15D.6 ATLANTIC HURRICANE FORECAST AND CLIMATOLOGICAL IMPACTS: A STATISTICAL ANALYSIS

Siamak Daneshvaran* and Maryam Haji
Impact Forecasting LLC, Aon Benfield, Chicago, Illinois, USA

1. INTRODUCTION

A better prediction of future hurricane activity, in general, leads to better estimation of future insured catastrophe losses. Catastrophe models employ complex mathematical techniques to generate forecasts of both loss frequency and severity. Even with the availability of powerful computers, these applications (mathematical models) can be time consuming and quite expensive. However, no matter how reliable and intelligent these individual prediction models are, they are subject to inherent and epistemic uncertainty that has a significant impact on the estimation of event losses. As Powers (2006) indicates, the problem with forecasting catastrophes is the absence of extensive data validation, given the short length of record with high-quality data.

In the first part of this study we consider one of the most publicly discussed forecasting models introduced by Gray (1984) and compare it to a simple forecasting alternative (similar to the study done by Powers, 2006). Gray (1984) generated a complex methodological and statistical model of global weather patterns to predict the number of named tropical storms and hurricanes originated in the Atlantic Basin. Prior to the start of each Atlantic Hurricane Season, Gray's model makes seasonal forecasts in early December, April, June and August (As of December 2011, Colorado State University no longer issues December hurricane forecasts for the upcoming Atlantic Hurricane Season).

For the purpose of this study, we only consider and compare the August prediction numbers from Gray's analysis for the time interval of 1990 to 2011. In our simple alternative, the historical hurricane record from 1900 to 1989 is taken from the NHC data-set and is employed to predict the hurricane activity for years between 1990 and 2011 using technical analysis. An ARIMA technique is used to model the behavior of the Atlantic hurricanes. As shown in Figure 1, the hurricane frequency series (1900 – 2011) is nonstationary, where the mean is slightly increasing over time. Autocorrelation functions are used to specify the orders of the ARIMA model. The results of this analysis are presented in section 3.

Over the last 100 years, Principal Component Analysis (PCA) has been employed in many different fields of study (Daneshvaran and Morden, 1998, Ramsay 1988, Holmes 1990) to identify the dominant variables and mechanisms that describe and control the structure and processes underlying a specific data-set. This method has been widely used in meteorological and oceanic data analysis (Anderson and Gyakum 1989, Lee and Cornillon 1995, Xie et al. 2005). Oftentimes, the expectation is that a few underlying variables are able to express essential structure embedded within the original data and by specifying those variables, we could achieve important insights.

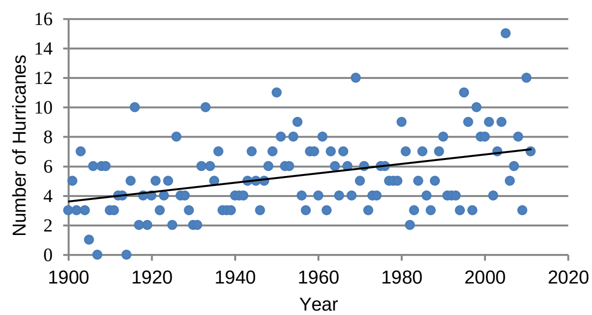

Figure 1: North Atlantic Hurricane Frequency

The advantage of PCA is that it describes the data variability in terms of orthogonal functions or mode shapes. If there are substantial correlations among the variables in the original data and the available information is redundant, PCA identifies the dominant mode shapes and reduces a data-set containing a large number of variables to a smaller set of data which represents a large fraction of the variability contained in the real phenomenon. The remarkable property of principal components (mode shapes) is that they are uncorrelated.

Global climate patterns and atmospheric conditions are linked to the vitality of hurricanes and they are used in empirical models to forecast hurricane activity several months in advance (Elsner and Schmertmann 1993, Gray et al. 1992). An extensive number of studies have

* *Corresponding author address*: Siamak Daneshvaran, Impact Forecasting LLC, Aon Benfield, Chicago, Illinois, 60601; e-mail: Siamak.deneshvaran@aonbenfield.com.

been focused on the effect of one or more of these climate factors (Hoyos et al. 2006, Mann and Emanuel 2006, Webster et al. 2005, Trenberth 2005, Godenberg et al. 2001, and Klotzbach 2006). In the second part of this work, anomalies associated with six climate indices obtained from the National Oceanic and Atmospheric Administration (NOAA) are examined in an orthogonal space using PCA. Depending on the degree of the redundancy, the variances of the orthogonal system indicate the order of contribution of each mode shape to the total response. The minimum number of modes, without discarding important information carried in the original data, is used to construct a multiple linear regression model to predict the annual number of Atlantic hurricanes.

2. DATA

The data used in this analysis is briefly discussed in the following section. The Atlantic hurricanes' data are used in our analysis came from the HURDAT data-set (Jarvinen et al. 1984 and Jarrell et al. 1992). For the first part of the paper, hurricane records between 1900 and 2011 are used. In the second part, climate indices from 1951 to 2010 were obtained from the NOAA's Earth System Research Laboratory, Physical Sciences Division. This information is available at http://www.esrl.noaa.gov/psd/data/climateindices/list/.

Atlantic Meridional Mode (AMM) is the measure of sea surface temperatures and zonal and meridional components of the 10-meter wind field over the region 21S-32N and 74W-15E of the Atlantic Ocean.

Atlantic Multidecadal Oscillation (AMO) is a weighted average of sea surface temperatures in the North Atlantic, roughly from 0 to 70N.

El Niño - Southern Oscillation (ENSO) is the cycle of the year-to-year variations in sea surface temperatures, convective rainfall, surface air pressure, and atmospheric circulation that occur across the equatorial Pacific Ocean. El Niño and La Niña represent warm (positive) and cool (negative) extremes, respectively, in the ENSO cycle. ENSO directly influences the Atlantic hurricane season based on the changes in the vertical wind shear that accompanies ENSO phases.

North Atlantic Oscillation (NAO) is a measure of the pressure anomaly over the North Atlantic. The NAO loading pattern projection to the daily anomaly 500 millibar height field over 0-90°N is used to obtain the NAO Index. The negative phase of the NAO reflects above-normal heights and pressure across the high latitudes of the North Atlantic and below-normal heights and pressure over the central North Atlantic, the eastern United States and Western Europe. The positive phase of the NAO reflects an opposite pattern of height and pressure anomalies over these regions.

Tropical North Atlantic (TNA) is the anomaly of the average of the monthly sea surface temperatures from 5.5N to 23.5N and 15W to 57.5W of the Atlantic Ocean.

Caribbean SST Index (CAR) analyzes SST anomalies averaged over the Caribbean. These anomalies were smoothed by three months running mean procedure and projected onto 20 leading empirical orthogonal functions. In the next two sections of this paper, we focus on the analysis of these data.

3. AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA)

Figure 1 suggests that the mean of the Atlantic hurricane time series indicates a nonstationary behavior. The stationarity of a time series can also be interpreted based on an autocorrelation plot. Autocorrelation is the correlation of a data set with itself, lagged by 1, 2 or more periods. Autocorrelation function (ACF) is a set of autocorrelations with lags $1, 2, \cdots, k$. If the autocorrelation has a moderately high correlation for several time lags and it gradually decreases, the plot suggests that the time series is nonstationary. However, the autocorrelation plot of a stationary data drops to zero relatively quickly.

Based on the observation from Figure 2, ACF shows a significant correlation in lag 1 and drops quickly after that time. Although the hurricane frequency (1900 – 2011) in Figure 1 indicates that the time series is nonstationary, the behavior of ACF suggests that the time series is white noise and is reasonably stationary.

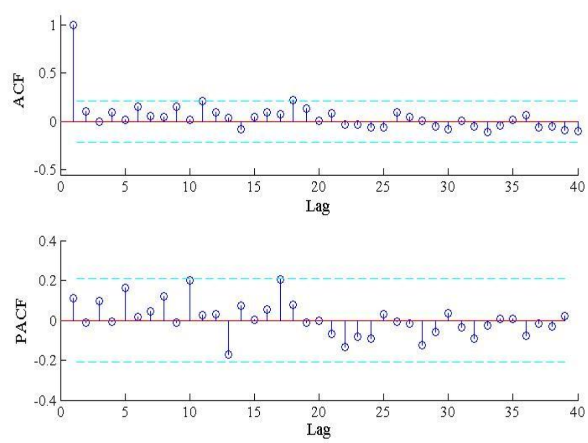

Figure 2: ACF and PACF of Atlantic hurricanes (original data) time series

Partial autocorrelation function (PACF) is a conditional correlation between two variables (Y_t and Y_{t+k}) when the effect of other time lags are removed

(Makridakis et al., 1998). PACF could determine the appropriate order of autoregressive in the ARIMA model. PACF in Figure 2 shows that the correlation measurements lie within the 95% confidence interval and suggests that the order of autoregressive is zero, $AR(0)$. However, Akaike information criterion (AIC) – a measure of the relative goodness of fit of a statistical model (Makridakis et al., 1998) – identifies $AR(1)$ as a better model compared to $AR(0)$. Given a set of candidate models for the data, the preferred model is the one with the minimum AIC value.

The pattern of ACF indicates that $MA(1)$ is an appropriate model. Therefore, based on ACF and PACF patterns $ARIMA(1,0,1)$ can be a good fit to the data-set. Figure 1 suggests using first order difference to account for nonstationarity. Therefore, $ARIMA(1,1,1)$ was also considered as an appropriate model to be fitted to the original data. Our analysis shows that the $ARIMA(1,1,1)$ has a smaller log-likelihood and AIC compared to $ARIMA(1,0,1)$. The three-parameter equation for $ARIMA(1,1,1)$ is as follows:

$$(Y_t - Y_{t-1}) - \varphi(Y_{t-1} - Y_{t-2}) = c + e_t - \theta e_{t-1}$$

Y_t denotes the number of hurricanes in year *t*, e_t is the estimated residual at year *t*, and *c*, *θ*, and *φ* are the three parameters of the model that must be re-estimated as the number of hurricanes of each successive year becomes available. Since a small difference (2 or less) in the AIC value is not substantial, using $ARIMA(1,1,1)$ suggests a better fitted model to the same data-set. Table 1 presents a comparison of $ARIMA(1,1,1)$ with Gray's forecast and the actual number of Atlantic hurricanes from NHC records.

Estimation of $ARIMA(1,1,1)$ parameters are done based on two different assumptions: 1) using all of the data from 1900 to 1989 and 2) separating warm phase (1942 – 1964) and cool phase (1965 – 1989) using AMO Index. The warm/cool phase intervals are defined based on analysis from Elsner et al. (2000).

Comparing the results of the first assumption and Gray's forecast; for each year that both forecast values are not equally distant from the observed value, the forecast value which is closest to observed value is highlighted. Considering the highlighted years it can be seen that, out of 19 years, 16 years Gray's forecasts show a better estimate of the hurricane activity than the ARIMA model. This suggests that a forecasting methodology which integrates meteorological impacts better captures the real phenomenon. It must be noted that engaging atmospheric conditions and increasing the number of explanatory variables in a forecasting model will increase the complexity of the model. If we compare the results from the first and second assumptions, out of 11 years that the predicted values are different, the second assumption provides better estimates in 7 years. Although Gray's forecast presents a better estimate overall.

Table 1: Atlantic hurricane original records and forecasts

Year	Actual Number	Gray's Forecast	$ARIMA(1,1,1)$	
			1st assumption	2nd assumption
1990	8	6	5	5
1991	4	3	6	6
1992	4	4	5	5
1993	4	6	5	5
1994	3	4	5	5
1995	11	9	5	6
1996	9	7	6	7
1997	3	6	6	7
1998	10	6	5	6
1999	8	9	6	7
2000	8	7	6	7
2001	9	7	6	7
2002	4	4	6	7
2003	7	8	6	6
2004	9	7	6	7
2005	15	10	7	7
2006	5	7	8	9
2007	6	8	7	8
2008	8	9	7	7
2009	3	4	7	7
2010	12	10	7	7
2011	7	9	7	7

4. PRINCIPAL COMPONENT ANALYSIS (PCA)

As shown in Figure 3, the annual Atlantic Basin hurricane observations are positively correlated with AMM, AMO, TNA, and CAR (0.49, 0.53, 0.55, and 0.40, respectively) and have negative correlations with ENSO ($R = -0.29$) and NAO ($R = -0.18$). The analysis shows that the number of hurricanes is poorly correlated with NAO. This is relevant since the NAO pattern generally has a greater influence on the location of the jet stream on the North Atlantic and also on hurricane tracks as opposed to being a hurricane development mechanism (Elsner et al. 2000).

Correlation analysis of climate indices shows that these climate indices have some degree of dependency

(Vimont and Kossin, 2007). AMM is influenced by a number of local climate conditions that all influence hurricane activity in Atlantic Basin. AMM is characterized by an SST gradient (Xie and Philander, 1994; Chang et al. 1997) and according to Chiang et al. (2002), Czaja et al. (2002) and references therein, the AMM can be excited by variation in NAO and ENSO. AMO is another candidate that affects the excitation of decadal AMM variability. The SST variability patterns in North Atlantic climate models suggest that a warm phase of the AMO strengthens Atlantic hurricane activity. AMM also correlates strongly with Atlantic hurricane activity on decadal and interannual time scales (Vimont and Kossin, 2007). Weaker AMM coincide with negative AMO years (Rumpf et al. 2010). The TNA Index can be considered as a smaller subset of the AMO and since the sea surface temperatures also influences air pressures, it is correlated with NAO. Approximately about 25% of the variation in the TNA can be explained by ENSO fluctuations (Hastenrath et al. 1987, Hameed et al. 1993, and Enfield and Mayer, 1997).

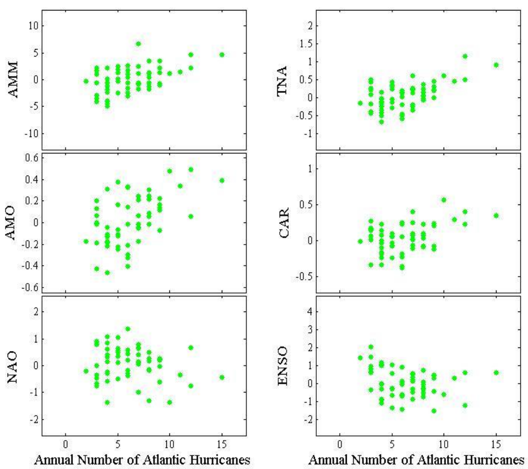

Figure 3: Climate indices vs. historical Atlantic hurricanes

If we assume that the annual number of hurricanes is a function of the above-mentioned six indices, any dependency between the indices exhibits a multicollinearity problem (Belsley, 1976). Multicollinearity does not actually bias results but if there are any other problems which could introduce bias, multicollinearity can increase the order of magnitude of the effects of that bias. In order to remove the redundancy among the indices, PCA is used to identify the skill of the combined set of indices in predicting the annual number of hurricanes. In the next section, such an approach is explained and the degree of success of prediction of basin events is discussed and compared.

PCA is based on the analysis of the covariance matrix, which the variances are diagonal elements of the matrix and the covariance values are off-diagonal terms. By dividing the covariance matrix by the variances the correlation matrix will be obtained which is the covariance matrix of the normalized variables. The eigenvalues and eigenvectors of the covariance matrix are then computed. Unlike the original data vectors, the eigenvectors are uncorrelated and orthogonal. The projection of the original data vectors onto the eigenvectors space yields the principal components. In general, fewer eigenvectors are required to sufficiently represent the data. This quantity of principal component analysis allows for data simplification and reduction.

In this paper, a climate matrix is generated considering six climate indices: AMM, AMO, NAO, ENSO, CAR, and TNA that influence the hurricane activity in the Atlantic Basin. The PCA is conducted to transform the climate matrix $[CI]$ ($N \times M$ matrix) to the normalized orthogonal space. *M* is the number of climate indices and *N* represents the number of observations. This transformation is achieved by performing a standard eigenvalue analysis on the covariance matrix as described above using the climate indices. The covariance matrix is defined by

$$[Cov] = \frac{1}{N-1}[CI]^T[CI]$$

If $[E]$ represents the normalized eigenvectors of $[Cov]$, then the coefficient matrix can be obtained by:

$$[A] = [E]^T[CI]^T$$

$[A]$ is an $M \times N$ matrix. The first element of each column of the coefficient matrix describes the contribution of the first mode; the second element defines the participation of second mode, and so on. The degree of importance of an eigenvector is related to the size of its eigenvalue and the larger the eigenvalue the more important the corresponding mode shape (eigenvector).

Given that substantial correlation exists among the oceanic/atmospheric data (indices), the information in the original data-set are redundant. In general, the first few eigenvectors of the covariance matrix will capture most of the variations. Figure 4 shows the relationship between each climate index and each principal component (mode shape) in normalized space. The AMM, AMO, TNA, and CAR are highly correlated with first principal component (0.80, 0.91, 0.95, and 0.87, respectively). ENSO is highly negatively correlated with the second mode (-0.92) as the NAO is with the third mode (-0.79).

For each climate index, anomalies during pre-hurricane-season months – from May through July – are used. The proportion of the total joint variation in the data represented by the first L modes can be computed from the eigenvalues using the following equation:

$$proportion\ of\ total\ variance = \frac{\sum_{m=1}^{L} \lambda_m}{\sum_{m=1}^{M} \lambda_m}$$

Where $\lambda_m$is the eigenvalue for mode m. The first three principal components define 88% of the joint variability in the climate indices in the normalized space. Figure 5 shows the percentage of the total variability as described by each principal component.

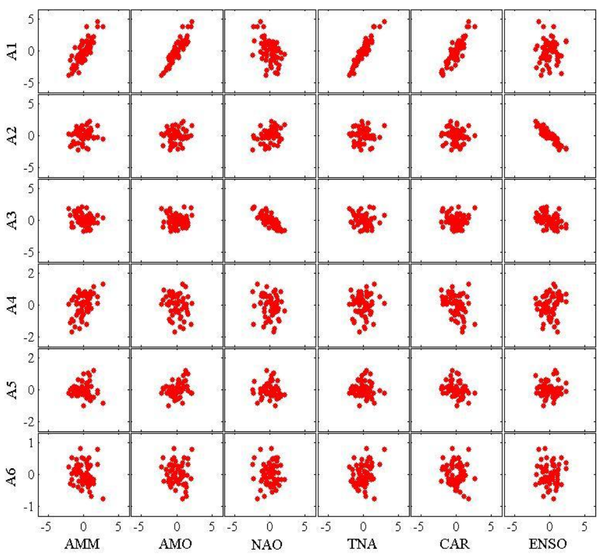

Figure 4: Correlation between climate indices and principal components (normalized space).

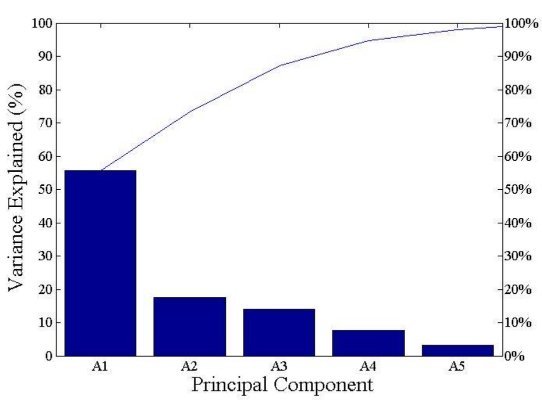

Figure 5: The proportion of total variance in the data represented by each mode

In the next step, the coefficient matrix associated with the first three principal components is used to estimate the annual number of hurricanes in the Atlantic Basin. As the principal components are uncorrelated, the first three modes which explain most of the variance are employed to perform multiple linear regression analysis.

$$Y = \beta_0 + \beta_1 A_1 + \beta_2 A_2 + \beta_3 A_3 + e$$

Where $Y = [Y_t]$ and Y_t is the number of hurricanes in year t. e (residual) is a random variable that is normally distributed with mean zero and a variance of σ_e^2 ($\sigma_e = 2.01$). In order to estimate the parameters of the regression function, "the ordinary least squares" method is used. This method obtains the parameters in which the sum of squared residuals is minimized.

$$SSE = \sum_{i=1}^{N} e_i^2$$

Figure 6 shows the correlation between the observed and predicted number of hurricanes for years between 1951 and 2010. The predicted number of hurricanes using the best fit regression model is relatively well correlated with the observed numbers (R = 0.64).

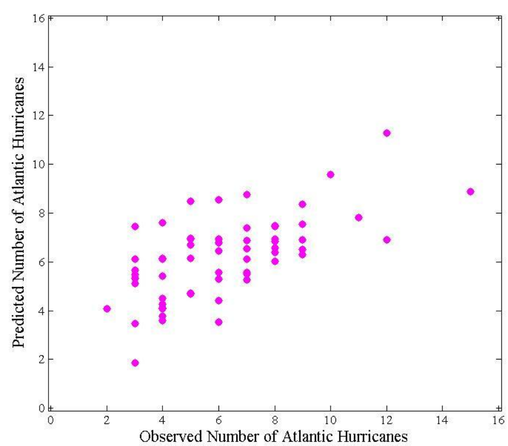

Figure 6: Correlation of observed and predicted number of Atlantic hurricanes

In order to compare the predicted results obtained from PCA using the first three principal components with Gray's forecasts as described in Table 1, we used the data-set from 1951 to 1989 to specify the regression parameters. The β_i $(i = 0, 1, 2, 3)$ parameters is re-estimated as the number of hurricanes of each successive year becomes available (least squared analysis has been applied dynamically following 1990).

The actual number of hurricanes, Gray's August forecasts and the predicted number of hurricanes (using the above regression model) from 1990 to 2010 are presented in Table 2. Out of 15 years in which one of the forecast methods did better than the other one, the PCA provides better estimates in 8 of those years. The corresponding number for Gray's is 7. The correlations between the actual and predicted hurricane numbers are shown in Figure 7. Considering the data from 1990 to 2010, the correlation coefficients are 0.79 using the PCA and 0.62 based on Gray's forecasts. 2011 prediction is not added as some of the indices used for PCA-based model is not currently available.

Table 2: PCA and Gray's forecast comparison

Year	Actual Number	Gray's Forecast	PCA
1990	8	6	6
1991	4	3	4
1992	4	4	5
1993	4	6	4
1994	3	4	4
1995	11	9	7
1996	9	7	6
1997	3	6	5
1998	10	6	9
1999	8	9	7
2000	8	7	6
2001	9	7	6
2002	4	4	4
2003	7	8	7
2004	9	7	7
2005	15	10	9
2006	5	7	7
2007	6	8	7
2008	8	9	8
2009	3	4	6
2010	12	10	11

5. CONCLUSIONS

In the first part of this paper, the Atlantic hurricane forecast results from $ARIMA(1,1,1)$ were compared to the forecast from Gray's model for years between 1990 and 2011. This analysis suggests that Gray's methodology demonstrates clear superiority compare to $ARIMA(1,1,1)$. This finding leads us to the second part of the paper which examined the viability of the new approach. Considering the contribution of climate fluctuations and their influence on hurricane activity in North Atlantic Basin and the possibility of redundancy among oceanic/atmospheric indices, we used a PCA-based approach along with multiple linear regressions.

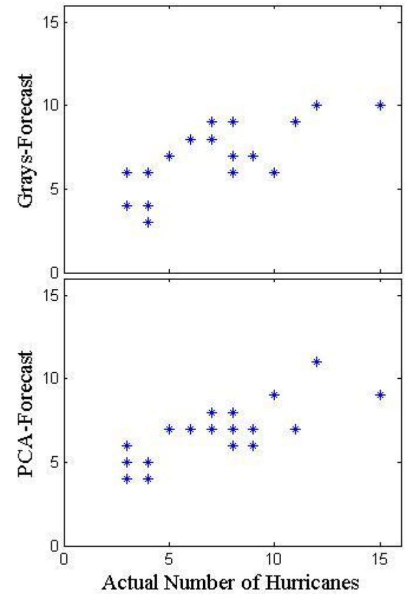

Figure 7: Correlation between Actual number of Atlantic hurricanes and estimated number of hurricanes using Gray's forecast and PCA methodologies.

The PCA indicates that the first three principal components can reasonably explain the variability existing in the original data. These three components were used to design a multiple linear regression model in order to estimate the hurricane numbers per year. The predicted number of hurricanes using the best fitted regression model designed based on the PCA shows a better correlation with historical records compared to Gray's forecasts.

REFERENCES

Anderson, J.R., and J.R. Gyakum, 1989: A diagnostic study of Pacific Basin circulation regimes as determined from extratropical cyclone Tracks. Mon. Wea. Rev., 117, 2672-268.

Belsley, D.A., 1976: Multicollinearity: Diagnosing its presence and assessing the potential damage it causes least squares estimation, working paper 154, Computer Research Center, National Bureau of Economic Research, Cambridge, Mass.

Chang P., L. Ji, and H. Li, 1997: A decadal climate variation in the tropical Atlantic Ocean from thermodynamic air-sea interactions, Nature, 385, 516-518.

Chiang, J. C. H., Y. Kushnir, and A. Giannini, 2002: Deconstructing Atlantic ITCZ variability: Influence of the local cross-equatorial SST gradient, and remote forcing from the eastern equatorial Pacific, J. Geophys. Res., 107(D1), 4004, doi:10.1029/2000JD000307.

Czaja, A., P. van der Vaart, and J. Marshall, 2002: A diagnostic study of the role of remote forcing in tropical Atlantic variability, J. Clim., 15, 3280-3290.

Daneshvaran S. and R. E. Morden, 1998: Fast hurricane wind analysis using the orthogonal decomposition method, Journal of Wind Engineering and Industrial Aerodynamics, 77 & 78, 703-714.

Elsner J. B., and C. P. Schmertmann, 1993: Improving extended-range seasonal predictions of intense Atlantic hurricane activity, Weather & Forecasting, 8, 345- 351.

Elsner, J.B., T. Jagger and X.F. Niu, 2000: Changes in the rates of North Atlantic major hurricane activity during the 20th century, Geophysical Research Letters, 27, 1743-1746.

Elsner, J.B., K.B. Liu and B. Kocher, 2000: Spatial variations in major U.S. hurricane activity: statistics and a physical mechanism, Journal of Climate, 13 (13), 2293-2305.

Enfield, D.B., and D.A. Mayer, 1997: Tropical Atlantic SST variability and its relation to El Niño-Southern Oscillation. J. Geophys. Res., 102, 929-945.

Goldenberg, S.B., Landsea, C.W., Mestas-Nuñez, A.M. and Gray, W.M., 2001: The recent increase in atlantic hurricane activity: causes and implications, Science, 293 (5529), 474-479.

Gray, W. M., 1984: Atlatnic seasonal hurricane activity. Parti: El Niño and 30-mb Quasi-Biennial Oscillation influences. Mon. Wea. Rev., 112, 1649-1668.

Gray, W. M., W. L. Christopher, P. W. Mielke Jr., and K. J. Berry, 1992: Predicting Atlantic seasonal hurricane activity 6–11 months in advance. Wea. Forecasting, 7, 440-455.

Hameed, S., K. R. Sperber, and A. Meinster, 1993: Teleconnections of the Southern Oscillation in the tropical Atlantic sector in the OSU coupled upper ocean-atmosphere GCM. J. Climate, 6, 487-498.

Hastenrath, S., L. C. de Castro, and P. Aceituno, 1987: The Southern Oscillation in the Atlantic sector. Contrib. Atmos. Phys., 60, 447-463.

Holmes, J. D., 1990: Analysis and synthesis of pressure fluctuations on Bluff bodies using eigonveltors, J. Ind. Aerodyn. , 33, 219-230.

Hoyos, C.D., Agudelo, P.A., Webster, P.J. and Curry, J.A., 2006: Deconvolution of the factors contributing to the increase in global hurricane intensity, Science, 312 (5770), 94-97.

Klotzbach, P.J., 2006: Trends in global tropical cyclone activity over the past twenty years (1986-2005), Geophysical Research Letters, 33, L10 805. (doi:10.1029/2006GL025881).

Lee, T. and P. Cornillon, 1995: Temporal variation of meandering intensity and domain-wide lateral oscillations of the Gulf Stream. J. Geophys. Res., 100, 13603-13613.

Mann, M.E. and Emanuel, K.A., 2006: Atlantic hurricane trends linked to climate change, EOS, Transactions American Geophysical Union, 87 (24), 233-244.

Powers, M. R., 2006: Catastrophe Forecasting: seeing “Gray” among the “Black Boxes”, Journal of Risk Finance, 7 (5), 458-462.

Ramsay, S. R., 1988: The interaction of a 2D turbulent wake with a bluff body, PhD Thesis, The University of Cambridge.

Rumpf J., H. Weindl, E. Faust, and V. Schmidt, 2010: Structural variation in genesis and landfall locations of North Atlantic tropical cyclones related to SST, Tellus, 62A, 243-255.

Trenberth, K.E., 2005: Uncertainty in hurricanes and global warming, Science, 308 (5729), 1753-1754.

Vimont D. J. and J. P. Kossin, 2007: The Atlantic meridional mode and hurricane activity, Geophysical research Letters, 34, L07709, (doi:10.1029/2007GL029683).

Webster, P.J., Holland, G.J., Curry, J.A. and Chang, H.R., 2005: Change in tropical cyclone number, duration and intensity in a warming environment, Science, 309 (5742), 1844-1846.

Xie S. P. and S.G. H. Philander, 1994: A coupled ocean-atmoshere model of relevance to the ITCZ in the eastern Pacific, Tellus, Ser. A, 46, 340-350.

Xie L., T. Yan, L. J. Pietrafesa, J. M. Morrison and T. Karl, 2005: Climatology and Interannual Variability of North Atlantic Hurricane Tracks, Journal of Climate, 18, 5370-5381.