13A.1 NATIONAL HURRICANE CENTER FORECAST VERIFICATION

James L. Franklin
NOAA/NWS/NCEP/National Hurricane Center, Miami, Florida

1. INTRODUCTION

For all operationally-designated tropical (or subtropical) cyclones in the Atlantic and eastern North Pacific basins, the National Hurricane Center (NHC) issues an "official" forecast of the cyclone's center location and maximum 1-min surface wind speed. Forecasts are issued every 6 hours, and contain projections valid 12, 24, 36, 48, 72, 96, and 120 h after the forecast's nominal initial time (0000, 0600, 1200, or 1800 UTC). At the conclusion of the season, forecasts are evaluated by comparing the projected positions and intensities to the corresponding post-storm derived "best track" positions and intensities for each cyclone. A forecast is included in the verification only if the system is classified in the final best track as a tropical or subtropical cyclone at both the forecast's initial time and at the projection's valid time. All other stages of development (e.g., tropical wave, [remnant] low, extratropical) are excluded. For verification purposes, forecasts associated with special advisories do not supersede the original forecast issued for that synoptic time; rather, the original forecast is retained.

It is important to distinguish between *forecast error* and *forecast skill*. Track forecast error is defined as the great-circle distance between a cyclone's forecast position and the best track position at the forecast verification time. Skill, on the other hand, represents a normalization of forecast error against some standard or baseline, and is positive when the forecast error is smaller than the error from the baseline. To assess the degree of skill in a set of track forecasts, the track forecast error can be compared with the error from CLIPER5, a climatology and persistence model that contains no information about the current state of the atmosphere (Neumann 1972, Aberson 1998). Errors from the CLIPER5 model are taken to represent a "no-skill" level of accuracy that can be used as a baseline for evaluating other forecasts. If CLIPER5 errors are unusually low during a given season, for example, it indicates that the year's storms were inherently "easier" to forecast than normal or otherwise unusually well behaved. The current version of CLIPER5 is based on developmental data from 1931-2004 for the Atlantic and from 1949-2004 for the eastern Pacific.

Particularly useful skill standards are those that do not require operational products or inputs, and can therefore be easily applied retrospectively to historical data. CLIPER5 satisfies this condition, since it can be run using persistence predictors (e.g., the storm's current motion) that are based on either operational or best track inputs. The best-track version of CLIPER5, which yields substantially lower errors than its operational counterpart, is generally used to analyze lengthy historical records for which operational inputs are unavailable. Forecasters, of course, see only the operational version of CLIPER5, and therefore this version is the more appropriate one for the verifications discussed below.

Forecast intensity error is defined as the absolute value of the difference between the forecast and best track intensity at the forecast verifying time. Skill in a set of intensity forecasts is assessed using Decay-SHIFOR5 (DSHIFOR5). The DSHIFOR5 forecast is obtained by initially running SHIFOR5, the climatology and persistence model for intensity that is analogous to the CLIPER5 model for track (Jarvinen and Neumann 1979, Knaff et al. 2003). The output from SHIFOR5 is then adjusted for land interaction by applying the decay rate of DeMaria et al. (2006). The application of the decay component requires a forecast track, which here is given by CLIPER5. On average, DSHIFOR5 errors are about 5-15% lower than SHIFOR5 in the Atlantic basin from 12-72 h, and about the same as SHIFOR5 at 96 and 120 h.

The verifications described in this report are based on forecast and best track data sets taken from the Automated Tropical Cyclone Forecast (ATCF) System on 29 January 2008. Here we briefly review the accuracy and skill of Atlantic basin NHC tropical cyclone forecasts in recent years, and present forecast trends, biases, and error distributions. A more complete review that includes a discussion of eastern North Pacific errors, guidance model performance, and a reference list, can be found in the NHC 2007 Verification Report, available at:

http://www.nhc.noaa.gov/verification/verify3.shtml?

2. FORECAST TRENDS

Figure 1 shows recent trends in NHC official Atlantic basin track errors and skill. Since 1990, 24-72 h track forecast errors have been reduced by a little more than 50%. Although the track error reductions appear to have been relatively constant over this period, there is some evidence for more significant drops in the mid-1990s, and again around 2003. The former period corresponds to the advent of the four primary dynamical models still in use today (GFDL, UKMET, AVN/GFS, and NOGAPS), while the latter occurred when the use of consensus techniques became more formalized (Goerss 2008, personal communication). The most abrupt skill increases occurred from 2000-2, with little change in skill since then.

Corresponding author address: James L. Franklin, NOAA National Hurricane Center, 11691 SW 17th Street, Miami, FL 33165; e-mail: James.Franklin@noaa.gov.

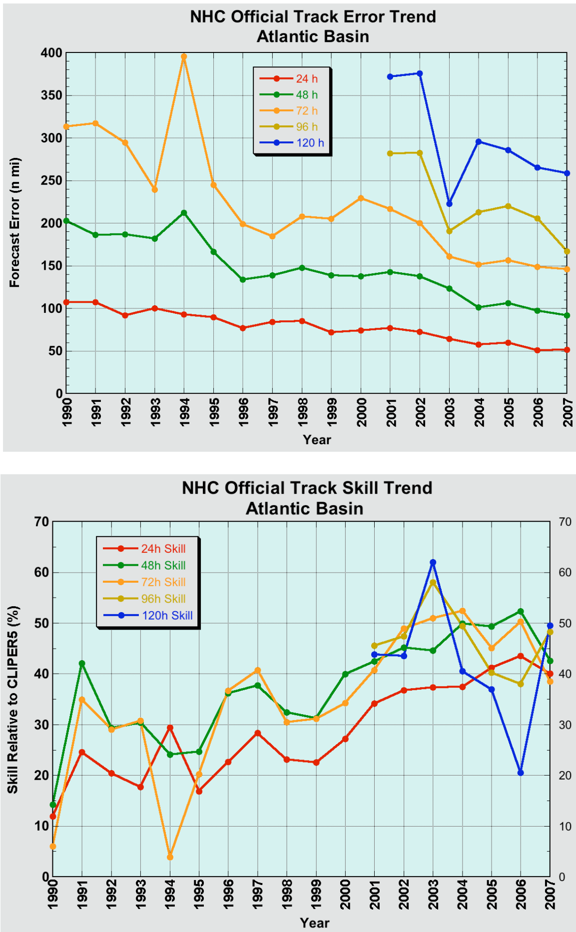

Figure 1. Atlantic basin annual mean track forecast error (top) and skill (bottom).

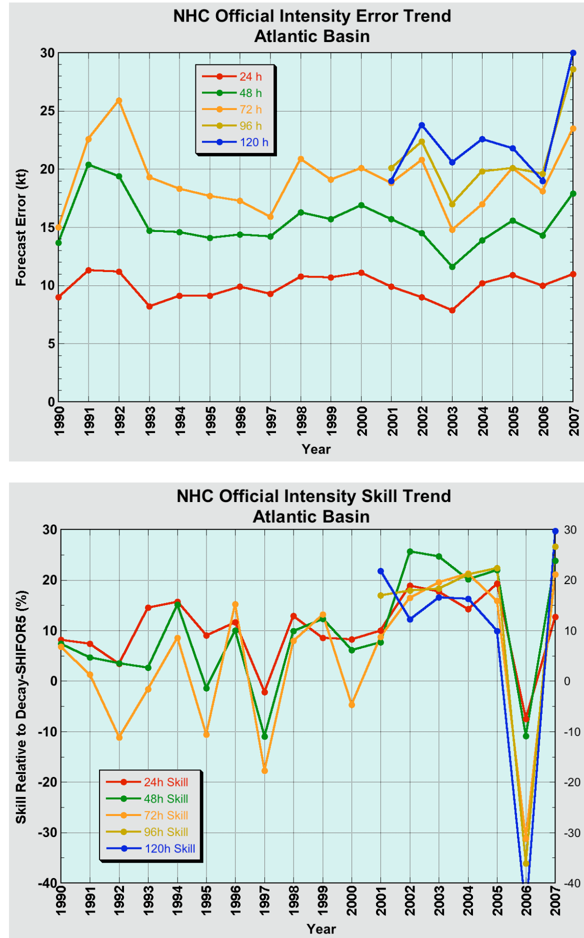

Figure 2 Atlantic basin annual mean intensity forecast error (top) and skill (bottom).

While track forecasts have shown clear improvements in recent years, the same cannot be said for tropical cyclone intensity forecasts (Fig. 2). There has been no detectable reduction in error over the period, although forecast skill does appear to have increased some in recent years. The difference between the error and skill trends is attributed to a increase in the number of intense and rapidly intensifying tropical cyclones over the past few seasons (except in 2006). These systems, which are associated with larger Decay-SHIFOR5 errors, are inherently harder to forecast. Indeed, there was some evidence of a downward error during the period ending in 2003.

3. 5-YEAR STATISTICS

Table 1 presents Atlantic basin summary statistics for official forecasts for the most recent 5-yr period 2003-7. Mean track and intensity errors are plotted in Fig. 3. Forecast track errors are seen to increase in a nearly linear fashion with forecast interval, at roughly 55 n mi/day. Because the range of possible intensity errors is relatively small, intensity errors tend to increase more slowly at the longer forecast periods. Forecast skill is considerably larger for track than it is for intensity (e.g., 49% and 14%, respectively, at 72 h. Interestingly, however, forecast biases for track are larger (as a fraction of the mean error) than their intensity counterparts, at least through 96 h. NHC official track forecasts have a bias generally directed to the northwest (i.e., forecast position tends to lie to the northwest of the verifying position). Intensity biases are essentially negligible through 72 h; modest negative biases (i.e., forecast intensity too low) are seen at 96 and 120 h (although they are still smaller than the 5 kt forecast increment).

Figure 4 shows the cumulative frequency distribution of the Atlantic basin track errors. This analysis is used to determine the size of the NHC track forecast "cone" that is a prominent feature on NHC web page storm graphics. The cone is meant to represent the probable track of the center of a tropical cyclone, and is formed by enclosing the area swept out by a set of imaginary circles placed along the forecast track. The size of each circle is set so that two-thirds of historical official forecast errors over the most recent 5-year sample fall within the circle. The figure lists the sizes of the circles that will form the cone in 2008.

Table 1. NHC official forecast errors for all tropical cyclones in the Atlantic basin for the period 2003-7. Verification is homogeneous with the climatology and persistence skill benchmarks CLIPER5 (for track) and Decay-SHIFOR5 (for intensity).

	Forecast Period (h)						
	12	24	36	48	72	96	120
Mean official track error (n mi)	34.0	58.2	82.3	106.2	154.2	207.5	272.5
Mean CLIPER5 track error (n mi)	46.6	96.6	152.6	205.9	301.0	393.1	480.2
Mean official track skill relative to CLIPER5 (%)	27	40	46	48	49	47	43
Mean official track bias vector (°/n mi)	307/7	312/15	316/23	320/32	317/33	328/29	001/38
Mean official along-track error (n mi)	23.1	39.0	55.3	72.3	105.9	139.9	179.8
Mean official cross-track error (n mi)	20.1	35.1	49.0	62.8	87.7	121.1	164.7
Mean official intensity error (kt)	6.7	10.0	12.3	14.3	18.2	19.7	21.8
Mean Decay-SHIFOR5 intensity error (kt)	8.0	11.7	14.9	17.7	21.2	23.9	24.5
Mean official intensity skill relative to Decay-SHIFOR5 (%)	16	15	17	19	14	18	11
Mean official intensity bias (kt)	0.0	0.1	-0.5	-1.2	-2.2	-3.9	-4.8
Number of cases	1742	1574	1407	1254	996	787	627

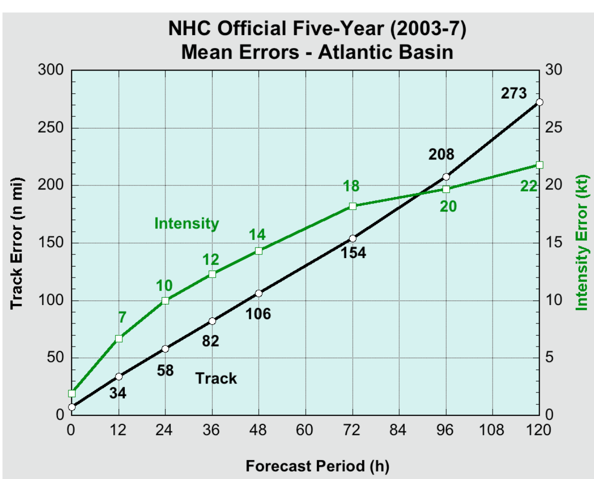

Figure 3 Average NHC official Atlantic basin track and intensity errors for the period 2003-7.

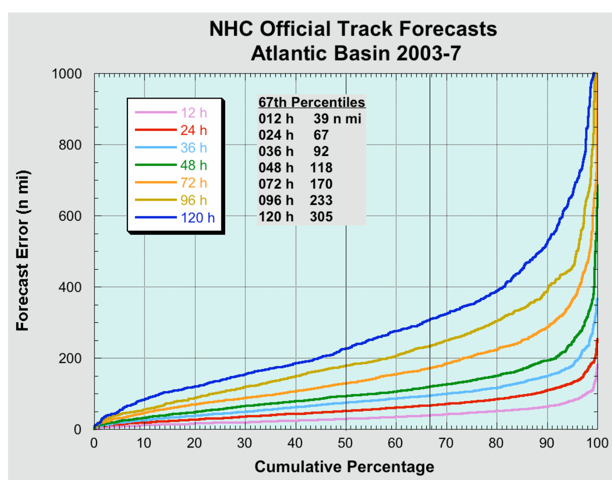

Figure 4. Cumulative frequency distribution for NHC official Atlantic basin track errors for the period 2003-7.

4. REFERENCES

Aberson, S. D., 1998: Five-day tropical cyclone track forecasts in the North Atlantic basin. *Wea. Forecasting,* **13**, 1005-1015.

DeMaria, M., J. A. Knaff, and J. Kaplan, 2006: On the decay of tropical cyclone winds crossing narrow landmasses, *J. Appl. Meteor.*, **45**, 491-499.

Jarvinen, B. R., and C. J. Neumann, 1979: Statistical forecasts of tropical cyclone intensity for the North Atlantic basin. NOAA Tech. Memo. NWS NHC-10, 22 pp.

Knaff, J.A., M. DeMaria, B. Sampson, and J.M. Gross, 2003: Statistical, five-day tropical cyclone intensity forecasts derived from climatology and persistence. *Wea. Forecasting*, **18**, 80-92.

Neumann, C. B., 1972: An alternate to the HURRAN (hurricane analog) tropical cyclone forecast system. NOAA Tech. Memo. NWS SR-62, 24 pp.