NOAA Joint Hurricane Testbed (JHT)
Final Report

Date: January 22, 2016
Project Title: Upgrades to the Operational Monte Carlo Wind Speed Probability Program
Principal Investigators: Andrea Schumacher
Affiliation: CIRA / Colorado State University
Award Period: September 2013 - August 2015
Reporting Period: September 2013 - August 2015

1. PROJECT OVERVIEW AND OBJECTIVES

This project sought to complete a number of upgrades to the current Monte Carlo wind speed probability model (hereafter MC model), many of which are based on NHC feedback over the past few hurricane seasons. Three improvements to the MC model algorithm itself were made including replacing the linear forecast interpolation scheme with a more precise spline fit scheme, applying a bias correction to the model track error statistics to provide consistency between NHC's uncertainty products, and applying a bias correction to the radii-CLIPER used by the MC model to improve the accuracy of the wind speed probabilities for exceptionally small or large tropical cyclones. Three additions/enhancements were also developed, which included estimates of the arrival and departure times of 34/50/64-kt winds, an integrated GPCE parameter, and wind speed probabilities to 7 days. Finally, the error statistic generation code and scripts were consolidated to simplify annual product updates.

2. ACCOMPLISHMENTS

All upgrades were completed during the project period. A brief description and evaluation of each upgrade is given below. Additional time was given to us to continue real-time evaluation through the end of the 2015 Atlantic and N.E. Pacific hurricane seasons. Seasonal verification for 2015 can be found in Section 2f. A summary and developer recommendations can be found in Section 3, in addition to suggestions for future work in this area.

a. Potential improvements to the MC model algorithm

Three potential improvements were made to the MC model algorithm, each of which is meant to address a specific issue with MC model that was noted by either developers or users at NHC (often during special cases such recurving landfall cases and extremely large storms like Hurricane Sandy). A brief overview of each potential improvement is given below in Sections (i)-(iii) and impacts on MC model performance are shown in Section (iv).

(i) Improved time interpolation scheme

The MC model currently uses linear interpolation to interpolate between forecast times. As result, errors have been found to be larger for the times between the NHC forecast points and an eastward bias is introduced for re-curving cyclones. Also, realization tracks have an unrealistic “jagged” appearance. To address these issues associated with linear interpolation, a spline fit interpolation approach was added as an option in the MC model. Figure 1 shows the wind speed probabilities using linear interpolation (top left) and the spline fit (top right) for Hurricane Earl when it was forecast to recurve along the U.S. east coast. The spline fit methodology appears to be correcting the eastward bias in this case, providing a more realistic interpolated track forecast after 48 hours. The corresponding 34-kt wind speed probabilities along the North Carolina coast increase from 5-10% (easiest to see in difference plot in Fig. 1 bottom left) where tropical storm force winds were indeed observed (Fig. 1 bottom right). Also, the spline fit makes realization tracks appear somewhat smoother and more realistic (Figure 2).

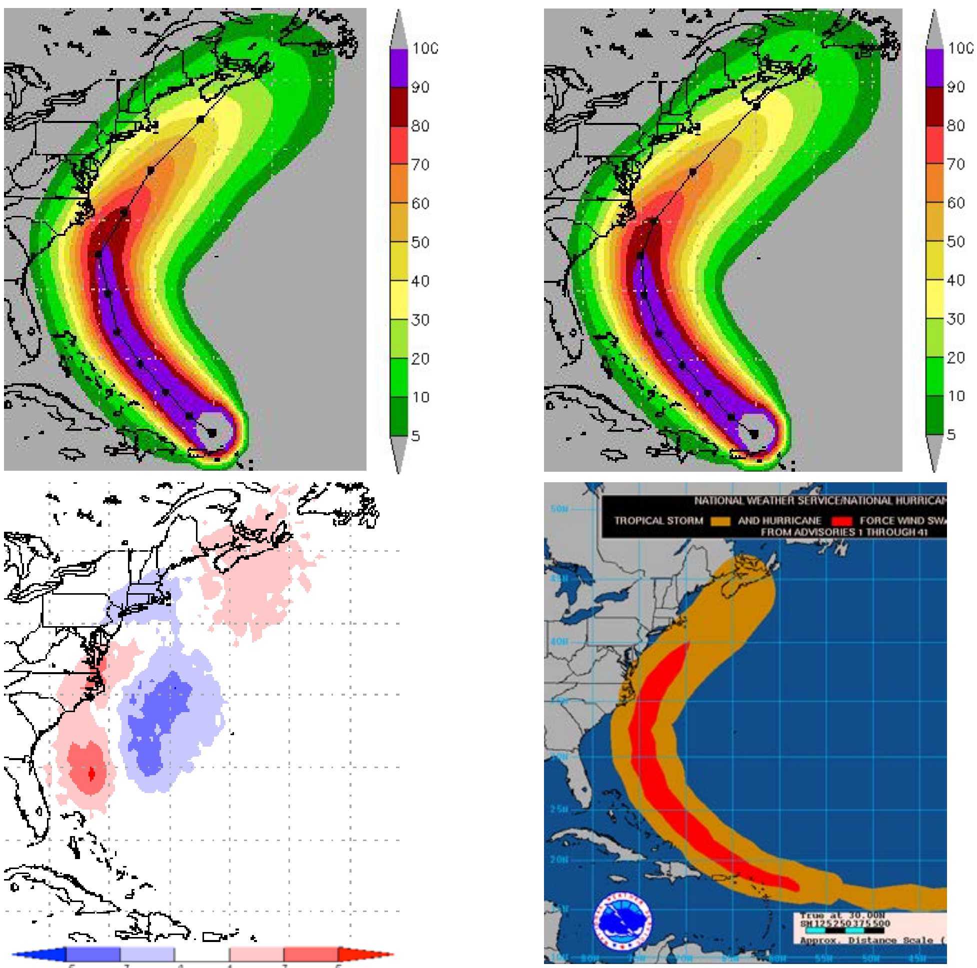

Figure 1. 34-kt wind speed probabilities using linear interpolation (top left), spline fit (top right), wind speed probability differences (spline - linear, bottom left), and NHC wind history (bottom right) for Hurricane Earl on 31 Aug 2010 0000 UTC.

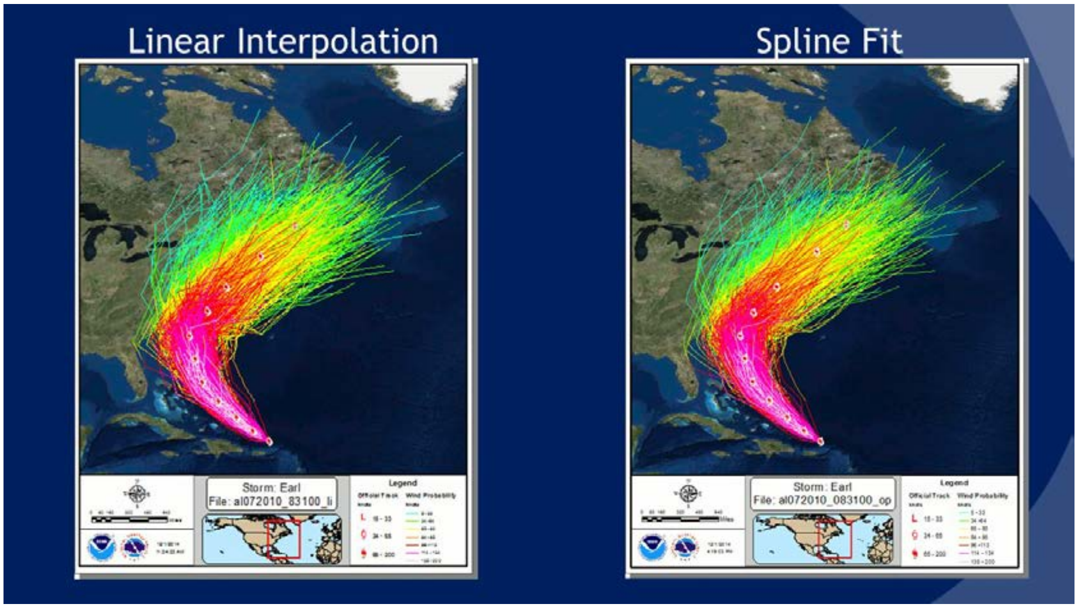

Figure 2. Realization tracks using linear interpolation (left) and spline fit (right) for Hurricane Early on 31 Aug 2010 0000 UTC (courtesy of C. Ogden).

(ii) Bias correction to track error statistics

NHC routinely provides the cone of uncertainty based on the 67th percentile of the previous 5-year average track errors as part of their advisory package. In principle, this same information could be determined from the MC model track realizations. Preliminary tests of this procedure showed that the cones did not exactly match, partially due to differing samples used to create the MC model error statistics, which include extra-tropical cases, and a small bias introduced by the serial correlation correction in the MC model. To fix this inconsistency new error statistics were derived using only cases NHC uses in their annual verification (i.e., excluding extra-tropical cases). Also, the MC model was adapted to compute the difference between the 67^{th} percentile cone size derived from model-sample track errors that do not incorporate GPCE and the official cone size. This difference, or estimate of bias introduced by the serial correlation, is then used to correct the MC model-sampled track errors with GPCE information incorporated. The advantage to obtaining the cone from the MC model is that it would impose consistency between NHC's uncertainty products, and would increase or reduce the size of the cone based on the track forecast confidence that is obtained from the GPCE parameter, should this capability be desired.

(iii) Wind radii bias correction

The current MC model uses a climatology and persistence model to estimate the 34-kt, 50-kt, and 64-kt wind radii for its realizations, which has been found to greatly over (under) estimate radii for exceptionally small (large) tropical cyclones. To address this issue, the MC model was adapted to use the official R34, R50, and R64 radii forecasts. Since official radii forecasts are only made out to 36-72 hours, radii at times after the last available radius forecast are relaxed back to the climatology-persistence method at an e-folding time of 32 hours. This radii bias correction was added as an option that can be turned on in the main MC model code.

Figure 3 shows an example of the 34-kt WSPs for Hurricane Sandy using the current method (left) and the bias-corrected method (center) in comparison to the actual observed 34-kt winds (right). As expected, this bias-correction method gives much more realistic-looking WSPs for this large storm out to 120 hours.

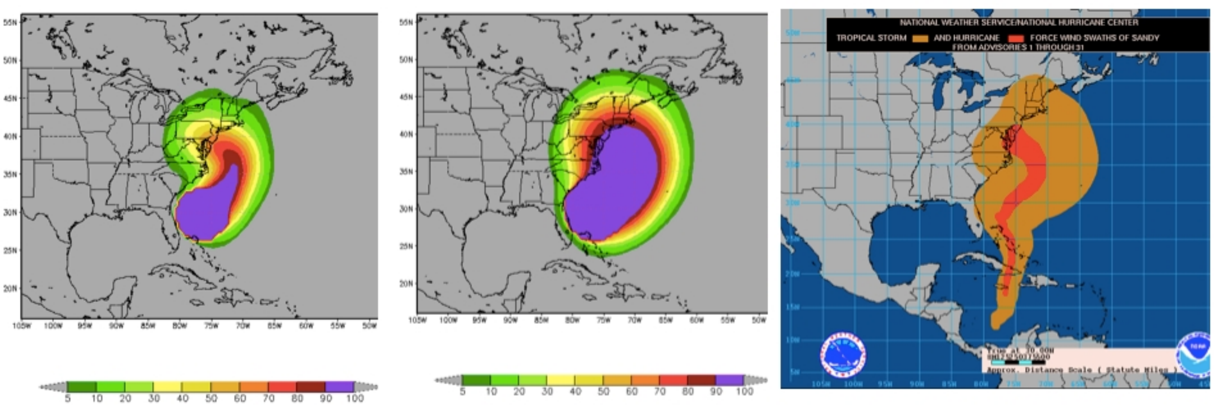

Figure 3. 34-kt wind speed probabilities for Hurricane Sandy before (left) and after (center) radii bias correction. Observed 34-kt winds are shown for comparison (right, in orange).

(iv) Impacts of potential improvements on basin-wide MC model performance

To assess the extent to which each algorithm change impacts the performance of the MC model, experimental versions of the MC model incorporating each potential improvement *individually* were developed and run over the Atlantic and N.E. Pacific basins from 2011-2014. Brier scores (Figure 4) and reliability (Figure 5) were calculated for each experimental MC model and a control version that is nearly identical to the 2015 operational MC model to provide a quantitative look at how each potential improvement actually improves (or degrades) MC model performance.

Summary of 2011-2014 verification (for cumulative probabilities):

- Spline interpolation: Impact on Brier scores and reliability is very small for all wind thresholds in both basins (as expected).

- Error statistics bias correction: Degrades Atlantic Brier scores by ~2% for 34-kt WSPs but improves 50-kt and 64-kt by a similar amount. Degrades all N.E. Pacific Brier scores by 3-4% for all wind thresholds. Worsens the slight tendency towards overprediction in the Atlantic for 34 and 50-kt WSPs but has very small impact on reliability of Atlantic 64-kt and N.E. Pacific WSPs.

- Radii bias correction: Improves Atlantic 34-kt and 50-kt Brier scores up to 5%. Improves N.E. Pacific 34-kt Brier scores up to 8% and has very small impact on N.E. Pacific 50-kt Brier scores. Degrades 64-kt Brier scores up to 5% in both basins. Improves reliability of N.E. Pacific WSPs for all wind thresholds (reduces consistent overprediction in control WSPs to nearly perfect). Has very little impact on reliability in the Atlantic, except for 64-kt WSPs where it introduces an underprediction.

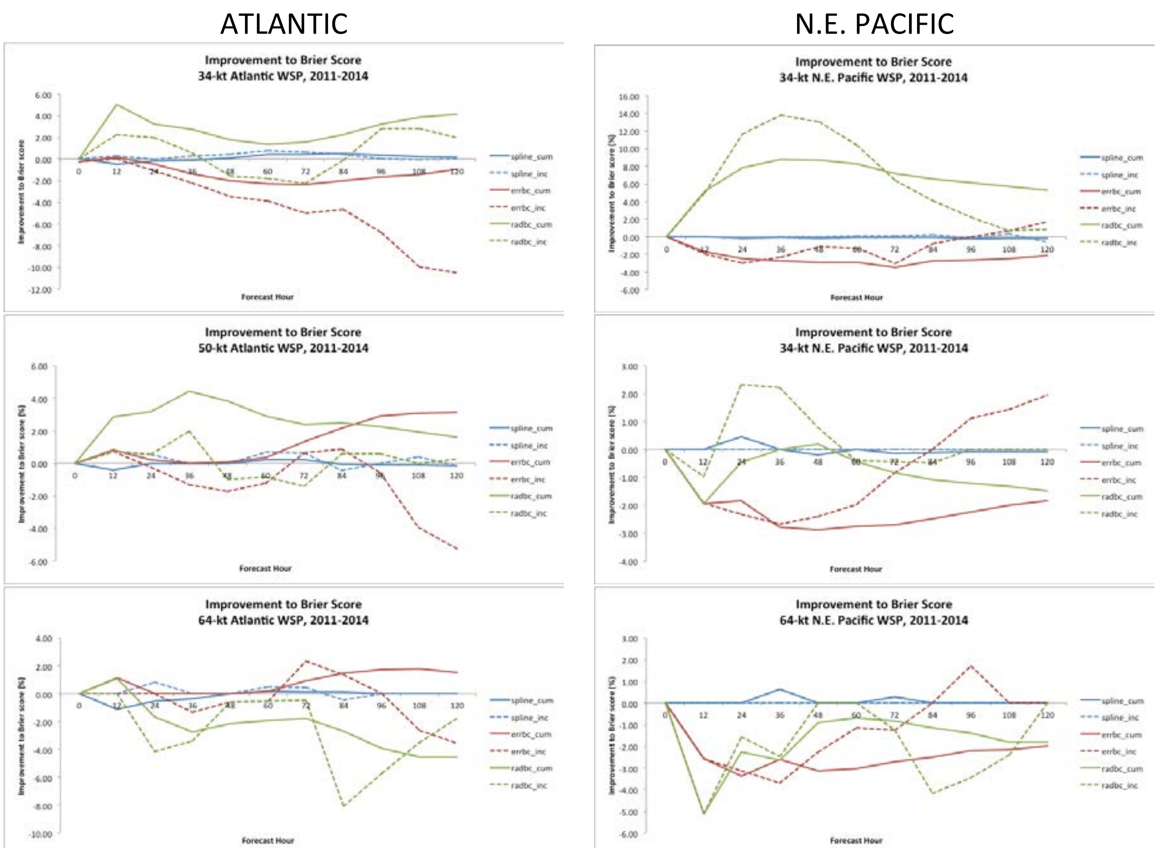

Figure 4. Percent improvement in Brier scores for 34-kt (top), 50-kt (center), and 64-kt (bottom) wind speed probabilities in the Atlantic (left) and N.E. Pacific from 2011-2014 for each potential improvement; spline interpolation (blue), error statistic bias correction (red), and radii bias correction (green) versions of the MC model. Improvement is defined as a reduction in Brier scores from the control MC model over the same sample. Brier score improvements for cumulative (incremental) wind speed probabilities are solid (dashed).

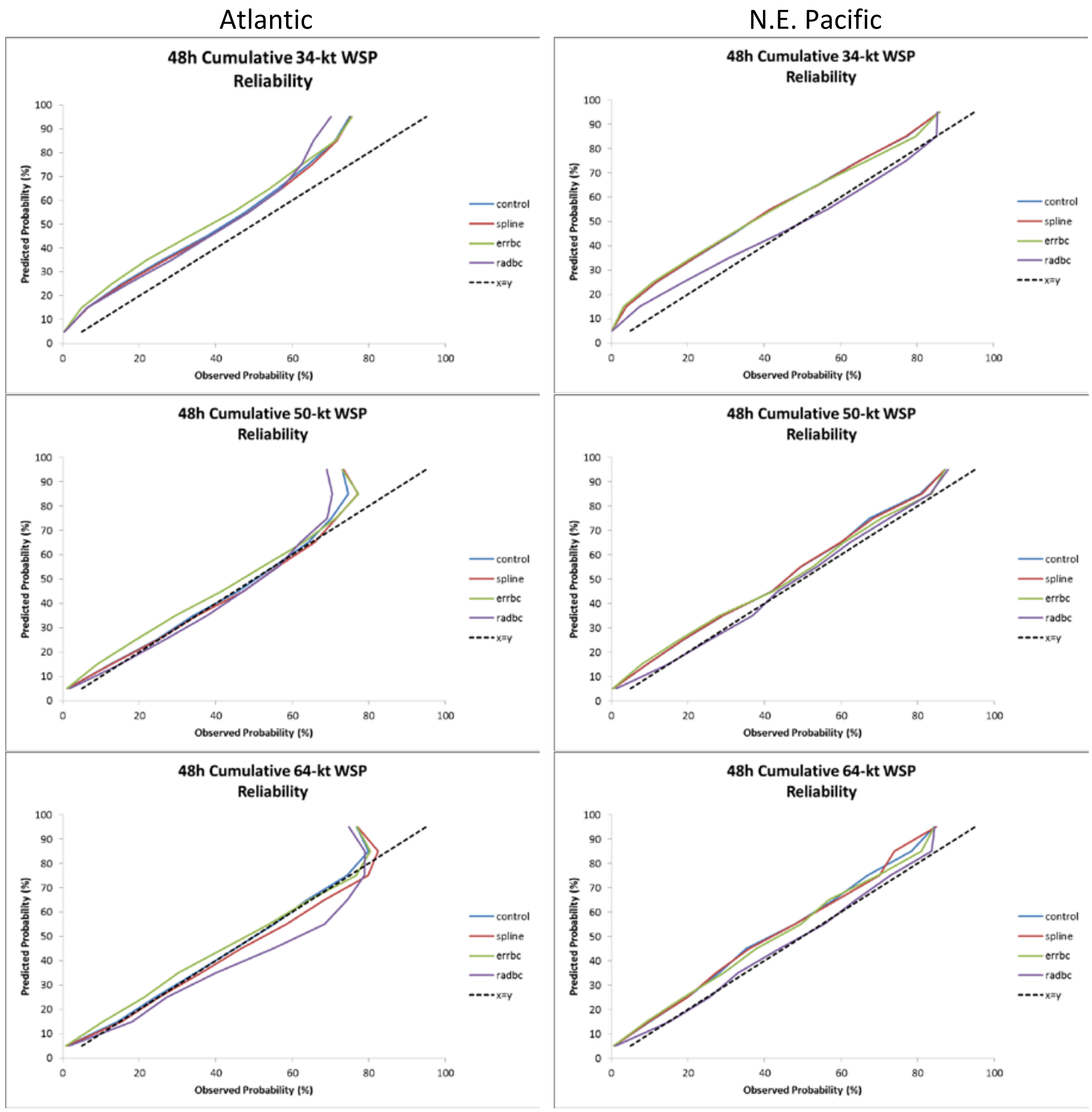

Figure 5. Reliability diagrams for 34-kt (top), 50-kt (center), and 64-kt (bottom) cumulative wind speed probabilities in the Atlantic (left) and N.E. Pacific from 2011-2014 for each potential improvement; spline interpolation (red), error statistic bias correction (green), and radii bias correction (purple). For reference, reliability for the control (blue) and the x=y line (black dashed) are shown.

b. Time of arrival and departure estimates

One of the applications that can be generated with the Hurricane Landfall Probability Application (HuLPA) is the times of arrival and departure of 34, 50 and 64-kt winds at a user-specified point. NHC feedback has indicated that this information would be extremely useful for emergency managers, but generating these values for a wide range of coastal points using HuLPA's post-processing GUI would be very inefficient. This timing information has now been incorporated directly into the MC model. The MC model code was adapted by NHC focal points to estimate the time of arrival/departure of 34-, 50-, and 64-kt winds using various probability thresholds (e.g., Figure 6). The version of the MC model code that includes the arrival/departure estimates was given to CIRA in Year 2 and, with additional support from Hurricane Sandy Supplemental funds, was upgraded to Fortran 90 and prepared for operational implementation. The clean, modularized, F90-compliant code was delivered to Mark DeMaria in June 2015. The code was installed, compiled, and tested on WCOSS and the Automated Tropical Cyclone Forecast (ATCF) system by TSB staff. The test case provided ran successfully on both systems with approximate run times of 2 minutes.

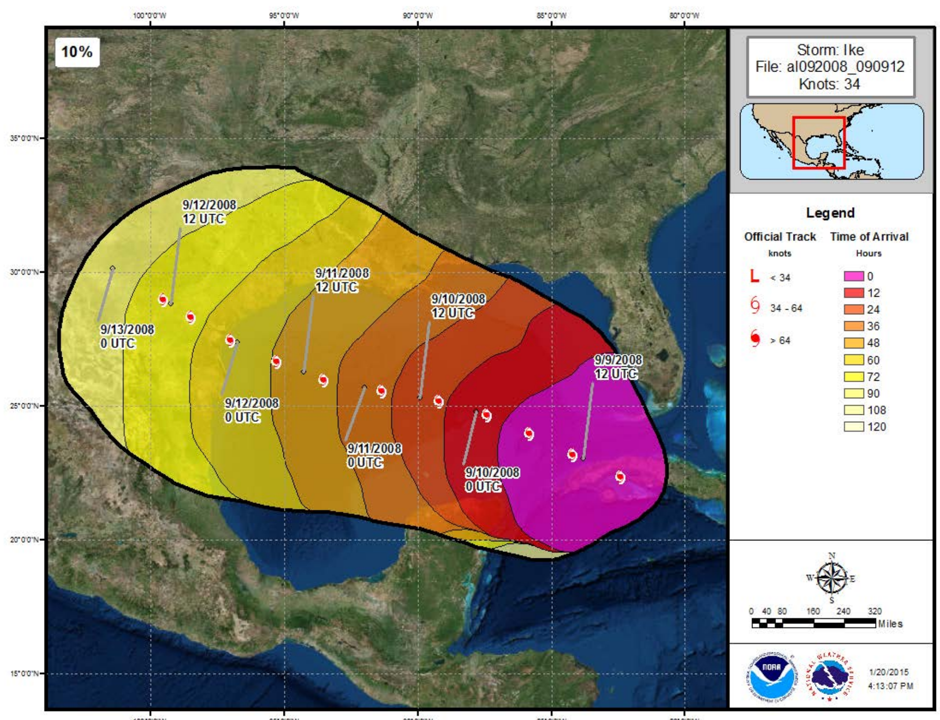

Figure 6. Time of arrival estimates (based on 10th percentile) for Hurricane Ike on 18 September 2008 at 18 UTC generated by the newly upgraded MC model. Image courtesy of Casey Ogden, NHC.

c. Integrated GPCE parameter

It has been shown that past NHC track forecast errors can be separated into terciles based on their corresponding GPCE value, and that track forecast errors in the low (high) terciles tend to correspond to less (more) spread in forecast errors (DeMaria et al. 2013). This finding motivated the use of a GPCE parameter in the MC model. At present, this GPCE parameter determines the track error statistics used by the MC model at each forecast time to estimate wind speed probabilities, but the GPCE categories (low, medium, high) are not output directly.

It is trivial to output the GPCE categories form the MC model, but it would be desireable to do so in a way the provides a single GPCE estimate valid for the entire forecast. As such, a time integrated measure of the GPCE, called time-average normalized GPCE (TANG), was developed from GPCE information currently used in the MC model. This integrated uncertainty measure is obtained by normalizing GPCE at each forecast time by the mean and standard deviation of the last 5 years of GPCE values and then averaging over all forecast times. The same method was applied to track forecast errors to calculate time-average normalized track errors (TANTE).

After discussion with NHC POCs in late year 1, we decided to investigate several different methods of developing a categorical GPCE estimate. Two different thresholding schemes were tested; partitioning GPCE values into equally sized terciles (33%/33%/33%) and a scheme with smaller numbers of high and low values (10%/80%/10%) in order to emphasize extremes. NHC POCs expressed an interest in having integrated GPCE estimates for both the early part of the forecast (days 1 and 2) and the later part (days 3, 4, and 5). 2009-2013 Atlantic TANTE distributions for low (blue), medium (red), and high (green) GPCE values using the two different thresholding definitions are shown in Figure 7 for all 3 forecast times examined (entire forecast, early, and late). These distributions suggest that all TANG thresholds being tested provide a separation between low and high track forecast errors, with low (high) TANG corresponding to lower (higher) errors and less (more) spread. To further investigate the skill of the different thresholds, several cases were chosen from the development sample that have low GPCE values early in the forecast and high GPCE values later on (and vice versa, Figure 8). For this test, we found that the tercile thresholds did a better job of correctly classifying cases at all forecast intervals.

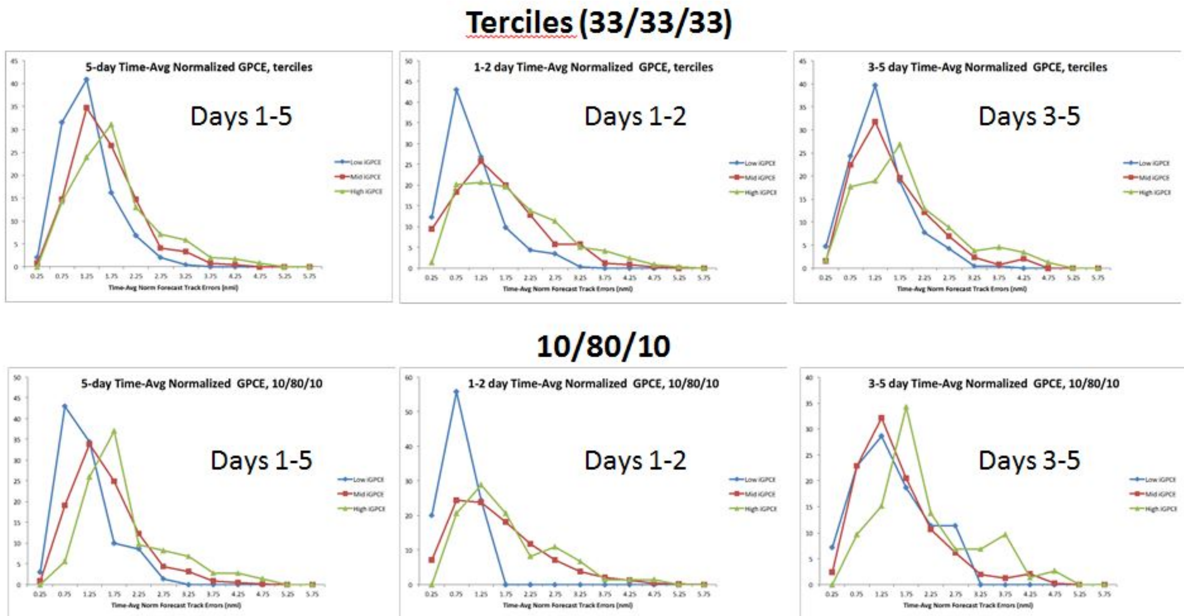

Figure 7. Distributions of 2009-2013 Atlantic time-average normalized track errors for low (blue), medium (red), and high (green) time-average normalized GPCE values, using terciles (top) and 10%/80%/10% (bottom) partitioning to define GPCE categories.

Low GPCE Days 1-2 / High GPCE Days 3-5											TANG 1-2day		TANG 3-5day		TANG 0-5day	
Fcst Hr	12	24	36	48	60	72	84	96	108	120	10/80/10	33/33/33	10/80/10	33/33/33	10/80/10	33/33/33
AL172012 101512	33	54	69	86	130	174	240	306	382	458	M	L	M	H	M	H
AL182011 102606	23	42	66	97	154	211	333	455	607	759	L	L	H	H	H	H
AL162011 093006	26	46	71	100	148.5	197	279	361	427.5	494	M	L	H	H	M	H
AL062010 082706	21	42	65	100	137.5	175	226	277	328	379	L	L	M	H	M	M

High GPCE Days 1-2 / Low GPCE Days 3-5											TANG 1-2day		TANG 3-5day		TANG 0-5day	
Fcst Hr	12	24	36	48	60	72	84	96	108	120	10/80/10	33/33/33	10/80/10	33/33/33	10/80/10	33/33/33
AL032012 062100	41	66	122	154	153	152	177.5	203	229.5	256	H	H	M	L	M	H
AL162011 092112	42	70	96	122	127.5	133	162.5	192	230.5	269	M	H	M	L	M	M
AL052011 080706	52	97	149	165	152.5	140	166.5	193	216	239	H	H	M	L	M	H
AL062010 082118	63	112	146	168	160	152	168	184	189	194	H	H	M	L	M	H

Figure 8. Categorization test results for 4 cases with low GPCE at early forecast times and high GPCE for late forecast times (top) and 4 cases with high GPCE at early forecast times and low GPCE at late forecast times (bottom).

d. Extend wind speed probabilities to 7 days

NHC began making in-house track and intensity forecasts beyond 5 days in 2011. Currently, 148h and 166h forecasts are available in the official a-decks available on the NHC ftp site. Using these extended forecasts, the 6-day and 7-day track and intensity forecast errors were calculated and a 7-day version of the MC model was developed. An example of the cumulative 34-kt wind speed probabilities for Hurricane Isaac out to 5 days (left) and out to 7 days (right) is shown in Figure 9. Brier skill scores were calculated over the 2011-2014 sample cases the Atlantic and N.E. Pacific through 7 days and are shown in Figure 10. The Brier skill scores for the incremental wind speed probabilities drop off quickly to zero after 120 hours, indicating the wind speed probabilities have little skill after 5 days.

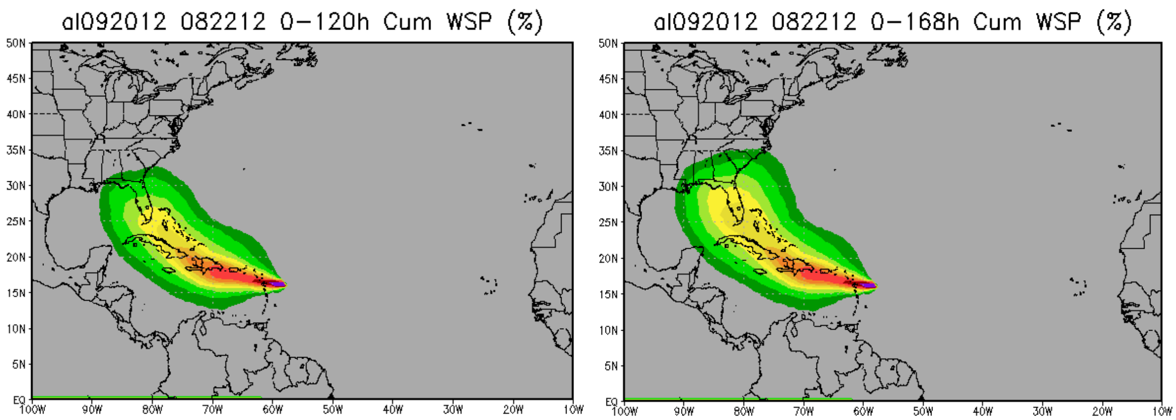

Figure 9. Example of cumulative 5-day (left) and 7-day (right) 34-kt wind speed probabilities for Hurricane Isaac on 22 August 2012 at 12Z.

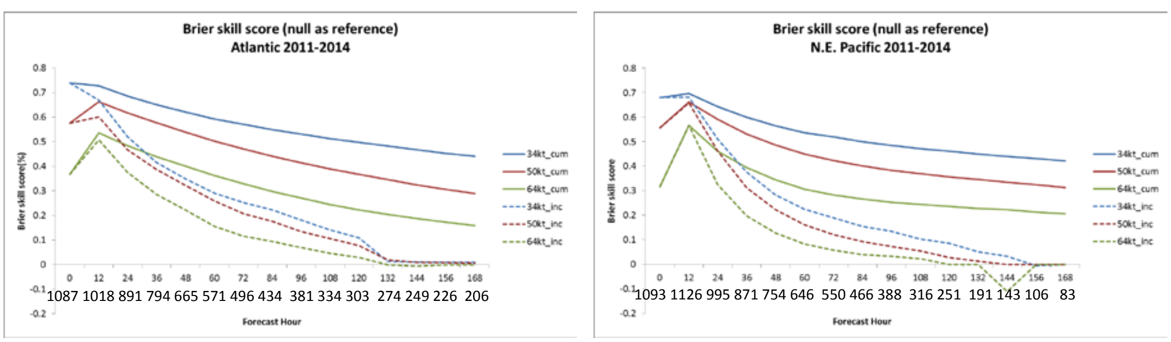

Figure 10. Brier skill scores (using null forecast as reference) for the Atlantic and N.E. Pacific 2011-2014 samples for 7-day wind speed probabilities.

e. Software improvements

Updating the MC model each year is a multi-step process requiring the execution of several scripts and Fortran programs. To streamline this process, the various pieces of script and code have been consolidated and wrapped with a master update script. Annual updates to the MC model now only require the execution of the single script with a run time of approximately 1 minute per basin. This simplification makes it possible to turn over the task of annual updates to NHC, should that be desired.

f. Real-time product demonstration and 2015 verification

Experimental versions of the MC model were developed to incorporate each of the 3 improvements listed above. These 3 experimental versions were run in parallel at CIRA in near-real-time (available approximately 3 hours after synoptic time) during the 2014 and 2015 Atlantic and N.E. Pacific hurricane seasons.

In 2014, plots of the 34-, 50, and 64-kt 120-hr cumulative wind speed probabilities were generated at each synoptic time reflecting each potential improvement listed in 2a. A control version that is nearly identical to the operational wind speed probability product was also run for comparison purposes. These images were made available to NHC POCs for evaluation via an ftp site. Feedback from 2014 suggested that difference plots be added to the demonstration so forecasters could more easily see the effects each updates has on the wind speed probabilities. In 2015, difference plots were added to the parallel run demonstration and the output was moved to the CIRA/RAMMB TC Realtime webpage (http://rammb.cira.colostate.edu/products/tc_realtime/). Both the wind speed probability plots and difference plots are displayed (Figure 11).

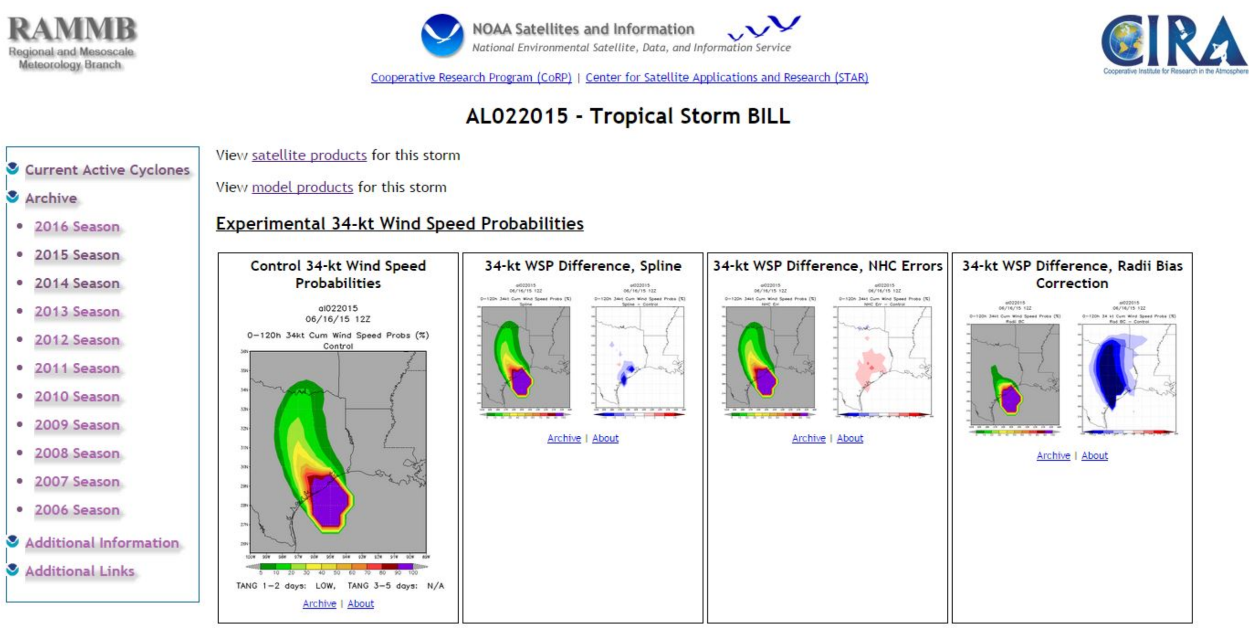

Figure 11. Example of experimental 34-kt wind speed probabilities for tropical storm Bill on the CIRA/RAMMB TC Realtime webpage.

Brier score improvements and reliability for 2015 are shown below in Figures 12 and 13, respectively. Overall, these verification values are quite similar to those for the 2011-2014 sample (Figures 4 and 5). The only notable difference is the negative impact the radii bias correction had on Atlantic 34-kt wind speed probability Brier scores. The impact of the radii bias correction was positive for the 2011-2014 sample. It is possible that the absence of cases much larger or smaller than climatology in 2015 in the Atlantic reduced the opportunities this change had to make a positive impact on Brier scores. More testing of the radii bias correction may be needed to ensure it does not degrade the MC model during average years.

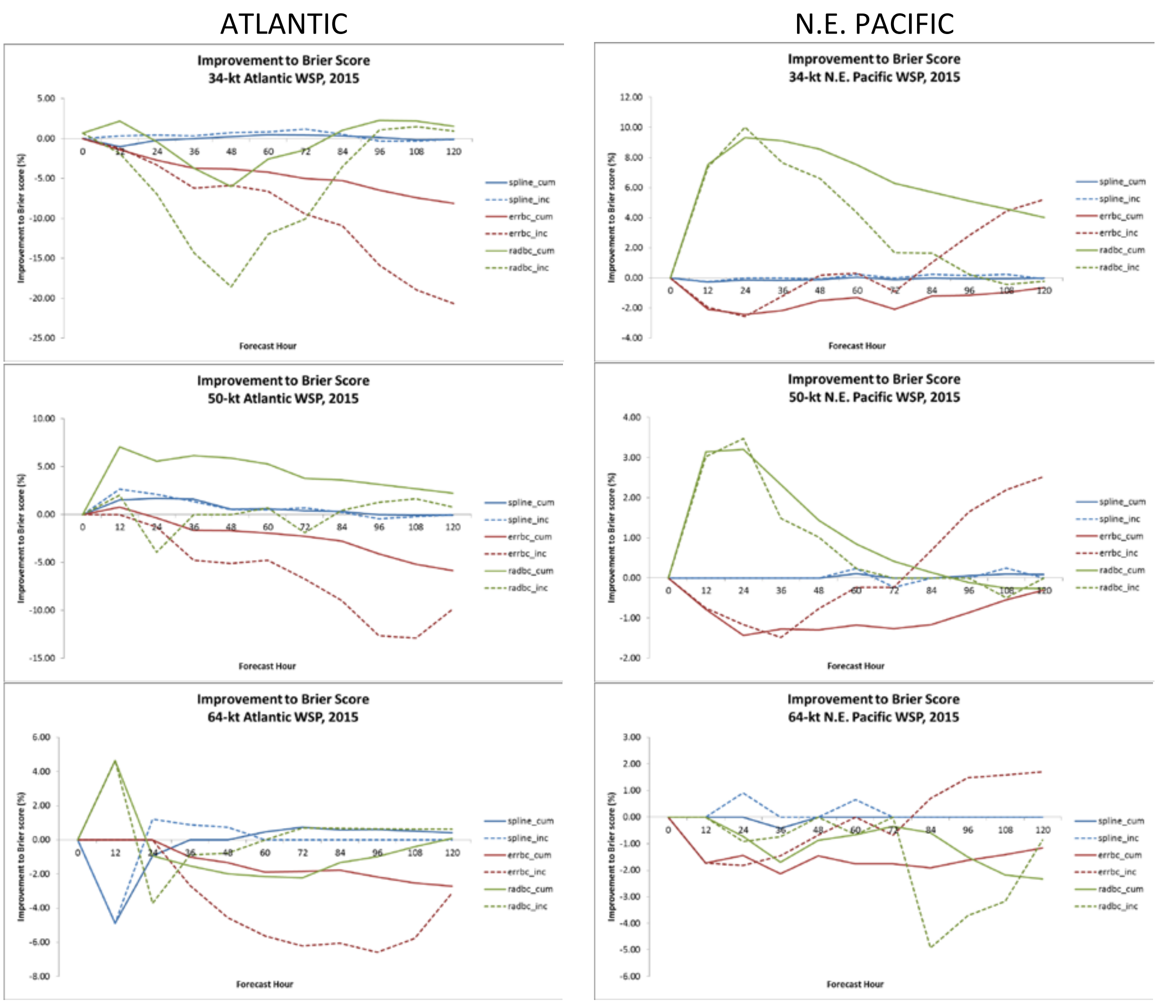

Figure 12. Percent improvement in Brier scores for 34-kt (top), 50-kt (center), and 64-kt (bottom) wind speed probabilities in the Atlantic (left) and N.E. Pacific in 2015 for each potential improvement; spline interpolation (blue), error statistic bias correction (red), and radii bias correction (green) versions of the MC model. Improvement is defined as a reduction in Brier scores from the control MC model over the same sample. Brier score improvements for cumulative (incremental) wind speed probabilities are solid (dashed).

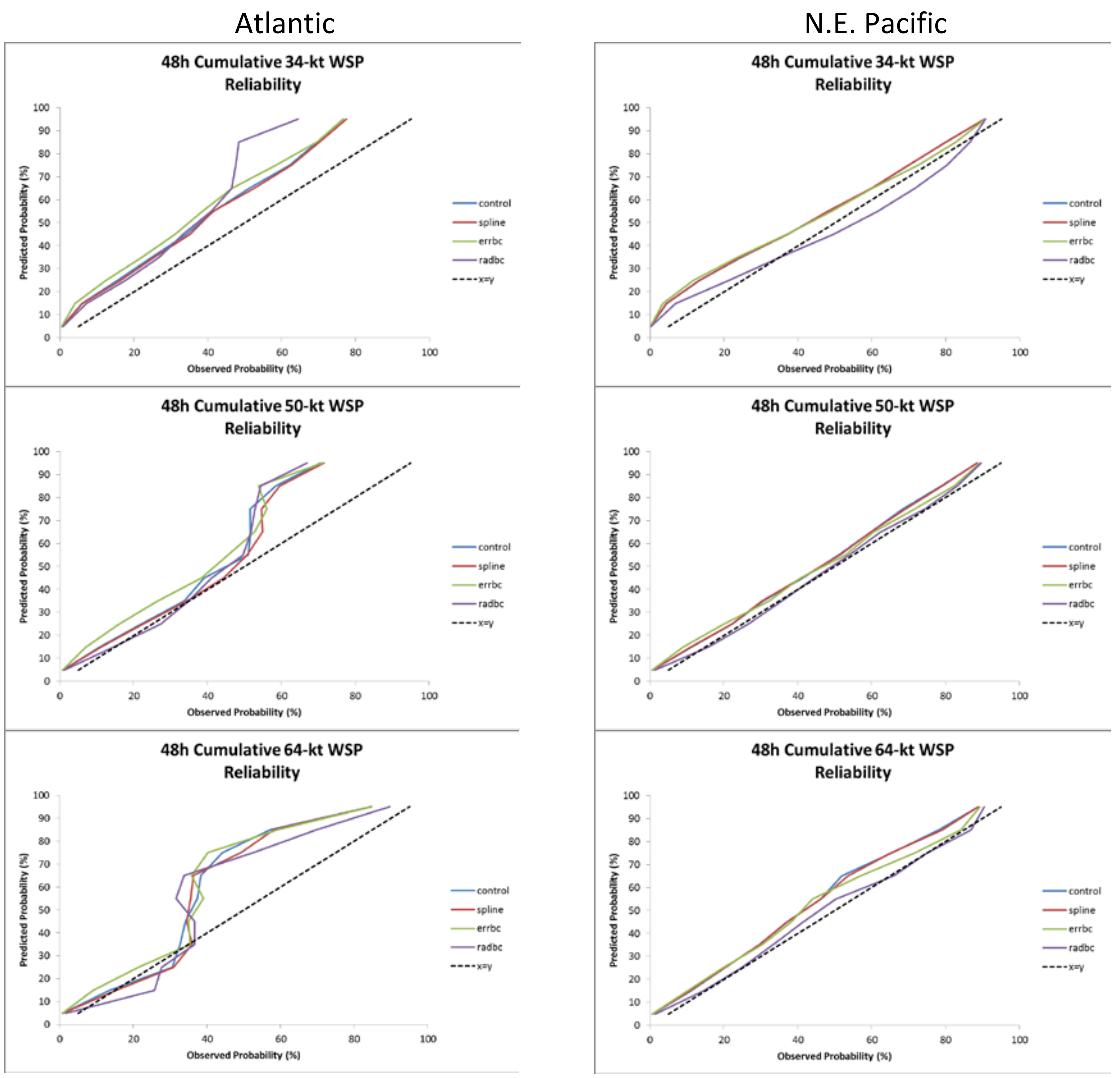

Figure 13. Reliability diagrams for 34-kt (top), 50-kt (center), and 64-kt (bottom) cumulative wind speed probabilities in the Atlantic (left) and N.E. Pacific in 2015 for each potential improvement; spline interpolation (red), error statistic bias correction (green), and radii bias correction (purple). For reference, reliability for the control (blue) and the x=y line (black dashed) are shown.

3. SUMMARY, DEVELOPER RECOMMENDATIONS, AND FUTURE WORK

a. Spline interpolation

Using the spline interpolation scheme instead of linear interpolation for both the official track forecast and the realizations has a negligible impact on both Brier scores and reliability, yet generates more realistic-looking tracks and improves product performance for recurving tropical cyclones.

Developer recommendation: Implementation of this improvement presents little risk of product degradation and should be considered.

b. Error statistics bias correction

Error statistics were standardized and a bias correction was added to the MC model to allow it to exactly reproduce the NHC cone of uncertainty from its sampled track realization errors. The impact of this correction on 2011-2015 Brier scores was mixed but generally small (~±2-4%). While this correction improved reliability in the N.E. Pacific for all wind thresholds it degraded reliability and introduced underprediction for Atlantic 64-kt wind speed probabilities. One benefit of this bias correction is to provide consistency between NHC's uncertainty products. Another potential future benefit would be to enable to ability to obtain the cone of uncertainty directly from the MC model. The advantage to obtaining the cone in this manner is that its size would increase and decrease with forecast confidence obtained from the GPCE parameter.

Developer recommendation: Given the mixed impacts on Brier scores and reliability, more real-time testing of this bias correction may be necessary to verify that the bias correction is working as it should for all cases.

c. Radii bias correction

This bias correction essentially "nudges" the realization radii towards the official R34, R50, and R64 forecast when they are available. This correction made a significant improvement in the appearance of the wind speed probabilities for extremely large and small cases (e.g., Sandy) and improved Brier scores in both basins for 34-kt and 50-kt wind speed probabilities. However, the 64-kt wind speed probability Brier scores were degraded by ~5% in both the Atlantic and N.E. Pacific.

Developer recommendation: Consider a conditional implementation of this improvement to just the 34-kt and possibly 50-kt wind speed probabilities or wait and investigate other bias correction methods that work effectively for all wind thresholds.

d. Time of arrival / departure of 34, 50, and 64-kt winds

This product addition was completed primarily by NHC collaborator Mark DeMaria. The MC model now has a flag that, when set, will calculate the time of arrival of the 34, 50, and 64-kt winds at various probability thresholds and write them to an output text file. This addition of these calculations does increase the run time of the code, but it still runs in a short enough time (under 2 minutes) that should be adequate for operations.

Developer recommendation: The information provided by this addition will be very helpful to users such as emergency managers. Work should continue on determining the best way to display this information for users.

e. Integrated GPCE parameter

An integrated GPCE parameter has been developed and thresholds have been defined to provide integrated GPCE estimates of low, medium, or high for the early (days 1 and 2) and late (days 3, 4, and 5) official track forecasts. Partitioning time-average normalized GPCE values into terciles provided the best discrimination between low and high track errors and hence was chosen as our prototype GPCE categorization algorithm. GPCE categories are calculated as part of the regular MC model processing and can be easily added to the main output text file.

Developer recommendation: This is a no-cost implementation that will provide forecasters condensed GPCE information. Future work should be done to determine its utility in applications such as the generation of GPCE-based cone of uncertainty. As a start, we attempted some preliminary work - see section g below).

f. 7-day wind speed probabilities

The MC model was extended beyond 5 days using 144 hour and 168 hour official forecasts made available post-season in the a-decks. Only 4 years of 6-day and 7-day forecasts were available and there were a relatively small number of cases upon which to develop error statistics. Brier skill scores showed that wind speed probabilities beyond 120 hours had little to no forecast skill over a null forecast.

Developer recommendation: Given this very limited skill it is not recommended that the MC model be extended beyond 5 days at this time. Other methods, such as the incorporation of numerical model guidance and error statistics, could be explored to see if a skill can be improved at long range forecast times.

g. Implementation status and future work

Upgrades a, b, c, d, and e have all been added to the most current operational version of the MC model code as options with flags to turn them on and off. A new version of the code had to be developed to extend the MC model beyond 5 days (upgrade f). Since the beginning of this work, the MC model code has been updated and modularized through a Hurricane Sandy Supplemental project. As such, all upgrades that are accepted for implementation will have to be incorporated into the modernized model code. However, since A. Schumacher was involved with both projects this task should not take a significant amount of extra time or effort.

As a first effort to see how the integrated GPCE parameter developed in this study could be used to incorporate GPCE information into the size of the cone of uncertainty, prototype cone sizes for 2014 were created using the integrated GPCE thresholds derived for the entire forecast time (Table 1) and those created for the early and late forecast periods (Table 2). LOW and HIGH cone sizes are based on the 67th percentile of track errors for cases with the lowest and highest TANG values, respectively. The 2014 official cone sizes are shown for comparison. Using the TANG thresholds developed for the entire forecast period (Table 1) provides a cone that is smaller when TANG is smaller and larger when TANG is larger for all forecast times except 120 hours. The same is true when TANG thresholds developed for just the early (days 1 and 2) and late (days 3, 4, and 5) are used (Table 2), although LOW cone size < official cone size < high cone size is true at all forecast times using this method. This work is preliminary and just meant to demonstrate a single method that could be used for creating different cone sizes based on GPCE (i.e., track uncertainty). Future work could be done to test this method, as well as other methods such as the generation of a cone directly from the MC model, in the future.

Forecast Hr	2014 Cone (nmi)	LOW TANG (Lowest 33%)	HIGH TANG (Highest 33%)
12	32	28	42
24	54	46	68
36	71	70	93
48	91	82	113
72	124	102	139
96	169	157	186
120	223	215	219

Table 1. Cone sizes determined by integrated GPCE categories low and high for the entire forecast period.

Days 1 & 2			
Forecast Hr	2014 Cone	LOW TANG (Lowest 33%)	HIGH TANG (Highest 33%)
12	32	21	41
24	54	35	68
36	71	52	85
48	91	69	110
Days 3, 4, & 5			
72	124	109	139
96	169	159	196
120	223	215	235

Table 2. Cone sizes determined by integrated GPCE categories low and high for the early (days 1 and 2) and late (days 3, 4, and 5) forecast periods.