15D.1 AN INTRASEASONAL MODEL FOR THE PREDICTION OF ATLANTIC AND EAST PACIFIC TROPICAL CYCLONE GENESIS

Stephanie A. Slade* and Eric D. Maloney
Colorado State University, Fort Collins, Colorado

1. INTRODUCTION

Based on the success of the Leroy and Wheeler (2008) intraseasonal prediction model of the Southern Hemisphere, a statistical model is developed via multiple logistic regression for the prediction of weekly tropical cyclone (TC) genesis over the East Pacific (90˚W-120˚W, 5˚N-25˚N) and Atlantic Ocean (15˚W-100˚W, 5˚N-50˚N) regions using data from 1975 to 2009 (excluding 1978). The predictors used in the model include a climatology of tropical cyclone genesis for each ocean basin, two indices representing the propagating Madden-Julian Oscillation (MJO), and an El Niño-Southern Oscillation (ENSO) index derived from the first principal component of sea surface temperature over the Equatorial Pacific region. Prediction models are generated out to a week 7 forecast lead for each basin. This paper briefly describes the results of the intraseasonal statistical models.

2. DATA AND PREDICTOR DEVELOPMENT

Tropical cyclogenesis data is obtained from the National Hurricane Center Hurricane best track (HURDAT) TC archives (Jarvinen et al. 1984). For the East Pacific basin, the official hurricane season of May 15-November 30 is used. For the Atlantic basin, a more restrictive subset of the official hurricane season, July 1-October 31, is used. In the case of the Atlantic, better results were yielded using a more active subset. To exclude weaker systems, only storms that reach a minimum of 34 knots, or tropical storm strength, are considered. The model of choice, discussed further in section 3, allows input observations to be dichotomous; since the desired probabilities are weekly, a 0 is assigned for any week where no storm formed and a 1 if at least one storm formed.

The Wheeler and Hendon (2004) real-time multivariate MJO (RMM) indices are used as predictors in this study to represent the eastward propagation of the MJO. These indices are derived from the first two empirical orthogonal functions (EOFs) of the near-equatorially averaged and normalized 200-mb and 850-mb zonal wind fields and satellite outgoing longwave radiation. The daily observed data is projected onto the computed EOFs to yield the two principal components (PCs), RMM1 and RMM2. The year 1978 is excluded from all analysis in this study due to missing RMM data, and hereafter assumed in any mention of the period 1975-2009.

* *Corresponding author address:* Stephanie A. Slade, Colorado State Univ., Dept. of Atmos. Sci., Fort Collins, CO 80523; e-mail: sslade1@atmos.colostate.edu.

The Hadley Centre Sea Ice and Sea Surface Temperature (HadISST) re-analysis 1˚ gridded dataset (Rayner et al. 2003) is used to develop an ENSO index. To formulate the predictor index, a 3-point running mean average is applied to the monthly SST data to reduce intraseasonal influence. The annual cycle and long term mean are removed and the data is detrended. The monthly data is then linearly interpolated to daily resolution. The first EOF is calculated via the covariance matrix for 30˚S- 30˚N, 70˚W-110˚E. The PC is standardized and used as the predictor index which represents the temporal variability of ENSO.

2.1 CLIMATOLOGY

A climatology predictor is generated for each basin based on the weekly stratified dichotomous genesis observations by averaging each week over all years. This results in a weekly probability that is then multiplied by 100 to depict a climatological percent probability of genesis. The raw climatology is smoothed by applying a weighted running average with a 1-2-1 filter four times for the Atlantic and twice for the East Pacific.

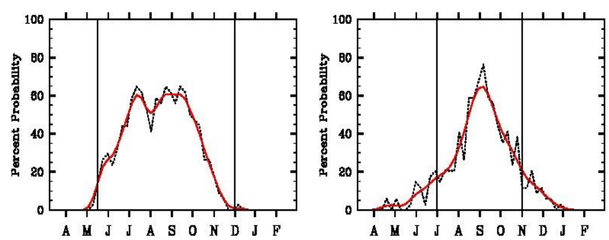

Figure 1: Climatology of TC genesis probability (%) for the East Pacific (left), and Atlantic (right). The black curve is a raw weekly stratified climatology. The red curve is a smoothed climatology, smoothed with a 1-2-1 filter. The x-axis ranges from March-February. Vertical lines represent time range boundaries used in the data.

3. LOGISTIC REGRESSION

The prediction model developed in this study employs multiple logistic regression, formulated as follows:

$$\hat{P}(x) = P(Y|x) = \frac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_m x_m}}{1+e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_m x_m}} \qquad (1)$$

Equation 1 represents the conditional probability that the dichotomous variable Y occurs given the independent variables, x_m , where m is the number of independent variables. The regression coefficients are represented by β_m. This form of regression is unique in that its output yields a probability between 0 and 1 (Hosmer and Lemeshow, 2000). Probability hindcasts using equation 1 are generated independently for each year from 1975-2009. Hindcasts are a form of cross-validation, utilizing regression coefficients calculated using the full time range while excluding the hindcasted years (Elsner and Schmertmann, 1994). In order to eliminate any possible bias in the model with a storm forming within a certain day of the week, a week is defined beginning on every day, resulting in overlapping weekly probabilities.

4. PREDICTOR SELECTION

Stepwise selection schemes are commonly used in regression as a basis for determining the importance of a variable, including it only if significant given a specific criterion. In this study a sequential forward selection scheme is used. Variables are sequentially included to an initially empty set until the addition of further variables no longer improves prediction, meaning it no longer decreases the criterion. In the model developed here, deviance is used as the criterion. The deviance is analogous to the residual sum of squares in linear regression and follows a chi-square distribution (Hosmer and Lemeshow, 2000).

An advantage of using a forward selection scheme is that it ranks the selected predictors by level of “importance” based on a significance criterion. This provides information on which predictors have the greatest statistical influence on tropical cyclogenesis in each basin for each lead. Using the full data set, table 1 lists the ranks given by the forward selection scheme for each basin at every lead.

Table 1: Predictor selection rank according to the forward selection scheme for each basin from a zero to a seven week forecast lead. Data from 1975-2009 is used. A “1” designates the first predictor chosen by the selection scheme, and so on. Spaces left blank indicate the predictor did not meet the criteria and therefore was not chosen.

Basin	Week Lead	Clim	ENSO	RMM1	RMM2
East Pacific	0	1	4	2	3
	1	1	4	3	2
	2	1	3	2	
	3	1	3	2	
	4	1	2	3	4
	5	1	2		
	6	1	2		
	7	1	2		
Atlantic	0	1	2	4	3
	1	1	2	3	4
	2	1	2	3	4
	3	1	2	3	
	4	1	2		
	5	1	2		3
	6	1	2	3	4
	7	1	2		

5. HINDCASTS

Tropical storm strength cyclogenesis probability curves are hindcasted and an example for each basin is shown out to a 2 week forecast lead (figure 2). Shown are 2002 for the East Pacific, characterized by strong MJO variability and moderate El Niño conditions, and 1975 for the Atlantic, characterized by strong MJO variability and moderate La Niña conditions. The RMM indices provide the short term variability, in the order of days to a week, in the probability hindcasts. ENSO, on the other hand, is slow varying on the order of months; it has the tendency to shift most of the full season hindcast curve. When the hindcasted probability has the shape of climatology (with either a different or the same amplitude), there is no small order variability, meaning neither RMM index was selected. This is fairly common in the longer forecast leads.

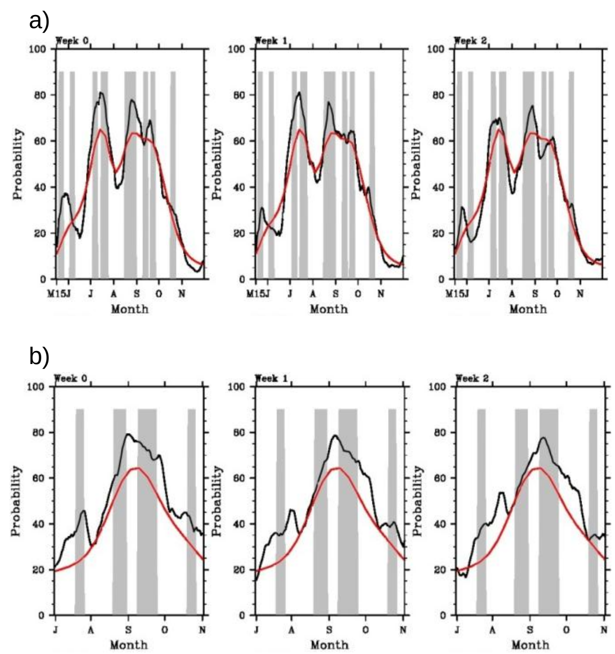

Figure 2: Percent probability hindcast (black curve) for a) the East Pacific 2002 season from May 15 - Nov 30, and b) the Atlantic 1975 season from July 1 - Oct 31. The climatology of each basin is overlaid in red for comparison. Gray bars are shown for weeks that underwent tropical storm (≥34 kt) cyclogenesis.

Most of the hindcasted variability in the 2002 East Pacific season occurs during the high amplitude MJO variability. The week 1 forecast lead shows a probability increase of ~20% in mid-July as the MJO approaches phase 8 based on the RMM indices of Wheeler and Hendon (2004). Observations (gray bars) tend to coincide with high peaks in probability. The moderate El Niño had little effect on the probability curves.

A moderate La Niña event in 1975 caused a 5-10% increase in genesis probabilities throughout the season for the Atlantic basin. MJO activity is responsible for the shorter time scale variability in probabilities, increasing mid-late July probabilities by an additional 7% during forecast lead weeks 1 and 2. This increase is associated with phase 2 of the MJO, known to create favorable conditions for cyclogenesis (Klotzbach, 2010). Also observed are probability increases in the month of September and in late October corresponding to phases 1 and 2 of the MJO.

6. BRIER SKILL SCORES: TROPICAL STORMS

The most common method for verification of dichotomous events is the Brier score (BS), defined by the equation:

$$BS = \frac{1}{n}\sum_{i=1}^{n}(y_i - o_i)^2 \qquad (2)$$

where y_i represents the forecasted probability between 0 and 1 for event i out of a total of n events (Wilks, 2006). The observations, o_i, are dichotomous where $o_i = 1$ if the event occurred and $o_i = 0$ if it did not occur. The Brier score calculation in essence calculates the mean-squared error of the forecasted probability. Since both the forecasts and observations are bounded by 0 and 1, so is the Brier score. A Brier score of 0 denotes a perfect forecast, meaning o_i always equals y_i. In contrast, a Brier score of 1 indicates that the forecast is wrong for every event. Using the Brier score calculated, a Brier skill score is then computed:

$$BSS = \frac{BS - BS_{ref}}{0 - BS_{ref}} = 1 - \frac{BS}{BS_{ref}} \qquad (3)$$

where BS is the Brier score calculated from the forecasts and BS_{ref} represents a reference forecast Brier score. This study utilizes a seasonal mean climatology to calculate the reference Brier score. The Brier skill scores are multiplied by 100 to denote a percentage and are shown in figure 3. This is ideal as it allows the Brier skill score to represent a percent decrease in mean squared error (by the forecasts) over a mean seasonal climatology.

The East Pacific climatology Brier skill score alone shows over a 15% improvement over using a mean seasonal climatology. Including ENSO as a predictor adds only an additional small fraction of a percent improvement without much change over week lead. Including the MJO to climatology improves the skill by almost an additional 1.5% during weeks 1-2, with further slight improvement given the predictor selection scheme; these tend to decrease towards the climatology model skill over longer forecast leads. Overall, including all selected predictors improves the skill of the model by almost 17% at the shortest leads.

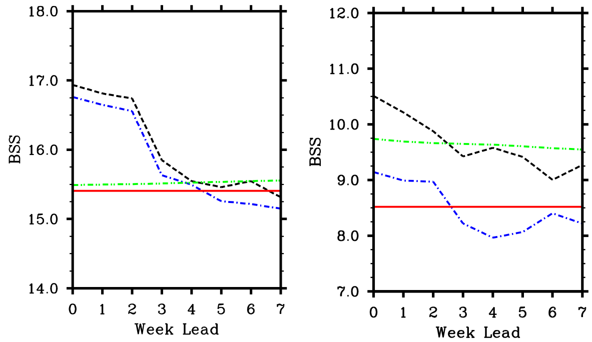

Figure 3: Brier skill scores (%) for the East Pacific (left) and Atlantic (right). Shown are Brier skill scores for the stepwise scheme selected predictors (dashed black), MJO + climatology (dashed blue), ENSO + climatology (dashed green), and climatology only (solid red). Skill scores are calculated using a seasonal mean reference climatology.

The Atlantic climatology Brier skill score shows an improvement of approximately 8.5% over the model using a mean seasonal climatology. The addition of the MJO to climatology shows a skill increase out to forecast lead week 2 of around 0.5%. The Brier skill score of the model using ENSO + climatology generates and improvement greater than 1% over climatology alone, decreasing only slightly at the longest lead times. Including all selected predictors, there is a skill improvement of almost 2% over the model using only climatology, this skill generally decreasing over forecast lead. Overall, the model of the selected predictors improves the skill of the model by almost 10.5% at the shortest leads.

7. RELIABILITY DIAGRAMS

Testing the reliability of a model is commonly done via reliability diagrams. This is done by binning the dichotomous genesis observations and the hindcasted probabilities according to the hindcasted probability. For the East Pacific, observations and hindcasted probabilities are binned into 20 groups of approximately 340 values each. For the Atlantic the bins consist of 17 groups of roughly 246 values each. For each group the bins are averaged and plotted to form the reliability curve. When the reliability curve lies above (below) the perfect diagonal, the forecast is underestimated (overestimated). Reliability curves are highly influenced by the tendency of tropical cyclone observations to follow climatology, while the other predictors work to increase skill and "tighten" the reliability curve about the perfect forecast.

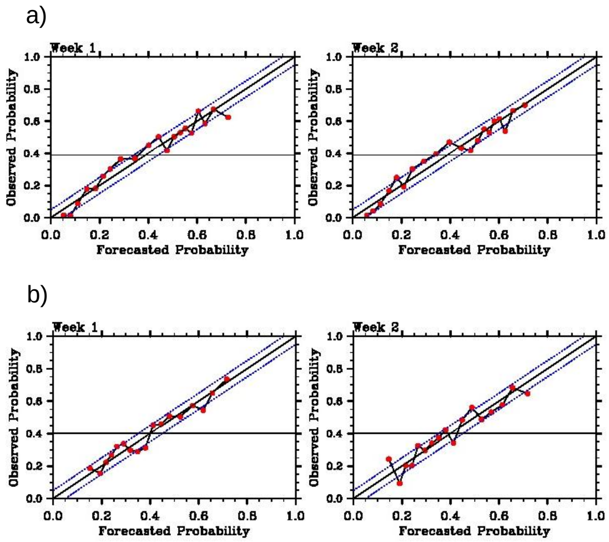

Figure 4: a) East Pacific and b) Atlantic reliability diagrams for forecast weeks 1 and 2. Binned averages are portrayed by red circles. The line connecting the circles forms the reliability curve. The perfect forecast and a 10% interval about the perfect forecast is shown by the solid black and blue dotted diagonal lines, respectively. The solid horizontal line indicates the average observed probability.

8. CONCLUSIONS AND FUTURE WORK

This study proposes an intraseasonal prediction model for tropical cyclone genesis in the East Pacific and Atlantic Ocean basins based on multiple logistic regression. Predictors used include ENSO, the MJO, and a climatology of TC genesis for each basin. A prediction model for each basin is generated out to a week 7 forecast lead. After undergoing a forward selection scheme process, the predictors selected generate regression coefficients which are used to produce hindcasts for each year from 1975-2009.

Brier skill scores and reliability diagrams were generated to determine the skill and dependability of the models. Results show an increase in model skill at predicting tropical cyclogenesis by the inclusion of the MJO out to a three week forecast lead for the East Pacific and a two week forecast lead for the Atlantic. When only considering storms that reach hurricane strength (not shown), the inclusion of the MJO in the Atlantic models show further increase in skill out to a 3 week forecast lead, with similar skill improvements above a time-varying climatology as ENSO for a week 1 forecast lead. Including ENSO increased the skill of the tropical storm Atlantic model significantly out to a 7 week forecast lead, while only slightly improving the skill of the East Pacific model.

Future work will focus on improving forecasting skill and forecast lead times, along with the inclusion of other predictors. Predictors under consideration include an index of West African monsoon intraseasonal variability, the North Atlantic Oscillation, the Atlantic Multi-decadal Oscillation, the QBO, and the Pacific Decadal Oscillation.

Ultimately, the goal of this study and the future work discussed is a successful operational real-time forecasting model. Once additional predictors have been implemented in the model, further work may also get at a better understanding of the physical basis behind some of the forecast relationships shown.

9. REFERENCES

Elsner, J. B., and C. P. Schmertmann, 1994: Assessing forecast skill through cross validation. *Wea. Forecasting*, **9**, 619-624.

Hosmer, D. W. and S. Lemeshow, 2000: Applied Logistic Regression. 2nd ed. Wiley, 373 pp.

Jarvinen, B.R., C.J. Neumann, and M.A.S. Davis, 1984: A tropical cyclone data tape for the North Atlantic basin, 1886-1983. NOAA Tech. Memo. NWS-NHC 22, 21 pp.

Klotzbach, P. J., 2010: On the Madden–Julian Oscillation–Atlantic Hurricane Relationship. *J. Climate*, **23**, 282–293.

Leroy, A., and M. C. Wheeler, 2008: Statistical Prediction of Weekly Tropical Cyclone Activity in the Southern Hemisphere. *Mon. Wea. Rev.*, **136**, 3637–3654.

Rayner, N. A.; Parker, D. E.; Horton, E. B.; Folland, C. K.; Alexander, L. V.; Rowell, D. P.; Kent, E. C.; Kaplan, A., 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, *J. Geophys.* Res, **108**, No. D14, 4407.

Wheeler, M. C., and H. H. Hendon, 2004: An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. *Mon. Wea. Rev.*, **132**, 1917-1932.

Wilks, D. S., 2006: Statistical methods in the atmospheric sciences. Academic Press, 648 pp.