J10.2 THE DETERMINATION OF OPTIMAL THRESHOLDS OF TROPICAL CYCLONE INCREMENTAL WIND SPEED PROBABILITIES TO SUPPORT EXPRESSIONS OF UNCERTAINTY WITHIN TEXT FORECASTS: AN UPDATE

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

1. INTRODUCTION

As tropical cyclone events unfold, decision-makers require a meteorologist's most likely deterministic wind speed forecast along with an accompanying expression of uncertainty. Both are necessary to effectively manage preparations for life-threatening weather events. Inherent forecast uncertainty reveals the obvious limitation of deterministic-only wind speed information as provided within the current Zone Forecast Product and Coastal Waters Forecast issued by National Weather Service (NWS) Weather Forecast Offices (WFOs). To address the problem, WFOs Miami and Melbourne have developed a means to consistently and coherently incorporate uncertainty information within 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, NHC wind probabilities are operationally produced in several forms, which 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 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. It is then applied to 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 (e.g., triggers) are a critical component to the formatter logic, it is necessary to validate them. Initial validation efforts for coastal locations along the east coast of the U.S. were reported by Santos et al. (2009) using Reliability Diagrams to determine whether the probabilities themselves are fundamentally 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 scored using two different metrics: relative operating characteristics (ROC) diagrams and threat scores (TS) (Santos et al., 2009).

This paper expands the initial validation study to further include inland points (in addition to coastal points) and hence represents an update to Santos et al., 2009. Based on preliminary results, the probabilities were deemed reliable for the intended use, and able to yield optimal trigger thresholds for both coastal and inland locations according to geographic region.

---

[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.

[3] *Corresponding author address*: **David Sharp**, NOAA/National Weather Service, 421 Croton Rd, Melbourne, FL, 32935; e-mail: david.sharp@noaa.gov.

2. METHODOLOGY

To compute the validation scores, 19 storms from the 2004-2008 seasons 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 Fig. 1a with the tracks and coastal and inland points for which the verification was performed shown in Fig. 1b. All probability runs starting 3 days before the first warning issuance were included in the analysis for a total of 400 forecast cases. The probabilities were evaluated at 343 U.S. coastal breakpoints and 286 inland points. NHC 6 hourly best track positions and wind radii (interpolated down to an hourly verification grid) were used to determine points with observed tropical storm and/or hurricane force winds. The validation scores were computed for all points combined, as well as coastal and inland points separately, and within regional groups 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); and all regions combined.

As previously stated, validation efforts were comprised of reliability diagrams, ROC diagrams, and TS analyses. A reliability diagram consists of a plot of observed frequencies of an event versus the forecast probability. For simplicity, a reliability diagram was created 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 (trigger) thresholds as a function of time period that yield the maximum hit ratio (also known as probability of detection or POD) but also minimizes false alarm ratio (FAR). The main intent is to find those probability thresholds that best discriminate between events and non-events while accounting for uncertainty. To accomplish this, probability thresholds were evaluated using: 1) relative operating characteristic (ROC) diagrams and 2) threat scores (TS).

For evaluation, the probabilities were 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 said scoring determined from the following contingency table:

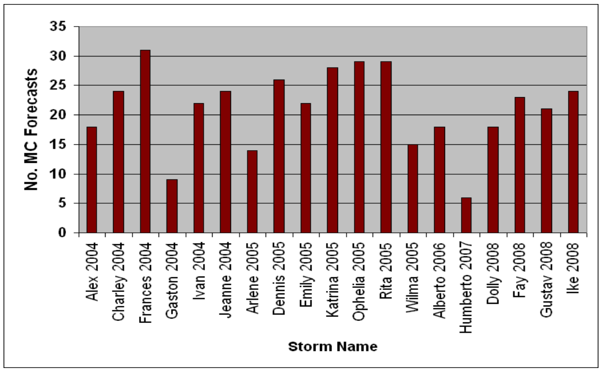


Figure 1a. Storm cases used in this study for the validation of the incremental wind speed probability forecasts using the 2008 version of the Monte Carlo (MC) model (DeMaria *et al.*, 2009; Knaff and DeMaria 2005, 2007). A total of 400 MC model runs were used derived from these 19 storms.

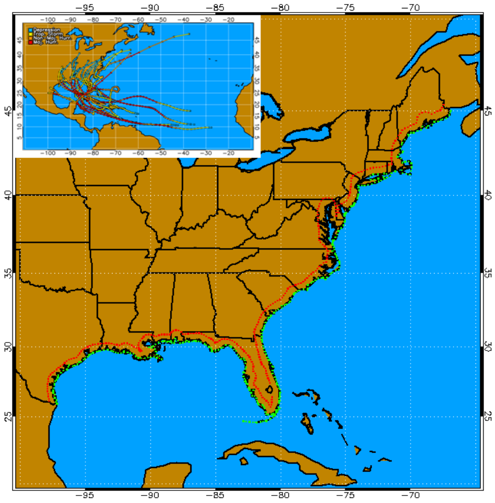


Figure 1b. Tracks of storm cases used in this study (insert), as well as coastal and inland points for which verification was conducted. Inland points (286) are in red and coastal points (343) in green. Inland points are about 50 km from the coast.

		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 (coastal, inland, combined) for each forecast period. These thresholds were then plotted in a 2-D plane of POD versus FAR. The threshold closest to the upper left corner of the diagram represents the value with the highest POD and the smallest FAR, thus indicating the optimal threshold for any given period.

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. On the other hand, if a = b =c, then the score is 0.33. The primary difference between the POD and the TS is that the POD scoring does not penalize for over-forecasting while the TS does. TS values 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 derived for the incremental wind speed probabilities for all cases and time periods combined and for coastal, inland, and all break points combined. The 64-kt probabilities have somewhat of a low bias, meaning, they tend to under-predict events, particularly for coastal points, but still by not 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 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. Overall, the incremental probabilities appear to be reliable and behave within an acceptable tolerance as to be useful for establishing trigger thresholds.

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 analyses as a function of forecast period for coastal points only. Additionally, the thresholds used during the 2006-2008 experimental periods based on a raw histogram analysis as discussed in section 3 of Santos *et al.* 2008 are shown for reference and perspective. It is clear that the thresholds identified by the ROC diagrams are very low particularly in the early periods compared to the thresholds from the two other methods. The TS values and histogram-based thresholds are much more consistent with each other. The seemingly low thresholds for the ROC-based values are likely the result of having a

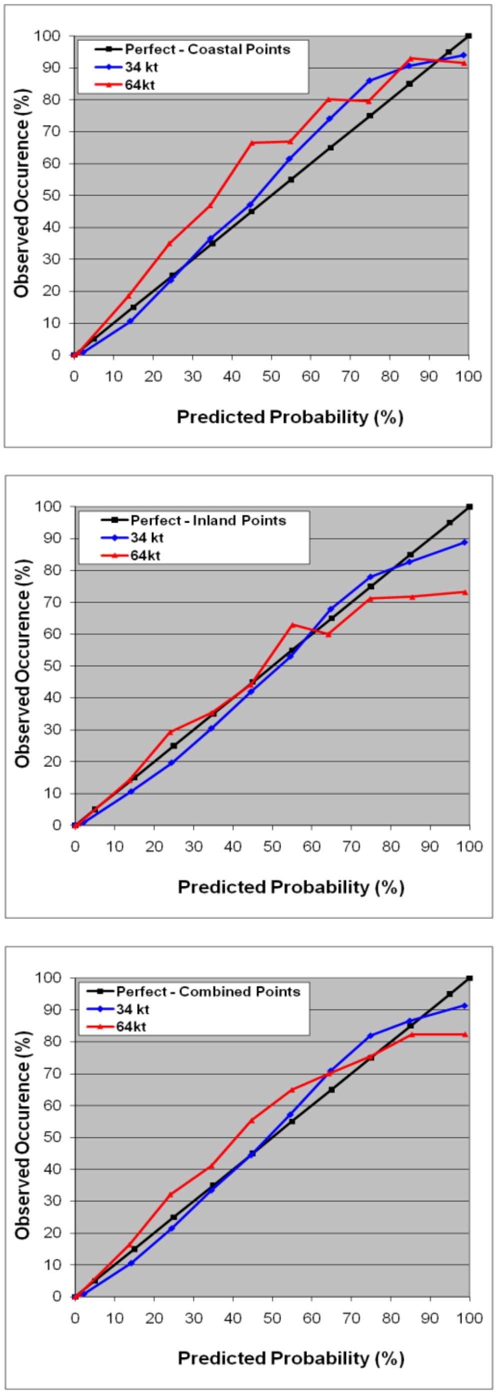


Figure 2. Observed frequency of events versus predicted incremental wind speed probabilities from the MC model for all 400 cases evaluated for coastal (top), inland (middle), and both points combined (bottom). 34-kt incremental wind speed probabilities in blue and 64-kt in red.

large portion of coastal points located where tropical storm or hurricane conditions did 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 that is intuitive in its behavior with time. The histogram approach depends only on verified events, and so, like the TS, does not have the problem with a lower than expected optimal threshold (Santos *et al.* 2008). Focusing then on the TS scores, it is clear from Fig. 3 that in the extended periods the thresholds used during the experiments of 2006-2008 as described in Santos *et al.* 2008 were rather good. However, in the shorter ranges, the thresholds experimentally used appear to be lower than optimal. This implies that expressions of uncertainty were being invoked more often than they should. It is also evident that the main adjustments needed are related to the hurricane wind speed probability thresholds. This is consistent with feedback received from many 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. This was indicated for both the zone version of their forecasts, as well as the point and click version.

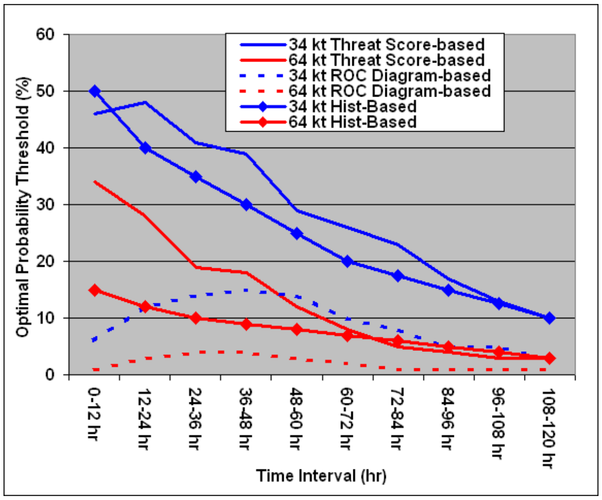


Figure 3. Optimal incremental wind speed probability thresholds as a function of forecast period based on threat scores, ROC diagrams, and raw histogram analysis (section 3 Santos *et al.* 2008). These thresholds are from all coastal points combined used in this study along the U.S. Gulf of Mexico and Atlantic coasts for the 400 storm cases analyzed during the 2004 to 2008 seasons.

The results shown in Fig. 3 also show that optimal probability thresholds decay almost exponentially with time (as expected). This highlights the notion that different probability thresholds, from a skill perspective, have different significance depending on how far out the event is forecast to happen from the present. Hence, a 10% probability forecast for 34-kt wind event forecast to happen in the first period can be interpreted as meaning it is very unlikely to occur. Yet, the same 10% probability forecast out at day 5 would require a heightened state of awareness given the reasonable possibility for the event to actually occur.

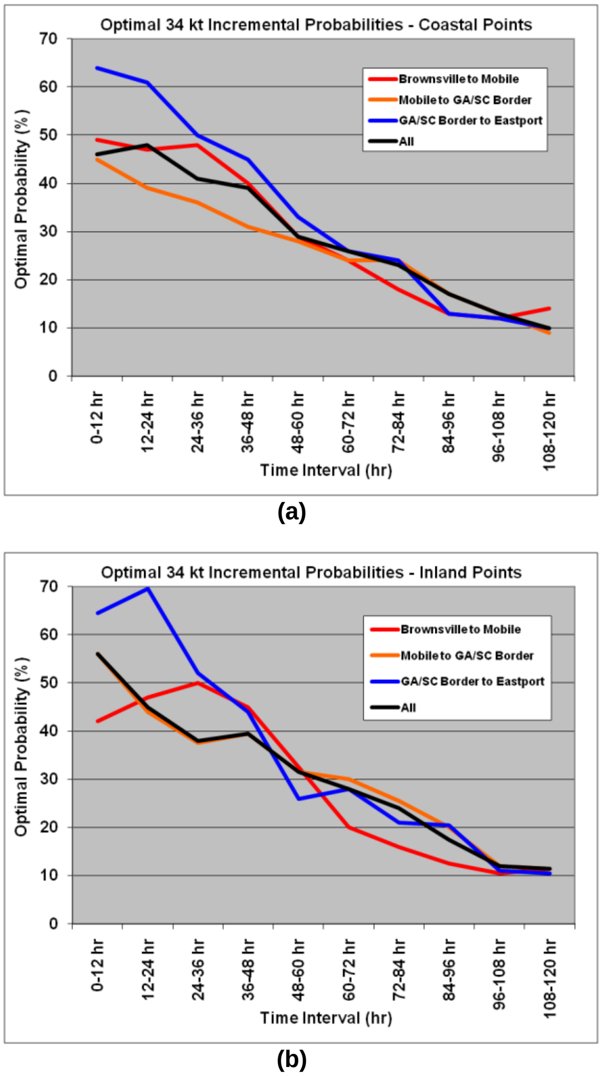
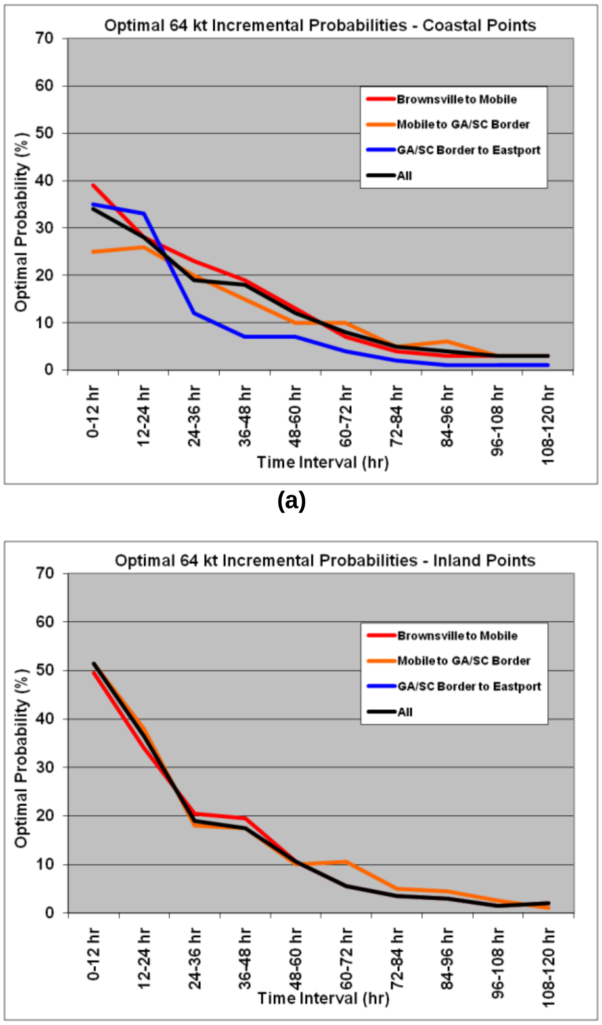


Figure 4. Threat Scores based optimal incremental wind speed probability thresholds for 34-kt for coastal (top) and inland (bottom) points versus forecast period and stratified by regions shown. Shown in black are all regions combined.

Figure 4 shows the TS-based optimal 34-kt incremental wind speed probability thresholds for coastal (a) and inland (b) points as a function of time period also, but this time stratified by region as well. The same observations as in Fig.3 hold true in this case. Additionally, the results in this figure do highlight that the optimal thresholds are regionally dependant, especially in the short ranges. The inland points appear to have slightly higher probabilistic thresholds particularly in the short ranges when compared to the coastal points.

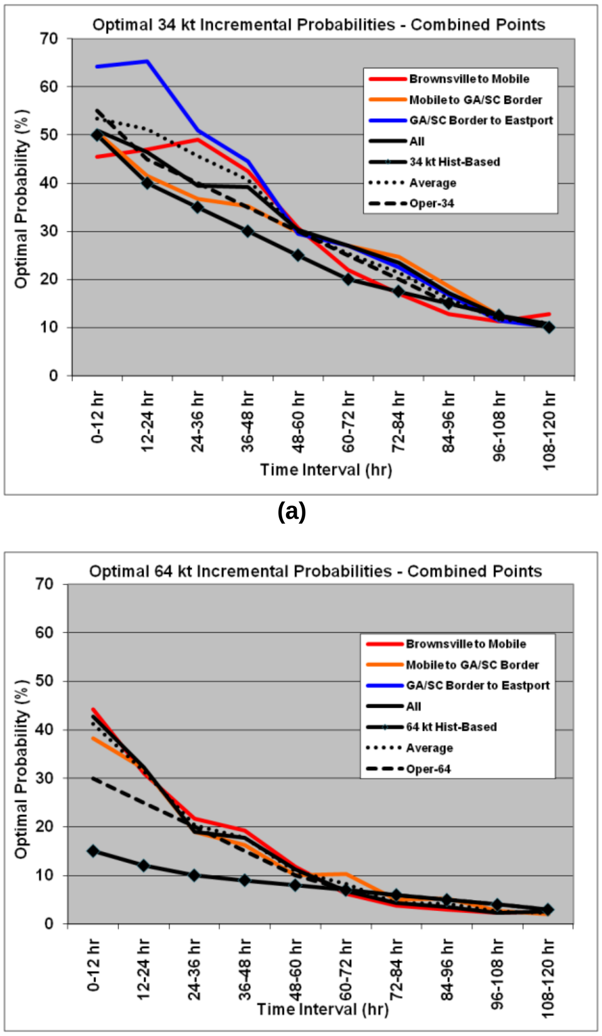


(a)

(b)

Figure 5. As in Fig. 4 but for the 64-kt incremental wind speed probabilities.

Figure 5 shows the same statistics as in Fig. 4, but for the 64-kt incremental wind speed probabilities. Similar observations are made, except that the overall regional dependence appears less pronounced, and virtually negligible for inland points. However, the inland points still stand out as having larger optimal trigger thresholds than coastal points, particularly in the short ranges. Notice that no probability thresholds are plotted for the mid-Atlantic and northeast coasts for inland points. None of the 400 derived cases from 19 separate storms resulted in observed hurricane force winds for any of the inland points evaluated in this study.



(a)

(b)

Figure 6. Optimal incremental wind speed probability thresholds for 34-kt (top) and 64-kt (bottom) versus forecast period for all points (coastal plus inland) combined. Shown in black with diamond shaped markers are the histogram analysis based results shown in Fig. 3. In black solid line are the TS scores based optimal thresholds calculated for all regions combined. The black dotted line is the average of the results obtained for each region, the red line are the results for the Gulf of Mexico region, orange for the Florida peninsula, and blue the mid Atlantic and northeast coast. The black dash line represents the operational thresholds values being proposed as of this writing based on these results.

Figure 6 is similar to the previous figures, 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 Fig. 3 are shown again. As before, the regional dependence is evident in these results also, but mainly for the 34 kt thresholds and mainly for the short range forecast periods. The overall results also highlight the highest thresholds in terms of a particular region are for the mid-Atlantic and northeast coast, and also in the short ranges. This is evident in all figures prompting the authors to speculate that it is perhaps related to the fact that many of the storms often affecting these areas are moving fast across the region and/or curving out to sea. Regardless, it is evident that relative to the empirically-derived histogram-based thresholds used experimentally during the 2006-2008 tropical seasons, the results in this study clearly indicate the optimal thresholds need to be adjusted upward, particularly in the short ranges, and more so for the 64-kt incremental wind speed probabilities. It is also evident that regional dependence needs to be considered.

4. SUMMARY AND CONCLUSIONS

This paper has presented 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 2006-2008 seasons, in experimental mode, need upward adjustments, especially in the short range periods. Particularly for hurricane conditions, the results illustrate that the thresholds used in the first 4 to 5 forecast periods are too low and need considerable upward adjustment.

The results also indicate that for purposes of this application regional, and even coastal versus inland, dependencies of the incremental wind speed probability thresholds need to be considered. However, operational sensitivity regarding consistency and/or continuity of message from forecast cycle to forecast cycle, region to region, or coastal to inland, led up to the authors proposing a single set of thresholds to be used for all regions, inland and coastal locations combined until more cases are gathered and evaluated. Table 1 shows the proposed set of operational threshold values. In fact these values were already implemented during the 2009 season and will be again employed for the 2010 season (and possibly beyond) until more storms can be evaluated.

The thresholds shown in Table 1 are those depicted by the black dash line shown in Fig. 6. Notice that the 64-kt thresholds shown are less than what the TS-based results suggest as optimal in the short range. This was done to accommodate the low bias discussed in Fig. 2, and also out of practical considerations of trying not to map areas for which there is still a sizeable risk too close to the actual forecast wind radii. This also helps mitigate any potential lack of consistency from forecast cycle to forecast cycle, mainly in cases involving ill-bahaved storms where forecast flip-flops become detrimental.

Period	PWS64	PWS34
00-12 hr	30%	55%
13-24 hr	25%	45%
25-36 hr	20%	40%
37-48 hr	15%	35%
49-60 hr	10%	30%
61-72 hr	7%	25%
73-84 hr	6%	20%
85-96 hr	5%	15%
97-108 hr	4%	12.5%
109-120 hr	3%	10%

Table 1. Single set of incremental wind speed probability thresholds as a function of forecast period proposed based on results of this study for 64-kt (second column) and 34-kt (third column).

The determination of these optimal thresholds will help fine tune the use of the expressions of uncertainty in the forecast application described by Sharp et al. (2006) and Santos et al. (2008). In the bigger picture, it helps identify probabilistic thresholds that would represent a decision (trigger) point for more advanced users such as emergency managers, but based not just on deterministic numbers but that also accounts for uncertainty.

Future work includes updating these results via an enhanced version of the Monte Carlo model used to generate the tropical cyclone wind speed probabilities that also incorporates track guidance spread (DeMaria et al. 2009). The enhanced model will be experimental during the 2010 season and likely to become operational during the 2011 season. Also, the authors plan to generate corresponding validation scores for the 50-kt incremental wind speed probability thresholds as they may become useful in the evolution of new forecast products. The authors also intend to explore the most effective means of implementing the regional-, and inland- versus coastal-, dependant thresholds within an operational environment. For now, the thresholds shown in Table 1 will continue to be evaluated during the 2010 season and beyond.

5. DISCLAIMER

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,* **24**, 1573-1591.

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, 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]

Knaff, J. A., and M. DeMaria, 2005: Improvements in deterministic and probabilistic tropical cyclone by the bsurface 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.]

Santos, P., D. Sharp, M. DeMaria, and S. Kiser, 2009: The Determination of Optimal Thresholds of Tropical Cyclone Incremental Wind Speed Probabilities to Support Expressions of Uncertainty in Text Forecasts. Preprints, *Symposium on Urban High Impact Weather/89th AMS Annual Meeting*, P1.5, Phoenix, AZ. [Available online http://ams.confex.com/ams/pdfpapers/147019.pdf]

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.