12A.2 VERIFICATION OF THE NATIONAL HURRICANE CENTER'S EXPERIMENTAL PROBABILISTIC TROPICAL CYCLONE GENESIS FORECASTS

Daniel P. Brown, James L. Franklin and Jamie R. Rhome
NOAA/NWS/NCEP/Tropical Prediction Center, Miami, Florida

1. INTRODUCTION

The National Hurricane Center (NHC) routinely issues Tropical Weather Outlooks (TWO) for both the Atlantic and eastern North Pacific basins. The TWOs are text products that discuss areas of disturbed weather and their potential for tropical cyclone development during the next 48 hours. In 2007, the NHC began producing in-house experimental probabilistic tropical cyclone genesis forecasts. Forecasters subjectively assigned a probability of genesis (0 to 100%, in 10% increments) to each area of disturbed weather described in the TWO. These forecasts were verified based on the time of genesis determined from NHC "best-track" data.

2. BACKGROUND AND METHODOLOGY

Pasch et al. (2002) noted that recent advancements in operational numerical modeling have led to increased skill in the prediction of tropical cyclone (TC) formation. These advancements led the Tropical Analysis and Forecast Branch (TAFB) of the Tropical Prediction Center (TPC) to begin issuing TC genesis forecasts in textual and graphical marine forecasts during the 2001 and 2002 hurricane seasons, respectively. TAFB forecasters would routinely collaborate with the NHC duty hurricane specialist to determine areas of "possible" formation within the next 36 h. The results of these forecasts (2002-2006) were quite favorable, with the probability of detection 80% or higher for both the Atlantic and eastern Pacific basins. The false alarm rates ranged from 30% for the eastern Pacific to 38% for the Atlantic.

The encouraging results led to ideas of enhancing the NHC TWO. One proposed enhancement included a graphical version of the text TWO (Rhome and Brown 2007, Rhome et al. 2008) with the possibility of providing categorical (low, medium, and high) forecasts of TC development as part of the graphic. To determine the feasibility of issuing categorical genesis forecasts, a verification study was preformed on the 2005-06 NHC TWOs (Brown and Rhome 2007). For that study a low, medium, or high chance of development was subjectively assigned to each disturbance mentioned in every TWO issued during the 2005 and 2006 hurricane seasons. The categorical assignments were based on wording used in the text TWO (e.g., low=development not expected, medium=some development possible, high=tropical depression or cyclone could from). The results from this study (Tables 1 and 2) provided evidence that a three-tiered categorical approach could succeed.

Corresponding author address: Daniel P. Brown, NOAA/NWS/NCEP/TPC/, 11691 SW 17th Street, Miami, FL 33165; e-mail: Daniel.P.Brown@noaa.gov.

Table 1. 2005-06 Atlantic basin TWO verification results. [Note the large separation between categories.] (From Brown & Rhome 2007)

	# Develop	# Forecasts	Percent
Low	20	753	2.7%
Medium	111	544	20.4%
High	129	243	53.1%

Table 2. 2005-06 east Pacific TWO verification results. [Note the large separation between categories.] (From Brown & Rhome 2007)

	# Develop	# Forecasts	Percent
Low	4	188	2.2%
Medium	108	379	28.5%
High	149	235	63.4%

In an attempt to validate these findings in an operational setting, the NHC produced experimental probabilistic genesis forecasts during the 2007 hurricane season. The NHC Hurricane Specialists subjectively assigned a probability of TC genesis from 0 to 100% in 10% increments, to each area of disturbed weather described in the TWO. The assigned probabilities represented the NHC forecaster's subjective determination of the chance of TC formation during the ensuing 48 h period. To assign probabilities, forecasters used their experience and knowledge of TC genesis, along with numerical model output, Dvorak (1975, 1984) satellite intensity estimates, and in-situ and remotely sensed observations. Additional products such as the Tropical Cyclone Formation Probability Product from the National Environmental Satellite, Data and Information Service (NESDIS) (Demaria et al. 2001), which provides a probability of genesis within 24 h in 5° by 5° latitude/longitude areas, and the Statistical Hurricane Intensity Prediction Scheme (SHIPS) were also used as TC genesis guidance.

The forecasts were verified based on TC development from the final NHC best-track data (Jarvinen et al. 1984).

3. ACCURACY OF THE 2007 GENESIS FORECASTS

Figure 1 is a reliability diagram showing the forecast genesis percentage versus the observed frequency for the Atlantic basin. The sample size for the Atlantic basin ranges from 197 at 10% to 1 at 100%. The lead time shown in the diagram is the average time between forecast issuance and genesis for all systems mentioned in the TWO that eventually become a tropical cyclone. The longest lead time for an Atlantic system was about

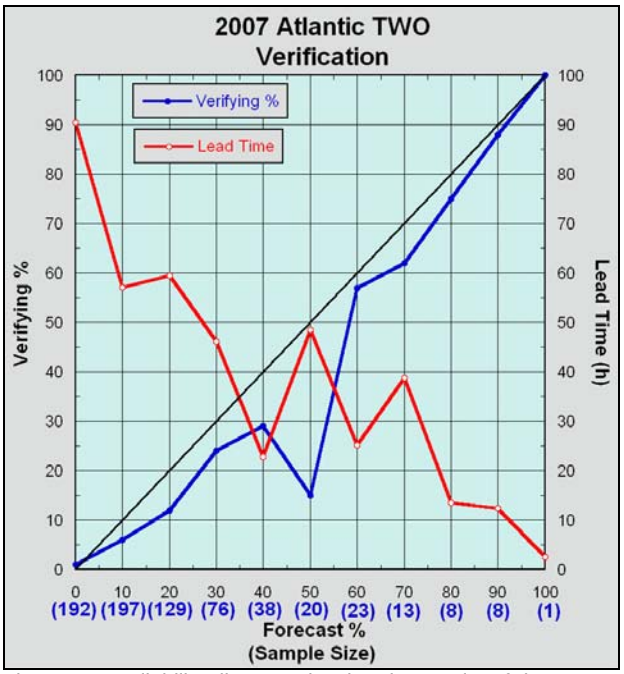

Figure 1. Reliability diagram showing the results of the 2007 NHC Atlantic basin probabilistic TC genesis forecasts. The blue line show the verification of the probabilistic forecasts, while the black diagonal line represents a perfect (x=y) relationship. The red line depicts the average lead time for forecasts made on systems that eventually became TCs.

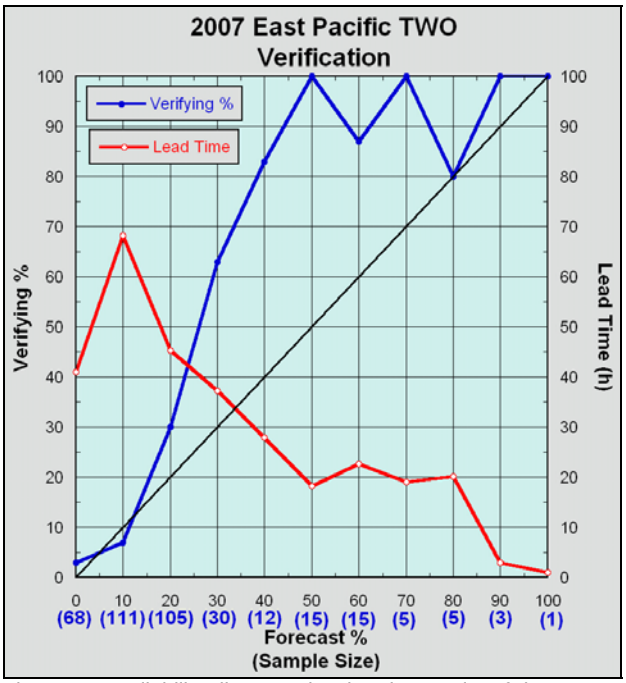

Figure 2. Reliability diagram showing the results of the 2007 NHC Eastern Pacific basin probabilistic TC genesis forecasts. The blue line show the verification of the probabilistic forecasts, while the black diagonal line represents a perfect (x=y) relationship. The red line depicts the average lead time for forecasts made on systems that eventually became TCs.

147 h for a forecast of 0% on the disturbance that became Erin on 15 August. The average lead time ranged from 2.5 h for 100% to 90 h for 0%.

The Atlantic basin results are very favorable, with only a small over-forecast bias. For example, when forecasters predicted a 30% chance of TC formation, genesis occurred within 48 h, 24% of the time. The bias is quite small for forecasts of 60% or greater. The forecast percentage that yields the greatest difference between the forecast percentage and observed frequency is 50%. The observed frequency of development for the 50% forecasts is 15%. The sample size at this percentage is relatively small compared to the surrounding bins (40% and 60%) and should be considered when assessing the poor results.

Figure 2 is the corresponding reliability diagram for the eastern North Pacific basin. The sample size for this basin ranges from 111 forecasts at 10% to 1 at 100%. The average lead time of verifying forecasts ranged from 1 h at 100% to 68 h at 10 %. The longest lead time for a system was 149 h for a 30% forecast for Flossie. The forecasts for the eastern Pacific exhibited a significant under-forecast bias. The observed frequencies of TC development were nearly twice as large as the forecast percentages predicted, for forecasts of 30% or greater. Only 5 forecasts of 40% or greater did not become a TC within 48 h, which resulted in an observed frequency of 91% for forecast probabilities of 40% or higher. The reason for the under-forecast bias is being investigated.

The results suggest that NHC has skill in predicting genesis in the Atlantic, but less-skill in the eastern Pacific. Since the sample sizes were relatively small and the forecast performance in the eastern Pacific was low, an attempt was made to bin the forecasts into three categories. A couple of different binning options were examined to determine which combination would result in a distinct three-tiered (low, medium, and high) categorization that could be used to enhance the NHC TWO.

Since the Atlantic forecast results were favorable across the spectrum, subtle changes to the bin boundaries did not appreciably change the three-tiered categorization results. In the east Pacific, however, slight changes in the binning led to significant differences in the verifying frequency of occurrence. The binning that provided the most distinct three-tiered categorization was as follows: 0-10% (low), 20-50% (medium), 60-100% (high). The results for this binning scheme are shown in Tables 3 and 4 for the Atlantic and east Pacific, respectively. The tables include both the expected percentage (mean of all the forecast probabilities within the bin) and the verifying percentage (frequency of occurrence).

Even though the binning produced very different verifying frequencies between the basins, the binning provides a clear distinction between categories in each basin. The low category verified less than 10% of the time in both basins and the high category verified more than

60% of the time for both basins. The largest difference between the basins occurred in the medium category, which in the Atlantic verified just below the lower threshold of the category (20-50%) while in the eastern Pacific it verified near the upper-end of the 20-50% range.

Table 3. Atlantic basin binned forecast reliability table. The expected percentage is an average of all the forecasts probabilities within the bin range.

	Expected %	Verifying %	Number of Fcsts.
0-10% (Low)	5	3	389
20-50 % (Med.)	28	18	263
60-100% (High)	71	66	53

Table 4. East Pacific basin binned forecast reliability table. The expected percentage is an average of all the forecasts probabilities within the bin range.

	Expected %	Verifying %	Number of Fcsts.
0-10% (Low)	6	6	179
20-50 % (Med.)	26	47	162
60-100% (High)	70	90	29

4. Color-Coding of bins for Graphical TWO

The large distinction in verification rates between the categorical bins has resulted in plans to enhance the experimental graphical TWO. In 2008, NHC forecasters are planning to again produce in-house probabilistic genesis forecasts. However, based on the 2007 binned results, the NHC is planning to release experimental categorical (low, medium, and high) forecasts to the public. The categorical forecast will be based on the bins shown in Tables 3 and 4. The low, medium, and high forecast categories will be represented by yellow, orange, and red, respectively on the graphical TWO (Rhome et al. 2008). To illustrate the color binning, Figures 3 and 4 show examples of the 2007 probabilistic forecasts within 48 h of tropical cyclone formation in the Atlantic (Table 5) and east Pacific (Table 6). The actual probabilistic values are shown in the table, while the shaded color represents the categorical forecast (low, medium or high). Green fill indicates that no forecast was made (i.e., the system was not mentioned in the TWO) and squares that are not filled indicated that no TWO was issued, since it was outside the hurricane season (except as discussed in Fig. 3).

The average forecast percentage about 12 h before genesis was 68% in the Atlantic and 39% in the east Pacific. About 24 and 48 h prior to genesis the percentages were 40% and 34%, respectively for the Atlantic and 35% and 25% for the east Pacific. The average percentage increased in each basin for each 6 h period as genesis approached. One motivation to examine the binned categorical forecasts and create Figures 3 and 4 is to determine the amount of "color-flashing" or forecast to forecast consistency. From the tables, one can conclude that this was not a significant problem. In most cases, a natural progression from low to medium or medium to high occurred and there were only a few examples of forecasts that "flashed" or progressed out of sequence.

	-44.5	-38.5	-32.5	-26.5	-20.5	-14.5	-8.5	-2.5
Andrea						DSA	DSA	DSA
Barry					DSA			20
Chantal	30	30	30	30	30	40	50	50
Dean		20	40	60	60	80	70	60
Erin	30	30	60	70	70	70	90	90
Felix	30	30	30	30	40	40	60	90
Gabrielle	60	50	20	20	20	20	90	90
Humberto	30	40	40	40	30	40	40	80
Ingrid	0	10	20	10	20	20	30	30
Ten	70	70	70	80	70	80	90	90
Jerry	10	10	10	0	10	10	10	40
Karen	40	60	60	60	60	80	80	100
Lorenzo	20	20	20	30	60	60	60	80
Melissa		20	20	20	20	10	10	10
Fifteen								
Noel	50	50	40	30	30	30	80	80
Olga				DSA	DSA	DSA	DSA	DSA
Avg.	**34**	**34**	**35**	**37**	**40**	**45**	**58**	**65**

Figure 3. The NHC experimental probabilistic forecasts within 48 h of genesis for the 2007 Atlantic basin TCs. The columns represent the approximate number of hours prior to TC genesis. The color-fill is based on the three-tiered binning planned for 2008 [Low (<20%)= yellow, Orange (20-50%)= medium, high (>50%)= red]. Note that Andrea and Olga developed outside of the hurricane season when TWOs are not issued. In some of these cases a Special Tropical Disturbance Statement (DSA) was issued to discuss the system. Barry developed 12 h after the season began, therefore only one forecast was issued. The dark green fill indicates that the system's precursor disturbance was not mentioned in the TWO.

	-43	-37	-31	-25	-19	-13	-7	-1
Alvin	30	40	50	70	60	60	50	90
Barbara	20	20	10	30	30	30	50	100
Three-E	10	0	0	10	30	30	30	30
Four-E	60	70	80	70	50	40	30	30
Five-E	10	40	30	30	40	60	60	60
Cosme	10	0	20	20	20	20	20	20
Dalila			10	10	20	20	40	70
Erick	30	20	20	20	20	20	20	20
Flossie	50	50	50	60	40	30	30	50
Gil	20	20	20	20	30	40	50	60
Henriette			20	20	20	30	50	50
Ivo	20	40	50	50	50	40	50	80
Thirteen		10	20	20	20	20	30	30
Juliette		10	20	40	60	80	90	90
Kiko	20	20	60	60	60	60	70	80
Avg.	25	26	31	35	37	39	45	57

Figure 4. As in Figure 3, except for the eastern North Pacific.

6. CONCLUSIONS

Advancements in numerical modeling have led to increased skill in the prediction of TC genesis. These advancements have allowed TPC to enhance their operational TC genesis products during the past several years. In 2007, NHC began producing experimental

probabilistic TC genesis forecasts and verification of these forecasts provided some promising results. The Atlantic basin forecasts were more successful than those for the eastern North Pacific, where a significant under-forecast bias was noted. Binning the probabilistic forecasts provided three distinct categories (low, medium, and high). The differences in genesis rates shown between the bins will allow NHC to begin issuing these categorical forecasts to the public in 2008 as part of the experimental graphical TWO.

7. REFERENCES

Brown, D. P., and J. R. Rhome, 2007: Verification of the National Hurricane Center's Atlantic and Eastern Pacific Tropical Weather Outlooks. 61st I*nterdepartmental Hurr. Conf.*, New Orleans.

Dvorak, V.F., 1975: Tropical cyclone intensity analysis and forecasting from satellite imagery. *Mon. Wea. Rev.,* **103**, 420-462.

-----, 1984: Tropical cyclone intensity analysis using satellite data. NOAA Tech. Rep. NESDIS 11, 47 pp.

DeMaria, M., J. A. Knaff and B. H. Connell, 2001: A Tropical Cyclone Genesis Parameter for the Tropical Atlantic. *Wea. and Forecasting*, **16**, 219-233.

Jarvinen, B.R., and Charles Neumann, and Mary Davis, 1984: A Tropical Cyclone Data Tape for the North Atlantic Basin 1886-1983: Contents, Limitations, and Uses. NOAA Tech. Mem. NWS NHC 22.

Pasch, R.J., J-G Jiing, F.M. Horsfall, H-L Pan, and N. Surgi, 2002: Forecasting Tropical Cyclogenesis in the NCEP Global Model. *Preprings, 25th Conf. Hurr. Trop. Meteor.*, San Diego, Amer. Meteor. Soc., 178-179.

Rhome, J.R. and D.P. Brown, 2007: Exploring the Concept of a Graphical Tropical Weather Outlook. *61st Interdepartmental Hurr. Conf.*, New Orleans.

Rhome, J.R. et al., 2008: The Experimental Graphical Tropical Weather Outlook: 2007 Results and Changes for 2008. *Preprints 28th Conf. Hurr. Trop. Meteor.*, Orlando, Amer. Meteor. Soc.