Atlantic Tropical Cyclone Intensification Regression Model (ATCIRM)

Bruno Platero Huarcaya, Lexi Comstock, Daniel Halperin, Timothy Smith

Embry Riddle Aeronautical University

INTRODUCTION & MOTIVATION

Even though operational track forecasts of Tropical Cyclones (TC) have improved a lot in recent years and are highly accurate, intensity forecasts are still relatively poor, especially for rapidly intensifying storms. An increase in intensity forecast accuracy would help give more credibility to TC forecasts as well as tremendously help authorities in their decision making to prevent loss of life and property. The purpose of this project is to develop a statistical linear regression model and determine if it can better predict TC intensification over water. This project was motivated by the poor performance of the intensity forecast for Hurricane Michael in 2018.

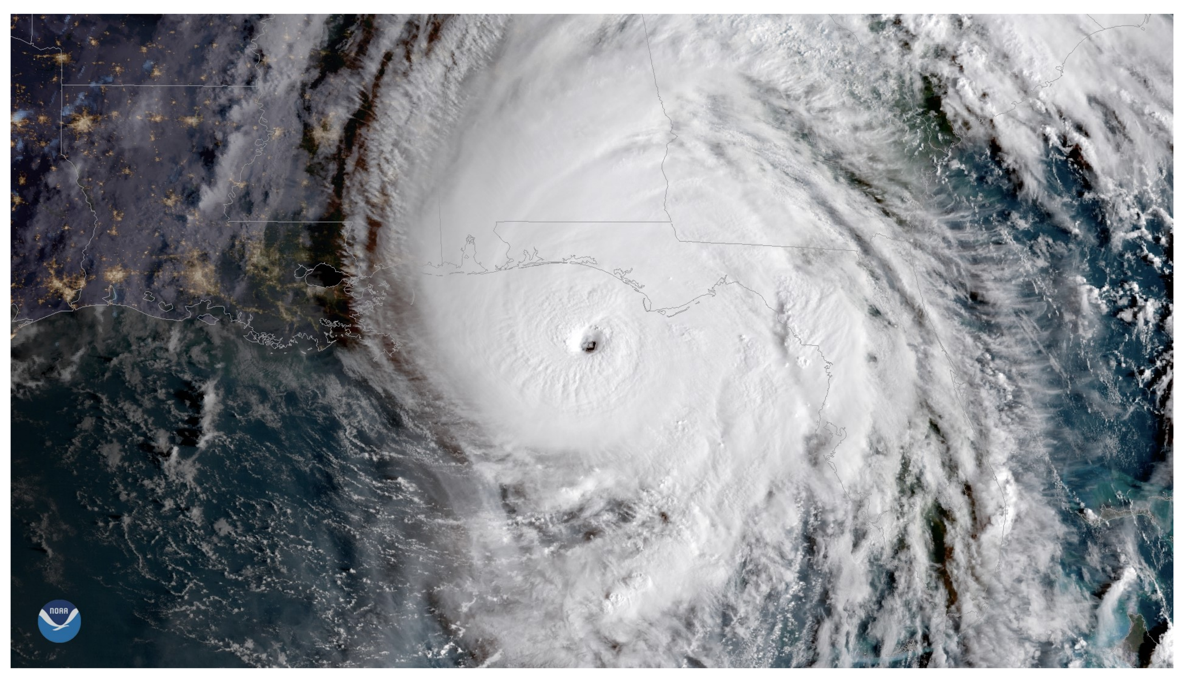

Figure 1. Hurricane Michael. Retrieved from NOAA Environmental Visualization Laboratory (2018).

DATA

Two separate data sets were used to develop the regression model. The predictors were obtained from the Automated Tropical Cyclone Forecasting System (ATCF) SHIPS archive, and the predictand (the change in maximum sustained windspeed in the 24-hour period) were obtained from the Hurricane Data 2[nd] generation (HurDat 2) Best Track files also stored in the ATCF archives. The predictor variables used in the model were: 700-500mb relative humidity (**RHMD**), 200-800km disk average windshear magnitude at the 850-200mb (**SHRD**)and the 850-500mb level (**SHRS**), Reynolds Sea surface temperatures (**RSST**), and 200mb divergence (**D200**). Prior research indicated these predictors are important for TC intensification. The data were limited to Atlantic basin storms from 2011 – 2017 that reached the Tropical Depression threshold over open water.

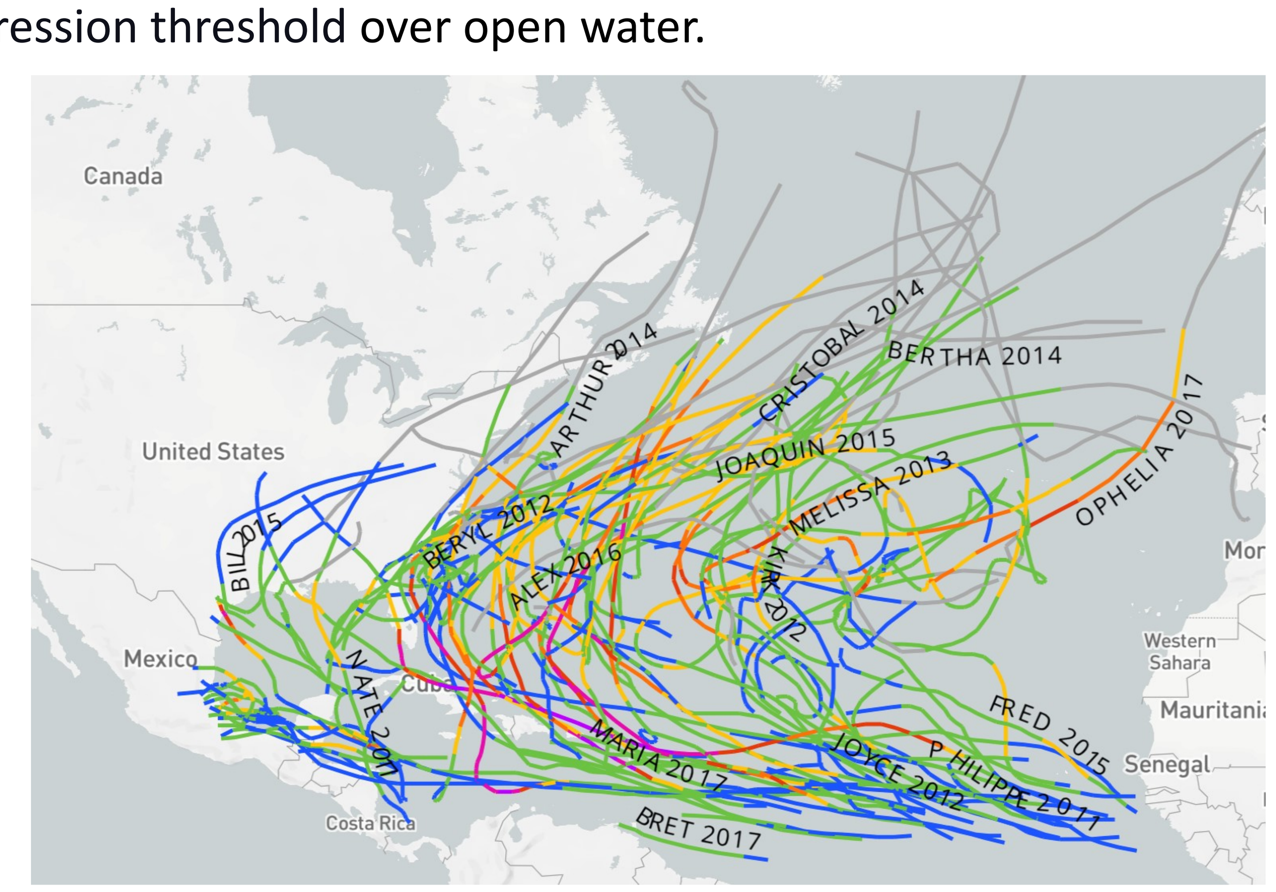

Figure 2. 2011-2017 Atlantic Basin hurricane tracks. Retrieved from NOAA's Historical Hurricane Tracks tool.

METHODS

Once the data were properly formatted in Excel with the 1674 rows containing the 5 predictors and the corresponding predictand, the initial linear regression model was created. This was done by using Excel's built in Data Analysis tool. A hypothesis test was performed to determine the validity of this initial model and determine whether it is statistically significant. A variance inflation factors (VIF) test was performed on those predictors that had multicollinearity, indicated by high correlation values between them. Those predictors with high VIF values were removed from the model. Once the inadequate predictors had been removed, a second linear regression model was created with the remaining variables and a new hypothesis test was performed. Finally, the strength, validity, and accuracy of the model were tested.

RESULTS

The final optimized linear regression model is described by the following equation:

$$Y = 1.091699x_1 + 0.197384x_2 - 0.234427x_3 + 0.041933x_4 - 34.7660$$

The model's regression statistics, coefficients, and correlation values are shown on table 1, 2, and 3 respectively.

Regression Statistics	
Multiple R	0.434381
R Square	0.188686853
Adjusted R Square	0.186743588
Standard Error	13.50821026
Observations	1675

Table 1. Regression Statistics.

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	-34.76604278	4.015060201	-8.65891	1.1E-17
SSRT	1.091699207	0.121252051	9.003553	5.8E-19
RHMD	0.197384214	0.039214347	5.033469	5.34E-07
SHRD	-0.234427057	0.036201942	-6.47554	1.24E-10
D200	0.041933712	0.010329946	4.059432	5.15E-05

Table 2. Model coefficients, Standard Errors, t Stats, and P-values.

Correlation	*SSRT*	*RHMD*	*SHRD*	*D200*	*VMAX*
SSRT	1				
RHMD	0.239313244	1			
SHRD	-0.504465183	-0.249368434	1		
D200	0.009221567	0.462893565	0.117104	1	
VMAX	0.352138758	0.281268242	-0.3108	0.148903	1

Table 3. Correlation Table of the Final Model.

Both R2 and the adjusted R2 are small and therefore the model's fit is not ideal. However, R2 and adjusted R2 are very similar, suggesting that the model is generalizable. The standard error of 13.5 shows that the data points don't fall that close to the model's regression line. The significance F is smaller than 0.05, and the F statistic is much greater than the critical F statistic, validating the model because it shows that it is statistically significant. The predictor's standard errors are fairly low, and the absolute value of their t Stats were much greater than **1.96**.

When using the model to predict Atlantic TC intensification of 2018 storms and then testing the error by comparing the predictions to the observations 24 hours later, the mean absolute error of the whole season was **10.43 knots**. The minimum error was 0.02 knots and the maximum error of 54.88 knots. The 2018 Official NHC forecast mean intensity error at 24 hours is approximately **8 knots** (Cangialosi 2019).

To determine if the model can be improved, residual plots were created for all the predictors (figure 3, 4, 5, and 6).

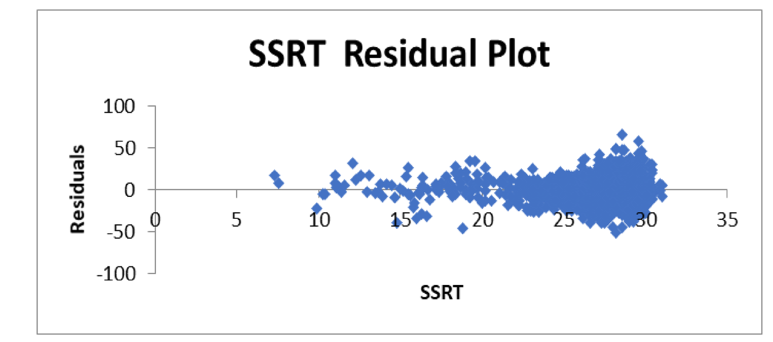

Figure 3. Residual plot for the SSRT predictor.

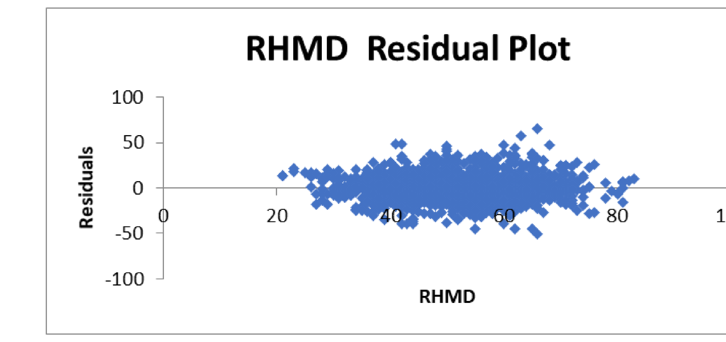

Figure 4. Residual plot for the RHMD predictor.

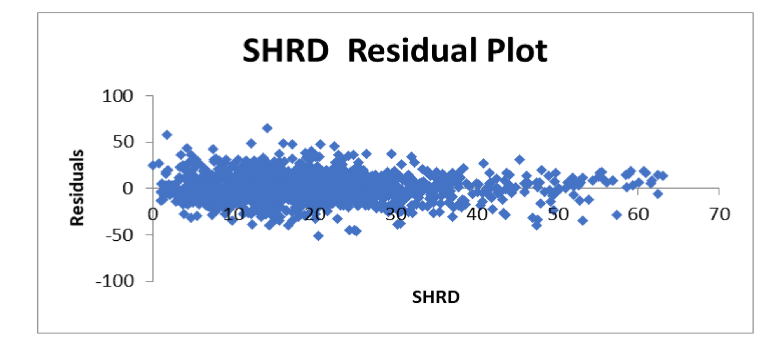

Figure 5. Residual plot for the SHRD predictor.

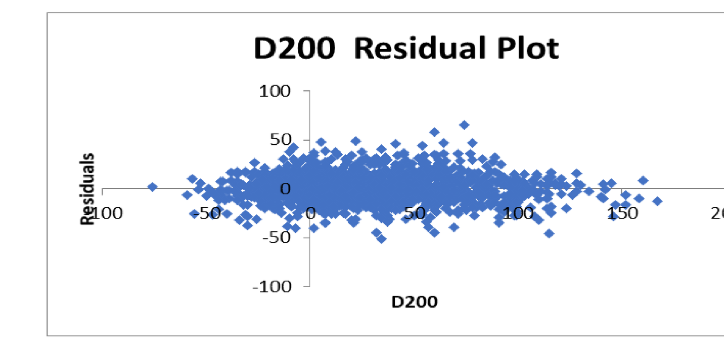

Figure 6. Residual plot for the D200 predictor.

The residual plots for RHMD and D200 look randomly dispersed, so this linear model is probably most appropriate for them. However, the residual plots for SSRT and SHRD show some heteroscedasticity, indicating that a transformation is necessary or that a variable is missing in the model. An exponential transformation was tested on all the variables, but it was unsuccessful: The heteroscedasticity grew and shifted to the other side. Therefore, to improve the model in the future and come to more concrete conclusions as to the influence of each predictor on TC intensification, further research is necessary.

CONCLUSION

Reynolds sea surface temperature was the most determinist predictor, having the largest coefficient and test stat. This is consistent with what would be expected meteorologically. The next most deterministic predictors were both RHMD and SHRD. The relative humidity and windshear had a similar level of influence on TC intensification. Moisture and latent heat release are necessary at the core of the storm for a TC to intensify. A high windshear and low relative humidity environment are detrimental to both, which coincides with our project results. Divergence at 200 mb would have been expected to have a stronger impact on TC intensification. Physically, divergence aloft should intensify surface low pressure systems, but this was not clearly shown by the model: D200 was the least deterministic predictor.

Even though a lot more research is needed to improve the model and finish this project, a decent model was developed, valuable results were obtained, and important conclusions were made. No ground-breaking or shocking results were produced, but they still help to confirm the work of many experts in the meteorology field: Sea surface temperature most strongly affects TC intensification and therefore there is a need for a higher density of surface weather observing systems in the ocean and higher quality temperature measurements.

References

[1] Cangialosi, J. P., (2019). National Hurricane Center forecast verification report. National Hurricane Center (NHC), 79.

[2] NOAA. (2018, October 10). Hurricane Michael Strengthens to a 'Potentially Catastrophic' Category 4 Storm. Retrieved from https://www.nesdis.noaa.gov/content/hurricane-michael-strengthens-'potentially-catastrophic'-category-4-storm

[3] NOAA. (2016, April 20). NOAA Historical Hurricane Tracks. Retrieved from https://oceanservice.noaa.gov/news/historical-hurricanes/