P1.5 THE DETERMINATION OF OPTIMAL THRESHOLDS OF TROPICAL CYCLONE INCREMENTAL WIND SPEED PROBABILITIES TO SUPPORT EXPRESSIONS OF UNCERTAINTY IN TEXT FORECASTS

Pablo Santos[1], Mark DeMaria[2], David Sharp[3], Scott Kiser

1. INTRODUCTION

As tropical cyclone events unfold, decision-makers require a meteorologist's most likely wind speed forecast, along with an accompanying expression of uncertainty. Both are necessary to effectively manage preparations for life-threatening weather events. The inherent uncertainty in tropical cyclone forecasts reveals the limitation of deterministic-only wind speed forecasts such as those found within the current Zone Forecast Product and Coastal Waters Forecast issued by National Weather Service (NWS) Weather Forecast Offices (WFOs). To address this limitation, WFOs Miami and Melbourne have developed a means to consistently and coherently incorporate uncertainty information in these text products through the creative use of the National Hurricane Center's incremental wind speed probabilities.

Since 2006, the National Hurricane Center (NHC) has produced operational gridded tropical cyclone wind speed probabilities for 34-, 50-, and 64-kt winds through 120 hours during operational forecast cycles for active systems in the Atlantic and Pacific Basins. The probabilities are centered about NHC's official track, intensity, and wind radii forecast, and incorporate error distributions over recent years for those variables (Gross *et al.,* 2004; Knaff and DeMaria 2005, 2007, DeMaria *et al.*, 2009). Since probability information is often designed to answer specific questions, these probabilities are produced in several forms to include the *cumulative* (the probability that wind speeds will reach or exceed 34/50/64 kt between the 00 and HH hour forecast), *interval* (the probability that 34/50/64 kt winds or greater will begin during the 12 hour forecast period ending at hour HH), and *incremental* (the probability that 34/50/64 kt winds or greater will occur during the 12 hour period between forecast hours HH-12 and HH) forms for each successive period of the forecast. Then, for the application described by Santos *et al.* 2008, the incremental probabilities are configured locally, in gridded form, to match the traditional time increments of the textual public and marine forecasts. Together with gridded hazard information (e.g., tropical storm/hurricane watches or warnings) and gridded deterministic wind speed information, the tropical cyclone incremental wind speed probabilities trigger enhanced wording, which conveys the situational uncertainty for successive forecast periods. The logic has been encoded within tropical cyclone versions of the respective text formatters which invoke the prescribed expressions. This includes the zone-based forecasts which are used to generate the legacy products, and the dynamic point-n-click (point-based) versions found on WFO web sites.

The aforementioned concept has been applied experimentally in selected offices since the 2006 season and it has depended, in part, on the exceedance of preliminary incremental wind speed probability thresholds as a function of time (e.g., forecast period) for uncertainty involving tropical storm force winds and hurricane force winds. Testing has yielded encouraging results (Sharp *et al*. 2006; Santos *et al*. 2008). Yet, since the incremental wind speed probability thresholds are a critical component to the formatter logic, it is necessary to validate them. Initial validation efforts were performed using Reliability Diagrams to determine whether the probabilities themselves are operationally reliable. Then, to formally identify those values which tend to maximize the responsible detection of a potential event while minimizing false alarms (that is, the thresholds that most closely separate the events from the non-events while accounting for uncertainty) optimal thresholds were calculated using two different metrics: relative operating characteristics (ROC) diagrams and threat scores (TS). This paper presents the validation results of the stated incremental wind speed probability thresholds, identifying those optimal values as a function of time for coastal locations stretching across the U.S. Gulf of Mexico and Atlantic coasts.

2. METHODOLOGY

To compute the validation metrics, 2004-2008 cases with tropical storm and/or hurricane warnings along the U.S. Gulf of Mexico and Atlantic coasts were used. The cases used are illustrated in Figure 1. All probability runs starting 3 days before the first warning was issued were included in the analysis for a total of 401 forecast cases. The probabilities were evaluated at 343 U.S. coastal breakpoints. NHC best track was used to determine points with observed tropical storm and/or hurricane force winds. The validation metrics were computed for all points combined as well as for sub-

[1] *Corresponding author address*: **Pablo Santos**, NOAA/National Weather Service, 11691 SW 17th Street, Miami, FL, 33165; e-mail: pablo.santos@noaa.gov.
[2] *Corresponding author address*: **Mark DeMaria**, NOAA/NESDIS Regional and Mesoscale Meteorology Branch; e-mail: Mark.DeMaria@noaa.gov.
NOAA/National Weather Service, 11691 SW 17th Street, Miami, FL, 33165; e-mail: pablo.santos@noaa.gov.
[3] *Corresponding author address*: **David Sharp**, NOAA/National Weather Service, 421 Croton Rd, Melbourne, FL, 32935; e-mail: david.sharp@noaa.gov.

groups of those points representing the following regions: *Gulf of Mexico* (from Brownsville, TX to Mobile, AL); *Florida Peninsula* (from Mobile, AL to the GA/SC border); *Mid-Atlantic and Northeast* (from the GA/SC border to Eastport, ME).

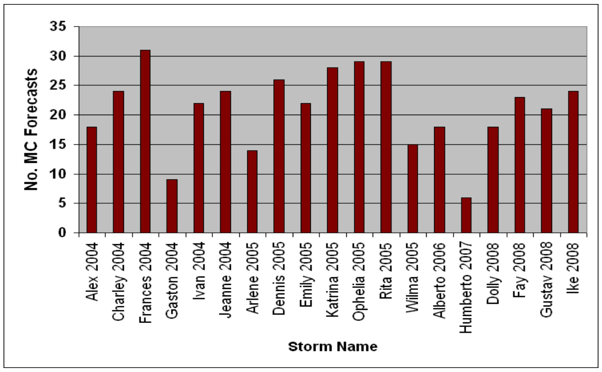


Figure 1. Storm cases used in this study for the validation of the incremental wind speed probability forecasts using the 2008 version of the Monte Carlo (MC) model (DeMaria *et al.*, 2009; Knaff and DeMaria 2005, 2007). A total of 401 MC model runs were used.

As previously stated, the validation metrics used in this analysis consisted of reliability diagrams, relative operating characteristic (ROC) diagrams and threat scores. A reliability diagram consists of a plot of observed frequencies of an event versus the forecast probability. For simplicity this diagram was created in this study using all forecast cases and periods combined. In practical terms, probability forecasts become more reliable as they approach a one to one (i.e., perfect) relationship between the observed frequencies and predicted probabilities.

The other set of validation metrics aimed at finding probability thresholds as a function of time period that yield the maximum hit ratio (also known as probability of detection or POD) and minimizes false alarm ratio or FAR. The main intent is to find those probability thresholds that discriminate between events and non events while accounting for uncertainty. To accomplish this, probability thresholds were evaluated using two metrics: 1) relative operating characteristic (ROC) diagrams and 2) threat scores.

Evaluation of these particular metrics required probabilities being converted to Yes-No forecasts for any given threshold by:

- Picking a probability threshold (Pt)
- If P ≥ Pt, forecast Yes
- If P ≤ Pt, forecast No

with the said metrics determined from the following contingency table:

:

		Observed Yes	Observed No
Forecast	Yes	a	b
	No	c	d

POD = a/(a + c)

FAR = b/(b+d)

Threat Score (TS) = a/(a + b + c)

To generate the ROC diagrams, POD and FAR were calculated for every probability threshold from 0 to 100%, with a 1% increment for all storm cases evaluated and all breakpoints for each forecast period. These thresholds were then plotted in a 2-D plane of POD versus FAR, called a ROC diagram. The threshold closest to the upper left corner of the diagram represented the one with the highest POD and the smallest FAR thus indicating the optimal threshold for any given period based on the ROC diagram.

Threat score (TS) is useful for low probability events since it does not include the No Forecast-No Observed events (Wilks 2006). It is a number that ranges from 0 to 1 with 1 being the best score. For perspective, if every forecast is a Yes forecast and they are all observed, then b = c = 0 and the score is 1. If a = b =c, then the score is 0.33. The primary difference between the POD and the TS is that the POD does not penalize for over forecasting while the TS does. TS scores were calculated for the same range of probabilities as for the ROC diagrams, with the Pt yielding the maximum TS for a given forecast period representing the optimal threshold for that period.

3. RESULTS AND DISCUSSION

Figure 2 illustrates the reliability diagram obtained for the incremental wind speed probabilities for all cases, locations, and time periods combined. The 64 kt probabilities have somewhat of a low bias, meaning, they tend to under predict the events somewhat, but still by not much more than 10% to 20%. This might be caused by the fact that the 2004-2008 U.S. landfalling cases included a greater than average fraction of very large storms (e.g., Frances, Ivan, Katrina, Rita, Wilma and Ike). The Monte Carlo (MC) model uses NHC wind radii at t=0, but relaxes towards climatological radii after about 24 hr because not all of the required radii are available from the NHC deterministic forecast (DeMaria *et al*., 2009). Verification of a larger sample of forecasts over the Atlantic showed much smaller biases for the 64 kt probabilities (Knaff and DeMaria, 2007). The 34 kt probabilities exhibit a rather close one to one relationship to the observed frequencies hence the results are deemed quite good. Overall, the incremental probabilities appear to be reliable and behave within an acceptable tolerance as to be useful for establishing trigger thresholds.

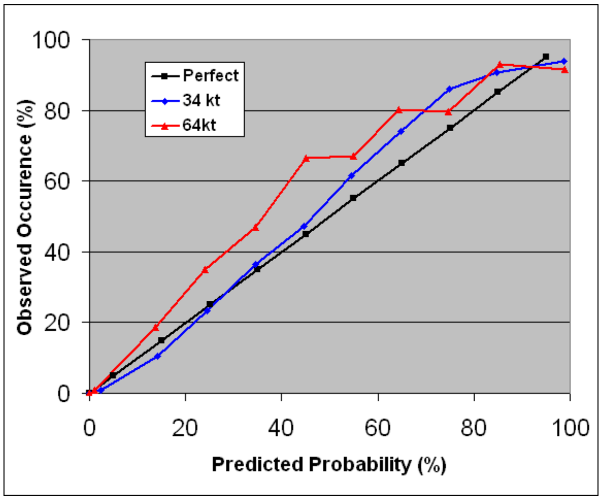


Figure 2. Observed frequency of events versus predicted incremental wind speed probabilities from the MC model for all 401 cases evaluated.

Figure 3 shows the optimal incremental wind speed probability thresholds determined for the 34 and 64 kt probabilities based on the TS and ROC diagram metrics as a function of forecast period. Additionally, the thresholds used during the 2006 to 2008 experimental periods based on a raw histogram analysis as discussed in section 3 of Santos *et al.* 2008 are shown for reference and validation purposes. It is clear from these diagrams that the thresholds based on the ROC diagrams are very low particularly in the early periods compared to the thresholds from the two other metrics. The TS scores and histogram based thresholds are much more consistent with each other. The low thresholds for the ROC-based optimal values are due to the fact that for a large fraction of the coastal breakpoints, tropical storm or hurricane conditions do not occur. Therefore, the "No-No" cases dominate the statistics and the optimal probability value is fairly low. In contrast, the TS does not depend on the cases where "No" is forecast and observed, so that maximizing the TS results in a higher optimal probability threshold. The histogram approach depends only on verified events, and so, like the TS, does not have the problem with a low optimal value like the ROC diagram does (Santos *et al.* 2008). Focusing then on the TS scores, it is clear that in the extended periods, the thresholds used during the experiments of 2006 to 2008 as described in Santos *et al.* 2008 were rather good. However, in the short range, the thresholds used appear to be lower than they should be. This implies expressions of uncertainty were being invoked more often than they should. It is also evident that the main adjustments needed are on the hurricane wind speed probability thresholds. This is consistent with feedback received from other coastal WFOs participating in the experiment, and highlights the fact that the expressions of uncertainty appear to be triggered more often than they would expect for hurricane conditions but that for tropical conditions they appear to be reasonable in both the zone version of their forecasts as well as the point and click version.

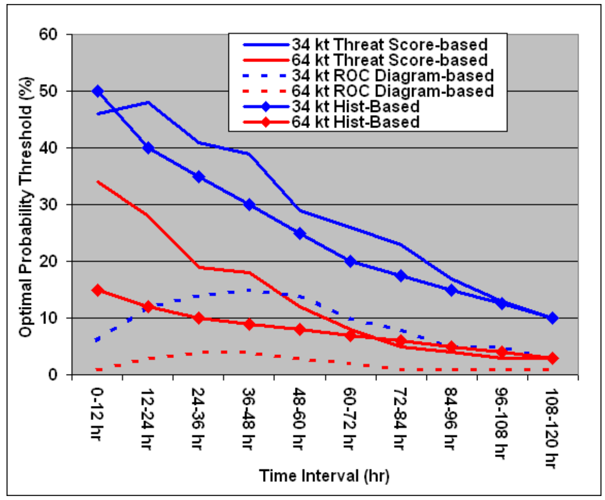
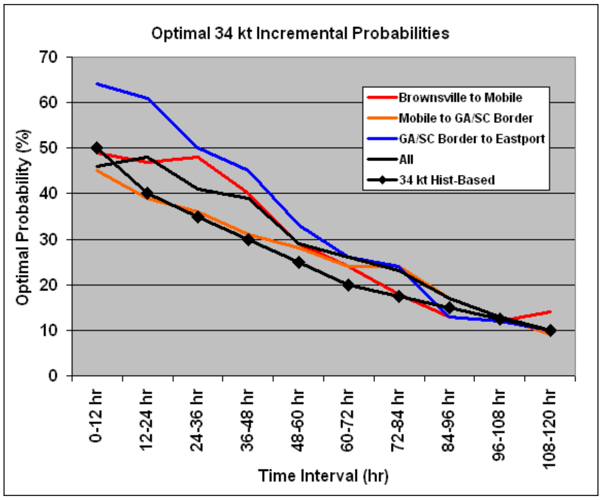
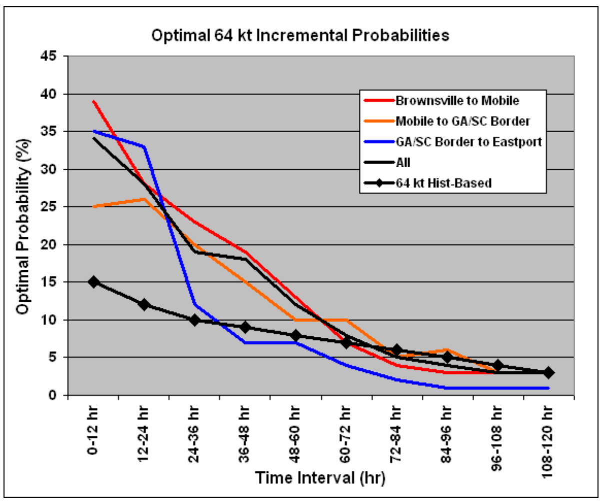


Figure 3. Incremental wind speed probability thresholds as a function of forecast period based on threat scores, ROC diagrams, and raw histogram analysis (section 3 Santos *et al.* 2008). These thresholds are from all breakpoints combined used in this study along the U.S. Gulf of Mexico and Atlantic coasts for the 401 storm cases analyzed during the 2004 to 2008 seasons.

Figure 4 is similar to Figure 3 but for the 34 kt (a) and 64 kt (b) threat score based optimal incremental wind speed probability thresholds for all points combined as well as broken down by the regions shown. Additionally, the histogram based optimal thresholds shown in Figure 3 are shown again, remembering those apply to all points combined. The same observations made in Figure 3 apply here too. What is different is the fact that there is a regional variation also to the optimal thresholds. For example, for the mid Atlantic and northeast coasts, higher optimal probability thresholds are apparent in the short range periods but lower in the extended periods. The authors believe this is perhaps related to the fact that many of the storms affecting these areas are moving fast across the region or curving out to sea.



(a)



(b)

Figure 4. Optimal incremental wind speed probability thresholds for 34 kt (top) and 64 kt (bottom) versus forecast period. Shown in black are the TS score and histogram analysis based results shown in Figure 3 for all break points combined. In color are the TS scores based optimal thresholds by region.

4. SUMMARY

This paper presents the validation of tropical cyclone incremental wind speed probability thresholds used in NWS Forecast Offices to generate expressions of uncertainty within their public and marine forecasts as shown by Sharp *et al.* 2006 and Santos *et al*. 2008. The results indicate that the probability thresholds used during the 2006 to 2008 seasons in experimental mode for tropical storms are quite good and need only minor upward adjustments, particularly in the short range periods. For hurricane conditions, the results illustrate that the thresholds used particularly in the first 4 to 5 forecast periods are too low and need considerable upward adjustment. The determination of these optimal thresholds will help fine tune the use of the expressions of uncertainty. In the bigger picture, it helps pin down probabilistic thresholds that would represent a decision point for more advanced users such as emergency managers but based not just on one deterministic number but one that also accounts for uncertainty in the forecast.

Future work includes the validation of the optimal thresholds found in this study by using other metrics such as Peirce Scores (Wilks 2006). Also, the authors plan to generate the validation metrics for the 50 kt incremental wind speed probability thresholds as they might become useful in the evolution of new forecast products or for other sophisticated users. The authors also plan to test and compare 2004 to 2008 cases using the current histogram based threshold values and derived optimal thresholds based on this study for ALL break points as well as the three different regions shown in Figure 4. This should help pin down and test the optimal thresholds used as the NWS considers going official with the use of tropical cyclone expressions of uncertainty in the public and marine forecast products.

5. ACKNOWLEDGMENTS

The views, opinions, and findings in this report are those of the authors and should not be construed as an official NOAA or U. S. government position, policy, or decision.

6. REFERENCES

DeMaria, M., J. A. Knaff, R. Knabb, C. Lauer, C. R. Sampson, and R. T. DeMaria, 2009: A new method for estimating tropical cyclone wind speed probabilities. *Wea. Forecasting,* submitted.

Gross, J. M., M. DeMaria, J. A. Knaff, and C. R. Sampson, 2004: A new method for determining tropical cyclone wind forecast probabilities. *Preprints, 26th Conference on Hurricanes and Tropical Meteorology*, Amer. Met. Soc., Miami, FL, pp 425-426.

Knaff, J. A., and M. DeMaria, 2005: Improvements in deterministic and probabilistic tropical cyclone surface wind predictions. *Final Report, Joint NOAA/Navy/NASA Hurricane Testbed, United States Weather Research Program* [Available from NOAA/Tropical Prediction Center, 11691 SW 17th Street, Miami, FL, 33165.]

Knaff, J. A. and M. DeMaria, 2007: Verification of the Monte Carlo Tropical Cyclone Wind Speed Monte Carlo Tropical Cyclone Wind Speed Probabilities. 61st Interdepartmental Hurricane Conference. New Orleans, LA. [Available online http://www.nhc.noaa.gov/jht/ihc_07/s8-04Knaff.ppt]

Santos, P., D. W. Sharp, M. Volkmer, and G. Rader, 2008: Employing Hurricane Wind Probabilities to Convey Forecast Uncertainty and Potential Impact

Through NWS Field Office Forecast Products. Preprints, *Tropical Meteorology Special Symposium/19th Conf. on Prob. And Statistics*, JP1.3, New Orleans, LA. [Available online http://ams.confex.com/ams/pdfpapers/131748.pdf]

Sharp, D. W., P. Santos, M. Volkmer, G. Rader, and M. Sardi, 2006: Employing Hurricane Wind Probabilities to Enhance Local Forecasts and Improve Guidance for Decision-Makers. Preprints, *27th Conf. on Hurricanes and Tropical Meteorology*, 9A.3, Monterey, CA. [Available online http://ams.confex.com/ams/pdfpapers/108409.pdf]

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*, Second Edition, Academic Press, 467 pp.