Hurricane Trajectory Prediction

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

April 15, 2021

Abstract

The unpredictable paths traced by hurricanes have challenged forecasters for many years. Because dynamic storm systems were thought to be too complex to accurately model, early regression models focused on storm momentum without incorporating regional atmospheric conditions. Later models have attempted to account for environmental influences on the storm path. The challenge of identifying a universal model for storm path prediction is multifaceted. Each region has unique oceanic currents and differences in the atmospheric jet stream. Additionally, the effects of seasonality and even differing concentrations of pollutants can affect storm development. It is difficult to create one model that could predict all future hurricanes. The application of various models is presented, including a recurrent neural network and various vector-valued time series methods. Each method proves difficult to generalize and produced inconsistent results. Currently available implementations of vector-valued time series models fall short of producing a universal model which may be applied to multiple storms; retraining on each individual storm proved necessary. Current vector-valued time series implementations handle high-dimensional data poorly. Simplifying the model to predict only on location (rather than considering atmospheric conditions, too) proved more accurate. While the RNN could easily incorporate the breadth of reported data, the predictions produced by our regional models are inaccurate.

1 Problem Statement and Motivation

The U.S. Department of Commerce reports that tropical cyclones and hurricanes have been responsible for nearly a trillion dollars in damage in the United States since 1980. On average these dangerous storms each cause more than twenty-one billion dollars of damage [**1**]. The ability to accurately predict a hurricane's route saves lives, money, and property by allowing states and cities to take 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 trajectory, 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, unnecessary evacuations could be avoided, airports and seaports could better coordinate departures and arrivals, and property could be better protected against the impending landfall.

The NOAA currently employs models dating back to the 1980s to predict a hurricane's path of motion. Iterations of these original models still underscore storm trajectory predictions today. The CLIPER (CLImatology and PERsistence), serves as a baseline against which NOAA measures newer models. This "no-skill" forecast method relies on a non-linear regression in which errors accumulate by a factor of e about every two and half days [**2**]. While the CLIPER model tracks a storm's momentum alone, newer models, such as the BAM (Beta and Advection Model), consider a more holistic description of storm structure and local atmospheric conditions. The BAM model relies on rapidly-estimated solutions to systems of PDEs and results in greatly increased accuracy over CLIPER modelling alone [**3**].

2 Data

The NOAA provides historical hurricane data for hurricanes in both the Atlantic and Pacific oceans [**4**]. Because the underlying mechanisms (currents, atmospheric conditions, etc.) that determine the behavior of hurricanes are different for these two oceans, we treat the oceans separately throughout the paper.

We note that the NOAA's measurements have improved with time. We discarded data that was too old to have good measurements. Specifically, although we have some data going back as far as 1851, the NOAA has only recorded wind radii max extent (the distance between the center of the storm and its strongest winds) since 2003, so we dropped all hurricane data previous to 2003 from both the Atlantic and Pacific datasets. The hurricane data we use includes information recorded every six hours on the location, maximum winds, central pressure, and size of each hurricane.

We chose to 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.

Finally, we manually corrected a few errors in the data. 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.

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

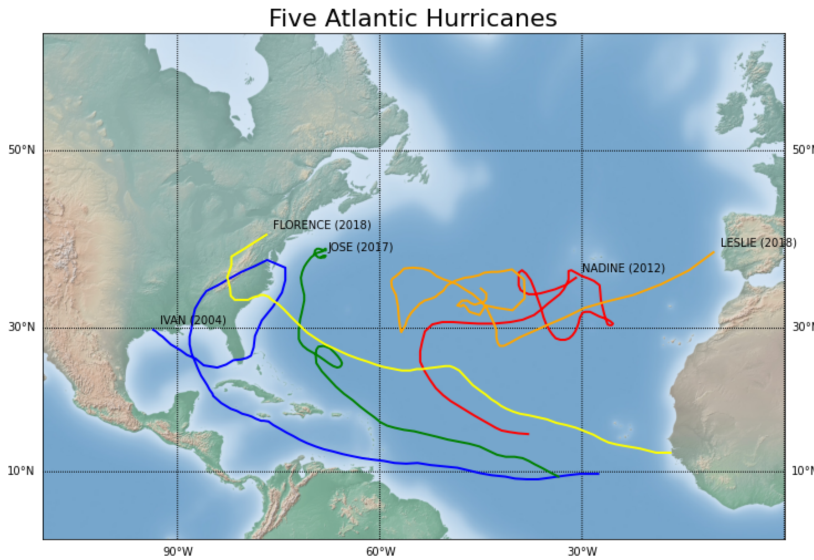

Figure 1: Example trajectories of five Atlantic hurricanes.

3 Methods

3.1 LSTM-based RNN

Recurrent neural networks are often used in natural language processing, where the input is a string of text and the output is the next predicted word in the sequence. This pattern, a series as input and a single value as output, varies from our implementation in that our output is multidimensional (latitude and longitude). Nevertheless, an RNN was appropriate here because of the time series structure of the data. With the added flexibility afforded by a neural network structure, we could still incorporate multidimensional data and produce multidimensional output. For many tests, we also included the features of wind speed and minimum pressure in the data, increasing the output dimensionality to four. Ultimately, we chose the model that used only latitude and longitude because including more features in the training data did not improve predictive performance.

Although we considered many system architectures in the deep learning

API Keras [**5**]-[**7**], the best variation consisted of five LSTM (long short-term memory) layers, each followed by a dropout layer with probability of 0.2, followed by a fully-connected layer and ReLU activation, which produced the two-dimensional output prediction.

For our final version, using Keras, we trained two RNNs with this architecture, one for the Pacific hurricanes and one for the Atlantic hurricanes. We followed a supervised training regimen to train the models: the sources were $n = 4$ time steps of hurricane data (latitude and longitude only) and the targets were the next time step's latitude and longitude. We trained other models with the number time steps in the input ranging from two to six, but four gave the best predictions.

3.2 Vector Autoregression (VAR)

Vector autoregression (VAR) is a state space method which can be used to predict future values for time series Y_t with multiple parameters. This multivariate process estimates relationships in the time series with lagged values in the following way:

$$Y_t = v + A_1 Y_{t-1} + ... + A_p Y_{t-p} + u_t$$

$$\text{where } u_t \sim \text{Normal}(0, \Sigma)$$

We used the Statsmodels implementation of vector autoregression (`statsmodels.tsa.vector_ar.var_model.VAR`). This model struggled with the high-dimensional data that included information about the size, pressure, and wind speeds of the storms, so we decided to train using only the latitude and longitude data in six-hour time steps reported by NOAA. Although we ignored a lot of information by restricting ourselves to only positional data, the quality of our predictions actually improved dramatically. This increase in accuracy indicates that currently-available state space models struggle when regression parameters (atmospheric data) are included.

After fitting the model, a forecast can be generated. While the process is very similar to a univariate autoregression model, currently available VAR implementations are not as robust. VAR models are relatively new and are difficult to apply correctly due to poor documentation. While using the Statsmodels implementation proved more difficult than we had anticipated, the VAR model produces decently accurate predictions.

3.3 Vector ARMA (VARMA)

We also attempted to predict hurricane trajectory with a vector-valued autoregressive moving average (VARMA) model. We used the Statsmodels implementation (`statsmodels.tsa.statespace.varmax.VARMAX`). Like the VAR model, this method uses lagged values to predict, but it also predicts on randomized error terms, in an attempt to account for noise in time series data. This leads to the following equation for our time series:

$$Y_t = v + A_1 Y_{t-1} + ... + A_p Y_{t-p} + B_1 \epsilon_{t-1} + ... + B_q \epsilon_{t-q}$$

$$\text{where each } \epsilon_i \sim \text{Normal}(0, \Sigma)$$

As we can see above, p and q correspond to the number of time steps considered in the autoregression and in the moving average terms, respectively. While we experimented with various values for p and q, in general we got our best predictions using values of either 2 or 3 for both. We used $p = 3$ and $q = 2$ for our final models.

To examine the specifics of the models we used, please reference our GitHub repository.

4 Results

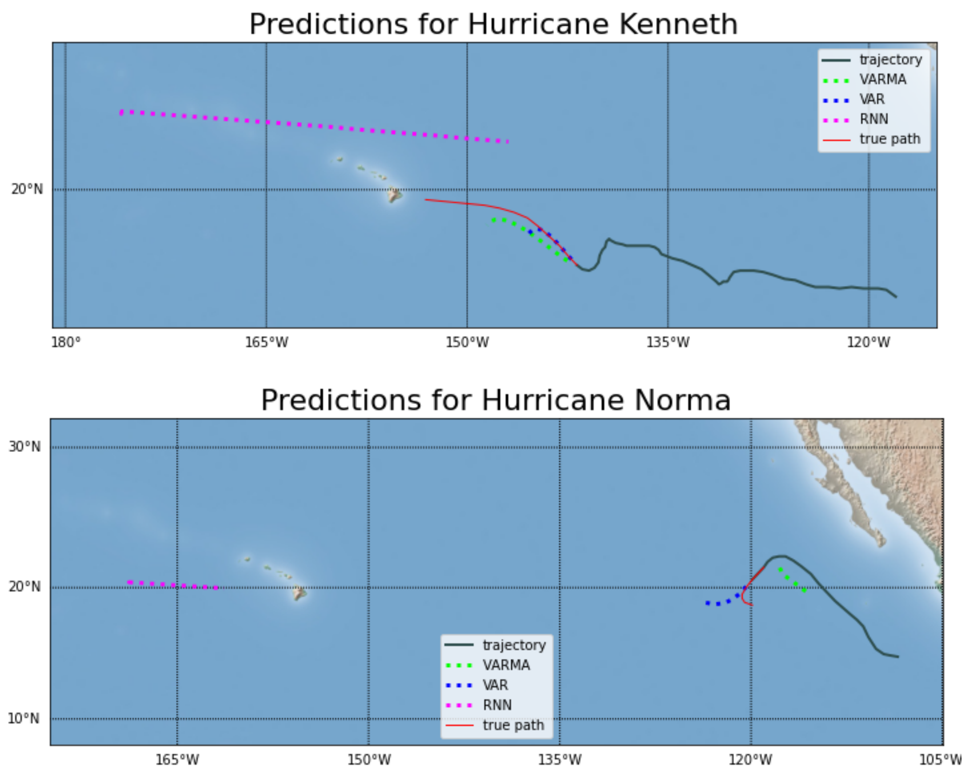

Figure 2: Comparison of predicted trajectories for two Pacific hurricanes, Hurricane Kenneth (2005) and Hurricane Norma (1981). Predicted trajectories are shown for all of our models (VARMA in green, VAR in blue, and RNN in purple), alongside the known past trajectory (grey) that was fed into the models and the true future trajectory (red) that we attempt to predict.

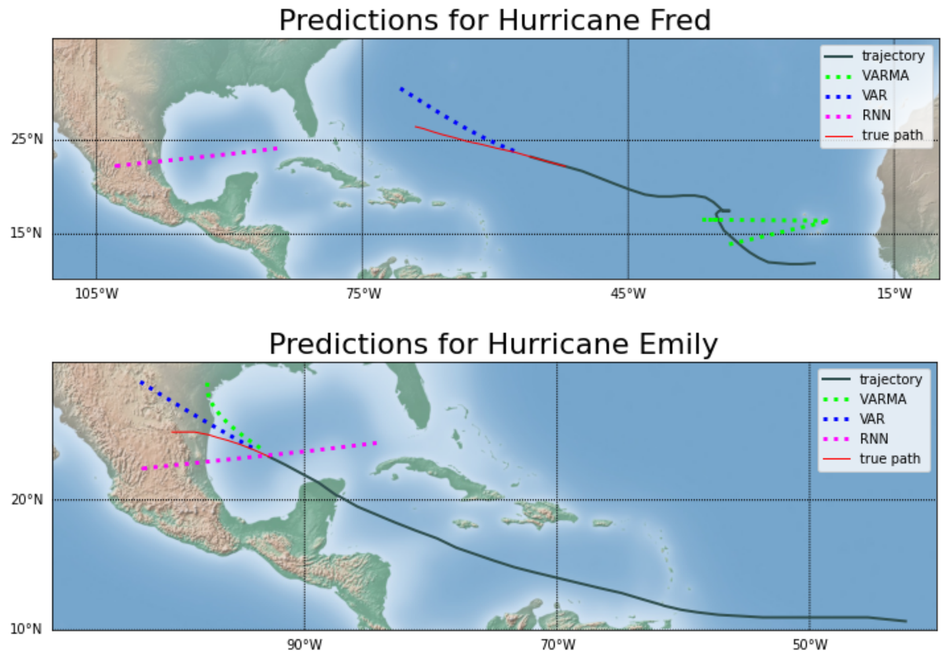

Figure 3: Comparison of predicted trajectories for two Atlantic hurricanes, Hurricane Fred (2015) and Hurricane Emily (2005). Predicted trajectories are shown for all of our models (VARMA in green, VAR in blue, and RNN in purple), alongside the known past trajectory (grey) that was fed into the models and the true future trajectory (red) that we attempt to predict.

4.1 LSTM-based RNN

The results show very little mastery of the behavior of hurricanes. The LSTM-based RNN shows very little resemblance to the true path. While the prediction (purple) stays within the general region of the hurricane path in some instances, no actionable insight is gained through the use of this model in predicting trajectories. Furthermore, the RNN's predictions were discontinuous—the prediction does not connect to the end of the training data.

When trained on four or more consecutive time steps, the RNN consistently predicted linear trajectories. When fewer time steps were used, the model produced nearly identical paths for all hurricanes—seemingly ignoring initial paths.

4.2 Vector Autoregression (VAR)

We were able to train the model on a $T \times P$ matrix where T represents the number of entries in the time series and P represents the number of parameters. It is evident that the model did not predict the exact trajectory. It started fairly close to the actual path, but as time increased it veered farther and farther off from the true trajectory.

That VAR was the most accurate of the three methods is interesting because we had to train a separate model on each individual hurricane. The RNN was trained on the data for all of the hurricanes in a given ocean, and we had anticipated that incorporating more information would produce better results. The VAR results tend to hold up quite well, although for more complicated hurricane trajectories, VAR still struggles to produce meaningful answers. Still, for an average hurricane, VAR does a surprisingly good job of predicting a path that is relatively close to the true trajectory.

4.3 Vector ARMA (VARMA)

Unfortunately, the Statsmodels implementation we were able to find is very new, and it tends to give unreliable results. Similar to our VAR results, we were unable to produce a generalized VARMA model. Each separate hurricane was treated as a single time series requiring its own state space model.

For many hurricanes, including Kenneth (Figure 2) and Emily (Figure 3), we are able to get reasonable results, but the moving average terms in the VARMA model seem to make it more unstable than vector autoregression alone. Instead of giving it more predictive power, the more complicated VARMA model tends to curve or veer off from the true trajectory, as opposed to the VAR model that does a much better job of predicting the true trajectories.

We also note that the VARMA model performed especially poorly on more complicated hurricane trajectories. The model often produced predictions which were far away from the true trajectory in the wrong direction. Furthermore, this model fails to compile for many hurricanes, especially when experimenting with larger values of p and q. This inconsistency prevented us from identifying a generalizable VARMA model which could be applied to several hurricanes in our dataset. Overall, our VARMA predictions tend to be wildly inaccurate for a significant proportion of the hurricanes and the model performed far worse than VAR.

5 Analysis

While training the state space models, we noticed that fewer features yielded better predictions. To improve our predictions, we removed all other features in the state space models except for latitude and longitude. Too many parameters made training these models difficult, and there simply wasn't enough data to allow us to effectively train on the original higher dimensional data, especially since we had to train a separate state space model for each hurricane. Perhaps if we had more data points for each hurricane, we could more effectively train the state space models to understand the underlying functions that determine each hurricane's path.

Our RNN models did not face this problem in quite the same way. While we could train our neural networks with all of the recorded data (not just position), the models trained on all the data performed very poorly. The models trained only on position consistently produced better results. Although the RNN models considered the data for an entire ocean collectively, the NOAA dataset proved insufficient to accurately train each RNN. There is simply too much variance between hurricanes for accurate training to occur without more frequent time steps and a larger corpus of storms. Due to constraints in publicly available data, the state space models that were trained separately on each individual hurricane consistently outperformed our RNN.

For some hurricanes, both state space models performed comparably. For example, both VAR and VARMA produced relatively good predictions for hurricanes Kenneth (Figure 2) and Emily (Figure 3). Neither is totally accurate, but both seem to follow the right general direction or path, with error accumulating at each time step. We also included hurricanes Norma (Figure 2) and Fred (Figure 3) to demonstrate the weaknesses of our models. The VAR model still performed fairly well, but the VARMA model produced terribly inaccurate predictions, despite Statsmodels confirming that it had converged.

Perhaps the moving averages used in the VARMA models decreased our accuracy because VARMA is not able to learn the right parameters to properly use the information gained by the moving averages. It is likely that the VARMA models also have other underlying issues in their implementation, as evidenced by the prolific warnings produced by the Statsmodels package (every time the `VARMAX` class is used, Statsmodels prints a warning notifying the user that its prediction methods are highly unstable). Even though current VARMA implementations do not model storm trajectories well, we believe a state space model could be made to accurately model the problem.

The RNN seems to predict nearly the same trajectory each time. Likely, this is because we trained each RNN on the respective ocean's entire dataset, wherein the variance from one hurricane to another is large enough to prevent accurate predictions, given the relatively small number of hurricanes in each dataset. This seems to indicate that hurricanes differ enough to make it extremely difficult for an overarching model to forecast on arbitrary storms. Our individualized state space VAR models do a much better job on each hurricane than our generalized RNNs. The VAR methods are the most reliable models at present.

6 Ethical Implications

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

7 Conclusion

Due to the complex relationships between the underlying variables present in hurricane trajectories, our models were unable to consistently generate viable predictions. Neural networks struggled to predict the future path of hurricanes, especially in the case of data limited to four daily measurements. State space models trained on position data did much better, but current vector-valued implementations tended to be unstable and failed to make good use of extra regression parameters like wind speed, hurricane size, and minimum pressure.

Future work could depart from our methods by training multiple generalized models, each using data for all available hurricanes from more specific regions of each ocean. These regional models could be combined to make weighted predictions accounting for regional differences.

Additional weather data not included in the NOAA hurricane database could have also helped in predicting hurricane trajectory, coupled with more complex models. Future work could use other meteorological data not local to the immediate radius of the hurricane in question, augmenting the data used in this study.

References

[1] Office for Coastal Management, “Hurricane Costs,” Noaa.gov, 2017. https://coast.noaa.gov/states/fast-facts/hurricane-costs.html.

[2] S. D. Aberson, “Five-day tropical cyclone track forecasts in the North Atlantic basin,” *Weather and Forecasting*, vol. 13, no. 4, pp.1005-1015, Dec. 1998, doi: 10.1175/1520-0434(1998)013¡1005:FDTCTF¿2.0.CO;2

[3] D. G. Marks, “The Beta and Advection Model for Hurricane Track Forecasting,” National Oceanic and Atmospheric Administration, Washington, D.C., USA, NOAA Technical Memorandum NWS NMC 70, 1992.

[4] C. W. Landsea and J. L. Franklin, ”Atlantic Hurricane Database Uncertainty and Presentation of a New Database Format.” (2013). Distributed by National Oceanic and Atmospheric Administration. Mon. Wea. Rev., 141, 3576-3592.

[5] M. Kordmahalleh, M. Sefidmazgi, A. Homaifar. “A Sparse Recurrent Neural Network for Trajectory Prediction of Atlantic Hurricanes” (2016)

[6] S. Alemany. ”Predicting Hurricane Trajectories Using an RNN” July, 2018. [Online] Available: https://github.com/sheilaalemany/hurricane-rnn [Accessed Mar. 1, 2021]

[7] S. Alemany, D. Castaneda, A. Perez. “SparkDev AI Hurricane Trajectory Prediction” December, 2017. [Online] Available: https://github.com/sheilaalemany/hurricane-rnn [Accessed Mar. 1, 2021]