DEEP LAYER OF UPPER AIR AND MULTIVARIATE TIME SERIES MODELS TO PREDICT HURRICANE TRACKS

Joan Manuel Castro
Advisor: Dr. Nazario D. Ramirez Beltran
University of Puerto Rico

ABSTRACT

Scientists, mathematicians and engineers, had been studied the evolution of the hurricanes, during long time. One of the most interesting aspects that the hurricane researchers consider are the reasons that determine the hurricane tracks. The people who work on the meteorology field had used multiple tools to predict hurricane tracks for six (6) to seventy-two (72) hours in advance. The proposed method in this study is limited to 24 hours as a lead time. One the major problem that the hurricane experts are facing is to identify the appropriate observations to initialize hurricane tracks. A second important problem is to obtain the enough observations to model a hurricane that is under development. In this paper, we want to show some strategies to predict the hurricane displacement based on the first day of information and predict the hurricane tracks during its evolution,

1. INTRODUCTION

Univariate and Multivariate Time Series models are used to predict hurricane tracks in the North Atlantic basin. Two kinds of data sets are created to build the prediction operation. The first data set is used to identify a group of hurricanes with similar characteristics to the current storm. The first data set includes climatology, and persistence information during ninth-teen (19) years (1984 – 2002) and the data source is the National Hurricane Center Best Track. The Best Track provides historical information of the hurricanes that have occurred on the North Atlantic basin. Climatology and persistence data and a self organized neural network (NN) are used to identify the storm analogs, which are based on: Julian date, Eastward and Northward displacements, storm intensity and direction as well as sea surface temperature.

The second data set includes the steering and synoptic observations along the storm track. The second data set is organized to create a time series to predict the hurricane displacements. The second data set is obtained from NCEP Reanalysis, and Radiosonde. The NCEP reanalysis data provides seven-teen (17) vertical pressure levels of climatological variables. NCEP Reanalysis and Radiosonde receiver provide five steering and synoptic variables on a horizontal area covering 21 x 21 degrees (relative humidity, geopotential height, air temperature, and wind speed components). Figure 1 shows the gridded information from NCEP and Radiosonde. The NCEP data comes on 2.5x2.5 degrees and Radiosonde was interpolated into a 1x1 degree to facilitate the comparison of the two data sets. The objective comparison of these data indicates that there are not enough evidences to reject the hypothesis that the data are similar.

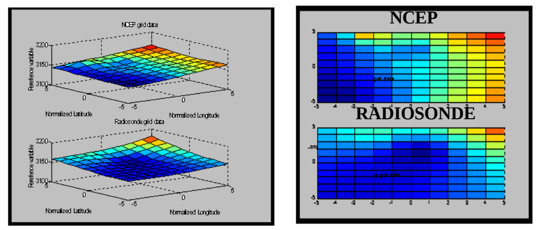

Figure 1: Geopotential Height conditions around the hurricane eye on September 14, 1995 (a) Information receive by NCEP (b) Information receive by Radiosonde (1 x 1 degree)

A spatial interpolation algorithm is used to obtain estimations at one degree of resolution on the horizontal. A deep layer for each variable is computed to reduce the vertical layers into a single layer and derive a robust estimator. Dimensionality reduction is accomplished without loosing important information by using the first 10 principal components, which represents more than 90% of the total variance of the considered variables.

An algorithm is used to identify the optimal lags and the variables that best explain the hurricane displacement. The best three steering and synoptic variables are selected from the dimensionality reduction process. The Hooke and Jeeves algorithm is used to identify the structure of the Univariate and Multivariate Time Series Models and the hurricane displacement is predicted. Different prediction estimations are built at every 6 hours, implementing the described methodology. Aircraft reconnaissance and satellite data are under exploration to be included in the data sets to improve prediction capabilities.

2. METHODOLOGY

Our suggested methodology to predict hurricane tracks includes three fundamental steps: 1) We want to identify the upper air information to generate data to initialize time series prediction model, 2) our model develops upper air time series along the storm track, and finally 3) our algorithm predicts the hurricane displacement.

2.1 Initialize the algorithm to predict hurricane tracks

During the first 24 hours of a tropical storm, we obtain five consecutives observations, every 6 hours per sampled observation. The first day hurricane data are used to estimate the initial upper air data that the algorithm needs to develop a multivariate time series model. The persistence variables are used to identify historical hurricanes with similar conditions than current storm to create a simulated hurricane track, before 24 hours of a current storm track. An analog hurricane is a storm that has similar meteorological behavior to the current tropical storm. Similarities are identified in term of the following variables: the Julian date and observations taken at the center of the tropical storm such as: latitude, longitude, barometric pressure, maximum sustained wind, and sea surface temperature.

A competitive artificial neural network (ANN) is used to classify the analog hurricanes to the current storm. The process to select analog storms is accomplished by implementing the following steps. The first step consists of selecting 15 neurons with a competitive ANN. A competitive ANN algorithm classifies each hurricane based on particular characteristics in each observation. Each historical storm is assigned a code number from one (1) to fifth-teen (15). The hurricanes belong the same code group to the current storm indicate strong similarities and are named potential analogs. The potential analogs are selected from a file that contains 100 hurricanes, data from 1984 to 2002. The second step consists of processing the potential analogs by a second ANN algorithm to analyze the five observations of each hurricane at the same time. The second ANN algorithm uses five (5) neurons to evaluate the strongest characteristics of potential analogs and classifies the hurricanes into five (5) categories. The hurricanes that belong into the same category of the current storm define the analog hurricanes. The third step consists on filtering the analog hurricanes by a second classification algorithm whose criterion consists on find the minimum Euclidian distance between the current storm and the potential analogs. The storm with the minimum distance is called the analog hurricane.

The location of the first five points of the current storm and the selection of the analog storm (as was explained in the last paragraph) are used to generate twelve (12) simulated hurricane track points. The new information will be used to initialize time series model. The analog storm, the upper air NCEP data, and the twelve (12) simulated hurricane positions plus the first five (5) points of the current hurricane are used to obtain upper air estimates for the current storm. Upper air data is extracted from NCEP at the following levels: 1000, 850, 700, 500, 400, 300, 250, 200, 150 and 100 mb. These data are centered on the storm position with 10 degrees measurements into the axial directions to East, West, South, and North. Thus, the extracted number of gridded points at each time interval, around the square of 21 x 21 degrees is 441 pixels.

The information provided by ten pressure levels is used to create the deep layer (Neumann, 1988). The values of each variable at each level are multiplied by its corresponding coefficient. The used weights indicate more relevance to 700 and 500 mb levels. The deep layer at each time interval and the location of the storm of the 17 points are defined as the minimum information to initialize the time series model and to predict at 6, 12, 18 and 24 hours in advance.

2.2 Develop the upper air time series

The time series for the current tropical storm is delivered at every six hours a soon as more information is available. Assuming that the location of the storm is known at point 18 and it is expected to predict at 6, 12, 18, and 24 hours in advance. The first step consists of estimating the upper air at point 18. The upper air is estimated based on selecting the appropriate analog hurricane. At this step the algorithm identifies an analog hurricane. The strategy for estimating the upper air is similar to the initial point. This includes three steps: 1) the algorithm uses only the last observation to estimate the analog storms, which is different from the initial point, 2) classification algorithm is used to identify the analog and 3) the classifier algorithm select the upper air for the current storm and at the current position.

A competitive ANN with 5 neurons is used to identify the analog hurricane based only on the last observation of the current storm. Neural network algorithm used six persistence variables (as in the previous section) to identify the analog. The synoptic variables that were added to the persistent and climatological variables are the following: Maximum Possible Intensity (MPI), Eastward Wind Speed, Northward Wind Speed, and Vertical Wind Shear. The classification algorithm is based on minimizing the Euclidian distance between the analog storms and the current. The historical storm selected from this process is called the analog hurricane. Data of the analog hurricane obtained from NCEP is used to generate gridded deep layer as explained in the previous section. The described process is applied at every six hours, as soon as new information is available.

2.3 Predicting hurricane displacement

Univariate and bivariate time series models were identified depending on the length of the available information (Brockwell, and Davis, 2002). Typically, the bivariate time series model requires more observation than the univarite model to build a time series model. Thus, at the early stages of the storm only the univariate models were identified, one for modeling zonal and the other for meridional displacements. The bivariate model expressed simultaneously the two displacements.

The best three predictors and the Hook and Jeeves algorithm are used to identify the structure of the time series model and the hurricane displacement was predicted (Reklaities, 1983). A different prediction model is built at every 6 hours and predictions are computed for 6, 12, 18 and 24 hours in advance.

3. RESULTS

The proposed methodology was implemented to asses the performance of the prediction algorithm. In a preliminary study three hurricanes were selected. The selected hurricanes occurred on the North Atlantic Ocean during the period of 1995 to 2002. The major characteristics used to select a hurricane were the following: erratic displacement and a large hurricane's life. The hurricanes that fall in that category are: 1) Hurricane Marilyn from September 12th to October 1st, 1995; 2) The second selected storm is Hurricane Danielle from August 24th to September 3rd, 1998 and the third selected storm is Hurricane Kyle form September 20th to October 12th, 2002.

Performance of the prediction scheme is evaluated in terms of the size of the squared prediction error. The prediction error in nautical mails is obtained after computing the distance between the observed and predicted location. The prediction error is computed by using the following equations:

$$e_t = 60\cos^{-1}\left\{\sin La_{o,t}\sin La_{f,t} + \cos La_{o,t}\cos La_{f,t}\cos\left(Lo_{o,t} - Lo_{f,t}\right)\right\}$$

where e_t is the predicted error at time *t*. $La_{o,t}$, and $Lo_{o,t}$, $La_{f,t}$, and $Lo_{f,t}$ is the observed and predicted location of storm at time *t*.

3.1 Case 1: Hurricane Marilyn (September 1995)

For a 12 hours prediction interval, the Absolute Average Prediction Error (AAPE) of the hurricane was 34 nautical-miles. This average was computed using 38 observations, i.e., nine and half days of hurricane life time. The National Hurricane Center official Best Track reported for this hurricane during the same analyzed period an average prediction error of 38 nautical miles. Thus the proposed model over performed the official prediction, in the sense that our model reduces the prediction error in 4 nautical-miles. The prediction algorithm selected the following variables as the best predictors: the gradients of eastward of geopotential height and the eastward air temperature.

The AAPE for a 24 hours lead time was 81 nautical-miles with 35 forecasts evaluations. This result is thirteen (13) nautical-miles greater than the reported by the National Hurricane Center official Best Track, which was 68 nautical-miles. The prediction model selected as the best predictors the same predictors reported for the 12 hours lead time. The prediction performances for Hurricane Marilyn are shown in Figures 2 and 3. Figure 2 shows the comparison between the observed and predicted eastward and northward displacements, for both 12 and 24 hours prediction interval. Figure 3 has two lines one of the lines represents the observed track and the other one is the predicted track, this figure is also organized for a 12 and 24 hours prediction interval.

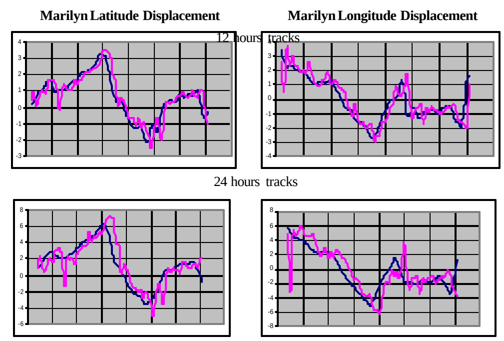

Figure 2 Hurricane Marilyn Tracks in 12 and 24 hours (blue – obs.) (red – estimated).

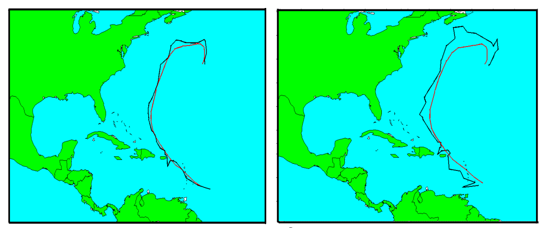
Figure 3 Hurricane Marilyn for 12 and 24 hours (blue – observed) (red – estimated)

3.2 Case 2: Hurricane Danielle (August 1998)

For a 12 hours prediction interval, the AAPE of the hurricane Danielle was 32 nautical-miles. This average was computed using 39 observations, i.e., nine days and three quarter of day of hurricane life time. The National Hurricane Center Official Best Track reported for this hurricane during the same analyzed period an average prediction error of 31 nautical miles. Thus the official prediction is better than the proposed algorithm. The prediction algorithm selected the following variables as the best predictors: the northward and eastward gradients of geopotential.

The average absolute prediction error for a 24 hours lead time was 64 nautical-miles with 35 forecasts evaluations. This result is thirteen (16) nautical-miles greater than the reported by the National Hurricane Center official Best Track, which was 48 nautical-miles. The prediction model selected the following predictors: northward and eastward gradients of geopotential height, and geopotential height variable. The prediction performances for Hurricane Danielle are shown in Figures 4 and 5. Figure 4 shows the comparison between the observed and predicted eastward and northward displacements, for both 12 and 24 hours prediction interval. Figure 5 has two lines one of the lines represents the observed track and the other one is the predicted track, this figure is also organized for a 12 and 24 hours prediction interval.

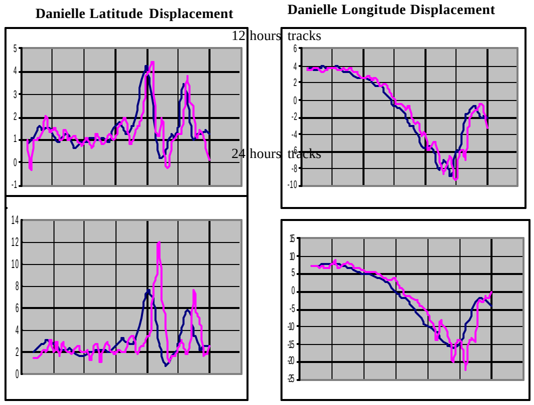

Figure 4 Hurricane Danielle Tracks in 12 and 24 hours (blue – obs. track) (red – estimated track)

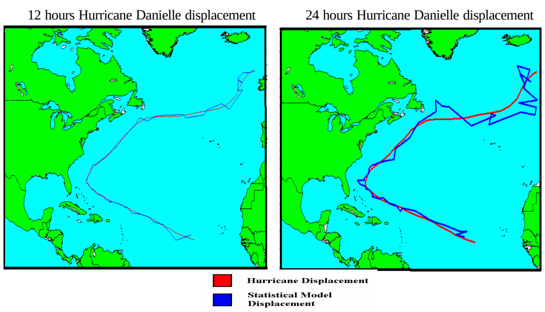

Figure 5 Hurricane Danielle Trajectory in 12 and 24 hours (blue – obs. track) (red – estimated track)

3.3 Case 3: Hurricane Kyle (September 2002)

Kyle is one of largest tropical storms that has occurred on the North Atlantic Basin and lasted about 14 days. The prediction model was implemented to perform 12 hours prediction interval and the AAPE was 40 nautical-miles. This average was computed using 57 observations. The National Hurricane Center Official Best Track reported for this hurricane during the same analyzed period an average prediction error of 33 nautical miles. Thus, the official prediction provides better results than proposed scheme. The prediction algorithm selected the following variables as the best predictors: the northward and eastward gradients of air temperature.

The average absolute prediction error for a 24 hours lead time was 69 nautical-miles with 35 observations. This result is ten (10) nautical-miles greater than the reported by the National Hurricane Center Official Best Track, which was 59 nautical-miles. The prediction model selected the following variables as the best predictors: the gradients of northward air temperature and eastward geopotential height. The prediction performances for Hurricane Kyle are shown in Figures 6 and 7. Figure 6 shows the comparison between the observed and predicted eastward and northward displacements, for both 12 and 24 hours prediction interval. Figure 7 has two lines one of the lines represents the observed track and the other one is the predicted track, this figure is also organized for a 12 and 24 hours prediction intervals.

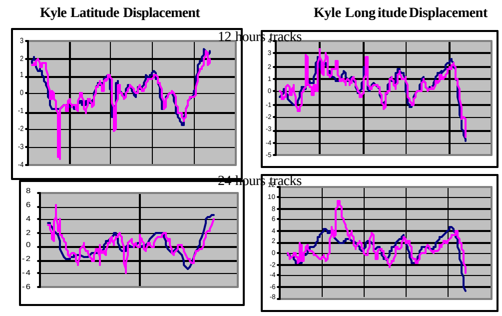

Figure 6 Hurricane Kyle Tracks in 12 and 24 hours (blue – obs. track) (red – estimated track)

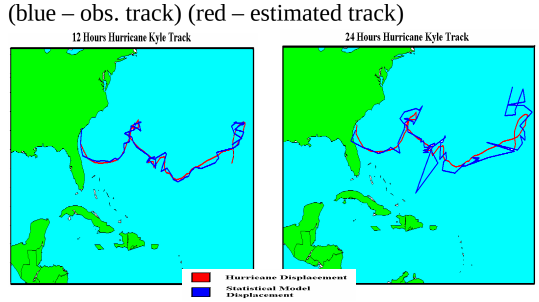

Figure 7 Hurricane Kyle Trajectory in 12 and 24 hours (blue – obs. track) (red – estimated track)

	Marilyn		Danielle		Kyle	
Models	Error (nm)	Cases	Error (nm)	Cases	Error (nm)	Cases
Official	38	39	31	39	33	56
Time Series Model	34	37	40.3	39	40.3	57

	Marilyn		Danielle		Kyle	
Models	Error (nm)	Cases	Error (nm)	Cases	Error (nm)	Cases
Official	71	39	47	37	59	51
Time Series Model	81.6	35	64.8	37	69.2	52

Table 1. Average Prediction Errors

4. CONCLUSION

According to preliminarily results obtained by validation of the three hurricane tracks, it has been shown that Time Series Models predict the storm tracks with small prediction errors, especially for the lead times of 12 and 24 hours. However, the proposed method obtains absolute average error greater than National Hurricane Center Official Best Track Prediction. It should be noted that the official prediction is the result of running several models and use the available information and the opinion of experts in hurricane tracking. Our target is be competitive with existing prediction models. Therefore, we need to perform more exploration to claim that the proposed method is better than the existing prediction scheme.

Another alternative to increase the prediction capabilities of the model is to include Aircraft reconnaissance data and use observation from the Advance Microwave Sounding Unit (AMSU).

Preliminary results show that the prediction scheme is a potential tool to increase the accuracy of predicting hurricane displacements.

5. ACKNOWLEDGEMENT

This research has been supported by NASA-EPSCoR grant NCC5-595, by the NOAA-CREST grant NA17AE1625 and also by the University of Puerto Rico at Mayagüez. Authors want to appreciate and recognize the funding support from these institutions.

6. REFERENCES

Brockwell, P.J, Davis, R.A (1996), *Introduction to Time Series and Forecasting*, Springer-Verlag New York.

Chen W.K, (1993) *Linear Networks and Systems* Belmont, CA: Wadsworth.

Cressie, Noel, A. C. (1991). *Statistics for Spatial Data*, New York: John Wiley & Sons.

Hagan, M.T, Demuth, H.B, Beale, M. (1996), *Neural Networks Design*. PWS, Publishing Co., Boston

Neumann, C.J (1988), *The National Hurricane Center NHC83 Model – NOAA Technical Memorandum NWS NHC 41*, National Hurricane Center, Coral Gables, FL, United States

Poor H., (1995) *An Introduction to Signal Detection and Estimation*, New York: Springer-Verlag,

World Meteorological Organization (1979), *Operation Techniques for Forecasting Tropical CycloneIntensity and Movement*, Secretariat of the World Meteorological Organization, Geneva –Switzerland, page II 3-9