To appear in: Communications in Applied Analysis **xx** (2009), no. N, xxx–xxx

LINEAR AND NON-LINEAR STATISTICAL MODELING OF HURRICANE FORCE WINDS

REBECCA D. WOOTEN[1] AND CHRIS P. TSOKOS[2]

[1]Department of Mathematics & Statistics, University of South Florida
Tampa, FL 33620 USA
E-mail: rwooten@cas.usf.edu

[2]Department of Mathematics & Statistics, University of South Florida
Tampa, FL 33620 USA
E-mail: profcpt@cas.usf.edu

ABSTRACT. In the present study, the primary aim concentrates on the modeling of hurricane force winds; that is, maximum sustained winds related to pressure, location and linear velocity. We were successful in modeling the wind speed within storm as a function of the contributing entities. In this study, we were able to re-evaluate the association between wind speed and pressure within storms. The analysis of wind speed versus pressure indicates that further analysis of the Saffir-Simpson Scale is necessary, as well as determining if pressure is an indicator or a consequence of a hurricane force wind speed.

AMS (MOS) Subject Classification. 99Z00.

1. INTRODUCTION

There are statistical models in forecasting the track of hurricanes, but how well do we understand the mechanics underlying the birth and pathway (track) of a tropical storm? What are the contributing variables and how do they rank in comparison; that is, what are the explanatory variables according to their contribution to the model? What are the significantly contributing interactions? To answer these questions, data gleaned from UNISYS Tropical Prediction Center was used which included five recent storms classified as category, see Table 1. Provision included: charts on the track of the storm, tracking information, position in latitude and longitude, maximum sustained winds in knots, and central pressure (hPa) and date including month, day and time as well as several other derived variables, see Table 2.

The phenomenon of hurricane force winds depends on the surrounding pressure as well as the latitude at which the circulations form. Hurricanes cannot form on

Received June 30, 2009

Year	Storms	Max Sustained Wind	Pressure	Color
2005	Wilma	150	882	Purple
2005	Rita	150	—	Red
2005	Katrina	150	902	Orange
2004	Ivan	145	910	Green
2003	Isabel	140	920	Blue

TABLE 1. Table of maximum hurricane force winds and their associated pressures for five recent storms in the Atlantic region

the equator thanks to the Coriolis Effect, Strobel [4]. These five storms will provide a glimpse into understanding the transitions between Category 0 (tropical storm) to Category 1, etc. as well as the relationship between gathered information.

Foremost, either wind speed or pressure could be considered as the response variable; however, the commonly held belief is the low pressures cause hurricanes to form, Heckert, Simiu, Fellow, ASCE and Whalen [3]. Therefore, in this paper we will treat the wind speed as the response variable and the pressure to be a contributing or explanatory variable, Emanuel [2].

Furthermore, the measurements of latitude and longitude are not uniformly scaled - they exist in a sphere; therefore latitudes are further apart near the equator and closer together near the poles. Hence, to model hurricanes into terms of their positions, these measurements first need conversion to Cartesian coordinates, where linear movements are a valid measure and therefore approximation linear velocities exist.

1.1. **Conversion for latitude and longitude into Cartesian coordinates.** To convert from the spherical coordinate system that is defined in terms of latitude and longitude into the Euclidian space, let $a = 6378137m$, $b = \frac{1}{298.25722563}$, $c^2 = 2b - b^2$, $h \approx 100m$ (height above geoids) and $\nu = \frac{a}{(1-b^2 sin^2 b)}$, then $x = (\nu + h) cosLAT cosLON$ and $y = (\nu + h) cosLAT sinLON$. Figure 1 and Figure 2 compare the graph in terms of Latitude and Longitude, Figure 1, versus the Cartesian coordinate, Figure 2, using the five storms outlined in Table1. This conversion will allow us to measure linear displacement, linear velocity and other mathematical operations.

In the paper, we will be interested in which parameter to include in the model. In terms of location, this is a question of latitude and longitude or the transformed x and y. Since x and y illustrate the real linear movement of the storm, this transformed information with be included in the following model.

2. DEVELOPED STATISTICAL MODEL

First, the available variables are ranked in order of maximum improvement in the coefficient of determination, R^2, see Table 3. Pressure is ranked number one;

w	Maximum sustained wind speed
P	Pressure at center
LAT	Longitude (in radians)
LON	Latitude (in radians)
x	Converted to Cartesian coordinates
y	Converted to Cartesian coordinates
Δx	The change in
Δy	The change in
$\delta = \sqrt{(\Delta x)^2 + (\Delta y)^2}$	The distance between movements
Δt	The change in time
$\nu = \lvert\vec{\nu}\rvert = \frac{\sqrt{(\Delta x)^2+(\Delta y)^2}}{\Delta t}$	The magnitude of the approximate linear velocity
	Here the approximate linear velocity is $\vec{\nu} = \frac{\Delta x}{\Delta t} i + \frac{\Delta y}{\Delta t} j$.
$D = \int \Delta t$	Duration (up to that point)
d	Day of Year
Y	Year

TABLE 2. Primary variables available with hurricane records

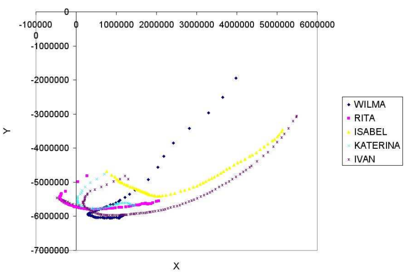

FIGURE 1. Figura 1. Scatter plot of latitude versus longitude

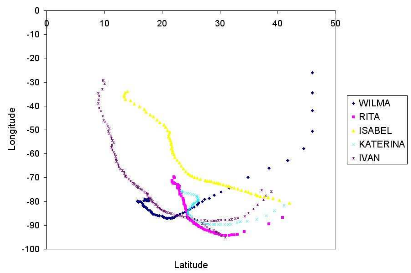

FIGURE 2. Scatter plot of converted latitude versus longitude into Cartesian coordinates x and y

however, this might be due to the fact that temperature, both water and atmospheric temperature, is not included in the data set and therefore can not be tested. However, preliminary studies show that temperature is a better indicator of wind speeds and wind gust. Unfortunately, this is information not easily measured in the eye of a storm.

First, we will consider the linear regression using all records within the five selected hurricanes regardless of hurricane status (category) and all parameters ranked in Table1.

$$\hat{w} = \hat{a}_0 + \hat{a}_1 P + \hat{a}_2 x + \hat{a}_3 y + \hat{a}_4 D + \hat{a}_5 d + \hat{a}_6 \nu + \hat{a}_8 dx + \hat{a}_9 dy + \hat{a}_{10} dt + \hat{a}_{11} \delta + \hat{a}_{12} Y \quad (2.1)$$

1.	Pressure P	90.72%	
2.	x	93.00%	
3.	y	94.03%	
4.	Duration D	94.34%	
5.	Day of Year d	94.55%	
6.	Velocity ν	94.79%	
7.	dx	94.88%	
8.	dy	95.07%	
9.	dt	95.17%	
10.	Distance δ	95.19%	
11.	Year Y	95.19%	(No improvement)

TABLE 3. Ranking of independent variables by maximum improvement in R^2

The developed statistical model found using regression and the data outlined above, we have the following information including the associated p-values, see Figure 3. There are several variables that are deemed insignificant; the linear velocity of the hurricane's track, the change in time (constant), the distance traveled over each (fixed) interval of time and the year. The first three insignificant variables are most likely due to the fact that the change in time is a constant; that is, storm readings are made every three hours. Furthermore, note that $\delta = \nu \Delta t$, therefore this model actually contains the interaction between velocities and the change in time and is insignificant. Other insignificant variables are the change in time and the year.

Dependent variable is: **Wind Speed**
402 total cases of which 5 are missing

R squared = 95.2% **R squared (adjusted)** = 95.1%
s = 8.727 with 397 - 12 = 385 degrees of freedom

Source	SS	df	Mean Square	F-ratio
Regression	580254	11	52750.4	693
Residual	29320.7	385	76.1576	

Variable	Coefficient	SE	t-ratio	*p-value*
Constant	-3079.32	3137	-0.982	0.3269
P	-1.12716	0.01549	-72.8	$\le$ 0.0001
x	9.41E-06	1.06E-06	8.84	$\le$ 0.0001
y	-1.32E-05	1.49E-06	-8.86	$\le$ 0.0001
D	0.672378	0.2182	3.08	0.0022
d	-0.0837455	0.02867	-2.92	0.0037
ν	6.58E-06	4.24E-06	1.55	0.1213
dx	-7.73E-05	1.41E-05	-5.48	$\le$ 0.0001
dy	6.81E-05	1.87E-05	3.64	0.0003
dt	-7.77545	12.08	-0.644	0.5201
δ	0.00E+00	2.84E-05	-0.196	0.8451
Y	2.08443	1.567	1.33	0.1843

FIGURE 3. Multiple regression of wind speed over pressure, time, location and other associated measures.

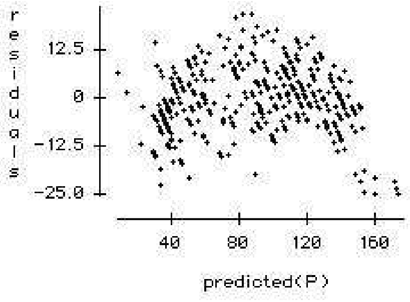

FIGURE 4. Residual plot for model outlined in Figure 7

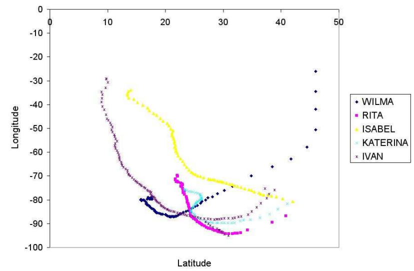

FIGURE 5. Normal probability plot for the residuals of the model outlined in Figure 7

Moreover, this model does explain 95.2% of the variation in the response and the residuals are normally distributed, see Figure 5; however, what is interesting to note is the residual plot in Figure 4. There is an obvious bowing of the data. Therefore, it illustrates the fact that there is at least one higher order term.

Consider the developed statistical model, equation 2.2: the full linear model with significant linear terms and the quadratic term pressure - squared, and its parameters including standard error are outline in Figure 6. All remaining variables are significant at the 10% level of significance; however, all but duration and linear velocity are significant at the 1% level of significance. While this statistical model only explains an additional 0.9% of the variance in the subject response, it does account for the curvature as illustrated by the residual plot in Figure 8.

$$\hat{w} = \hat{a}_0 + \hat{a}_1 P + \hat{a}_2 x + \hat{a}_3 y + \hat{a}_4 D + \hat{a}_5 d + \hat{a}_6 \nu + \hat{a}_8 dx + \hat{a}_9 dy + \hat{a}_{10} P^2 \tag{2.2}$$

Furthermore, this analysis yields $R^2 = 96.1\%$ and $R^2_{adj} = 96.0\%$, over the previous $R^2 = 95.2\%$ and $R^2_{adj} = 95.1\%$; both of which indicate that there are no extraneous variables, but that even linear velocity, duration and day of year should be considered a factor in the development of a tropical storm. Equation 2.3 gives the least-square regression model using five hurricanes which reached category five status as outlined in Table 1.

$$\hat{w} = \begin{cases} -2648.09 + 6.71895P + (8.12246 \times 10^{-6})x - (13.5435 \times 10^{-6})y \\ 0.207587D - 0.081476d + (3.5205 \times 10^{-6})\nu \\ -(58.0522 \times 10^{-6})\Delta x + (62.8245 \times 10^{-6})\Delta y \\ -0.00410053P^2 \end{cases} \tag{2.3}$$

To measure additional interaction between three most significant factors, that is, the interaction between location (latitude and longitude) and pressure, consider the

Dependent variable is: **Wind Speed**
402 total cases of which 5 are missing

R squared = 96.1% **R squared (adjusted)** = 96.0%
s = 7.819 with 397 - 10 = 387 degrees of freedom

Source	SSE	df	MSE	F-ratio
Regression	585915	9	65101.6	1060
Residual	23660.5	387	61.1383	

Variable	Coefficient	SE	t-ratio	*p-value*
Constant	-2648.09	365.9	-7.24	$\le$ 0.0001
P	6.72E+00	7.66E-01	8.77	$\le$ 0.0001
x	8.12E-06	5.82E-07	14	$\le$ 0.0001
y	-1.35435E-05	1.25E-06	-10.8	$\le$ 0.0001
D	0.207587	0.1056	1.97	0.05
d	-8.15E-02	2.43E-02	-3.35	0.0009
ν	3.52E-06	1.80E-06	1.96	0.0509
dx	-5.81E-05	1.05E-05	-5.55	$\le$ 0.0001
dy	6.28245E-05	1.47E-05	4.27	$\le$ 0.0001
P^2	-4.10E-03	4.00E-04	-10.2	$\le$ 0.0001

FIGURE 6. Multiple regression including significant linear terms and a single quadratic term for pressure.

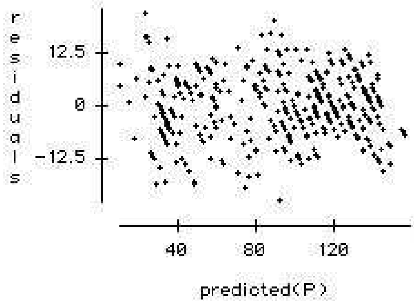

FIGURE 7. Residual plot for model outlined in Figure 6

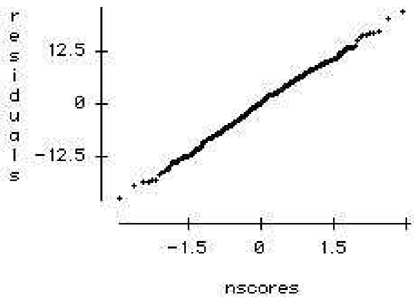

FIGURE 8. Normal probability plot for the residuals of the model outlined in Figure 7

statistical model outlined in equation 2.4.

$$\hat{w} = \begin{cases} \hat{a}_0 + \hat{a}_1 P + \hat{a}_2 x + \hat{a}_3 y + \hat{a}_4 D + \hat{a}_5 d + \hat{a}_6 \nu \\ +\hat{a}_8 dx + \hat{a}_9 dy + \hat{a}_{10} P^2 \\ +\hat{a}_{11} Px + \hat{a}_{12} Py + \hat{a}_{13} xy \end{cases} \tag{2.4}$$

This interactive model explains 97.1% of the variation in the wind speed as measured in knots. Furthermore, there is significant interaction between the vertical and horizontal axis; that is, the latitude and longitude as well as the pressure and primarily the vertical displacement in the converted rectangular coordinate system. Furthermore, the developed statistical model yields $R^2 = 97.1\%$ and $R^2_{adj} = 97.0\%$, with or without interaction between the pressure and the converted x value. Hence, invoking the law of parsimony, we will not include this interaction in our model. Thus, the developed statistical model including the significantly contributing interaction is given in equation 2.5. All variables included in this statistical model which includes both a quadratic term and interactions are significant at the 1% level of significance.

$$\hat{w} = \begin{cases} -727.895 + 3.911341P + 49.1989 \times 10^{-6}x + 210.721 \times 10^{-6}y \\ +0.650216D - 0.178422d + 11.9516 \times 10^{-6}\nu \\ -60.005 \times 10^{-6}dx + 94.4684 \times 10^{-6}dy \\ -0.00337569P^2 \\ -256.827 \times 10^{-9}Py + 7.71408 \times 10^{-12}xy \end{cases} \tag{2.5}$$

Applying this statistical model, equation 2.5, to category zero, that is, to tropical storms and depressions, the storm before it is classified as a hurricane. The pressure's interaction with the location is found to be insignificant as is the vertical displacement. Moreover, the explanatory power drops to 81.8%. With such low wind speeds, this model is less reliable; that is, the model when estimated using only data defined as a tropical storm or depression explains only 81.8% of the variation in the wind speed. Furthermore, the least square regression model developed indicates that the wind speed depends less on the latitude and longitude, and more on the change in latitude and longitude. Recall:$x = f(LAT, LON)$ and $y = g(LAT, LON)$.

Applying this statistical model, equation 2.5, to the storms when the status is hurricane category one, none of the variables are found to be statistically significant at the 1% level of significance. However, there is not enough information to determine this; that is, out of 397 readings for five different storms that reached category-five status, only 27 of these readings were when the storm was categorized as a hurricane category one. This may be indicative that once conditions are right for a hurricane to form, it will do so rather quickly. Furthermore, the most statistical significant term remaining at the 10% level of significance is the day of year and as previous studies indicate, this may be due to the drop in temperature that occurs as the storm progress to its eventual dispersion or death which would be negatively correlated to elapsing of time.

Applying this statistical model, equation 2.5, to the storms when the status is hurricane category two, none of the variables are found to be statistically significant. There were 52 readings in which the hurricane was classified as hurricane category two; this should be enough information, however there are twelve explanatory variables and these are more than likely confounding. Hence, these variables are contributing, but the explanatory variables are so intertwined that their contribution to the subject response can not be distinguished.

Applying this statistical model, equation 2.5, to the storms when the status is hurricane category three, the most significant variable is the day of year and third is the duration, which is how long the storm has been in formation; both related to the amount of energy expelled and temperature. The second variable that comes back into focus is location; first the interactive term, xy and then the horizontal displacement, x. Furthermore, the explanatory power increases to 67.1%; there is still a rather large discrepancy between the coefficient of determination and the associated adjusted statistics. However, this is with only 47.

Applying this statistical model, equation 2.5, to the storms when the status is hurricane category four, the most contributing variable is the horizontal displacement. Second is the linear velocity; third and fourth are the pressure's interaction with the vertical position and the vertical position, respectively. Furthermore, the explanatory power increases to 75.3% with an adjusted coefficient of 72.4%.

Applying this statistical model, equation 2.5, to the storms when the status is hurricane category five, this model again loses explanatory power. With 49 of the 397 readings, the variables appear confounded.

3. MODEL VALIDATION

First extend the data set to include all 110 storms tracked using 18,345 samplings recorded during the 1990s. The developed model given in equation 3.1 explains 90.7% of the variation in the subject response with no superfluous variables. Then using this model, the 111 storms were tracked using 11,610 samplings recorded from 2000 through 2006.

$$\hat{w} = \begin{cases} -5266.77 + 1.14195P + 3.73961 \times 10^{-6}x + 141.012 \times 10^{-6}y \\ +8.22624D - 0.0305525d - 2.03126 \times 10^{-6}\nu \\ -9.87403 \times 10^{-6}dx - 27.0026 \times 10^{-6}dy \\ -0.00615345P^2 \\ -133.384 \times 10^{-9}Py + 6.75367 \times 10^{-13}xy \end{cases} \tag{3.1}$$

Using the 1990s to calibrate the developed statistical model, then using this model to estimate the wind speeds during the 2000s, 90.8% of the variation in the subject response is explained. With a mean residual of -2.253 and a sample standard deviation of 8.529, this model does appear to handle prediction rather well with 50% of all the residual between -7.594 and 2.742. This is expected since wind speeds are recorded in multiples of five. However, the minimum residual is -37.991 and the maximum residual is 40.86. Compared to the residuals that exist between the original data set for the 1990s and the estimated wind speed according to the developed statistical model, the mean residual is 0.0124 and sample standard deviation of 7.58, with a minimum of -34.122 and a maximum of 34.384. This shows that while this statistically does explain a significant amount of the variance, there are additional contributing variables. Consider Hurricane Katrina, Figure 9; the blue indicates that actual records measured every six hours during Hurricane Katrina and the red indicates the estimated wind speed using equation 3.1.

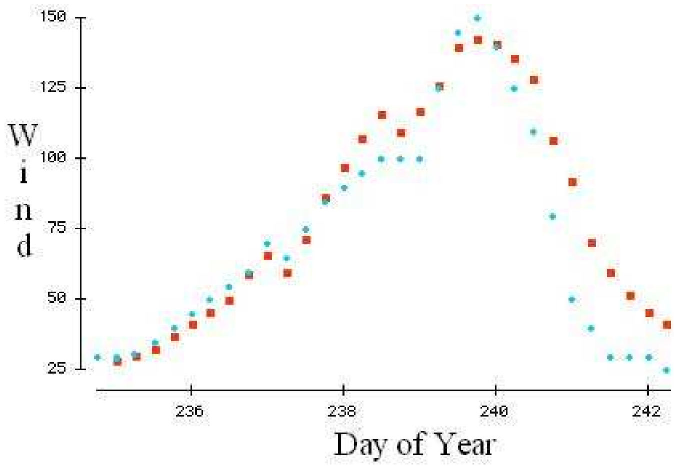

FIGURE 9. Time Series for Hurricane Katrina.

4. RESULTS AND INTERPRETATIONS (DISCUSSION)

Statistically, with just a few prior pieces of information, we can estimate with high degree of accuracy the associated wind speed; that is, our model explains 97.1% of the variation in the wind speed. Some of the secondary result, estimating the coefficients for the various categories may need to be re-evaluated since it has be shown that the Saffir-Simpson scale does not categorize hurricane force winds appropriately according to significant changes in the pressure. Reclassification of the categories might yield a better fitting model when regressed categorically. Furthermore, coupling physics with statistics should produce a much more reliable model; however, categories aside, the non-linear statistical model developed can still be used to accurately estimate the intensity of a storm.

5. CONCLUSION

With the present day technology and the historical data now readily available, hurricane prediction will become more accurate in the near future. While this model predicts the intensity of the storm, now we need to address the issues of direction and duration and how this relates to the intensity. The "spaghetti string" models, averaged and used to make the cone shaped predictions and forecast as new information is gathered, can be adjusted to be more accurate or simply replaced by stochastic systems developed by statisticians working with meteorologists, Darling [1] , Murnane, Barton, Collins, Donnelly, Elsner, Emanuel, Ginis, Howard, Landsea, Kam-liu, Malmquist, McKay, Michaels, Nelson, O'Brien, Scott, Webb [3].

REFERENCES

[1] R. W. R. Darling, Estimating Probabilities of Hurricane Wind Speeds Using a Large-Scale Empirical Model, Institute for Constructive Mathematics, University of South Florida, Tampa, 1991.

[2] K. A. Emanuel, The dependence of hurricane intensity on climate, *Nature* **326** (1987), 483–485.

[3] R. J. Murnane, C. Barton, E. Collins, J. Danelly, J. Elsner, K. Emanuel, I. Ginis, S. Howard, C. Landsea, Kam-Liu, D. Malmquist, M. McKay, A. Michaels, N. Nelson, J. O'Brien, D. Scott, T. Webb, Model Estimates Hurricane Wind Speed Probabilities, *American Geophysical Union* **81** (2000), 433–438.

[4] N. A. Hecker, E. Simiu, ASCE Fellow, and T. Whalen, Estimates of Hurricane Wind Speeds by PEAK OVER THRESHOLD Method, *Journal of Structural engineering* (1998), 445–449.

[5] N. Strobel, Astronomy Notes, Bakersfield College, Physical Science Department, Bakersfield, CA, 1998.