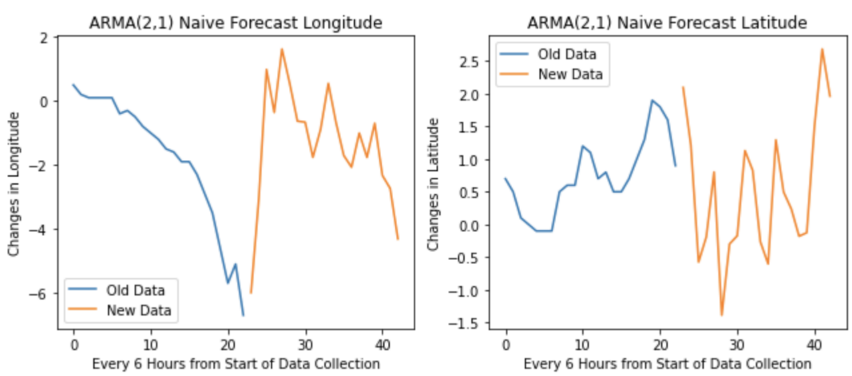

Figure 2: The initial projected latitude and longitude coordinates of Hurricane Alex

7 Results

Psh. It's still only (late) March.

8 Conclusion

Forthcoming.

9 References

Research and more sources forthcoming

A Sparse Recurrent Neural Network for Trajectory Prediction of Atlantic Hurricanes

SparkDevAI

Hurricane RNN