AMERICAN
METEOROLOGICAL
SOCIETY

Weather and Forecasting

The DOI for this manuscript is doi: 10.1175/2009WAF2222189.1

Statistical Extension of the National Hurricane Center 5-Day Forecasts

Daniel S. Wilks
Department of Earth and Atmospheric Sciences
Cornell University, Ithaca New York, 14853, USA
email: dsw5@cornell.edu

Charles J. Neumann, and Miles B. Lawrence
National Hurricane Center, retired

Submitted to *Weather and Forecasting*
Revised 5 January 2009

Abstract

U.S. National Hurricane Center (NHC) forecasts for tropical cyclone tracks and windspeeds are extended in time to produce spatially disaggregated probability forecasts for landfall location and intensity, using a weighted bootstrap procedure. Historical analogs, with respect to the forecast characteristics (location, heading and windspeed) of a current storm, are selected. These are resampled by translating their locations to random positions consistent with the current forecast, and recent NHC forecast accuracy statistics. The result is a large number of plausible Monte-Carlo realizations that jointly approximate a probability distribution for the future track and intensity of the storm. Performance of the resulting forecasts is assessed for U.S. tropical cyclone landfall probabilities during 1998-2006, and the forecasts are shown to be skillful and exhibit excellent reliability, even beyond the 120-hour forecast horizon of the NHC Advisory forecasts upon which they are based.

1. Introduction

The landfall location of a tropical cyclone is an important element of its damage potential, and forecasts of landfall location are critical components of information used by disaster preparedness officials and coastal residents to prepare for a threatened tropical cyclone strike, even though serious storm damage often also occurs well away from the landfall location of the storm center. The U.S. National Hurricane Center (NHC) issues Forecast Advisories for future tracks of tropical cyclones at lead times through 120 hours, and these are extended through day 7 by the Hydrometeorological Prediction Center (http://www.hpc.ncep.noaa.gov/medr/medr.shtml). However, these Advisory forecasts are not explicitly probabilistic, whereas probabilistic forecasts are inherently more valuable in decision making (e.g., Katz and Murphy 1997, Krzysztofowicz 1983), and tropical cyclone landfall probability forecasts could be used to enhance economic decision making in the face of an uncertain future storm path. Landfall forecasts for lead times beyond 5 days are typically subjective and qualitative extrapolations of the NHC 120-hour forecasts.

In a formerly (1983–2005) operational program, NHC computed landfall, or “strike,” probabilities by integrating the intersections of circles of radius 116 km centered near coastal points of interest, with a sequence of 3-hourly bivariate normal probability density functions representing the uncertainty in forecast storm positions (Sheets 1984). (This product has since been superseded by Monte-Carlo gridpoint probability forecasts of windspeeds above selected thresholds: Gross et al, 2004.) One strength of this approach was that these probability distributions for future storm positions, and thus also the strike probabilities, were based on all available information — dynamical, statistical, and subjective forecaster judgments — contributing to the official forecast, together with the archive of historical errors for those forecasts. A weakness was that the forecast calculations effectively assumed correct forecasts of the direction (i.e., heading) of storm movement, as distinct from forecast position.

Use of the bivariate normal distribution has been a longstanding practice for representation of position errors of tropical cyclone forecasts. For example, Hope and Neumann (1970) summarized the accuracy of HURRAN (HURRicane ANalog) forecasts with bivariate normal error distributions. The HURRAN forecasts were constructed by collecting historical analog storms having attributes similar to a current storm, translating the positions of these analogs to the current position of the storm to be forecast, and aggregating the downstream positions of this collection of analog tracks from that point into the future to form a distribution (i.e., an ensemble) of forecast positions whose dispersion could be characterized using a bivariate normal error distribution.

The present paper is concerned with extending the information in the official NHC forecasts for tropical cyclone position and track, in a way that draws on ideas from these earlier approaches. First, the HURRAN approach of extending downstream motions of historical analogs is adopted, but here these extensions are begun from a forecast future storm position, rather than from an initial observed position. Since these future positions are uncertain, they are characterized by bivariate normal error distributions, and random draws from these error distributions are used to initialize the subsequent downstream motions of the historical analogs. Rather than characterize the dispersion of a small sample of such analogs using a fitted distribution, the probability evaluations are achieved nonparametrically, by repeatedly drawing from the pool of available analogs, with replacement (i.e., bootstrapping — Efron and Tibshirani 1993). Relative frequencies of members of this distribution of bootstrapped storm tracks that cross coastal segments of interest are interpreted as the “strike” probabilities for those segments.

Section 2 describes the approach in detail, together with the data sources to be used. Section 3 presents two forecast cases in some detail, and Section 4 evaluates the probabilistic landfall forecasts overall. Section 5 relates the forecast method to the NHC "cone of uncertainty," which is part of a prominent, publicly disseminated graphic that communicates tropical cyclone track forecasts; and Section 6 concludes.

2. Data and approach

Probabilistic extensions of the NHC Forecast Advisories are computed here for tropical cyclone landfalls across the ten segments of the U.S. coastline indicated in Figure 1. These segments have been chosen somewhat arbitrarily, but in a way that yields roughly comparable relative hurricane strike probabilities (shown parenthetically in Figure 1). These relative risks are smoothed values that have been estimated using the HURDAT data set (Jarvinen et al. 1984, Landsea et al. 2004), 1851–2005, which is available at the website www.nhc.noaa.gov/tracks1851to2007_atl_reanal.txt.

The probability forecasts described here are computed using forecast track and windspeed information given in the NHC Forecast Advisories. These Advisories are produced at 6-hourly intervals when Atlantic tropical cyclones are present (historical archive available at http://www.nhc.noaa.gov/pastall.shtml). Among other information, they contain current and forecast storm positions and maximum sustained windspeeds, for 12-, 24-, 36-, 48-, 72-, 96-, and 120-hour lead times. Archived forecasts in the present format begin in 1998, and the 96- and 120-hour forecasts begin in 2003. In the following, Atlantic basin forecasts for 1998 through 2006 will be used.

The forecast position from the current Forecast Advisory that is furthest into the future, but has not yet crossed the continental coastline, is chosen to initialize the probabilistic forecast extension. The first step in this procedure is to choose historical analog storms having locations, headings, and windspeeds that are similar to these forecast characteristics for the current storm. The HURDAT data set provides estimates of storm location and maximum sustained wind at 6-hour intervals, so a given historical storm usually accounts for multiple entries in this database. There are 36989 such records (i.e., 6-hourly storm position entries) in the HURDAT data for 1880–2005 that are used in the following. However, only historical storms occurring in a year prior to a forecast storm may be used as analogs so that, for example, only the 32611 HURDAT entries for the years 1880–1997 are used to forecast 1998 storms, and the full 36989 records through 2005 are used to forecast 2006 storms.

A candidate historical storm position is chosen as an analog if three conditions are met:

1. The location of the candidate historical storm is within an elliptical region centered on the forecast location, which is elongated in the direction of forecast storm movement. The extent of this search ellipse is 6.67 latitude degrees (400 nm) in the along-track direction of the storm (i.e., 200 nm ahead and 200nm behind the forecast position), and half these distances in the perpendicular (cross-track) direction. These dimensions are consistent with the 300 km optimal radius for a circular search region derived by Hall and Jewson (2007). For perspective, the size of the initial search ellipse is roughly 50% larger than the island of Hispaniola.
2. The forecast storm direction and the direction of the candidate historical storm differ by no more than 20°, which is similar to the criterion used by Hope and Neumann (1970).

3. The maximum sustained windspeed for the candidate historical storm is at least 50%, and no more than 150% of the forecast maximum sustained wind from the NHC Advisory.

These criteria have been chosen subjectively, drawing on prior experience with similar techniques. Sensitivity tests (not shown) indicated relatively little effect on the forecast probabilities of varying them through reasonable ranges.

One of the problems with the original HURRAN model was that it failed to find sufficient analogs about 33% of the time (Neumann 1972). To address this problem we increase the size of the storm-oriented search ellipse if necessary until it intersects at least 20 historical storm positions. Generally, several consecutive positions for a particular historical storm are included among the analogs for a given forecast.

Hope and Neumann (1970) forecasted future movements of a current storm by translating positions of analog storms to the currently observed storm position, and then extended the historical paths of those storms from that point. A similar approach is adopted here, except that the analogs are chosen with respect to forecast rather than currently observed storm characteristics, and the historical paths of these analogs are extended from the vicinity of that forecast position. Because the future location of the storm being forecast is uncertain, these initial points are chosen as random draws from the circular bivariate normal distribution centered on the forecast position, and exhibiting dispersion consistent with the average NHC forecast position errors, at the appropriate lead time. These average position errors (2001–2005) were taken from the NHC website, and are reproduced in Table 1. Ten thousand random overwater positions are drawn from the appropriate error distribution to initialize downstream extensions of analogs. If a line connecting the NHC forecast position and an initial random position crosses a segment of coastline, that initial position is discarded and another is generated.

Individual analog storms are randomly selected from the pool of candidate analogs using a weighted bootstrap procedure. The probabilities (weights) with which the selections are made depend on the similarity between the forecast and analog maximum sustained windspeed, relative to the accuracy with which the windspeeds are forecast at the lead time in question. Specifically, analysis of NHC windspeed forecast errors for 2001-2005, the raw data for which were obtained from the NHC website, reveals that relative forecast windspeed errors (forecast windspeed divided by observed windspeed) at a given lead time follow approximately a Gaussian distribution, with mean 1 (i.e., they are basically unbiased), and a standard deviation that increases with lead time. These standard deviations are listed in Table 1, with the corresponding average position errors. The probabilities with which each of the candidate analogs are chosen for a given one of the 10000 simulations is based on relative likelihood, i.e.,

$$w_i = \frac{\phi[(u_i / u_f - 1)/\sigma]}{\sum_{i=1}^{n} \phi[(u_i / u_f - 1)/\sigma]} \quad , \quad i = 1, \ldots, n \ , \qquad (1)$$

where ϕ indicates the probability density function of the standard Gaussian distribution, u_{i} is the windspeed for the i^{th} candidate analog, u_{f} is the forecast windspeed, and σ is the standard deviation for the relevant forecast lead time from Table 1. Equation (1) assigns nearly equal weights to all candidate analogs for the longer lead times, because at those times the windspeed forecasts are relatively uncertain (σ is large), but at the shorter lead times it chooses the most similar analogs (in terms of windspeed) with much higher probability. In exploratory simulations (results not shown) it was found that weighting analogs according to windspeed similarity was necessary in order to compute reasonable results for the shorter ($\leq$36 h) lead times before landfall. Otherwise, lower-windspeed analogs were used too frequently, with the result that probabilities for (e.g.) hurricane-strength landfalls were too small.

Having displaced a candidate analog to a randomly chosen position near the NHC forecast, the remainder of its historical track is followed to see whether and where it intersects the U.S. coastline. For this purpose, the coastline is approximated using a collection of line segments corresponding to individual counties, as was also done by Hallegatte (2008), in order that intersections of the analog storm tracks with the coastline can be evaluated quickly and efficiently. Since storms typically lose intensity very rapidly after a landfall, post-landfall windspeeds in the historical database are set to their most recent overwater values, in order to reduce bias in the results. Landfall probabilities are then estimated as relative frequencies (among the 10000 simulated storms) of landfalls at each of the 10 coastal regions in Figure 1.

Figures 2 and 3 illustrate the procedure. Figure 2 is a graphical version of Forecast Advisory 17 for hurricane Emily (2005). At the time of issuance the storm was located in the eastern Caribbean Sea, and the final overwater forecast position is for the 120-hour lead time, when the storm was forecast to approach the Mexico/Texas border from the east-southeast. This case, exhibiting the 120-hour forecast position near the coastline, has been chosen for clarity of graphical exposition. However, notice that the method is applicable regardless of the distance between the terminal forecast point and the coastline, so that landfall probability forecasts with lead times substantially longer than 5 days can and will be produced.

Figure 3a shows the 120-hour forecast position (X), together with the 68 historical analog positions meeting the criteria listed above (black dots). These 68 analog positions are from 23 distinct storms, 1880–2004, and consecutive 6-hourly positions of a given analog storm within the ellipse are connected by the thin black lines. The size of the search ellipse is evident from the scatter of these points. Figure 3a also shows (grey dots and lines) the subsequent movements of three of the 23 storms: hurricanes Allen (1980), Gilbert (1988), and Bret (1999). Eleven of the 68 initial analog points (larger dots) locate earlier positions for these three storms.

The circle in Figure 3b indicates the 90% probability contour for the forecast position error distribution, consistent with the average position error of 303.3 nm at 120-hour lead time (Table 1). Also indicated (black dots) are twenty random positions drawn from this distribution. These positions are all over water, because any initial points randomly generated over land have been discarded and redrawn. Initial positions of historical analog storm tracks have been chosen randomly and with replacement from among the 11 black dots in Figure 3a for the three indicated storms, and translated to the random locations in Figure 3b.

Figure 3 shows extrapolations for only 3 storms for graphical clarity, but in an actual forecast, 10000 random positions would be generated for Figure 3b, and all of the 68 initial historical analog positions in Figure 3a, connecting to the subsequent tracks of all 23 historical analog storms, would be available to generate the distribution of simulated storm tracks. When this is done, 59.7% of the 10000 simulated tracks cross the Mexico coastline, 20.7% make landfall across coastal segment 1 (southern Texas, cf. Figure 1), 13.4% make landfall across coastal segment 2 (northern Texas), 4.4% make landfall across coastal segment 3 (western Louisiana), and 1.4% make landfall across coastal segment 4 (Mississippi delta region). These relative frequencies are then adopted as probability estimates for the respective events. In the ensuing days this storm persisted in moving, and continued to be forecast to move, steadily to the west-northwest.

It is also possible to forecast probabilities of landfall at or above a given storm intensity, by considering the maximum sustained wind associated with each analog storm. As noted above, because tropical cyclones typically weaken rapidly after landfall, historical post-landfall windspeeds have been set to their last overwater value. Probabilities for hurricane landfalls, for example, are estimated by counting as "hits" only those simulated storms that cross a coastline segment at or above hurricane strength. For the Hurricane Emily example in Figure 3, the resulting probabilities for U.S. coastal segments 1 through 4 are 19.5%, 12.5%, 3.7%, and 1.4%, respectively. These probabilities are only slightly smaller than the corresponding values for tropical cyclone landfall at any intensity because the 120-hour forecast maximum sustained wind was the relatively high value of 100 kt, so only analog storms with windspeeds between 50 and 150 kt have been chosen.

The example illustrated in Figures 2 and 3 is relatively straightforward, in that a U.S. landfall, if any, was likely to occur in Texas or Louisiana. However, if this storm and its forecast track had been located 5° further north the possibility of a landfall in south Florida would also need to be accounted for, even though the official track would not have intersected the Florida coast. In such cases, two random forecast simulations are undertaken. The second begins from the last overwater position, as described above, and the first is initiated from an earlier forecast position for which the forecast track is within 60° of a portion of the U.S. coastline, and at a distance that is closer (relative to the respective forecast position errors in Table 1) to the coast than the furthest-future overwater position from which a simulation will be initiated in any case.

For such forecasts, landfall probabilities from the two simulations are combined. Let f_1(i) be the relative frequency of hurricane landfall at coastal segment i, among the 10000 simulations initialized from the first (i.e., earlier lead time) of the two forecast positions, and let f_2(i) be the corresponding relative frequencies from simulations initialized at the later of the two lead times.

Denoting the event that the storm does not cross the U.S coastline by $i = 0$, the probability that the landfall will occur at segment i is estimated as

$$p(i) = f_1(i) + f_2(i) f_1(0) \quad . \tag{2}$$

Thus if the landfall probability associated with the earlier forecast position is very small, then $f_1(i) \approx 0$ and $f_1(0) \approx 1$, so Equation 2 will express probabilities derived from extrapolations of the final forecast position, $f_2(i)$. Conversely, if a landfall probability derived from the earlier forecast position is relatively high, then $f_1(0) \approx 0$ and $p(i) \approx f_1(\mathrm{i})$.

The purpose of this two-forecast procedure, in the minority of situations where it is invoked by the above-stated criteria, is to prevent what might otherwise be an extremely poor forecast resulting from an observed landfall associated with a very small or zero forecast probability. For the hypothetical example of the forecast track of hurricane Emily (Figure 2) displaced 5° northward, the official NHC forecast track would not cross the Florida Keys, but would be close enough for the storm to represent a substantial threat to that area. However, if only a single set of 10000 bootstrap simulations were to be initiated near the central Texas coast, few if any of the resulting analogs would have the opportunity to intersect the Keys, leading to a forecast of essentially zero landfall probability there. In principle, Equation 2 could be extended to combine analog simulations from all overwater storm positions in a given NHC Forecast Advisory (up to eight, as in Figure 2), and such extensions could be addressed in future work. However, the quality of forecasts derived from the present procedure, as detailed in Section 4, suggests that limiting the scope of Equation 2 to two storms at most is not a major deficiency of the procedure described here.

3. Two forecast examples

Table 2 illustrates the progression of the probability forecasts described in Section 2, for two cases. Table 2a shows forecasts at twice-daily intervals (alternate 6-hourly Forecast Advisories have been omitted for compactness) for Hurricane Emily (2005). This storm became a tropical depression on 11 July, when located in the Atlantic at about 11˚N and 43˚W, and moving to the west-northwest. At this time, more than a week from landfall, the eventual fate of this storm was quite uncertain: the first line of Table 2a shows nearly even chances that it will not make landfall in the U.S. (segment "0"), with the remaining probability spread across the entire Gulf and Eastern U.S. coastline, although with probability most concentrated in south Florida and the southeast U.S. coast (segments 7 and 8, cf. Figure 1). Over the ensuing 9 days, both the forecasts for, and the actual movement of this storm maintained this same general heading, so that the forecast landfall strike probabilities remain focused on Mexico and south Texas (segments "0" and 1), while probability becomes progressively more concentrated there as the storm approaches the coast. The probability of a U.S. landfall is nil as of 03Z on 20 July, approximately 9 h before landfall at a point about 70 nm south of the U.S. border, although even 24 h earlier the forecast probability of a U.S. landfall was quite small. A graphical loop of the sequence of NHC Advisories for this storm can be viewed at http://www.nhc.noaa.gov/archive/2005/EMILY_graphics.shtml.

A somewhat contrasting picture is presented by Tropical Storm Ernesto (2006), 12-hourly forecasts for which are shown in Table 2b. This storm first achieved tropical depression strength on 25 August, substantially further west (approximately 13˚N and 62˚W, in the eastern Caribbean Sea) of the initial NHC Advisory position for Hurricane Emily, at which time Ernesto's 5-day forecast position was near the northern tip of the Yucatan peninsula.

Accordingly, its initial landfall probabilities on the first line of Table 2b show very small values for the east coast of the U.S. (segments 8–10), and its 5-day forecast heading to the northwest yields less than a 10% chance of landfall in Mexico (by construction, these forecasts do not consider the possibility of a Yucatan landfall). Over the next two days the forecast track continued progressively further into the Gulf of Mexico, although with forecast positions that moved gradually to the east, and the progressive concentration of probability in the Gulf Coast segments 2–5 over this time period, with the focus of probabilities moving gradually east, reflects this. However, on 28 August the day 5 forecast position moved sharply eastward, with the result that the forecast probabilities shift almost entirely to Florida and the eastern U.S. coast (segments 5–10). In the lower half of Table 2b the probability assigned to segment "0" pertains to the possibility that Ernesto may recurve sufficiently to miss Florida entirely and move into the Atlantic. Over the period 28–29 August, forecast probability gradually concentrates in segment 7 (south Florida), and is approximately 90% there at 15Z on 29 August, approximately 12h ahead of landfall in the Florida Keys at 03Z on 30 August. A graphical loop of the sequence of NHC Advisories for this storm can be viewed at http://www.nhc.noaa.gov/archive/2006/graphics/al05/loop_5W.shtml.

4. Verification results

The forecast procedure described in Section 2 has been tested using the available NHC Forecast Advisories for 1998–2006. These have been abstracted from the NHC Tropical Cyclone Advisory Archive, available on the NHC website at www.nhc.noaa.gov/pastall.shtml. For 1998–2002, the Advisories forecast through the 72-hour lead time only. For 2003–2006 the maximum lead time is 120 hours. For landfalling storms, forecasts as described in Section 2 are initiated from advisories issued before the landfalls, only. For other storms, forecasts are initiated from all available advisories except the last. The result is that 3136 forecasts were initiated, pertaining to a total of 153 individual storms.

Storm landfall outcomes were taken from the respective NHC Tropical Cyclone Reports, which are also available at www.nhc.noaa.gov/pastall.shtml. During 1998–2006 there were 22 hurricanes, 28 tropical storms, and 5 tropical depressions making landfall on the portion of the U.S. coastline indicated in Figure 1. Included in these counts are seven storms with two U.S. landfalls sufficiently separated in time and space to be considered here as distinct: Georges (1998), Bertha (2002), Charley (2004), Frances (2004), Ivan (2004), Katrina (2005), and Ernesto (2006). Of these, only Charley and Katrina were at hurricane strength for both U.S. landfalls.

For each available Advisory, probability forecasts were computed for landfalls at each of the 10 coastal segments indicated in Figure 1, for a total of 31360 individual probability forecasts. Probability forecasts for U.S. tropical cyclone landfalls at any strength (hurricanes, tropical storms, and tropical depressions all considered "hits"), and landfalls at hurricane strength only, are considered separately. The forecasts are evaluated using reliability diagrams (e.g., Wilks 2006), which consist of plots of conditional event relative frequencies as a function of (binned) forecast probabilities (known as the calibration function), together with the frequencies-of-use of each of the forecast probabilities (the refinement distribution). The reliability diagram is a graphical representation of the joint frequency distribution of the forecasts and observations (Murphy and Winkler 1987), and so portrays the full information content of the verification data, allowing diagnosis of the key forecast performance attributes (Wilks 2006). Here, the forecast probabilities can take on any value consistent with the precision of the weighted bootstrap procedure, which is 10^{-4}. However, for purposes of plotting reliability diagrams these have been binned by rounding to the nearest tenth, with the horizontal positions of the plotted points reflecting the average forecast value within each of the eleven bins.

Figure 4 shows reliability diagrams for landfalls at the ten coastal segments in Figure 1, where landfall of a tropical cyclone at any strength is counted as a "hit." For landfalling storms, the lead time stratifications in panels (a) through (d) pertain to the time between the issuance of the advisory and the time of landfall according to the respective Tropical Cyclone Report. For other storms, the stratification is in terms of time until the final Forecast Advisory. The calibration functions (black dots connected by heavy lines) in Figure 4 indicate very good performance of these forecasts. The closeness of the calibration functions to the diagonal 1:1 line indicate that the forecasts "mean what they say," in the sense that event relative frequencies correspond well to the stated forecast probabilities. This remains true even at the longer lead times (panels c and d), which are necessarily greater than the longest (120-hour) lead time of the NHC Forecast Advisories on which they are based. Brier skill scores (e.g., Wilks 2006) range from 58.0% for the shorter lead times (Figure 4a) through 4.1% for the > 10-day lead times (Figure 4d).

The inset histograms in Figure 4 show frequencies of use of forecasts in the eleven probability bins. In all cases, a large majority of forecasts are smaller than 0.05, and so are placed in the "zero" bin. Most of these zero or near-zero probability forecasts reflect "easy" cases, for which the forecast storm track is well away from the U.S. coastline, for example toward southern Mexico or Central America, or eastward into the Atlantic Ocean. However, especially for the shorter lead times, there are substantial numbers of forecasts with relatively large U.S. landfall probabilities, which could potentially be valuable for decision making.

Figure 5 shows corresponding results for U.S. landfalling hurricanes, only. Again, these forecasts show generally good calibration, although the forecasts for the shorter lead times exhibit some underconfidence (Wilks, 2006), and so could probably be improved through recalibration. Brier skill scores in Figure 5 range from 46.3% at the shorter lead times (Figure 5a) through –1.4% for lead times greater than 10 days (Figure 5d). The inset histograms in Figure 5 all show larger fractions of zero and near-zero forecasts than their counterparts in Figure 4, which reflects the fact that hurricane landfalls are rarer than landfalls of tropical cyclones of all strengths.

5. Application to the "cone of uncertainty"

This section considers the landfall probability forecasts described above in relation to the "cone of uncertainty," portrayed as the white and white-shaded zone surrounding the official forecast track on NHC Forecast Advisory maps such as Figure 2. Public dissemination of this graphical device was initiated in 2002, and awareness of the cone among the public is widespread during tropical cyclone events (Broad et al. 2006). For the 2002–2006 forecasts investigated in this section, these cones were constructed as the union of tangents to circles centered at each of the forecast positions (black dots in Figure 2) with the intercepted outer arc of the final circle, where the circle radii correspond to average forecast position errors over the previous 5 years (James Franklin, personal communication, 2006; www.nhc.noaa.gov/aboutcone.shtml). These 5-year average errors are similar in magnitude to the 5-year average position error values in Table 1.

Consistent with considering forecasts for U.S. landfalling tropical cyclones above, only Forecast Advisory cones that fully intersect the U.S. coastline on both flanks have been analyzed here, yielding 214 cases. Table 3 stratifies these cases by lead time, and tabulates the percentages of forecasts in which the eventual landfall was within the intersection of the cone and the U.S. coastline. The eventual landfalls were within these forecast cones in roughly 90% of cases overall. The 95% confidence intervals for these percentages in Table 3 are bootstrap estimates which, because consecutive forecasts for a given storm are strongly correlated, have been computed using these sequences of same-storm forecasts as blocks in a block-bootstrap procedure (Efron and Tibshirani 1993, Wilks 2006).

The results in Table 3 are derived only from NHC data, and do not relate to the forecast procedure described in Section 2. Figure 6 shows distributions of forecast probabilities, computed as described in Section 2, for the event that an eventual landfall is within its respective cone, for the same 214 cases. The mean and median forecast probabilities are reasonably consistent with the relative frequencies in Table 3, in that typical forecast probabilities are near 90%, although appreciable case-to-case variability is evident.

This approximate 90% coverage probability within the cones for U.S. tropical cyclone landfalls is larger than might be expected. If the forecast position errors follow a circular bivariate normal distribution, the probability of a 2-dimensional position error smaller than the average error (i.e., the radii defining the cone widths for 2002–2006) is $1 - \exp(-1) = 0.632$ (for 2007 and beyond the NHC error radii have been extended slightly to yield a value of two-thirds for this probability). However, as noted on the NHC website in the discussion on this point (www.nhc.noaa.gov/aboutcone.shtml), the actual coverage probability should be larger because of possible forecast timing errors: a storm that is much faster or slower than forecast, but which nevertheless follows the forecast track reasonably closely, will still be counted as having remained within the cone, even though it might be outside the 63.2% error circle for one or more lead times.

Another contribution to the relatively large probabilities in Table 3 and Figure 6 may be the fact that these 214 cases compose a biased sample of the forecast error cones, in that all have necessarily been initialized from positions relatively near the U.S. coastline, and so have benefited from relatively more accurate observations of initial storm characteristics. Previous studies (Neumann and Pelissier 1981, Gray et al. 1991) have noted that such storms tend to be more accurately forecast. The decline in average and median within-cone forecast probabilities with increasing lead times in Figure 6 suggests that this effect also contributes to the relatively high probabilities subtended by the cones, shown in Table 3, for landfalls at the U.S. coastline.

6. Summary and conclusions

A method has been described to temporally extend and spatially disaggregate the NHC Advisory Forecasts for tropical cyclone tracks and intensities. The procedure begins with a process similar to that of the HURRAN forecasts (Hope and Neumann 1970), in which analog storms were translated spatially to the observed location of a current storm, and then extended forward in time according to their historical paths. However, here the analog storms are translated to random positions representative of a forecast future time, including explicit and quantitative accounting for errors in the forecast location. Thus, the method draws upon and extends the combined dynamical, statistical, and subjective human expertise on which the NHC Advisories are based. The resulting population of projected analog storms is used to estimate landfall probabilities by tabulating relative frequencies with which coastal segments of interest are crossed. Here the method is similar in spirit to the coastal “strike probabilities” that were formerly issued by NHC (Sheets 1984); except that the probability evaluation is nonparametric, and accounts for details of coastline geometry and its relationship to the currently forecast and local climatological storm tracks. Similarly, the nonparametric extrapolation of NHC Advisory forecasts here can be compared to the parametric approaches of Hall and Jewson (2007) and Regnier and Harr (2006), although neither of these papers initiate storm track extrapolations from forecast cyclone positions.

Forecast performance was evaluated for landfall probabilities at ten U.S. Gulf and Atlantic coastal segments. These probabilities were found to be very well calibrated (i.e., "reliable"), and to exhibit skill even beyond the maximum 120-hour lead time of the NHC Forecast Advisories upon which they are based. These positive attributes were exhibited by both forecasts for tropical cyclones of any intensity (essentially, storm track forecasts), as well as landfall probabilities for hurricanes only. Of course, at longer lead times the range of probabilities is restricted to the smaller values, consistent with the greater uncertainty more than a few days into the future, and the forecast skills are lower.

The capacity to produce well-calibrated forecasts at lead times substantially greater than 5 days is a significant attribute of the forecast procedure described here. These forecasts will be better (notably, sharper) than very long-lead tropical cyclone forecasts produced using a conditional climatological approach, such as those described by Bretschneider (2008) and Regnier and Harr (2006), which are similar in spirit to the HURRAN (Hope and Neumann, 1970) approach. In particular, in the present approach the forecasts for lead times longer than 5 days are produced by appending climatological information to the 5-day NHC track and intensity forecast, accounting explicitly for uncertainties in those forecasts, rather than by considering climatological distributions conditional on a currently observed tropical cyclone location.

In addition to forecasting tropical cyclone landfalls across fixed coastal segments, the performance of the method was also investigated in relation to the "cone of uncertainty," the geographical extent of which is specific to each Forecast Advisory. Here the distributions of forecast probabilities agreed well with the raw relative frequencies of "hits" within these cones for U.S. landfalling storms, yielding roughly a 90% coverage probability. This coverage probability is larger than might have been anticipated, and likely derives from a combination of certain forecast timing errors not being captured by the intersection of the cones with the coastline, together with the near-U.S. storms in this sample being generally better observed at the time of forecast initialization than storms occurring further from the U.S. coastline. The present method could probably be extended to include landfall timing in addition to geographical location of the landfalls, by including also the speed of storm movement as a criterion for analog selection.

The procedure for selection of analog storms includes several adjustable parameters, namely limits on similarity of analog storm location, intensity, and direction of movement. The values for these parameters have been chosen subjectively, but sensitivity tests showed little impact on the probability forecasts of varying them through reasonable ranges. Similarly, Neumann and Pelissier (1981) found that that NHC forecast position errors during the 1970s were elliptical, not circular, but choosing initial bootstrap forecast positions from elliptical bivariate normal distributions also yields only small changes in the resulting probability forecasts. Still, it is certainly possible that overall performance of the procedure could be improved through analysis of a comprehensive tuning exercise for choice and weighting of analog storms.

Finally, forecast verifications have been computed using the relatively wide coastline segments shown in Figure 1, because of the limited number of U.S. landfalling storms during the 1998–2006 period for which NHC Advisory Forecasts in the current format have been available. Having demonstrated that these forecasts are well calibrated and skillful, use of the method to evaluate landfall probabilities for smaller regions could be undertaken with some confidence. For example, landfall probabilities in cases where the NHC cone only partially intersects the coastline could easily be evaluated. Similarly, landfall probabilities for smaller coastal segments such as individual counties could be computed, although in this case bootstrap samples larger than 10,000 might be desirable. Conversely, coastal segments encompassing specific landfall probabilities could also be computed.

Acknowledgement

This research has been supported in part by Kenneth Horowitz and Weather Risk Solutions, LLC.

References

Brettschneider, B., 2008: Climatological hurricane landfall probability for the United States. *Journal of Applied Meteorology and Climatology*. **47**: 704-716.

Broad, K., A. Leiserowitz, J. Weinkle, and M. Steketee, 2007: Misinterpretations of the "cone of uncertainty" in Florida during the 2004 hurricane season. *Bulletin of the American Meteorological Society*, **88:** 651-667.

Efron, B., and R. Tibshirani, 1993 *An Introduction to the Bootstrap*, Chapman and Hall, 436 pp.

Gray, W.M., C.J. Neumann, and T.L. Tsui, 1991: Assessment of the role of aircraft reconnaissance on tropical cyclone analysis and forecasting. *Bulletin of the American Meteorological Society*, **72**: 1867-1883.

Gross, J.M., and M. DeMaria, 2004. A new method for determining tropical cyclone wind forecast probabilities. Preprints, *26th Conf. on Hurricanes and Tropical Meteorology*, Amer. Met. Soc., 425-426.

Hall, T.M., and S. Jewson, 2007: Statistical modelling of North Atlantic tropical cyclone tracks. *Tellus*, **59A**: 486-498.

Hallegatte, S., 2008: The use of synthetic hurricane tracks in risk analysis and climate change damage assessment. *Journal of Applied Meteorology and Climatology*, **46**: 1956-1966.

Hope, J.R. and C.J. Neumann, 1970: An operational technique for relating the movement of existing tropical cyclones to past tracks. *Monthly Weather Review*, **98**: 925-933.

Jarvinen, B.R., C.J. Neumann, and M.A.S. Davis, 1984: A tropical cyclone data tape for the North Atlantic basin, 1886-1983: Contents, limitations, and uses. NOAA Technical Memo. NWS NHC-22. 21 pp.

Katz, R.W., and A.H. Murphy, 1997: *Economic Value of Weather and Climate Forecasts*, Cambridge University Press, 222 pp.

Krzysztofowicz, R., 1983: Why should a forecaster and a decision maker use Bayes' theorem? *Water Resources Research*, **19**: 327-336.

Landsea, C.W., A. Anderson, N. Charles, G. Clark, J. Dunion, J. Fernandez-Partagas,, P. Hungerford, C.J. Neumann, and M. Zimmer, 2004: The Atlantic hurricane database re-analysis project: documentation for 1851-1910 alterations and additions to the HURDAT database. In: Murname, R.J., and K.B. Liu, eds, *Hurricanes and Typhoons: Past, Present and Future*, Columbia University Press, 177-221.

Murphy, A.H, and R.L. Winkler, 1987: A general framework for forecast verification. *Monthly Weather Review*, **115**, 1330-1338.

Neumann, C.J, 1972: An alternate to the HURRAN tropical cyclone forecast system. *NOAA Technical Memorandum SR-62*, National Weather Service, Southern Region, 24pp.

Neumann, C.J. and J.M. Pelissier, 1981: An analysis of Atlantic tropical cyclone forecast errors, 1970-1979. *Monthly Weather Review*, **109**: 1248-1266.

Regnier, E., and P.A. Harr, 2006. A dynamic decision model applied to hurricane landfall. *Weather and Forecasting*, **21**: 764-780.

Sheets, R.C., 1984: The National Weather Service hurricane probability program. NOAA Technical Report NWS 37, 70 pp.

Wilks, D.S., 2006: *Statistical Methods in the Atmospheric Sciences*, 2nd Edition. Academic Press, 627 pp.

Table 1. Mean absolute errors for forecast position, and standard deviations of relative errors in forecast maximum sustained windspeed, for NHC forecasts 2001-2005.

Lead time, h	MAE, nm	Relative error std. deviation
12	37.3	0.16
24	64.5	0.24
36	91.3	0.31
48	118.3	0.39
72	171.4	0.55
96	231.1	0.60
120	303.3	0.67

Table 2. Forecast landfall probabilities, shown at 12-h intervals, over each of the 10 coastal segments in Figure 1, for (a) Hurricane Emily (2005), and (b) Tropical Storm Ernesto (2006). Segment "0" indicates the event that no landfall occurs on the U.S. coastline.

a. Hurricane Emily (2005)

NHC Adv #	Time & Date	Probability of Landfall at Segment #										
		0	1	2	3	4	5	6	7	8	9	10
1	03Z 11 Jul	0.496	.008	.007	.011	.015	.034	.038	.145	.119	.055	.072
3	15Z 11 Jul	0.376	.045	.038	.034	.038	.059	.067	.194	.089	.033	.026
5	03Z 12 Jul	0.366	.011	.006	.014	.024	.065	.082	.205	.103	.054	.070
7	15Z 12 Jul	0.354	.073	.069	.091	.076	.090	.083	.121	.030	.008	.004
9	03Z 13 Jul	0.363	.124	.105	.092	.075	.090	.086	.056	.002	.003	.005
11	15Z 13 Jul	0.744	.064	.042	.046	.044	.037	.018	.005	.000	.000	.000
13	03Z 14 Jul	0.411	.139	.109	.096	.081	.067	.059	.035	.000	.001	.002
15	15Z 14 Jul	0.793	.123	.057	.015	.010	.003	.000	.000	.000	.000	.000
17*	03Z 15 Jul	0.597	.207	.134	.044	.014	.002	.002	.000	.000	.000	.000
19	15Z 15 Jul	0.597	.241	.118	.033	.008	.001	.001	.000	.000	.000	.000
21	03Z 16 Jul	0.677	.199	.069	.036	.009	.008	.002	.000	.000	.000	.000
23	15Z 16 Jul	0.759	.193	.044	.004	.000	.000	.000	.000	.000	.000	.000
25	03Z 17 Jul	0.871	.090	.031	.007	.000	.001	.000	.000	.000	.000	.000
27	15Z 17 Jul	0.903	.086	.010	.001	.000	.000	.000	.000	.000	.000	.000
29	03Z 18 Jul	0.793	.153	.022	.030	.002	.000	.000	.000	.000	.000	.000
31	15Z 18 Jul	0.779	.217	.005	.000	.000	.000	.000	.000	.000	.000	.000
33	03Z 19 Jul	0.982	.018	.000	.000	.000	.000	.000	.000	.000	.000	.000
35	15Z 19 Jul	0.996	.004	.000	.000	.000	.000	.000	.000	.000	.000	.000
38	03Z 20 Jul	1.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000

* Advisory #17 is shown in map form in Figure 2.

b. Tropical Storm Ernesto (2006)

NHC Adv #	Time & Date	Probability of Landfall at Segment #										
		0	1	2	3	4	5	6	7	8	9	10
2	03Z 25 Aug	0.088	.069	.100	.158	.162	.163	.121	.113	.012	.006	.007
4	15Z 25 Aug	0.085	.086	.103	.162	.169	.167	.122	.089	.002	.006	.009
6	03Z 26 Aug	0.215	.127	.186	.194	.158	.084	.034	.004	.000	.000	.000
8	15Z 26 Aug	0.018	.064	.090	.150	.178	.218	.151	.106	.001	.009	.015
10	03Z 27 Aug	0.116	.007	.031	.121	.214	.354	.143	.014	.000	.000	.000
12	15Z 27 Aug	0.105	.072	.057	.021	.056	.083	.086	.429	.037	.042	.014
14	03Z 28 Aug	0.134	.000	.000	.000	.004	.081	.184	.460	.074	.048	.014
16	15Z 28 Aug	0.202	.000	.000	.000	.000	.009	.026	.474	.184	.072	.033
18	03Z 29 Aug	0.117	.000	.000	.000	.000	.013	.018	.715	.100	.036	.001
20	15Z 29 Aug	0.013	.000	.000	.000	.000	.004	.007	.909	.038	.027	.002

Table 3. U.S. landfalling tropical cyclones, 2002–2006, in relation to NHC forecast error "cones". Fifth column presents block-bootstrap estimates of 95% confidence intervals for the percent coverage values in the fourth column.

Lead time, hours	# in cone	# forecast	% in cone	(95% C.I.)
≤ 12	34	39	87.2	(74.4% – 97.4%)
13 – 48	115	122	94.3	(85.5% – 98.4%)
49 – 120	43	53	81.1	(60.4% – 100.0%)
All	192	214	89.7	(83.2% – 95.3%)

Figure Captions

Figure 1. The Gulf and Atlantic U.S. coastline, divided into ten segments with approximately equal climatological probabilities of tropical cyclone landfalls. Relative hurricane strike probabilities, as estimated using data from Jarvinen et al. (1984) for the years 1851–2005, are shown parenthetically.

Figure 2. NHC Forecast Advisory 17 for hurricane Emily (2005). Final (120-hour) forecast position forms the basis for extension and probability disaggregation illustrated in Figure 3.

Figure 3. Illustration of the forecast procedure, using the 120-hour forecast position of hurricane Emily (2005) shown in Figure 2. (a) Forecast position (X) and locations (black dots) of the 68 historical analog positions. Consecutive positions of the same storm are connected by thin black lines. Grey dots and lines indicate subsequent tracks of three of the 23 storms. (b) 90% probability contour for the 120-hour position error distribution, and twenty random overwater locations drawn from this distribution. Each of these random points initializes extension of one of the eleven analog positions (larger dots) of the three storms identified in panel (a). An actual forecast would use 10,000 random initial points, and draw from all 68 initial positions of the 23 analog storms shown as black dots in panel (a).

Figure 4. Reliability diagrams for tropical cyclone (any intensity) landfall probabilities, at the ten coastal segments shown in Figure 1. Horizontal axes show average binned forecast probabilities, and vertical axes indicate corresponding event relative frequencies. Inset histograms show frequencies of use of the eleven rounded probability values, with only subsample sizes ≥10 plotted. (a) up to 2-day lead time, (b) 2- to 5-day lead time, (c) 5- to 10-day lead time, and (d), greater than 10-day lead time. Sample sizes (*n*), Brier scores (BS), and skills relative to the sample climatological relative frequencies (SS) are also indicated.

Figure 5. As Figure 4, for hurricane-strength landfalls only.

Figure 6. Histograms for probability forecasts of tropical cyclone landfalls within the NHC "cone of uncertainty," for cones entirely intersecting the U.S. coastline, 2002–2006, at lead times of (a) 12 hours or less, (b) 13 to 48 hours, and (c) 49 – 120 hours.

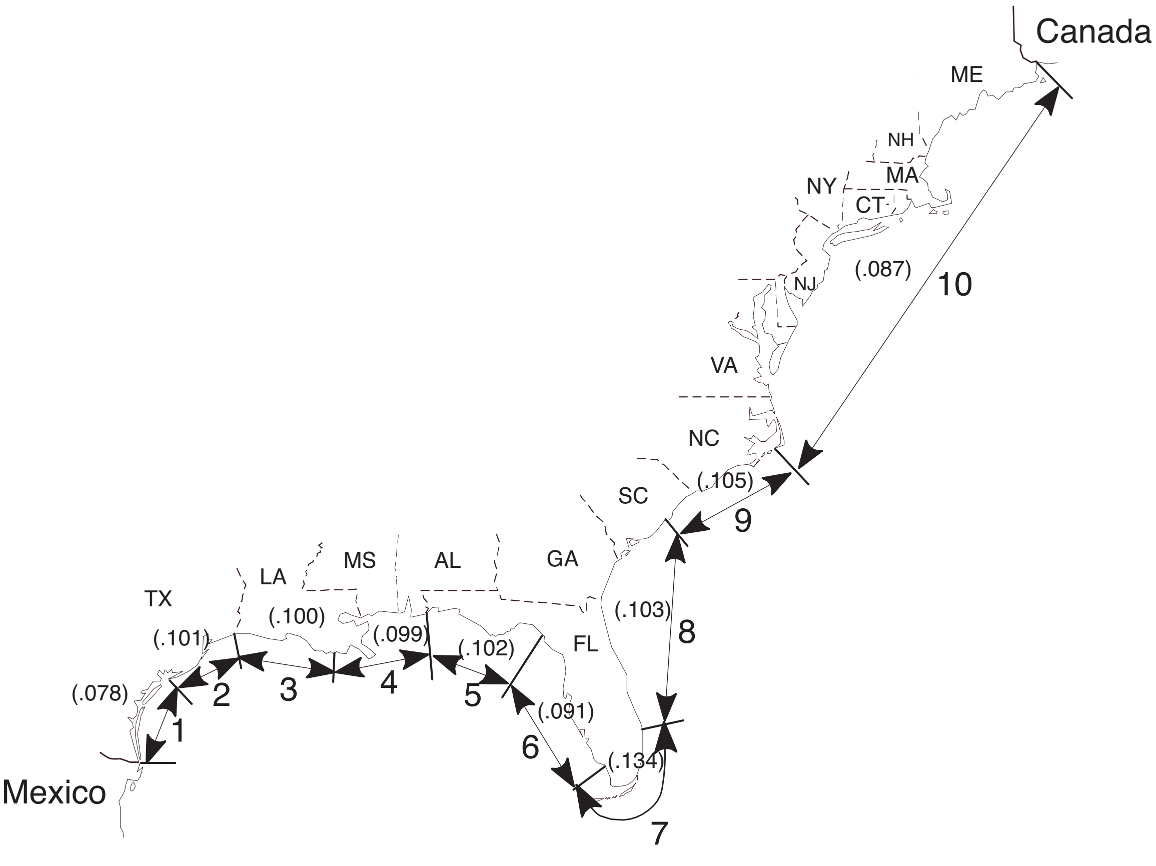

Figure 1. The Gulf and Atlantic U.S. coastline, divided into ten segments with approximately equal climatological probabilities of tropical cyclone landfalls. Relative hurricane strike probabilities, as estimated using data from Jarvinen et al. (1984) for the years 1851–2005, are shown parenthetically.

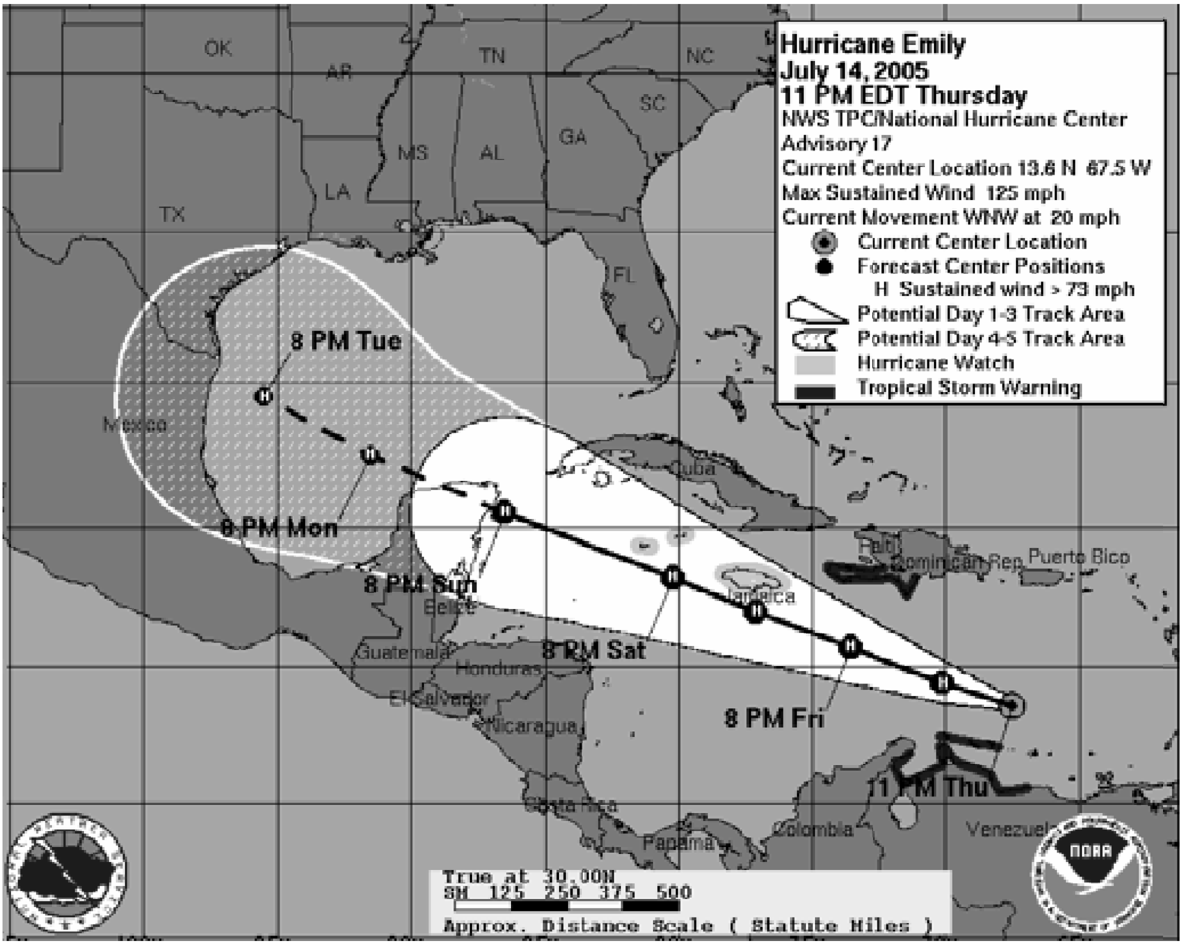

Figure 2. NHC Forecast Advisory 17 for hurricane Emily (2005). Final (120-hour) forecast position forms the basis for extension and probability disaggregation illustrated in Figure 3.

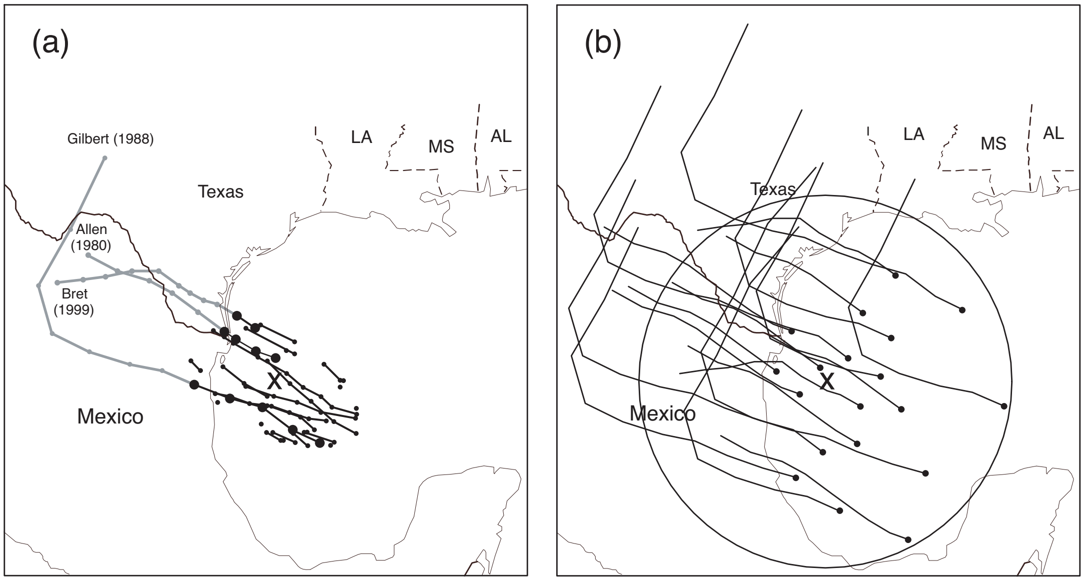

Figure 3. Illustration of the forecast procedure, using the 120-hour forecast position of hurricane Emily (2005) shown in Figure 2. (a) Forecast position (X) and locations (black dots) of the 68 historical analog positions. Consecutive positions of the same storm are connected by thin black lines. Grey dots and lines indicate subsequent tracks of three of the 23 storms. (b) 90% probability contour for the 120-hour position error distribution, and twenty random overwater locations drawn from this distribution. Each of these random points initializes extension of one of the eleven analog positions (larger dots) of the three storms identified in panel (a). An actual forecast would use 10,000 random initial points, and draw from all 68 initial positions of the 23 analog storms shown as black dots in panel (a).

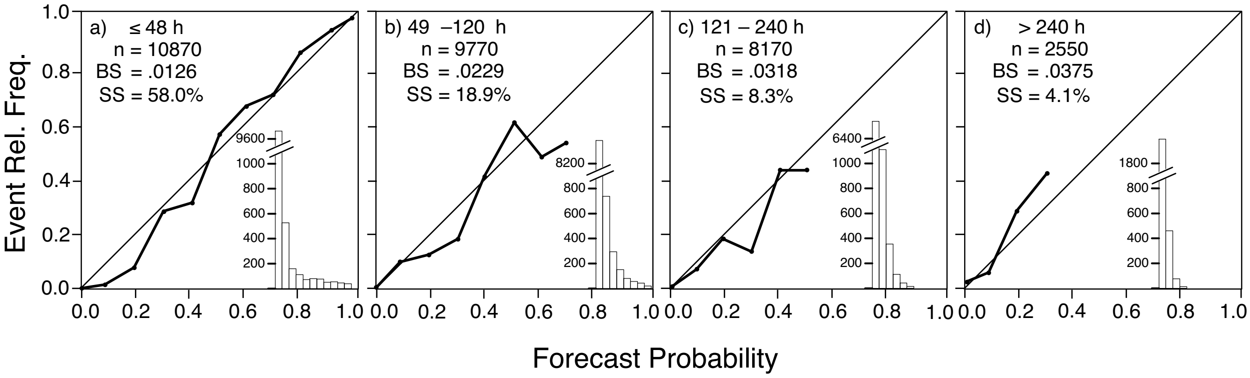

Figure 4. Reliability diagrams for tropical cyclone (any intensity) landfall probabilities, at the ten coastal segments shown in Figure 1. Horizontal axes show average binned forecast probabilities, and vertical axes indicate corresponding event relative frequencies. Inset histograms show frequencies of use of the eleven rounded probability values, with only subsample sizes ≥10 plotted. (a) up to 2-day lead time, (b) 2- to 5-day lead time, (c) 5- to 10-day lead time, and (d), greater than 10-day lead time. Sample sizes (n), Brier scores (BS), and skills relative to the sample climatological relative frequencies (SS) are also indicated.

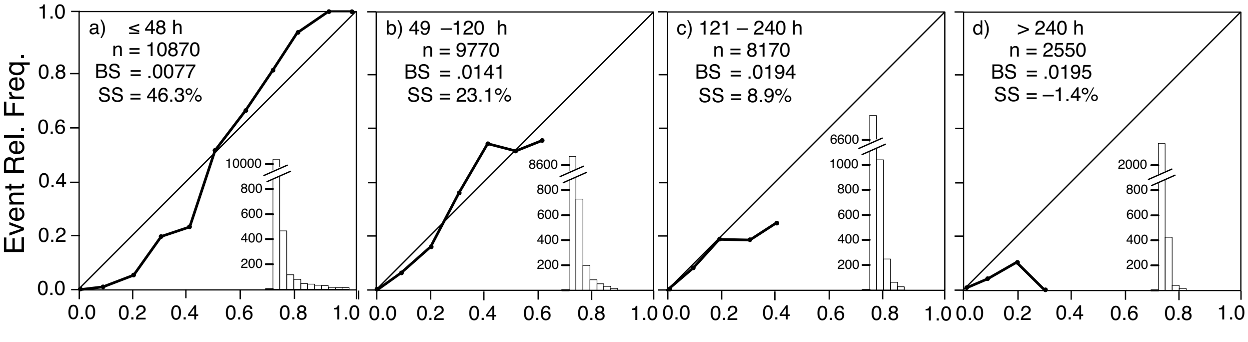

Figure 5. As Figure 4, for hurricane-strength landfalls only.

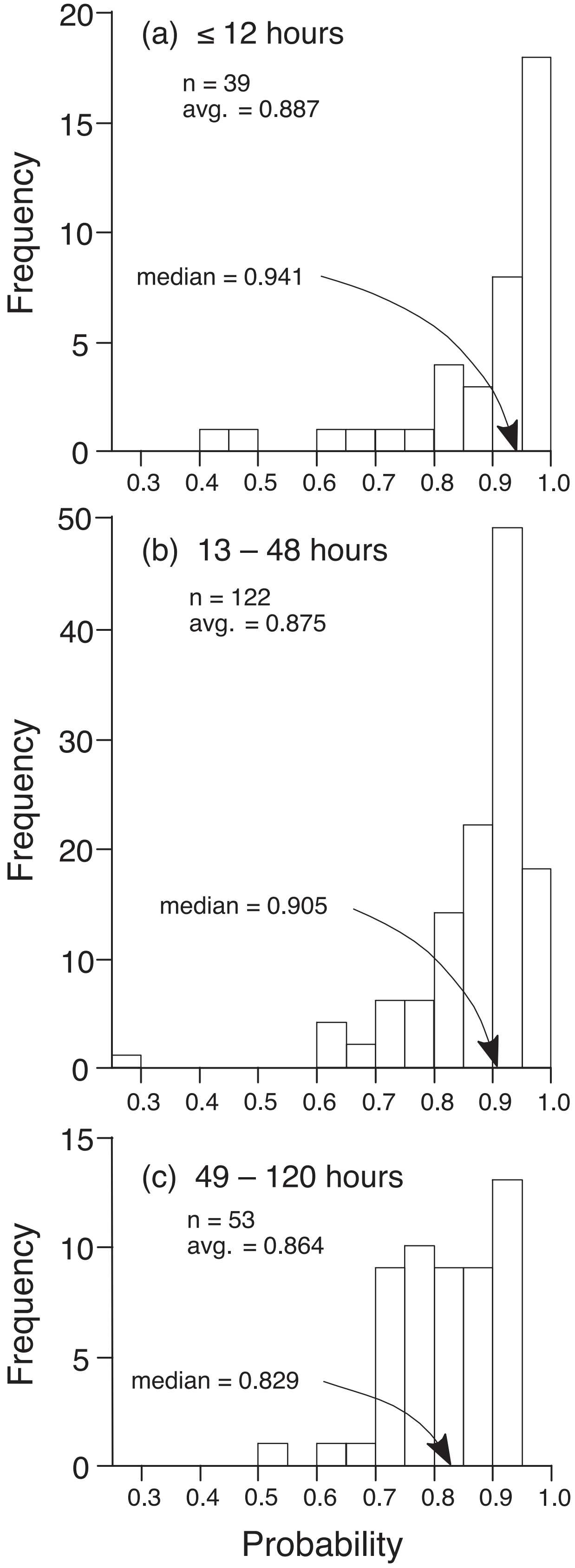

Figure 6. Histograms for probability forecasts of tropical cyclone landfalls within the NHC "cone of uncertainty," for cones entirely intersecting the U.S. coastline, 2002–2006, at lead times of (a) 12 hours or less, (b) 13 to 48 hours, and (c) 49 – 120 hours.