Improvements to the Operational Tropical Cyclone Wind Speed Probability Model

MARK DEMARIA,* JOHN A. KNAFF,* MICHAEL J. BRENNAN,+ DANIEL BROWN,+
RICHARD D. KNABB,+ ROBERT T. DEMARIA,# ANDREA SCHUMACHER,#
CHRISTOPHER A. LAUER,@ DAVID P. ROBERTS,& CHARLES R. SAMPSON,**
PABLO SANTOS,++ DAVID SHARP,## AND KATHERINE A. WINTERS@@

* *NOAA/NESDIS, Fort Collins, Colorado*
+ *NOAA/NWS/NCEP/NHC, Miami, Florida*
# *Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, Colorado*
@ *NOAA/NWS/NCEP/SWPC, Boulder, Colorado*
& *Fleet Weather Center, Norfolk, Virginia, and NOAA/NWS/NCEP/NHC, Miami, Florida*
** *NRL, Monterey, California*
++ *NWS, Miami, Florida*
NWS, Melbourne, Florida
@@ *45th Weather Squadron, Patrick Air Force Base, Florida*

(Manuscript received 18 October 2012, in final form 24 March 2013)

ABSTRACT

The National Hurricane Center Hurricane Probability Program, which estimated the probability of a tropical cyclone passing within a specific distance of a selected set of coastal stations, was replaced by the more general Tropical Cyclone Surface Wind Speed Probabilities in 2006. A Monte Carlo (MC) method is used to estimate the probabilities of 34-, 50-, and 64-kt ($1\,\text{kt} = 0.51\,\text{m s}^{-1}$) winds at multiple time periods through 120 h. Versions of the MC model are available for the Atlantic, the combined eastern and central North Pacific, and the western North Pacific. This paper presents a verification of the operational runs of the MC model for the period 2008–11 and describes model improvements since 2007. The most significant change occurred in 2010 with the inclusion of a method to take into account the uncertainty of the track forecasts on a case-by-case basis, which is estimated from the spread of a dynamical model ensemble and other parameters. The previous version represented the track uncertainty from the error distributions from the previous 5 yr of forecasts from the operational centers, with no case-to-case variability. Results show the MC model provides robust estimates of the wind speed probabilities using a number of standard verification metrics, and that the inclusion of the case-by-case measure of track uncertainty improved the probability estimates. Beginning in 2008, an older operational wind speed probability table product was modified to include information from the MC model. This development and a verification of the new version of the table are described.

1. Introduction

Scientific and technological advances during the past few decades have contributed to enhancements in operational tropical cyclone analyses and forecasts. This progress has occurred during an era when coastal population and development have continued to increase, in many cases increasing the lead time at which officials and residents of these areas must take certain actions in advance of the various hazards posed by an approaching tropical cyclone. Operational forecast errors are still sufficiently large such that the uncertainties in these forecasts must be taken into account to make sound preparedness decisions. While most users acknowledge that deterministic forecasts have uncertainties, it is challenging for them to account for those uncertainties without additional information from the forecasters. Products are needed that better convey forecast uncertainties and enhance users' decision-making and preparedness actions in response to a tropical cyclone.

The National Hurricane Center (NHC) began issuing "strike probabilities" with their forecasts of Hurricane Alicia in August of 1983 (Sheets 1985) in response to the need to convey forecast uncertainty. The original strike

Corresponding author address: Mark DeMaria, NOAA/NESDIS/STAR, CIRA/CSU, 1375 Campus Delivery, Fort Collins, CO 80523.
E-mail: mark.demaria@noaa.gov

DOI: 10.1175/WAF-D-12-00116.1

probability product only considered track forecast uncertainties estimated from bivariate normal distributions fitted to the recent history of NHC track forecast errors. A tropical cyclone strike was defined as the passing of the center of a tropical cyclone 50 n mi to the right or 75 n mi to the left of a given location, and probabilities were provided at selected locations from 12 to 72 h (the discrete lead times of the NHC deterministic forecasts prior to 2003). Except for periodic updating of the track error statistics, the operational strike probability product changed very little from 1983 through the 2005 hurricane season.

The strike probabilities were used by many emergency managers and other decision makers to account for tropical cyclone track forecast uncertainties, whereby the ''close'' passage of the cyclone center was used as a proxy for weather effects from the cyclone. However, the strike probabilities did not account for uncertainties in the forecast intensity or size of the tropical cyclone and the resulting probabilities did not convey information about specific weather hazards that could be experienced at a given location.

A new set of products, the tropical cyclone surface wind speed probabilities, replaced the strike probabilities in 2006 (DeMaria et al. 2009, hereafter D09) following an experimental phase during 2004–05 that was supported by the Joint Hurricane Testbed (JHT; Rappaport et al. 2012). As with the strike probabilities, these new products were primarily developed for more sophisticated users of forecast information, such as government officials and other decision makers to support cost–benefit analyses. The new technique uses a Monte Carlo (MC) method to estimate the probability of winds of at least 34, 50, and 64 kt (1 kt = 0.51 m s^{-1}) at specific locations within multiple time periods out to 120 h. Probabilities are estimated for a set of well-known locations near the coast as well as for a regularly spaced latitude–longitude grid covering a very large domain. Versions are available for the Atlantic, the combined eastern and central North Pacific (hereafter referred to as East Pacific) and the western North Pacific (hereafter West Pacific). D09 described the MC probability model and presented verification statistics for the 2006–07 seasons. In this paper, updated verification results through the 2011 season will be presented.

The original version of the MC model randomly sampled official track and intensity error distributions from the previous 5 yr, which were then added to the official track and intensity forecast to generate the 1000 realizations. This paper describes a new method implemented in 2010, where the track uncertainty is estimated on a case-by-case basis. Recent studies have shown that carefully designed ensemble forecast systems can provide information on the uncertainty of a number of forecast parameters, including tropical cyclone tracks (e.g., Hamill et al. 2011). However, one of the constraints of the operational MC model is that the probabilities need to be consistent with the official track and intensity forecast, making direct inclusion of ensemble information from models problematic. In addition, the sizes of operational global model forecast ensembles are typically much smaller than 1000, and the horizontal resolution is inadequate to provide unbiased estimates of the cyclone intensities and wind radii. Also, the error spread of the ensembles can sometimes be too small relative to the forecast errors (i.e., underdispersive).

For these reasons, a simpler approach was undertaken to include ensemble track information in the MC model. Goerss (2007) developed a method of statistically estimating the error of a consensus track forecast using the spread of a small set of operational track forecast models and several other parameters such as the forecast storm intensity. This method produces the Goerss predicted consensus error (GPCE), which is available in real time at all the U.S. tropical cyclone forecast centers [NHC, the Central Pacific Hurricane Center (CPHC) and the Joint Typhoon Warning Center (JTWC)]. For use in the MC model, the official track forecast errors were stratified into terciles by the GPCE values. This stratification resulted in a very consistent relationship between the GPCE tercile category and the track errors, with increasingly broad along- and cross-track error distributions for the terciles with the larger GPCE values. These stratified error distributions are then used to modify the real-time MC model runs. After testing in 2008 and 2009, the GPCE version of the MC model was implemented operationally beginning in 2010. This paper describes the incorporation of the GPCE information in the MC model and evaluates its impact.

This paper also describes the use of the MC model to replace an older NHC intensity probability product, which was designed to complement the strike probability product. Beginning in 1996, a wind speed probability table (WSPT) was provided for the East Pacific (but without the central Pacific part) and the Atlantic via the NHC web page (Rappaport et al. 2009). This product estimated the probability that the cyclone's maximum wind would lie in various intensity ranges (dissipated, tropical depression, tropical storm, hurricane, and the five categories on the Saffir–Simpson hurricane wind scale) out to 72 h. Beginning in 2008, this table was modified to use the intensity input from the 1000 realizations from the MC model, rather than from fixed probability distributions from historical NHC intensity errors. The table information was also extended from

72 to 120 h at that time. A verification of the new WSPT product for 2008–11 will be presented.

The MC model is reviewed in section 2, the GPCE version is described in section 3, verification results are presented in section 4, the WSPT and its verification are described in section 5, and conclusions are presented in section 6.

2. The Monte Carlo wind speed probability model

The MC model and the operational products derived from it were described in detail in D09. The MC model estimates probabilities of the magnitude of the wind vector (wind speed), but does not provide any information on wind direction. For brevity, the term ''wind'' is assumed to mean ''wind speed'' in the remainder of the paper. All of the products are determined from a set of 1000 plausible storm tracks generated by randomly sampling from the previous 5 yr of track errors from the operational forecast centers (NHC, CPHC, and JTWC). Versions are available for the Atlantic, and the East and West Pacific. Each realization has corresponding maximum wind estimates, which are also determined from the previous 5 yr of the operational forecast errors. Adjustments are applied to the intensity of a realization for times when the official track crossed land but the realization track did not, and vice versa. The wind structure along the track of each realization is estimated from a simple wind radii climatology and persistence (CLIPER) model (Knaff et al. 2007) and its error distributions, given the track and intensity. The structure is defined in terms of the radii of the 34-, 50-, and 64-kt winds in four directions (NE, SE, SW, and NW) relative to the storm center. These radii are azimuthally interpolated to provide a wind radii estimate at any given azimuth. The radii-CLIPER model starts with the $t = 0$ h radii estimates from the operational forecast centers, and then relaxes those toward climatological radii estimates. By about 36 h, the radii estimates are almost entirely from the climatological values.

Serial correlations of the track, intensity, and wind structure errors are taken into account. For example, the 12-h position in a realization is determined by randomly sampling from the along- and across-track error distributions of the operational center forecasts from the previous five years. The 24-h track errors are then predicted from the 12-h errors using a first-order autoregressive procedure to account for the serial correlation, and then a random component is added to that estimate, based on the residuals from the fit of the autoregressive estimate. Along- and across-track errors were utilized because the autoregressive estimate is more accurate compared to when the track errors are partitioned into geographic components. Similar procedures are used to account for serial correlations of the intensity and structure errors.

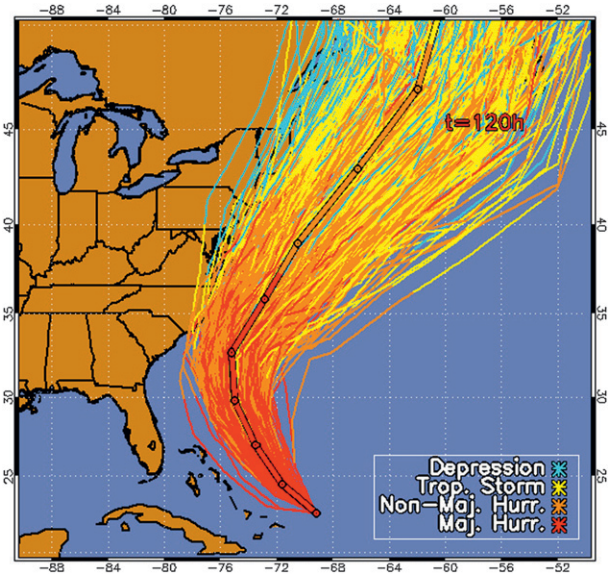

FIG. 1. The tracks of the 1000 realizations in the MC model for the forecast beginning at 0000 UTC 1 Sep 2010 for Hurricane Earl. The NHC official forecast is indicated by the thick line near the center of the 1000 realizations and the points on that track are at 12-h intervals. The intensities of the NHC official forecast and the realizations are indicated by the colors.

Figure 1 shows an example of the tracks and intensities of the 1000 realizations and the NHC official forecast for a case from Hurricane Earl during the 2010 Atlantic Hurricane season. The NHC official forecast for this case did not make landfall in North Carolina, but several of the realizations did. The impact of the land correction can be seen since most of the realizations that did cross land in North Carolina (to the left of the NHC forecast) have lower intensities than those of the official forecast and the realizations to the right of the NHC track.

Once the 1000 realizations are generated, the probabilities of 34-, 50-, or 64-kt winds at any given point are determined simply by counting the number of realizations where that point came within the area of 34-, 50-, or 64-kt winds during the time period of interest. Cumulative (0–6 h, 0–12 h, ...) and incremental (0–6 h, 6–12 h, ...) probabilities are provided out to 120 h. For some National Weather Service Weather Forecast Office (WFO) applications, probabilities in 12-h increments are needed. These can be calculated from the 6-h cumulative and incremental probabilities as described in D09.

The operational MC model was run in 2006–08 with no changes other than an update of the track and intensity error distributions at the beginning of each year to

include cases from the previous 5 yr. When data for a new year are added to the error distributions, the data from the oldest year are removed so the error distributions always include the most recent 5 yr. In 2009 a code optimization was applied, in addition to updating the error statistics.

In 2010, two other changes were made to the model. In the original model the underlying time step for the calculation was 2 h. However, for very small or very fast-moving storms, this time step was too large, making the probabilities unrealistically noisy when plotted as a function of latitude and longitude. For this reason, the time step was decreased to 1 h.

A more significant modification to the 2010 MC model was the inclusion of track error distributions stratified by the GPCE values. This version is described in the next section and was also run in 2011. The probability distributions were also updated to include the previous 5 yr for the 2011 season. However, due to difficulties associated with computer transitions and a small coding error identified during the first 2011 East Pacific cyclone (Adrian), the 2010 version was run in all basins in 2011 starting with the second East Pacific cyclone (Beatriz). The error distributions for the 2011 version are very similar to those from 2010, so this delay in updating the distributions was not a serious problem. The error distributions were updated for 2012.

A few other minor model changes have been implemented since 2007. An examination of the intensities for the realizations over land showed that the random perturbations sometimes resulted in maximum winds that were too high. A bias correction was implemented in 2009 that prevents the intensity in a realization from exceeding the observed maximum wind as a function of the distance inland, developed from a large sample of U.S. landfalling Atlantic storms (1967–2007). For example, for cyclones that are 500 km inland, the highest observed maximum wind was 40 kt, so the intensities in a realization cannot exceed that value when they are 500 km inland. In 2012 an improved method for estimating the inner radii of 34-, 50-, and 64-kt winds was added. As described in D09, the probabilities are determined by counting the number of realizations at a point that come between the inner (inside the radius of maximum wind) and outer wind radii for each threshold. The inner and outer radii are azimuthally interpolated to each grid point from values along four radial directions. However, for cases where the maximum wind is close to the wind threshold of interest, the inner radii at some of the four azimuths were zero, resulting in interpolated radii values that are too small. The new azimuthal interpolation method uses extrapolation from the nearest nonzero value rather than interpolation in the cases between nonzero and zero radii values. Also in 2009, a check was implemented to make sure the track error changes over 12-h intervals (e.g., the difference between the track error at 24 h and the track error at 12 h) from the random sampling never exceed the maximum change in the original official forecast track error distributions.

3. Inclusion of track error uncertainty

As described above, the tracks for the 1000 realizations are determined by randomly sampling from the previous 5 yr of operational track forecast errors. These error distributions are basin wide (Atlantic, East Pacific, or West Pacific), and so do not contain any information about a specific forecast case. Goerss (2007) developed a parameter called GPCE that estimates the track error of a consensus forecast based on the spread of the forecast tracks in the models that contributed to the consensus. The GPCE parameter is available in real time for all the basins where the MC model is run, and was used to provide forecast-specific information in the probability estimates.

When GPCE was first developed, the consensus model on which it was based, ''CONU'', included track forecasts from three global models and two regional models: the National Centers for Environmental Prediction (NCEP) Global Forecasting System (GFS), the Met Office global model (UKMet), the U.S. Navy Operational Global Atmospheric Prediction System (NOGAPS), the NCEP version of the regional coupled Geophysical Fluid Dynamics Laboratory (GFDL) hurricane model, and a version of the GFDL model run by the U.S. Navy (GFDN). Because none of these models is available by the time the forecast is issued, CONU uses the runs from the previous forecast cycle adjusted so that the initial position matches the position of the tropical cyclone at the beginning of the forecast period (these adjusted models are sometimes called the interpolated or early models). GPCE parameters for each forecast are available back to 2004, although the models included in the consensus on which GPCE is based have changed over the past few years (e.g., Cangialosi and Franklin 2011). The primary predictor of track error in the GPCE parameter is the spread of the multimodel consensus tracks. Other predictors are also included, such as the forecasted intensity.

As described above, the GPCE parameter was developed to provide an uncertainty measure of a dynamical model consensus track forecast. However, an analysis by Hauke (2006) showed that GPCE also provides uncertainty information about the NHC official track forecast, and suggested that it could be used to

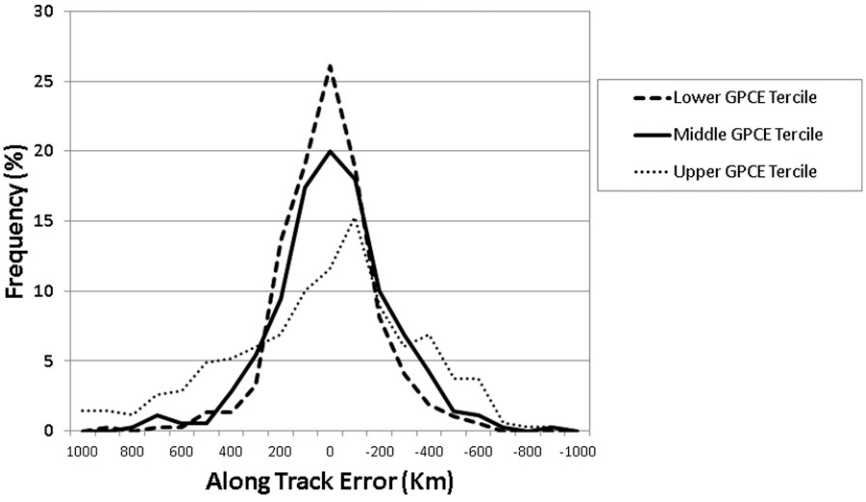

FIG. 2. Along-track error distributions for the 72-h Atlantic forecasts from NHC for 2005–09, stratified by the 72-h GPCE values.

improve the MC model. To confirm those results, the 5-yr samples of official forecast track errors for the three basins were divided into terciles based on the GPCE values. It turned out that there was a very consistent relationship between the GPCE values and the spread of the along- and across-track official forecast error distributions. Figure 2 shows an example for the 72-h Atlantic across-track errors from 2005 to 2009. The spread of the error distributions increases for each GPCE category, so that the track spread increases when the errors are sampled from the distributions with successively higher GPCE terciles. As a further test, the standard deviations of the across- and along-track errors distributions at 12–120 h were calculated for each tercile. These results (not shown) indicate that the standard deviations of both the across- and along-track errors increased monotonically with the GPCE tercile at nearly every forecast period in every basin, providing further confirmation that the GPCE parameter can be used to provide uncertainty estimates of the official track forecasts.

For the real-time MC model forecasts, the GPCE value at each forecast time is provided as model input. The corresponding tercile at each forecast time is then determined from the GPCE thresholds used to stratify the track errors, based on the previous 5-yr sample. The MC model then samples from the appropriate error distributions. It is fairly common for the GPCE category to change during the forecast period. For example, the value could be in the lower category for 12–48 h but then switch to the middle or upper category for the rest of the forecast. This does not create a problem in the simulations, however, because the method used to include serial correlation eliminates abrupt track changes that might result from suddenly sampling from a broader or narrower error distribution. Also, in practice, the GPCE categories vary fairly smoothly during the 120-h forecast period.

The impact of the GPCE input depends on the GPCE category. For cases where the GPCE values are in the middle tercile, the resulting probabilities are not much different than the version without the GPCE input. When the GPCE values are mostly in the upper (lower) tercile, the probabilities tend to decrease (increase) close to the official forecast track, but increase (decrease) away from that track. Figure 3 shows an example of the impact of the GPCE input on the 120-h cumulative probabilities for a case from Hurricane Gustav from the 2008 season. In this case, the GPCE values were nearly all in the lower tercile, with a few in the middle tercile near the end of the forecast. Figure 3 shows the cumulative probabilities with GPCE minus those without GPCE. The probabilities of 50-kt winds increase by up to about 10% close to the NHC track, and decrease by up to about 7% away from the track, resulting in a distribution more tightly clustered about the official forecast track.

Several tests of the version of the MC model with the error distributions stratified by the GPCE parameter were performed before that version was made operational in 2010. Postseason reruns of the GPCE version of the MC model were made for 2008 Atlantic storms within 1000 km of the U.S. coast and compared with the operational version. The Brier score and threat score (see section 4 below for details on these metrics) were used to compare the two versions of the model. Results

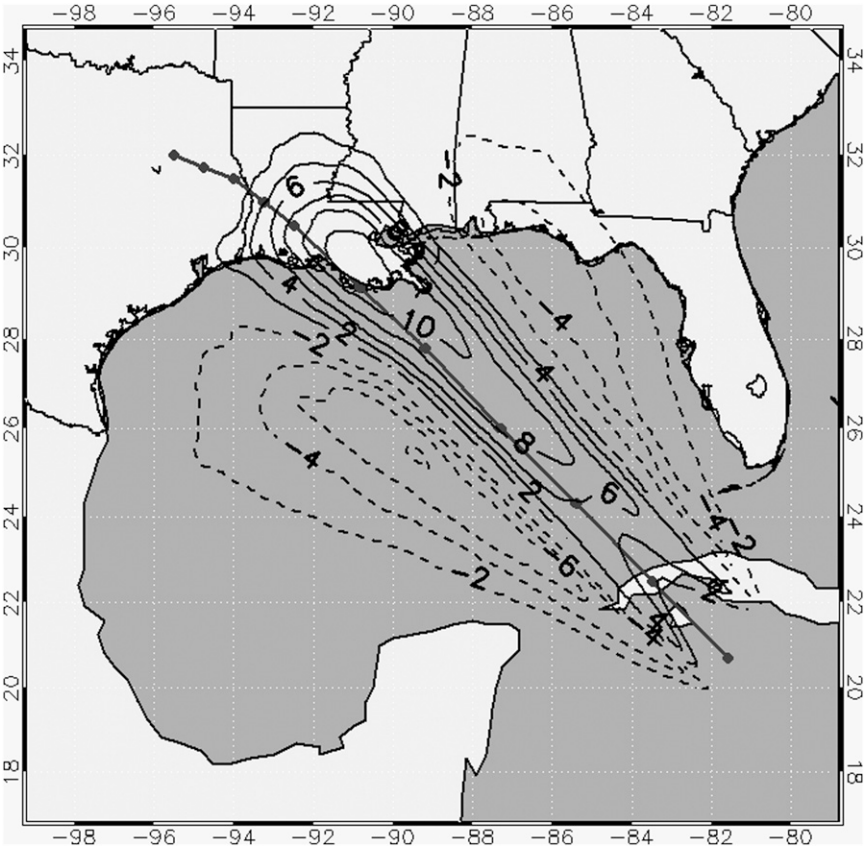

FIG. 3. The difference in the 0–120-h cumulative probability of 50-kt winds for the MC model run with and without the GPCE input for Hurricane Gustav initialized at 1200 UTC 30 Aug 2008. In this case, the GPCE values were nearly all in the lower tercile. Positive (negative) values indicate that the probabilities were higher (lower) with the GPCE input.

showed that the inclusion of GPCE improved the verification at the majority of time periods through 120 h. A more comprehensive test was performed in 2009, where the GPCE version was run in parallel for all cyclones in all three basins. Verification results showed improvements in the threat scores at all time periods in all three basins for all wind radii. The improvements were as high as 15%, with the largest in the Atlantic, and the smallest in the East Pacific. The Brier score was also improved for all forecast periods in the Atlantic and at most time periods in the East and West Pacific. Based on these results, NHC made the GPCE version operational beginning in 2010. An evaluation of the impact of GPCE on the operational runs of the MC model for the 2010 and 2011 seasons is described in section 4.

4. MC model verification

In this section the operational MC model probability forecasts are evaluated. D09 evaluated the forecasts for the years 2006 and 2007. Here, we will follow the same methodology, but concentrate our analysis on the years 2008–11, producing similar graphics to those discussed in D09 so that direct comparisons can more easily be made. It should be noted that only the 2010–11 models included the GPCE input as described in section 3.

a. Evaluation methodology

The MC model verification methodology is described in detail in section 4a of D09. Basically, the 6-h cumulative and incremental probabilities produced operationally on a 0.5° latitude–longitude grid are verified against the NHC, CPHC, and JTWC best-track data. The best-track positions and wind radii are used to create verification grids, where the probabilities at each point are assigned to be 1 (0) if the wind speed of interest did (did not) occur within the time period of the forecast. The entire best track is used for this purpose, including the tropical, subtropical, and extratropical stages of the cyclones. A bias correction is applied to the best-track wind radii to account for the fact that they represent the maximum in each quadrant (see D09 for details). The

probabilities in the three basins are verified individually. A verification of the combined basin forecasts is also provided.

All operational MC model runs from 2008 to 2011 were included in the verification. Although the three basins are verified separately, when multiple TCs are present at the same time, sometimes contributions from East Pacific storms appear in the Atlantic domain, and vice versa. Therefore, it was necessary to include every forecast period that had an Atlantic or East Pacific TC in the verification samples for those two basins. There was no overlap with the West Pacific domain, so those were verified only for those times with a TC in that basin. With these conditions, the Atlantic and East Pacific samples include 1595 forecast cases and the West Pacific sample includes 1988 cases. The combined basin sample includes 2278 cases.

As described above, the 12-h probabilities are used in some WFO applications, but most of the verification results in this section are for the 6-h values. This is not a problem for the cumulative values because the 12-h values are a subset of the 6-h values. For the incremental probabilities, the 12-h values (not shown) are generally a little smoother than the 6-h values and usually lie between the verification curves for the 6-h incremental and cumulative probabilities.

Hamill (1999) describes statistical significance testing for probabilistic forecasts. One of the difficulties is the estimation of the number of degrees of freedom. For example, the Atlantic basin sample described above contains 1595 MC model runs. The Atlantic model grid includes 21 681 probability values at each forecast time, so more than 30 million values are included in the verification statistics for each wind threshold. However, these are not all independent. Hamill (1999) indicates that a conservative method for estimating the degrees of freedom is to assign one degree of freedom per model field. That method is used in the significance testing described below. A common metric for evaluating probabilistic forecasts is the Brier score, which is the average of the squared difference between the probabilistic forecast and the observed value, where the observed value is 1 if the event occurred and 0 if it did not. A perfect forecast would have a Brier score of zero. Brier scores by themselves are somewhat difficult to interpret, especially for cases like the MC model where the probabilities are zero over large portions of the domain. The utility of Brier scores is enhanced when they are compared with Brier scores from a reference forecast, such as a simple climatological probability forecast. Then, the percent improvement over the Brier score of the reference model can be calculated to give a Brier skill score (BSS). Several different reference models are used in the evaluation results below. The use of the Brier score from a reference forecast also provides the basis for the statistical significance testing. Hamill (1999) showed that the simple paired t test is appropriate for comparing Brier scores between two models. In the comparisons described below, a Brier score difference is considered statistically significant if the null hypothesis that there was no difference between the Brier score from the MC model and the reference model could be rejected at the 95% level.

b. Evaluation results

The purpose of verification is to answer specific questions about the forecasts. To determine the gross calibration of the model, the multiplicative biases (hereafter referred to as bias) are calculated [using Eq. (9) of D09] as the ratio of the sum of all the probabilities for a given location and time interval to the actual number of events that occurred. If the bias is less than (greater than) one, then the forecast probabilities are too small (large), on average for that forecast period. The biases are shown in Fig. 4 as a function of forecast lead time for the Atlantic (1°–50°N, 110°–1°W), the East Pacific (1°–40°N, 180°–75°W), the West Pacific (1°–50°N, 100°E–180°), and the entire domain (combined; 1°–60°N, 100°E–1°W).

Figure 4 shows that the cumulative probabilities have relatively small biases for the entire domain. For the Atlantic, the 34- and 50-kt cumulative probabilities have very small biases, with somewhat of a high bias for the 64-kt probabilities. Positive biases are evident for all almost all radii and times in the East Pacific, with low biases in the West Pacific. The biases for the incremental probabilities tend to be higher than those for the cumulative probabilities, especially for the 34-kt probabilities in the East Pacific. These results are fairly similar to those presented in D09 with a high bias in the East Pacific, although the magnitude of the low bias in the West Pacific is a little larger for the more recent sample in Fig. 4. Nonetheless, the biases of the combined domain in Fig. 4 show good gross calibration, as was also seen in D09 for the 2006–07 sample.

The biases in Fig. 4 at the initial time can only be caused by biases in the operational intensity and radii estimates relative to the final best-track values. Verification of the official intensity forecasts showed that at $t = 0$, the Atlantic and East Pacific intensities biases were near zero, but the West Pacific biases were negative, consistent with the results in Fig. 4. At the later forecast times, the Atlantic and East Pacific official intensity forecasts had a high bias, again consistent with Fig. 4. The magnitudes of the Atlantic and East Pacific official intensity forecast biases were generally less than

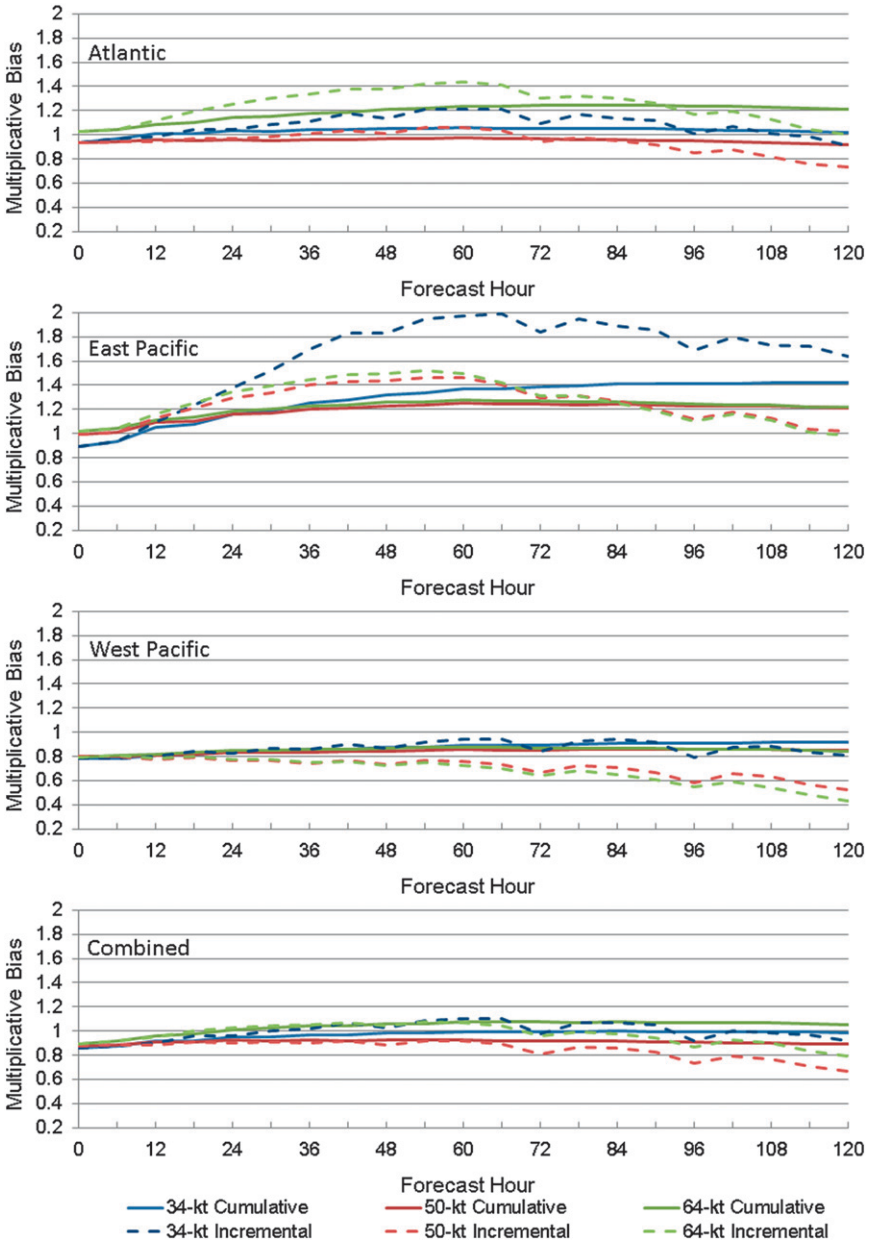

FIG. 4. The multiplicative biases associated with the 2008–11 MC model verification in the North Atlantic (1°–50°N, 110°–1°W), East Pacific (1°–40°N, 180°–75°W), West Pacific (1°–50°N, 100°E–180°), and the combined multibasin domain (1°–60°N, 100°E–1°W) are shown in the panels starting from the top, respectively. Biases for the cumulative probabilities are given by solid lines and for the incremental probabilities they are given by dashed lines. Blue, red, and green lines correspond to the biases associated with the 34-, 50-, and 64-kt wind probabilities, respectively.

about 5%, which is smaller than the biases in the probabilities seen in Fig. 4 at the longer forecast periods. However, the intensity biases are amplified by the MC model. For those cases with intensities right at the wind threshold of interest, a 1-kt increase in intensity in a realization would increase the corresponding wind radii from zero to a climatological value from the radii-CLIPER model.

For the West Pacific, the intensity biases of the official forecasts were small out to about 36 h and became positive beyond 36 h. Although the West Pacific biases in Fig. 4 increase slightly with time, they still remain

negative, so the behavior cannot be fully explained by the intensity biases of the official forecasts. Another source of bias during the forecast period is the radii-CLIPER model. Verification results showed that the radii-CLIPER model did have a low bias for the West Pacific for the 2008–11 sample, which helps to explain the low bias during the forecast period, despite a high bias in the official intensity forecasts. The radii-CLIPER model is also contributing to the high biases in the East Pacific for the longer forecast periods, since it did have a slight high bias for the 34-kt wind radii forecasts.

To determine whether the MC model forecasts have skill relative to the deterministic forecast, the BSS was computed using the deterministic forecast as the skill reference. A probability for the deterministic forecast is set to either 1 or 0, depending on whether that grid point came within the forecast radii for each wind threshold. Because the operational wind radii forecasts only extend to 72 h for 34 and 50 kt and to 36 h for 64 kt, the wind radii-CLIPER model was used to extend the official radii forecasts out to 120 h. The Brier score was calculated from the MC model probabilities and from the deterministic forecasts converted into a binary probability, as described above. A perfect Brier score is zero, so skill is measured by the percent reduction in the MC model Brier score relative to the deterministic forecast Brier score, where BSSs > 0 indicate skill. The results of this comparison (Fig. 5) depict a favorable interpretation of the MC model. The MC model forecasts are superior (BSS > 0) to the deterministic forecasts beyond 6 h in all regions and for all wind thresholds except the West Pacific, where skill is evident beyond 24 h. The skill generally increases with time for both the cumulative and incremental probabilities. This is because the deterministic forecast (converted to a binary probability) is a very unreliable measure of the uncertainty at the longer time periods. The paired t test showed that the improvement of the MC model over the deterministic forecast was statistically significant at every forecast time for the Atlantic and East Pacific, as well as all forecast times after 36 h for the West Pacific.

The next question addressed in the verification concerns the calibration of the MC model forecasts. A common way of assessing how well a probabilistic forecast is calibrated is through the use of reliability diagrams (sometimes referred to as calibration functions) and refinement distributions. The former display the forecast probabilities as a function of observed frequency and the latter provide the relative frequency of various probability forecasts. Figure 6 shows the reliability diagrams and refinement distributions associated with the MC model forecasts for 34-, 50-, and 64-kt winds at 36, 72, and 120 h for the combined model domain. The model is well calibrated because the reliability diagrams show nearly a perfect 1:1 correspondence along the 45° diagonal. The refinement distributions in Fig. 6 show that most of the forecasted probabilities are very low (noting that y axes of the inset figures are on a log scale) as expected due to the very large domain of the gridded product. However, the probability forecasts between about 0.25 and 0.95 are fairly uniformly distributed, indicating that the MC model can produce high probability predictions. Furthermore, Fig. 6 is directly comparable to Fig. 10 in D09 and generally shows that biases found in 2006–07 have lessened to some degree in the 2008–11 samples. This improvement is probably due to the larger sample sizes in the current verification, and to the improvements made to the model since 2007.

For real-life mitigation activities it is usually necessary to make a yes–no decision. For this type of application, a probability threshold is determined based on risk and lead time analysis, and an action would be triggered when the threshold was exceeded at a given location for the lead time when it was still possible to complete the action. To provide verification metrics for these applications, a probability threshold was specified and used to divide a yes from a no event at each grid point for each lead time and wind speed threshold. Two by two contingency tables were then generated, which contain counts of the number of cases when the event (the occurrence of 34-, 50-, or 64-kt winds during the time interval of interest) was forecast to occur and did occur, was forecast to occur but did not occur, was not forecast to occur but did occur, and was forecast not to occur and did not occur. These tables were generated for a range of threshold probabilities from 0% to 100% with an increment of 1%. Many forecast metrics can be calculated from these contingency tables (Wilks 2006). In D09, threat scores and relative operating characteristic (ROC) skill scores, which provide a measure of how well the MC model discriminates events from nonevents, were calculated. While not shown for succinctness, ROC scores for the 2008–11 samples showed substantial improvements when compared to those of 2006–07, particularly for the incremental probabilities.

The threat score (TS) can be interpreted as the ratio of the intersection of the area where an event was predicted to occur and the area where an event did occur to the union of those two areas. An advantage of the TS compared to the ROC score for the MC model verification is that the TS does not consider the number of correct nonevents (the event was not forecast to occur and did not occur). The number of nonevents is very large when the sample includes the entire forecast domain and can inflate the verification statistics. The TS

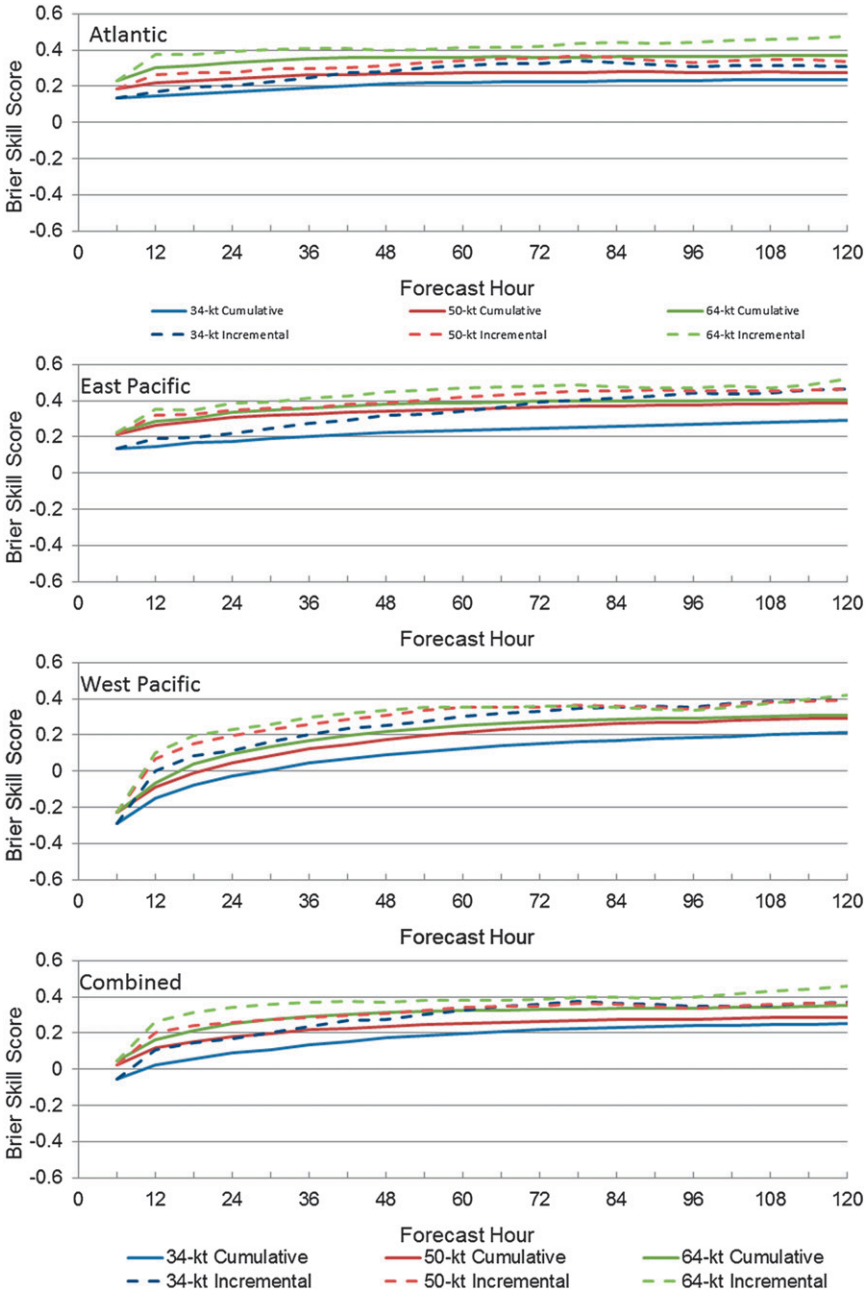

FIG. 5. The BSSs associated with the 2008–11 MC model verification in which the deterministic forecast is used as the reference for the North Atlantic (1°–50°N, 110°–1°W), East Pacific (1°–40°N, 180°–75°W), West Pacific (1°–50°N, 100°E–180°), and the combined multibasin domain (1°–60°N, 100°E–1°W) are shown in the panels starting from the top, respectively. Solid (dashed) lines indicate cumulative (incremental) probabilities. Blue, red, and green lines are for 34-, 50-, and 64-kt wind probabilities, respectively.

was calculated for each probability threshold from 0% to 100%. For simplicity, only those thresholds that maximized the TS are examined further.

Figure 7 shows the maximum TSs associated with the cumulative and incremental 34-, 50-, and 64-kt wind probabilities and Fig. 8 shows the threshold probabilities that maximize the TSs. The results in Figs. 7 and 8 for the 2008–11 sample generally show smoother temporal transitions, higher TSs, and slightly higher threshold probabilities overall, compared with the 2006–07

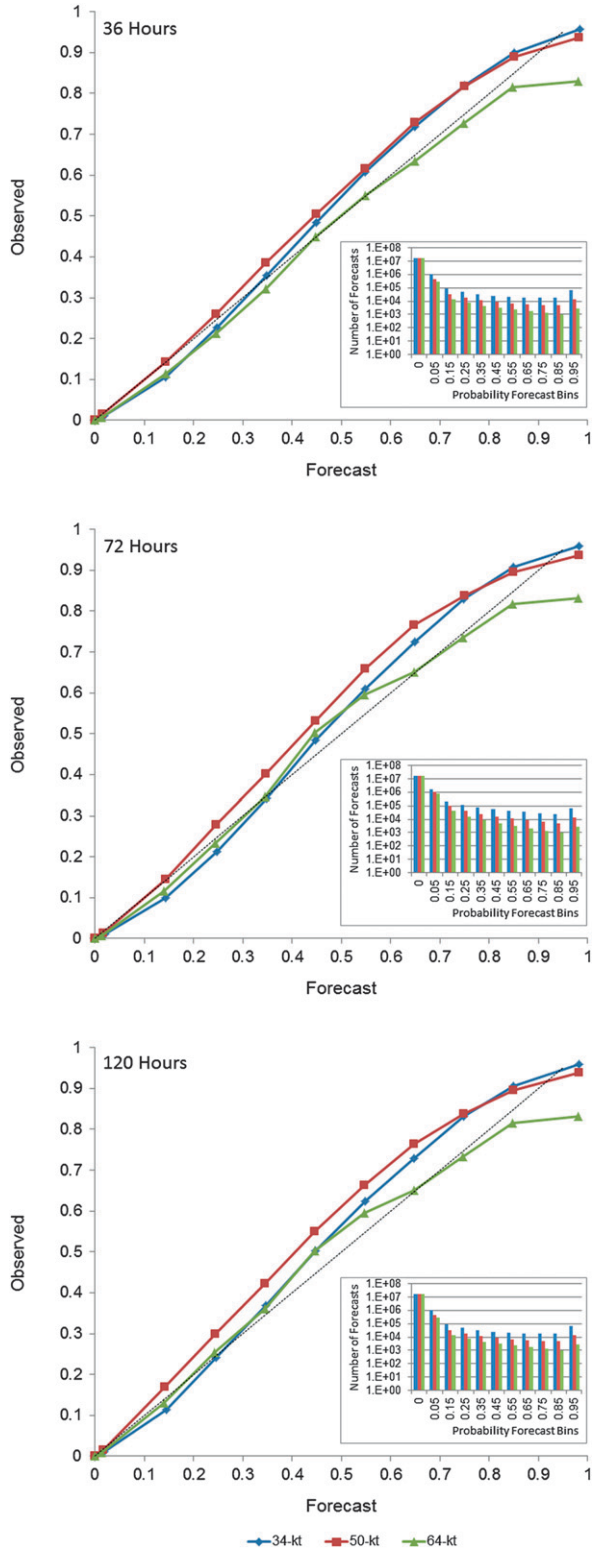

FIG. 6. Reliability diagrams and embedded refinement distributions for (top) 36, (middle) 72, and (bottom) 120 h associated with 34- (blue), 50- (red), and 64-kt (green) probability forecasts over the entire MC model domain (1°–60°N, 100°E–1°W).

sample from Figs. 11 and 12 in D09. This is not surprising given the continued maturation and improvements of the MC model, including the implementation of the GPCE version in 2010. It should be noted that the threshold probabilities in Fig. 8 provide a reasonable basis for a yes–no decision in cases where the cost of an incorrect ''no'' forecast is similar to that of an incorrect ''yes'' forecast, since they maximize the overlap of predicted and observed areas of occurrence. However, as described above, the real-world application of probability thresholds should account for factors such as loss and risk tolerance. For example, in high-cost or low-risk tolerance situations, lower-probability thresholds for yes–no decisions would likely be more appropriate than those in Fig. 8.

To get a better idea of the impact of the GPCE on the MC model performance, the forecasts for 2010 and 2011 for the Atlantic basin were rerun without the GPCE input. The sample included 820 forecast cases. This was straightforward because the error distributions for the full 5-yr samples are still calculated for use as a backup in case the GPCE parameter was not available in real time. This almost never occurred in real time so the reruns without the GPCE input provide a good benchmark for evaluating the impact on the real-time runs. The percent reduction in the Brier score with the GPCE input was calculated for the cumulative and incremental 34-, 50-, and 64-kt probabilities and the results showed an improvement of 1%–4% at all times from 24 to 120 h for all wind radii, with larger improvements for the incremental probabilities. There was a very slight degradation at 12 h for the 50- and 64-kt thresholds. The optimal TSs were also calculated for the no GPCE version of the model. Figure 9 shows the increase in the optimal TS for the GPCE version. Figure 9 shows that the TS improved at all forecast times for all wind radii for both the incremental and cumulative probabilities, confirming that the GPCE input is improving the performance of the operational MC model.

The statistical significance of the improvements of the GPCE over non-GPCE version of the MC model was evaluated using the paired t test described above. The improvements were significant at all forecast times for the cumulative and incremental 34-kt probabilities. For the 50-kt probabilities, the improvements in the incremental (cumulative) values were significant at 36–120 h (60–120 h). For the 64-kt incremental (cumulative) probabilities the improvements were significant at 24–120 h (48–120 h).

In summary, the verification results indicate that the MC model forecasts are skillful compared to the deterministic forecasts provided by NHC, CPHC, and JTWC,

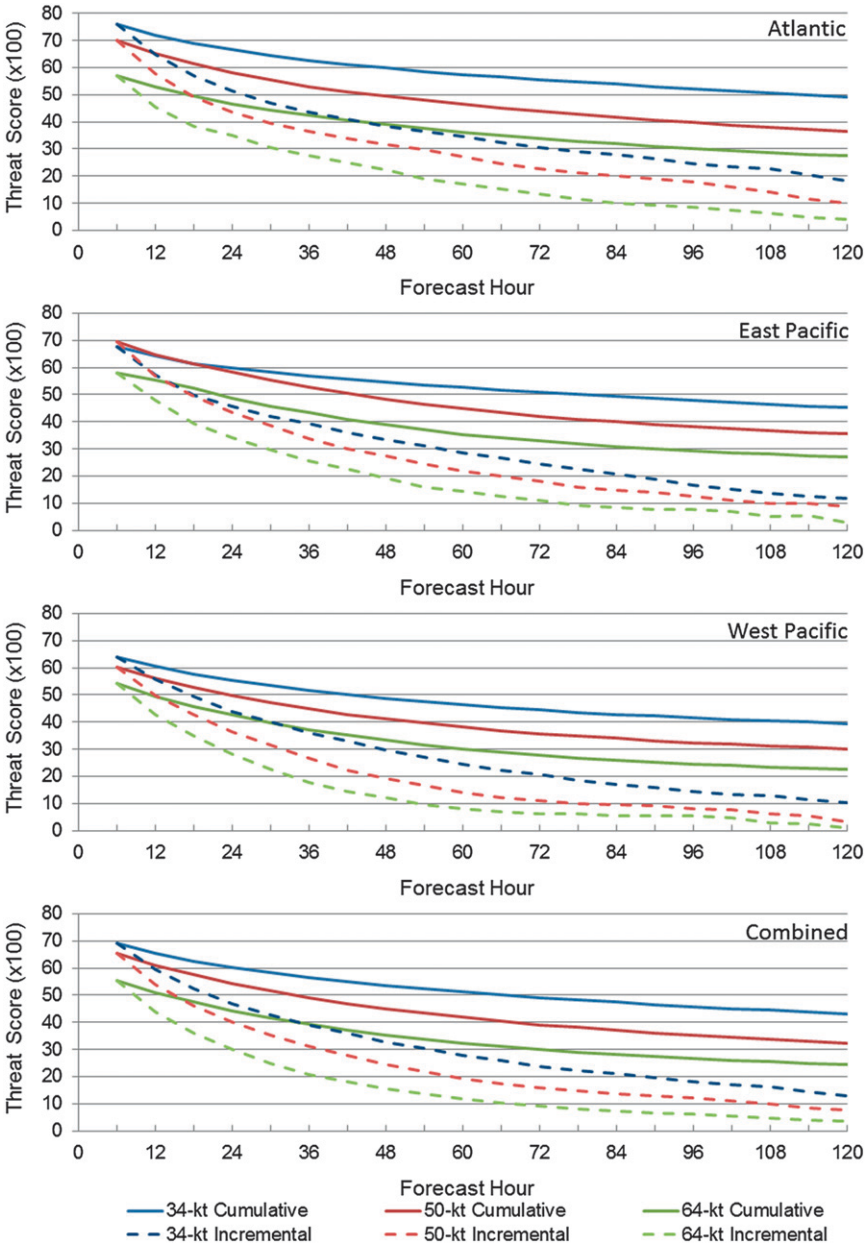

FIG. 7. The maximum conditional TSs (×100) associated with the 2008–11 MC model verification for the North Atlantic (1°–50°N, 110°–1°W), East Pacific (1°–40°N, 180°–75°W), West Pacific (1°–50°N, 100°E–180°), and the combined multibasin domain (1°–60°N, 100°E–1°W) are shown in the panels starting from the top, respectively. Solid (dashed) lines indicate cumulative (incremental) probabilities. Blue, red, and green lines are for 34-, 50-, and 64-kt wind probabilities, respectively.

are well fairly well calibrated to observed frequencies based on best-track information, and show skill in discriminating yes–no events based on the TS. Furthermore, the verification metrics indicate that the MC model continues to be a stable tool for providing tropical cyclone probability estimates for decision making. The verification shows that the probabilities do have some overall low and high biases for individual basins that are related to biases in the official intensity forecasts and the radii-CLIPER model, especially for the West Pacific, but are still much better than using the deterministic forecast as a binary probability. The GPCE version of the MC model shows significant improvement over the version without the GPCE input.

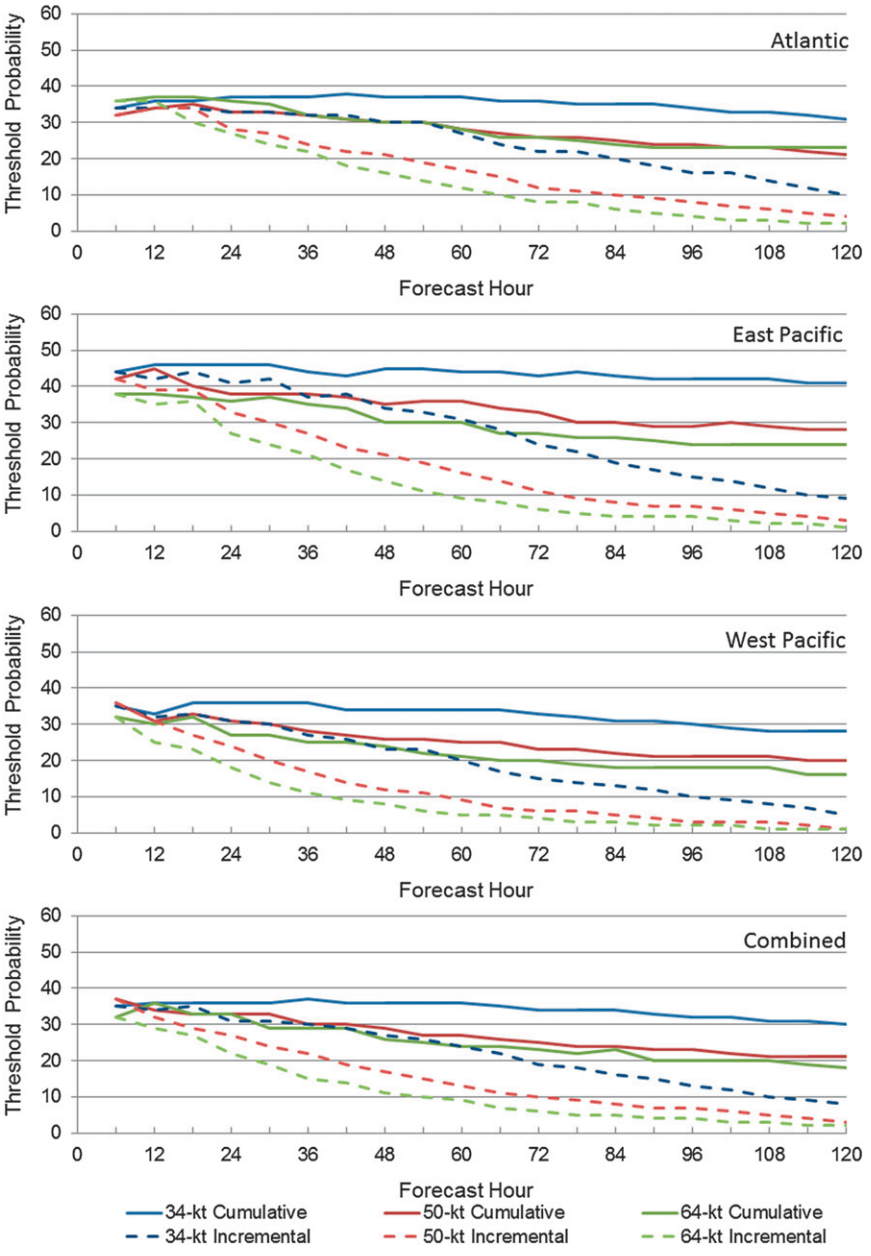

FIG. 8. The probability thresholds associated with the maximum conditional TSs shown in Fig. 7 and based on the 2008–11 MC model verification for the North Atlantic (1°–50°N, 110°–1°W), East Pacific (1°–40°N, 180°–75°W), West Pacific (1°–50°N, 100°E–180°), and the combined multibasin domain (1°–60°N, 100°E–1°W) are shown in the panels starting from the top, respectively. Solid (dashed) lines indicate cumulative (incremental) probabilities. Blue, red, and green lines are for 34-, 50-, and 64-kt wind probabilities, respectively.

5. The wind speed probability table

As described in the introduction, the operational WSPT product was modified in 2008 to utilize input from the MC model. This product is available for the Atlantic and East Pacific and estimates the probability that the cyclone intensity will be within each of nine categories (dissipated, tropical depression, tropical storm, hurricane, and the five Saffir–Simpson hurricane wind scale categories) at several forecast lead timed out to 5 days. The probability of the combined hurricane categories is also provided. The probabilities are calculated from the 1000 intensity realizations generated by the MC model. The previous version of the product

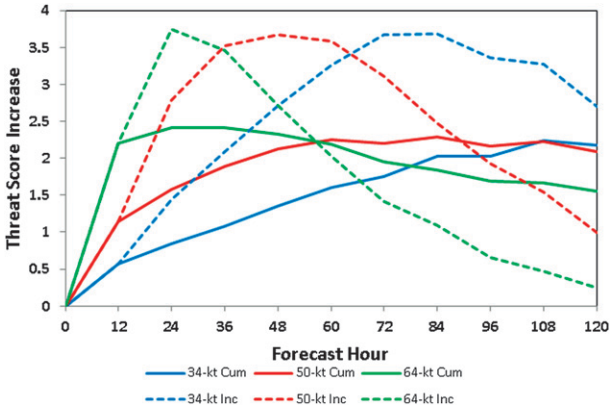

FIG. 9. The increase in the TS for the 2010–11 Atlantic MC model forecasts for the case when the GPCE input was included compared with the runs without the GPCE input.

estimated the probabilities from a fixed 10-yr period (1986–95), which had not been updated in more than 10 yr, and only included information out to 72 h. The generation of the WSPT from the MC model provided increased consistency between NHC probabilistic products, and ensured that the underlying error distributions were updated each year.

Figure 10 shows an example of the WSPT product for Tropical Storm Igor just before it became a hurricane. As described in D09, the MC model applies a bias correction to the official forecast that is a function of the forecast wind speed. For this reason, the highest probabilities are not always in the same category as the official forecasted intensity, which is shown along the bottom of the table. For example, at 72 h in Fig. 10, the NHC official forecasted intensity was 100 mi h^{-1} (category 2, 1 mi h^{-1} = 0.447 m s^{-1}), but the highest probability (37%) at 72 h was for a category 1 storm. The overland correction can also result in the highest probability occurring in a different category from that of the official forecast because some of the realizations make landfall at times when the official forecast does not, and vice versa.

Verification of the WSPT for all 2008–11 Atlantic and East Pacific tropical cyclones was performed using reliability diagrams. Forecast probabilities at each lead time (e.g., 12 h, 24 h, etc.) for each intensity category were cataloged. To increase the sample size, the forecast probabilities were grouped into bins at 10% intervals. The observed frequencies for each forecast probability bin, intensity category, and lead time were computed using the cyclone status from the final NHC best-track data.

Since the methodology for computing the intensity probabilities does not make a distinction between tropical cyclones and other types of cyclones in the best track (extratropical, remnant low, etc.), the verification was performed regardless of whether or not the system was a tropical cyclone at the verifying time, similar to the verification of the MC model described in section 4. For

Intensity (Maximum Wind Speed) Probability Table
Tropical Storm Igor Advisory Number 11
11:00 PM AST Sep 10 2010

	Forecast Time						
Wind Range (mph)	12 hour for 8 AM Sat	24 hour for 8 PM Sat	36 hour for 8 AM Sun	48 hour for 8 PM Sun	72 hour for 8 PM Mon	96 hour for 8 PM Tue	120 hour for 8 PM Wed
Dissipated	<1%	1%	<1%	<1%	<1%	1%	1%
Tropical Depression (<39)	3%	4%	3%	1%	2%	2%	2%
Tropical Storm (39-73)	90%	72%	45%	29%	20%	19%	17%
Hurricane (all categories)	8%	23%	52%	69%	78%	79%	81%
-- Category 1 (74-95)	7%	21%	41%	42%	37%	33%	31%
-- Category 2 (96-110)	1%	2%	7%	18%	22%	23%	23%
-- Category 3 (111-130)	1%	1%	3%	8%	14%	17%	20%
-- Category 4 (131-155)	<1%	<1%	1%	2%	4%	5%	6%
-- Category 5 (>155)	<1%	<1%	<1%	<1%	1%	1%	1%
Forecast Maximum Wind	60 mph	65 mph	75 mph	85 mph	100 mph	105 mph	115 mph

FIG. 10. An example of the wind speed probability table product for Advisory 11 from Atlantic Tropical Cyclone Igor.

example, a 20-kt remnant low would be counted in the tropical depression category, and a 50-kt extratropical cyclone would be counted as a tropical storm for our verification purposes. Dissipated forecasts were considered correct when a best-track point for that cyclone was not available at the verifying time. Probabilities issued with so-called special advisories were verified to help increase the sample size.

Despite verifying all WSPT forecasts for four seasons, the sample sizes for some forecast categories and bins were quite small, particularly for the individual Saffir–Simpson hurricane scale categories. For this reason, only the results for the tropical storm and combined hurricane intensity categories are discussed here. In addition, the probabilities for 12–48 and 72–120 h were combined in the reliability diagrams and BSS calculations. When combined in this way, the verification sample sizes range from 385 for the 72–120-h East Pacific hurricane cases to 2409 for the 12–48-h Atlantic tropical storm cases.

For the Atlantic, the probability forecasts of tropical storm intensity at lead times of 12–48 h show good reliability, as the observed frequency steadily increases with the forecast probability (Fig. 11, top). However, the technique somewhat overforecasts the occurrence of tropical storms at these lead times in the 5%–35% bins. Above the 35% bin, the technique works well for tropical storms, with the reliability curve very close to the one-to-one line. For the longer lead times, the tropical storm results in the top of Fig. 11 are similar to those for the shorter lead times, although there are no cases where the forecast probability was above 50%.

For the hurricane intensity category in the Atlantic (Fig. 11, top), the forecasts are fairly reliable for both the short and long lead times, with curves close to the one-to-one line. There is some tendency for the hurricane probabilities to be overestimated in the 5%–45% bins and underestimated in the 55%–95% bins.

As a bulk evaluation of the Atlantic reliability curves in Fig. 11, the average absolute value of the difference between the forecasted and observed probability for each bin was calculated and then the sample weighted average of the bin errors was calculated. For simplicity the short- and long-term forecast cases were combined but the calculation was performed separately for the tropical storm and hurricane cases. This calculation is referred to as mean absolute error (MAE) of the reliability diagrams, which can also be interpreted as the sample weighted-average distance of all the points on the reliability diagram from the one-to-one line. The MAE of the Atlantic reliability curves was 4.4% for tropical storms and 4.1% for the hurricanes. As a basis for comparison, the MAE calculations were repeated, but with the WSPT probability forecasts replaced by the climatological probabilities of a tropical cyclone being a tropical storm (40%) or hurricane (31%) for this same 2008–11 sample. Results showed that the MAE of the climatological probability forecasts was 21% for the tropical storms and 31% for hurricanes. The MAE of the WSPT is about a factor of 5 smaller than those using a climatological probability forecast for tropical storms, and a factor of 8 smaller for hurricanes. Thus, by this measure, the WSPT has considerable skill relative to climatology.

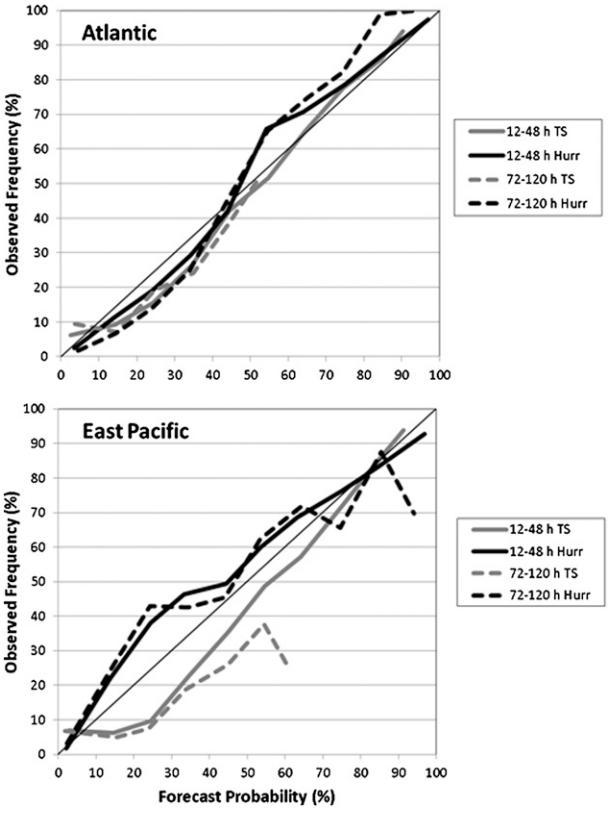

FIG. 11. Reliability diagram of 2008–11 tropical storm forecasts at lead times of 12–48 h (gray solid line), tropical storm forecasts at lead times of 72–120 h (dashed gray line), hurricane forecasts at lead times of 12–48 h (solid black line), and hurricane forecasts at lead times of 72–120 h (dashed black line) for the (top) Atlantic and (bottom) East Pacific. The forecast probability is shown along the *x* axis, and the observed frequency is shown along the *y* axis.

For the East Pacific (Fig. 11, bottom) the WSPT forecasts are less reliable than those in the Atlantic, although there is still a general tendency for the observed frequencies to increase with the forecasted probabilities. The exceptions are the tropical storm probabilities in the 5%–25% bins, the 95% bin for the longer-range hurricane probabilities, and the 65% bin for the long-range tropical storm probabilities. This result indicates that the WSPT product may be less useful for the East Pacific, especially for tropical storms with low probability

forecasts. The MAEs of the East Pacific reliability curves were 10.7% for tropical storms and 3.6% for hurricanes. Using climatological probabilities, the MAEs were 20% for tropical storms and 26% for hurricanes. Thus, although the reliability for East Pacific tropical storms is less than for the Atlantic, the MAEs are still a factor of 2 smaller than those from a climatological probability forecast for tropical storms and a factor of 7 smaller for hurricanes.

The WSPT product is designed to provide uncertainty information to complement the deterministic official intensity forecasts. Similar to the MC model verification results in section 4, the skill of the WSPT information can be determined by comparing it with the deterministic forecast converted to a binary probability in each category (0%–100%). For this purpose, the BSS was computed as the percent reduction of the Brier score from the WSPT relative to that for the NHC official forecast in probabilistic form. Figure 12 shows that for the Atlantic and East Pacific, the WSPT have skill at all lead times for both the tropical storm and hurricane categories. The tropical storm probabilities have greater skill than the hurricane probabilities at most forecast times.

The statistical significance of the WSPT skill relative to the deterministic forecast (the BSS values in Fig. 12) was evaluated using the paired t test. Results showed that skill was significant at the 95% level at every forecast time from 12 to 120 h in both basins.

In summary, the WSPT product is fairly reliable, especially for the Atlantic basin. In both the Atlantic and East Pacific, the probability estimates for the tropical storm and hurricane categories are accurate to within 4%–11% and have considerable skill relative to climatological probability forecasts. The main limitation is for low-probability forecasts of East Pacific tropical storms. Nevertheless, the WSPT probabilities improve significantly over determining the intensity category directly from the official intensity forecasts, as indicated by the BSS results shown in Fig. 12.

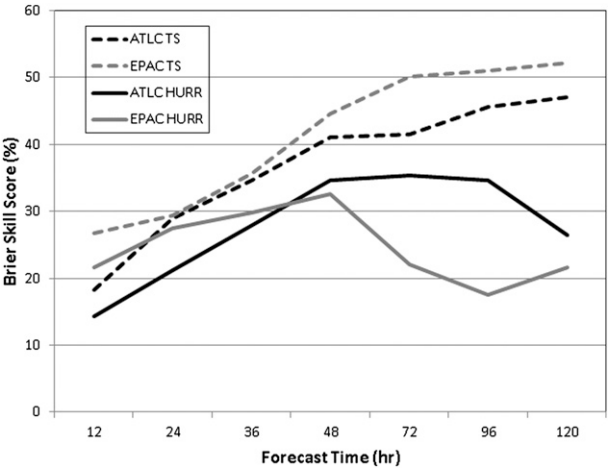

FIG. 12. Percent reduction in Brier score from the WSPT relative to the NHC official forecasts for Atlantic tropical storms (black dashed line), East Pacific tropical storms (gray dashed line), Atlantic hurricanes (black solid line), and East Pacific hurricanes (gray solid line) for 2008–11.

6. Concluding remarks

Modifications to the operational MC wind speed probability model since 2007 were described. These included updating the underlying error distributions prior to the start of each hurricane season, a change to the underlying model time step, intensity bias corrections for storms over land, a method for eliminating unrealistically large track perturbations, and the inclusion of ensemble information through the GPCE parameter beginning in 2010. The MC model was verified for 2008–11 using a number of standard metrics for probabilistic forecasts. Results showed that the wind speed probabilities add value to the deterministic forecasts, and are generally reliable in all three basins where the wind speed probability model is run (Atlantic, East Pacific, and West Pacific). To isolate the impact of the inclusion of the GPCE input, the Atlantic forecasts from 2010 and 2011 were rerun without the GPCE input. Results showed that the GPCE version had 1%–4% improvements in the Brier and threat scores.

The probabilities also showed some biases, with the largest high biases in the East Pacific and largest low biases in the West Pacific. These biases can be explained by biases in the official intensity forecasts for the specific time period of the verification, and to biases in the underlying wind radii-CLIPER model. It is expected that the biases would be reduced for a large sample, as was seen for the case where all the basins are combined. These results also suggest that the radii-CLIPER model should be updated to include a larger sample of cases.

The existing operational wind speed probability table (WSPT) product was modified in 2008 to use the intensity information from the MC model. A verification of the WSPT's ability to distinguish between tropical storms and hurricanes showed that it has skill in the Atlantic and East Pacific relative to climatology and the NHC official forecast. The reliability was generally greater in the Atlantic than the East Pacific.

It is anticipated that the MC model will continue to be run operationally for the next several years. One limitation of the current version of the model is that the horizontal grid is too coarse. As described in D09, the wind speed probabilities are calculated on an evenly

spaced 0.5° latitude–longitude grid, which is then interpolated onto a 5-km grid for use in the National Digital Forecast Database (NDFD). The probabilities are also calculated on a specified set of points near the coast for use in a text product. It was found that in regions with large spatial gradients, the probabilities in the text product can differ from those in the NDFD by up to 25% in an absolute sense, although the average errors due to the interpolation were much smaller. To alleviate this problem, plans are under way to decrease the spacing of that latitude–longitude calculation grid to reduce the interpolation errors on the finer NDFD grid, and make the gridded product more consistent with the text product.

The incorporation of track ensemble information through the GPCE parameter improved the skill of the MC model for the 2010–11 sample. This track uncertainty information is included in a very conservative way, where the official track errors are stratified into terciles based on the GPCE values. As ensemble systems become more mature, it should eventually be possible to replace the random sampling of track errors with a set of tracks generated directly from a dynamical ensemble modeling system. In the longer term, the intensity and wind structure information could also come from a dynamical model ensemble system. The relatively simple statistically based MC model probabilities and the validation system described in section 4 could be used as a baseline for the transition to a dynamical ensemble. Bias corrections might be needed to ensure consistency with the official forecast if dynamical models were used to generate the realizations.

As described above, the WSPT product was adapted to use input from the MC model. This change provides consistency between that product and those generated from the MC model. NHC also provides a graphical product that shows the cone of uncertainty about the official forecast track. The size of the cone depends on the 67th percentile of the NHC official track errors from the previous 5 yr. In principle, the 67th percentile of the track errors could be generated directly from the MC model output. This would make the error cone consistent with the probabilistic forecast products from the MC model, similar to the WSPT. Due to the inclusion of the GPCE parameter in the MC model, it would also make the size of the cone a function of the ensemble model spread, rather than a fixed size based on the previous 5 yr of track errors. Preliminary tests of this idea showed that the 67th percentiles from the MC model do not exactly match those used in the probability cone, even though both are based on the previous 5 yr of official track forecast errors. This discrepancy is partially due to the differing verification samples used to define the cone (the operational product includes only the tropical and subtropical storm stages while the MC model includes all cases regardless of the storm classification). Another factor is the method used to include the serial correlation of track errors in the MC model, which introduces a slight high bias in the track error distributions at the longer forecast intervals. Methods to reconcile these differences are being developed, and, if adequate solutions can be found, the NHC uncertainty cone might also be determined from the MC model output in the future.

Acknowledgments. This project is supported by the NOAA Joint Hurricane Testbed and the Hurricane Forecast Improvement Project under NOAA Grants NA17RJ1228 and NA09AANWG0149. Valuable comments were received from the three anonymous reviewers. The views, opinions, and findings contained in this report are those of the authors and should not be construed as an official National Oceanic and Atmospheric Administration or U.S. government position, policy, or decision.

REFERENCES

Cangialosi, J. P., and J. L. Franklin, 2011: 2010 National Hurricane Center forecast verification report. NHC, 77 pp. [Available online at http://www.nhc.noaa.gov/verification/pdfs/Verification_2010.pdf.]

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.

Goerss, J. S., 2007: Prediction of consensus tropical cyclone track forecast error. *Mon. Wea. Rev.,* **135,** 1985–1993.

Hamill, T. M., 1999: Hypothesis tests for evaluating numerical precipitation forecasts. *Wea. Forecasting,* **14,** 155–167.

——, J. S. Whitaker, M. Fiorino, and S. J. Benjamin, 2011: Global ensemble predictions of 2009's tropical cyclones initialized with an ensemble Kalman filter. *Mon. Wea. Rev.,* **139,** 668–688.

Hauke, M. D., 2006: Evaluating Atlantic tropical cyclone track error distributions based on forecast confidence. M.S. thesis, Dept. of Meteorology, Naval Postgraduate School, Monterey, CA, 105 pp.

Knaff, J. A., C. R. Sampson, M. DeMaria, T. P. Marchok, J. M. Gross, and C. J. McAdie, 2007: Statistical tropical cyclone wind radii prediction using climatology and persistence. *Wea. Forecasting,* **22,** 781–791.

Rappaport, E. N., and Coauthors, 2009: Advances and challenges at the National Hurricane Center. *Wea. Forecasting,* **24,** 395–419.

——, J.-G. Jiing, C. W. Landsea, S. T. Murillo, and J. L. Franklin, 2012: The Joint Hurricane Testbed: Its first decade of tropical cyclone research-to-operations activities reviewed. *Bull. Amer. Meteor. Soc.,* **93,** 371–380.

Sheets, R. C., 1985: The National Weather Service hurricane probability program. *Bull. Amer. Meteor. Soc.,* **66,** 4–13.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences.* 2nd ed. International Geophysics Series, Vol. 59, Academic Press, 627 pp.