17A.7 A METHOD FOR THE VERIFICATION OF PROBABILISTIC TROPICAL CYCLONE WIND SPEED FORECASTS

Derek Ortt*
ImpactWeather, Houston, Texas

1. INTRODUCTION

Probabilistic tropical cyclone forecasts have been provided by the National Hurricane Center (NHC) and private industry for the past several hurricane seasons. Wind speed probabilities were first created by the National Hurricane Center for the 2006 Atlantic hurricane season (DeMaria et al. 2009). These forecasts provide estimates of the likelihood of 34KT, 50KT and 64KT winds being observed at a certain location from a given tropical cyclone (TC). Private industry has also been producing these forecasts. To date, there has not been any comprehensive verification technique developed to allow for the verification of these forecasts using observations of an objective analysis. It is essential to have a robust verification technique for probabilistic TC wind speed forecasts so that interests threatened by a TC can understand the level of risk that they face. If there is a certain forecast percentage chance of winds of a give threshold, knowing how often those winds occur in reality would allow for better preparations and risk management.

Probabilistic verification has been conducted since at least 1950 when the Brier Score was derived (Brier, 1950). A lower Brier Score indicates a more skillful forecast. However, there are some problems with the Brier Score. The main problem is that forecasts indicating a near 50 percent chance of an event occurring will yield relatively high Brier Scores. Therefore, a different method of evaluating probabilistic forecasts is required. Kay and Brooks (2000) presented the Reliability Diagram, which has the forecast probability on the x-axis and the observed probability on the y axis. This allows for a qualitative evaluation of the probabilistic forecasts. For TC wind speed probabilities, the rare nature of TCs at a given location makes it difficult to use traditional techniques, such as used for evaluating probability of precipitation at a given location. Thus, a new technique is required.

DeMaria et al. (2009) created a technique to allow for the verification using the best track wind radii. Splitt et al. (2010) and Collins (2013) were able to conduct a limited verification of National Hurricane Center wind speed probabilities based upon observations. Specifically, the two studies evaluated the forecast probabilities using observations from coastal locations in the United States and the Caribbean. The short coming of this method, however, is that there are thousands of data points in a 5 day forecast swath. Thus, using coastal locations alone as the observed data set limits the verification to a very small fraction of the forecast. In addition, only systems that threaten land can be verified. TCs that remain over the water cannot be verified using this methodology.

This study introduces a new technique that will allow for a more robust verification of TC wind speed probabilities. This new method will allow for verification throughout the entire forecast swath, not just for coastal locations. This will allow for a

Corresponding author address: Derek Ortt, ImpactWeather Inc., 12650 N. Featherwood Dr. Ste. 140. Houston, TX 77034; e-mail: dortt@impactweather.com

better overall evaluation of the TC wind speed probabilities.

2. DATA

The TC wind speed probabilities that will be evaluated in this study are the 34 KT probabilities from the ImpactWeather Atlantic advisories for Hurricanes Gustav (2008, 9 forecasts), Ike (2008, 11 forecasts), Irene (2011, 6 forecasts), Isaac (2012, 7 forecasts), and Sandy (2012, 7 forecasts). These 5 day probabilistic forecasts will be verified against the Hurricane Research Division H-WIND analyses (Powell et al. 1998). These 5 storms were selected as these were the 5 storms that have multiple H-WIND analyses with a temporal resolution of at least 6 hours for 5 consecutive days, allowing for an interpolation into a swath similar to the forecast probabilities.

A brief discussion as to the difference between the ImpactWeather probabilities compared to the NHC probabilities is required. NHC uses a Monte Carlo sampling to generate their probabilities. ImpactWeather calculates the probabilities using a Gaussian distribution of the most recent five-year mean error. In addition, the size of the distribution is affected by forecast confidence, which is determined by the forecaster in real time. If the forecast confidence is higher, the distribution is narrowed. If there is low confidence, a wider distribution is used. The final difference is that the ImpactWeather probabilities are generated off of the ImpactWeather track, intensity, and wind radii forecasts. The horizontal resolution of the wind speed probabilities is .25^o. This creates several thousand data points for each 5 day forecast.

As previously stated, the H-WIND data have a temporal resolution of at least 6 hours. The spatial resolution varies. However, for this study, the H-WIND data was interpolated onto the same .25^o grid as were the forecast probabilities. Thus, verification of individual forecasts is now possible due to the fact that each forecast has several thousand data points, as well as verification of an entire storm, or a group of storms combined.

3. ANALYSIS METHOD

The method of analysis used for the study begins with the previously mentioned interpolating of the H-WIND analysis data onto the same grid as the forecast probabilities. This is to allow for a comparison at identical locations between the forecast probabilities and the observed winds. The forecast probabilities are then binned into the following probabilities: 5, 15, 25, 35, 45, 55, 65, 75, 85, 95, and 100 percent. Following the method used by Splitt et al. (2010), forecast probabilities of less than 1 percent are disregarded. This had the practical impact of limiting the analysis domain to areas in the vicinity of the forecast track.

After the probabilities are binned, the observed probabilities were calculated at the same grid points as the binned probabilities. Individual TC forecasts, all forecasts for the individual storms, and the forecasts for all 5 storms were then evaluated qualitatively using Reliability Diagrams. Taking into consideration the weakness of the Brier Score, quantitative verification of the forecasts was conducted using the Mean absolute error. This is defined as

$$MAE = (\sum n_i|p_i - o_i|) / N$$

where p_i is the predicted probability, o_i is the observed occurrence rate for the given predicted probability, N is the overall number of grid points, n_i is the number of grid points for a predicted probability, while i represents the predicted probability, ranging from .05 to 1.

4. RESULTS

To demonstrate the viability of this new method, verifications of an individual forecast for Sandy, one from Ike, the overall Sandy verification, and the overall verification for all 5 storms will be presented. The first example will be an individual Sandy forecast. Figure 1 shows the ImpactWeather 34 KT wind speed probabilities from 3 PM CDT October 24, 2012 along with the observed area of 34 KT winds.

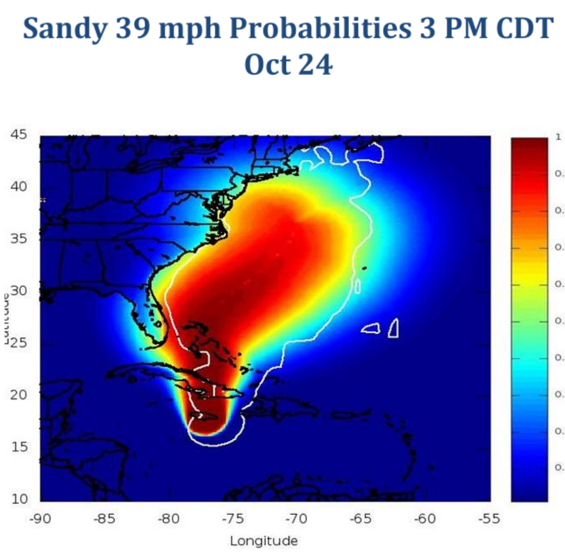

Figure 1: ImpactWeather 120 hour 34 KT wind speed probabilities for Hurricane Sandy (shaded) from 3 PM CDT October 24 along with the observed area of 34 KT (white contour) winds within the 120 hours.

For the above forecast, the observed 34 KT winds occurred in the areas where the forecast probabilities were high. Most areas where there were high forecast probabilities, those winds occurred. Areas with a low forecast probability generally did not receive 34 KT winds. As one would expect, the verification for the forecast would be reasonably good. This was in fact the case as shown in Figure 2.

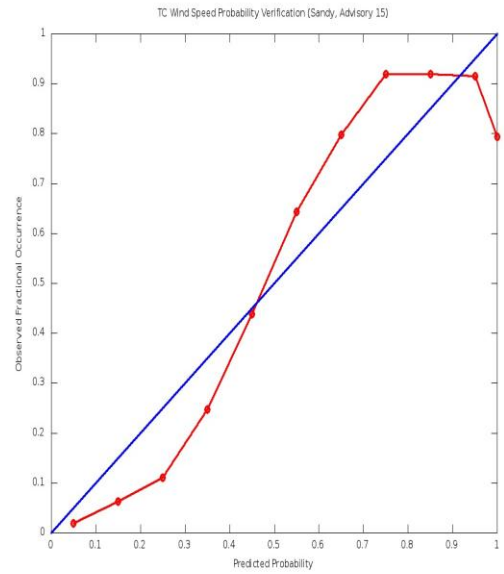

Figure 2: Reliability diagram for the Hurricane Sandy forecast from 3 PM CDT October 24. The red line represents the verification of the forecast, while the blue line represents a perfect forecast.

While there was a slight over forecast for the lower probabilities and a slight under forecast at the higher probabilities, the observed occurrence of winds is similar to the predicted probabilities. This leads to a low MAE for this forecast of .0742 with 11981 data points. In the case of Sandy, the track, intensity, and wind radii forecasts were reasonably accurate. For the individual TC case, if the track or wind radii forecasts are in error, this will lead to a larger MAE. In the case of a track error, as was the case with Isaac, the MAE was as high as .1118. However, results from Hurricane Ike indicate that errors in the forecast wind radii may have the greatest impact on the forecast wind speed probabilities. If the wind radii are too small, it creates a probability swath that is too small, leading to a large under prediction. Figure 3 shows the 34 KT probabilistic forecast from 4 AM CDT September 7, 2008 and the observed area of 34 KT winds. Figure 4 shows the reliability diagram.

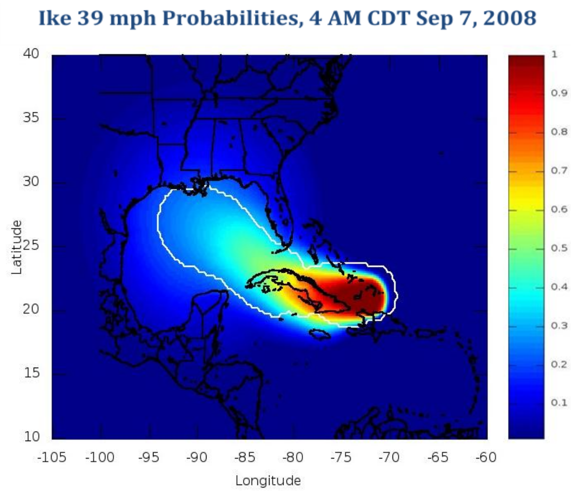

Figure 3: ImpactWeather 120 hour 34 KT wind speed probabilities for Hurricane Ike (shaded) from 4 AM CDT September 7 along with the observed area of 34 KT (white contour) winds within the 120 hours.

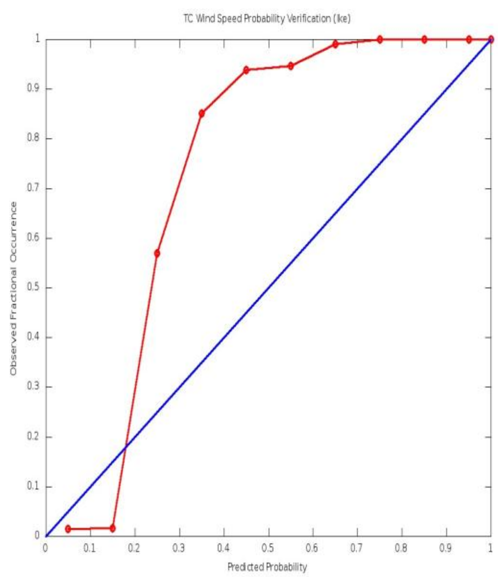

Figure 4: Reliability diagram for the Hurricane Ike forecast from 4 AM CDT September 27 The red line represents the verification of the forecast, while the blue line represents a perfect forecast.

For nearly all forecast probabilities, the actual occurrence of 34 KT winds was greater than what was predicted. The wind radii forecast error led to a large MAE of .1653 for 8034 data points. This indicates that errors in wind radii forecasts may have more impact on the wind speed probabilities than the track forecast errors.

The low probabilistic wind speed forecast errors were not limited to the previously presented Sandy forecast. All ImpactWeather Sandy forecasts had reasonably accurate 5 day track and wind radii forecasts. Thus, the overall Sandy verification results were similar as the results for the previously demonstrated forecast. Figure 5 shows the Reliability Diagram for all 7 Sandy forecasts.

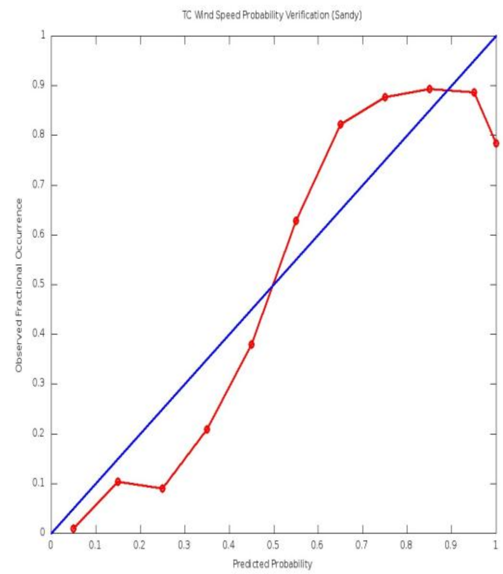

Figure 5: Reliability diagram for all Hurricane Sandy forecasts. The red line represents the verification of the forecast, while the blue line represents a perfect forecast.

When considering all Sandy forecasts, the previously shown signal of over prediction at the lower probabilities and the under prediction of the higher probabilities remained present. However, the observed occurrences lied reasonably close to the forecast probabilities for each threshold. This led to a MAE for Sandy of .0850 with 71925 data points. It is worth noting that other storms, especially Ike and Isaac had higher MAEs. The Ike MAE was .1402 with 103385 data points while the Isaac MAE was .0984 for 57339 data points.

This further demonstrates that while track errors will degrade the accuracy of the wind speed probabilities, the forecasts are more sensitive to wind radii.

The overall verification for all 5 storms is now presented. Figure 6 shows the overall verification. For the 5 storms, there is an under prediction between 25 and 75 percent. The overall MAE is .0995 with the verification conducted over 368880 data points. When considering the sample as a whole, individual storm biases such as the wind radii being too large or small and track errors may be cancelled out somewhat. Thus, it is possible that adding additional storms is required for a more robust and meaningful verification.

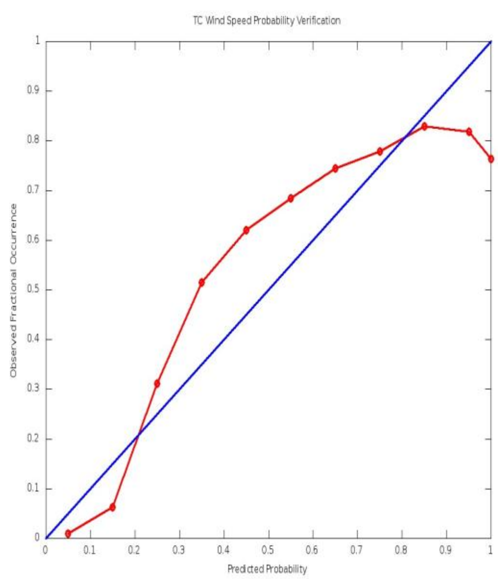

Figure 6: Reliability diagram for all forecasts from all 5 storms. The red line represents the verification of the forecast, while the blue line represents a perfect forecast.

5. SUMMARY AND FUTURE WORK

Using H-WIND analyses allows for the observation based verification of probabilistic TC wind speed forecasts. While it is possible to verify individual forecasts and individual storms, a larger sample size may be required to establish an overall validation as to how accurate probabilistic forecasts are. That said, the results indicate that the wind speed probabilities are very sensitive to wind radii forecasts, more so than to track forecasts.

The methodology presented here can be applied to any TC probabilistic wind speed forecast. For example, it could be used with the NHC forecasts or model ensemble forecasts. All that is required is a gridded probabilistic forecast and an objective observation based analysis.

In the future, the study may include analyses from using the multi-platform satellite analysis from NOAA. A significant advantage of using the multi-platform satellite analysis is that verification will be possible for any storm worldwide. H-WIND is only available in the western Atlantic basin where there is reconnaissance aircraft data. Thus, a much larger data set will be available by using the satellite data. This would allow for several seasons of forecasts to be evaluated over a long time period.

6. AKNOWLEGEMENTS

The –H-WIND data is courtesy of the Hurricane Research Division. I would also like to thank Devin Eyre and Bob Weinzapfel for providing the wind speed probabilistic forecasts.

7. REFERENCES

Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.*, **78,** 1-3.

Collins, S., 2013: Evaluation of the National Hurricane Center's 2010 Monte-Carlo wind speed probability forecast model. Masters Thesis, Florida Institute of Technology, 89 pp.

DeMaria, M., and co authors: A new method for estimating tropical cyclone wind speed probabilities. *Wea. and Forecasting*, **24,** 1573-1591.

Kay and Brooks, 2000: Verification of probabilistic severe storm forecasts at the SPC: *Available online at http://www.spc.noaa.gov/publications/mkay/probver/index.html*

Powell, M. D., S. H. Houston, L. R. Amat, and N. Morisseau-Leroy, 1998: The HRD real-time hurricane wind analysis system. *J. Wind Engineer. and Indust. Aerodyn.* 77&78, 53-64.

Splitt, Michael E., J. A. Shafer, and S. M. Lazarus, 2010: Evaluation of the National Hurricane Center's Tropical Cyclone Wind Speed Probability Forecast Product, *Wea. and Forecasting*, **25**, 511-525.