10C.2

Verification of National Hurricane Center Forecasts of Extratropical Transition

John L. Beven II
NOAA/NWS/NCEP/National Hurricane Center

I. Introduction

A great variety of cyclonic circulations exist in the atmosphere, each with its characteristic structure and driving energy (Beven 1997, Figure 1). Of particular interest is the tropical cyclone (TC), which has a warm-core non-frontal thermal structure generated by diabatic heat release from convective activity near the center. The strongest winds and heaviest rains in a tropical cyclone are typically within 100 km of the center with maximum sustained winds sometimes as high as 90 ms^{-1}.

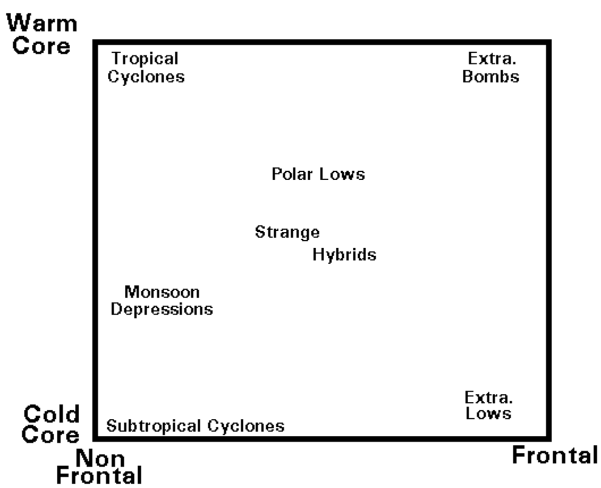

Figure 1. Conceptual model of cyclone frontal and thermal structures showing the general placement of various cyclone types. From Beven (1997).

Atmospheric cyclones can undergo significant structural and energetics changes during their lifetimes. Baroclinic frontal lows can become TCs in a process known as tropical transition (Davis and Bosart 2004). More commonly, TCs leaving the tropical environment interact with baroclinic systems in the westerlies. This causes the TCs to become frontal or extratropical cyclones with the driving energy derived from air mass contrast and the strongest winds typically more than 100 km from the center, accompanied by significantly changed precipitation patterns. This process is known as extratropical transition (ET).

There are several studies of ET which highlight the many ways a TC can interact with a baroclinic environment and the variety of resulting structures. These include the Thorncroft and Jones (2000) study of Hurricane Iris which became a powerful baroclinic cyclone with a warm-core structure, the Abraham et al. (2004) study of Hurricane Michael, and the Beven (2002) study of interrupted and failed transitions. These studies show the complexity of the process, which can pose a significant challenge to TC forecasters trying to predict ET.

A TC undergoing ET can also cause significant challenges from a warning and response standpoint. TC warning centers employ a well-developed set of forecast and warning processes with a proven track record of saving lives and protecting property. A different set of processes exist for extratropical cyclones, and these generally produce a different public response than those for TCs. Which warning processes are used is a binary decision superimposed on the continuous spectrum of the transition process. This can be problematic for cyclones undergoing ET that maintain their intensity and affect land, such as Noel in 2007 (Brown 2008), or for cyclones that are wrongly forecast to undergo ET.

The National Hurricane Center (NHC) has been including forecasts of ET in its products since 1960. However, unlike track and intensity

Corresponding author address: John L. Beven, NOAA/NWS/NCEP/TPC/, 11691 SW 17th Street, Miami, FL 33165-2149; e-mail: John.L.Beven@noaa.gov.

forecasts (Franklin 2008), there has been no systematic verification of these forecasts. The current work begins such a verification, examining how well forecasts of "will ET occur" verify, as well as the timing errors of the forecasts.

II. Methodology

The data for the study consists of the official NHC TC 12, 24, 36, 48, 72, 96, and 120-h forecasts for the period 1993-2007 in the Atlantic basin and the associated 6-hourly "best track" data. In the forecasts, the time the forecast designated the cyclone as "extratropical" was used as time of ET, based on the assumption that the transformation process was complete at that time. (Many forecasts used the term "becoming extratropical", which was assumed to show that the ET process was underway, but not complete. These forecasts were not used in the verification.) In the "best track" data, the time that the cyclone status was first listed as extratropical was used as the verifying time for the completion of ET. It should be noted that the "best track" had to include at least one 6-h position with the cyclone designated as extratropical. Cyclones that were absorbed by frontal systems leaving no traceable circulation being counted as dissipation rather than transition. This criterion has a notable impact on the skill scores described below.

The verification comes in two parts. The first part is examines the timing errors of the ET forecasts, looking at means and biases over the entire forecast period and for the individual forecast times.

The second part examines the categorical accuracy of the ET forecasts based on the two-dimensional contingency diagram methodology used by the National Weather Service for extreme event warning verification (Doswell et al. 1990). An example of such a table is shown in Table 1, with the associated statistical verification metrics given in Table 2. The metrics include: 1. The Critical Success Index (CSI), Probability of Detection (POD), False Alarm Ratio (FAR), and Event Bias, which use the cases where ET was forecast or occurred, 2. The Gilbert Skill Score, Heidke Skill Score, and True Skill Statistic, which

663 total forecasts	**ET Observed**	**ET Not Observed**
ET Forecast	**A - 155** **(successfully forecast ET)**	**B - 57** **(false alarms)**
ET Not Forecast	**C - 49** **(missed transitions)**	**D - 402** **(successfully forecast non-ET)**

Table 1. Example of the two-dimensional contingency diagram used for ET forecast verification showing data for the 2005 Atlantic hurricane season. The lettering A-D denotes the quantities used in the formulas in Table 2.

Skill Measure	**Formula**	**Range**	**Skillful Values**
Critical Success Index (CSI)	A/(A+B+C)	0.0 to 1.0	Close to 1.0
Probability of Detection (POD)	A/(A+C)	0.0 to1.0	Close to 1.0
False Alarm Ratio (FAR)	B/(A+B)	0.0 to1.0	Close to 0.0
Percentage Correct	(A+D)/ (A+B+C+D)	0.0 to1.0	Close to 1.0
Event Bias	(A+B)(A+C)	0.0 to ?	Close to 1.0
Chance Hits (CH)	(A+B)(A+C)/ (A+B+C+D)	N/A	N/A
Correct Random Forecasts (CRF)	[(A+B)(A+C)+ (B+D)(C+D)]/ (A+B+C+D)	N/A	N/A
Gilbert Skill Score	(A-CH)/ (A+B+C-CH)	-0.33 to 1.0	Close to 1.0
Heidke Skill Score	(A+D-CRF)/ (A+B+C+D-CRF)	-1.0 to1.0	Close to 1.0
True Skill Statistic	(AD-BC)/ (A+C)(B+D)	-1.0 to 1.0	Close to 1.0

Table 2. Formulae and score ranges for the skill metrics associated with the 2-D contingency table. The letters in the formulae are defined in Table 1.

all use the numerous cases where ET was neither forecast nor occurred, and 3. The percentage of the ET forecasts that are correct.

It should be noted the NHC extended its forecast period from three days to five days internally in 2001 and publicly in 2003. In order to maintain homogeneity in the ET forecast verification, separate verifications were created for the 3-day and 5-day forecasts.

Forecasts of position, intensity, and wind radii are not currently verified for the extratropical stage in normal NHC post-analysis, and this study does not verify those forecasts either.

III. Results

a. Timing errors

Figure 2 shows the mean absolute timing errors and biases for 5-day ET forecasts averaged over the entire forecast period. The average timing errors are between 6-18 h for most of the years, with an average 2007 error of less than 6 h. The associated timing biases are generally less than 12 h, with negative biases in 2001 gradually decreasing over time. This means that a tendency to forecast ET too late compared to the verifying best track time in 2001 has decreased over time. It should be noted that the minimum time between NHC forecast points is 12 h, and the tendency for the average timing errors to oscillate around 12 h

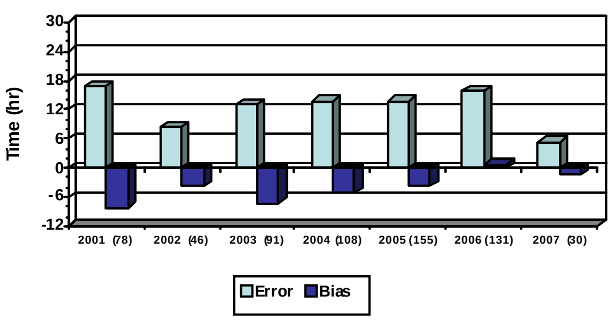

Figure 2. Five-day ET forecast timing errors and biases averaged over the entire forecast period. The number of forecast is listed in parentheses next to the year labels. Negative bias values indicate that ET was forecast to occur after the time of transition in the best track.

may be related to this rather coarse temporal resolution.

The timing errors and biases for the 5-day forecasts broken down for individual times are shown in Figure 3. As might be expected, the shorter-range (12-48 h) forecasts have smaller errors and biases than the longer-range (72-120 h) forecasts. The forecast errors are generally near or less than the time interval between the respective forecast (12 h for the short range and 24 h for the long range). There is a small upward trend in the errors from 2001-2006, most notably in the 36, 72, and 120-h forecasts. This trend stops in 2007. However, the 2007 sample is small (30 cases total), so the significance of this is not clear. The data also shows a trend of decreasing biases from 2001-2007.

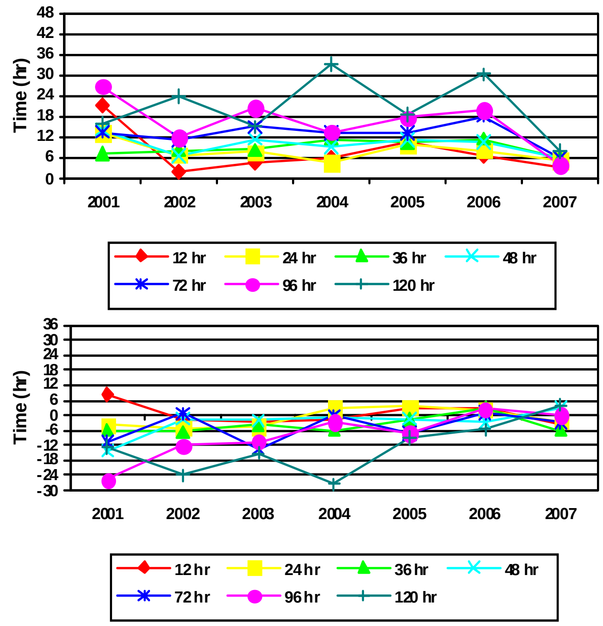

Figure 3. Five-day ET forecast timing errors (top) and biases (bottom) broken down by individual forecast times from 12-120 h. Negative bias values indicate that ET was forecast to occur after the time of transition in the best track.

It should be noted that ET forecast performance have significant intra-seasonal variation. Figure 4 shows the 5-day entire forecast errors and biases broken down by storm for the 2005 season. As can be seen, there is a large spread in both errors and biases from storm to storm. Some possible reasons for this will be discussed in the final section.

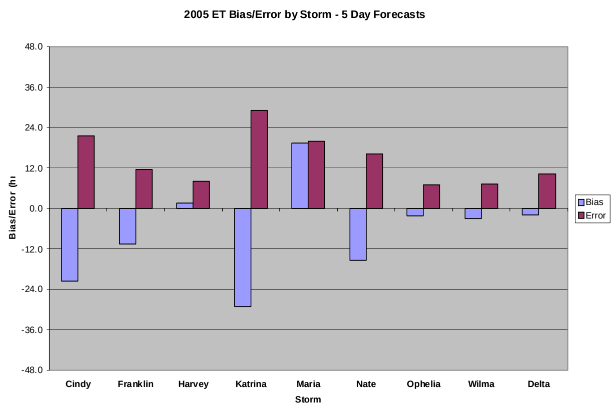

Figure 4. Five-day ET forecast timing errors and biases for the 2005 Atlantic hurricane season broken down by storm. Negative bias values indicate that ET was forecast to occur after the time of transition in the best track.

Figure 5 shows the mean absolute timing errors and biases for 3-day ET forecasts averaged over the entire forecast period. As might be expected, the errors and biases for these forecast are somewhat smaller than those for their 5-day counterparts from 2001-2007, although small upward trend seen in the 5-day data from 2001-2006 is apparent here. Large errors and biases are present prior to 2001, with notably smaller errors and biases after that time. Similar features are seen for the individual forecast times (Figure 6). Examination of individual forecasts reveals that some storms in this period had extremely large ET forecast timing errors. The most extreme example is for Hurricane Lili (1996), where the average 24-h ET forecast error was 99 h! This was due to the actual transition of Lili occurring far later than forecast.

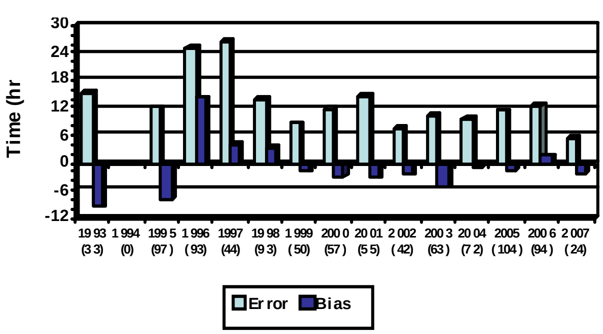

Figure 5. Three-day ET forecast timing errors and biases averaged over the entire forecast period. Otherwise same as Figure 2.

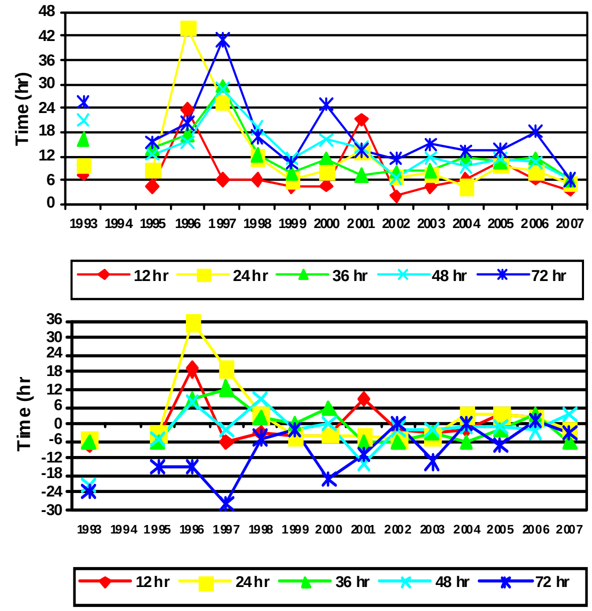

Figure 6. Three-day ET forecast timing errors (top) and biases (bottom) broken down by individual forecast times from 12-72 h. Otherwise same as Figure 3.

Note that there are no ET forecasts verifying in 1994. The best track data show while some cyclone that year did merge with frontal systems, there were no best track positions listed as extratropical. This lack of cases has a significant impact on the skill scores described next.

b. Skill scores

Figure 7 shows the 3-day ET forecast skill scores for the CSI, POD, FAR, and event bias. The CSI is generally in the range of 0.5-0.8, except for 1994 where it is zero due to no verifying ET cases. The POD is generally above 0.6, while the FAR is below 0.4 except for 1994 and 2002. These metrics indicate that the NHC ET forecasts have skill, as for the most part the scores are closer to skillful values (near 1.0) than to non-skillful values. The event bias, which for non-biased forecasts should be 1, generally fluctuates between 0.75 and 1.25, with the exception of the peak of greater than 2 in 2002. A closer examination shows that there were many forecasts of ET that did not verify that year, a result supported by the peak in the FAR score.

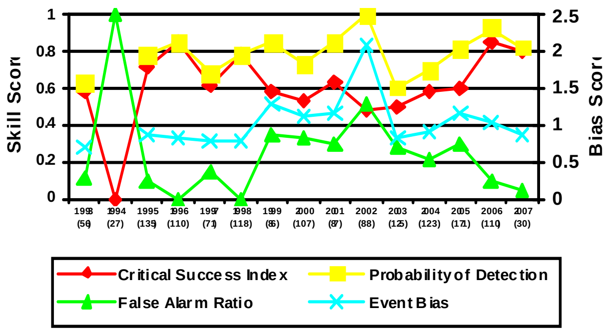

Figure 7. Three-day ET skill scores. The scale for the critical success index, probability of detection, and false alarm ratio is on the left. The scale for the event bias is on the right. The number of cases is shown in parentheses beneath the year labels. Event bias scores greater than one mean that ET is forecast more often than was observed.

Figure 8 shows the 3-day ET forecast skill scores including the percentage of correct forecasts, the Gilbert skill score, the Heidke skill score, and the true skill statistic. Again, the results show that the NHC forecasts have skill, with the score generally closer to skillful values than non-skillful values. The Gilbert skill score shows lower values than the Heidke or true skill scores, which appears to be due to the Gilbert score not accounting for correctly-forecast non-transitions. The 1994 case of no verifying ET cases causes problems with all the scores, resulting in the zero values for the Gilbert and Heidke skill scores, the undefined true skill score, and a decrease in the percentage correct.

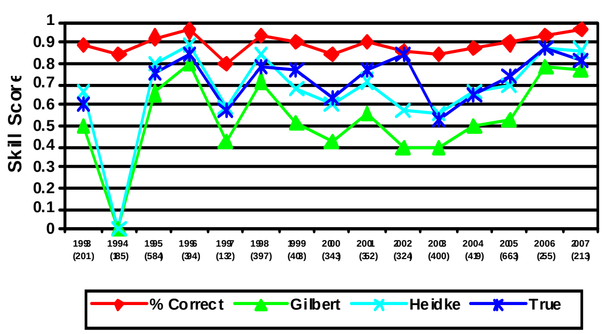

Figure 8. Three-day ET skill scores, including the percentage of correct forecasts, the Gilbert skill score, the Heidke skill score, and the true skill statistic. The number of cases is shown in parentheses beneath the year labels.

Figures 7 and 8 show a trend of decreasing CSI and Gilbert Skill Score for the period 1995-2002, accompanied by an increase in FAR. This is followed thereafter by a notable increase in the CSI, Gilbert, Heidke, and True skill scores with an accompanying decrease in FAR. Potential reasons for this are discussed in the next section.

The skill scores of the 5-day forecasts (not shown) are similar to, but a little less skillful, than the 3-day forecasts. This is not surprising, as it is expected that the 96 and 120-h forecasts would be less skillful than the shorter-ranged forecasts.

IV. Discussion

In terms of both timing and skill scores, the NHC forecasts of ET have generally improved from 1993 to the present, especially during the period from 2003 onward. The most likely causes of this are an improved basic understanding of the ET process (i. e. Klein et al. 2000, Evans and Hart 2003) and the development of the Cyclone Phase Space (CPS) tool (Hart 2003). The CPS became part of NHC operations in the early 2000's at about the time that the various skill metrics began the upward trends seen in Figures 7 and 8. This is likely not a coincidence – the CPS provided the methodology through which the NHC forecasters could interpret output from likely-improving numerical weather prediction models to make better ET forecasts.

Prior to the development of the CPS, the NHC had no formal methodology for either forecasting or analyzing ET. This suggests not only issues with ET forecasts, but also the possibility of inconsistencies in the NHC best tracks regarding when and if ET occurred. The verification statistics from 1994 and 2002 in particular may have resulted from a combination of forecast and best track uncertainties. Re-examination of ET events in the pre-CPS era will be part of the ongoing Atlantic Hurricane Re-Analysis Project (Landsea et al. 2003).

Meteorologically, there are many sources of error in the NHC ET forecasts. Perhaps the largest factor affecting the skill scores are TCs that are correctly forecast to merge with fronts, yet dissipate instead of continuing as extratropical lows. This happened several times during the 1993-2007 period. Another error source is TCs where poor track and intensity forecasts lead to poor ET forecasts. Cases of aborted transitions such as Alberto (Beven 2002) have reduced skill scores, while cases of slow or interrupted transitions have caused large timing errors. Cases of unclimatological transitions (low latitude) and unclimatological non-transitions (high latitude) also impact the skill scores. One final issue involves cases of ET forecasts over the United States, which sometimes have large timing errors. While the overall level of skill of the ET forecasts is good, the large number of error mechanisms suggests that the skill can be improved as the meteorological issues are addressed.

Future work includes extending the verification farther into the past, and automating the currently cumbersome manual verification process for future seasons. Other potential areas of research in ET forecasting include verifying the skill of numerical weather prediction models and developing a climatology-based skill metric.

References

Abraham, James , J. W. Strapp, C. Fogarty and M. Wolde. 2004: Extratropical transition of Hurricane Michael: An aircraft investigation. *Bulletin of the American Meteorological Society*, **85**, 1323-1339.

Beven, J. L., 1997: A study of three “hybrid” storms. Preprints: 22nd Conference on Hurricanes and Tropical Meteorology. American Meteorological Society, Ft. Collins, CO, 1997, 645-646.

Beven, J. L., 2002: Tropical cyclone extratropical transition: A reversible process? Preprints: 25th Conference on Hurricanes and Tropical Meteorology. American Meteorological Society, San Diego, CA, 2002, 495-496.

Brown, D. P., 2008: Tropical Cyclone Report, Hurricane Noel. Available from the web site of the National Hurricane Center at http://www.nhc.noaa.gov/pdf/TCR-AL162007_Noel.pdf.

Davis, Christopher A. and L. F. Bosart. 2004: The TT Problem: Forecasting the Tropical Transition of Cyclones. *Bulletin of the American Meteorological Society*, **85**, 1657-1662.

Doswell III, C.A., R.P. Davies-Jones, and D.L. Keller, 1990: On summary measures of skill in rare event forecasting based on contingency tables. *Wea. Forecasting*, **5**, 576-585.

Evans, J.L. and R. Hart, 2003: Objective indicators of the extratropical transition lifecycle of Atlantic tropical cyclones. *Mon. Wea. Rev.*, **131**, 909-925.

Franklin, J. L, 2008: 2007 National Hurricane Center forecast verification report. Available from the web site of the National Hurricane Center at http://www.nhc.noaa.gov/verification/pdfs/Verification_2007.pdf

Hart, R.,2003: A cyclone phase space derived from thermal wind and thermal asymmetry. *Mon. Wea. Rev.*, **131**, 585-616.

Klein, P. M., P. A. Harr and R. L. Elsberry, 2000: Extratropical transition of Western North Pacific tropical cyclones: An overview and conceptual model of the transformation stage. *Wea. Forecasting*, 15, 373-396.

Landsea,C.W. C.Anderson, N.Charles, G.Clark, J.Dunion, J.Partagas, P.Hungerford, C. Neumann, and M.Zimmer (2003) "The Atlantic hurricane database re-analysis project: Documentation for the 1851-1910 alterations and additions to the HURDAT database" Hurricanes and Typhoons: Past, Present, and Future R.J.Murnane and K.-B. Liu Editors, Columbia University Press pp.177-221

Thorncroft, C. and S. C. Jones, 2000: The extratropical transitions of Hurricanes Felix and Iris in 1995. *Mon. Wea. Rev.*, **128**, 947-972.