Predictability of the East Atlantic Tropical Cyclone Genesis in NCEP Global Ensemble: Case Studies During NASA African Monsoon Multidisciplinary Analyses

Andrew D. Snyder and Zhaoxia Pu
Department of Atmospheric Sciences, University of Utah
Salt Lake City, Utah

Yuejian Zhu
NCEP Environmental Modeling Center
Camp Springs, MD

Submitted to Weather and Forecasting
July 6, 2009

Corresponding Author: Dr. Zhaoxia Pu, Department of Atmospheric Sciences, University of Utah, 135 S 1460 E, Rm. 819, Salt Lake City, UT 84112. Email: Zhaoxia.Pu@utah.edu
Abstract

This study evaluates the NCEP global ensemble forecast system in its performance of predicting the genesis and evolution of tropical cyclones with five named tropical cyclones and two unnamed non-developing tropical systems during the NASA African Monsoon Multidisciplinary Analyses (NAMMA) between August and September 2006.

The effectiveness of tropical cyclone tracking methods has been firstly evaluated. Various tracking methods with different variables and vertical levels are compared. A manual tracking method with three variables (vorticity maxima, geopotential height and circulation) at the 850 hPa level is then adopted to track tropical cyclones in ensemble forecasts for both their pre-genesis and post-genesis phases. The overall skill of ensemble forecasts, ensemble means and spreads, and probabilities are verified relative to a genesis time defined to be the first designation of the tropical depression from the National Hurricane Center (NHC). Additional comparisons have also been made with high-resolution forecasts from the NCEP global forecast system (GFS).

It is found that the overall rate of accuracy for the genesis forecasts is slightly above 70% in the NCEP ensemble forecasts. The forecast uncertainty generally decreases with the reduction of the forecast leading time, while the short-range ensemble forecasts tend to be more accurate. In contrast, the skill of ensemble track forecasts is significantly improved for matured storms, mainly because of the implementation of the NCEP storm relocation scheme, which makes an accurate initial storm location for all ensemble members. Although the overall performance of the high resolution GFS forecasts is equal or better than the ensemble forecasting in almost half of the cases of tropical cyclone genuses, the ensemble forecasting added value in predicting the probability of the tropical cyclone genesis. In addition, the ensemble spread could be a good indication of the forecast track errors.
1. Introduction

Numerical weather prediction (NWP) has continuously improved since widespread use began half a century ago. However, forecast errors exist due to uncertainties in model initial conditions and imperfect physical parameterization schemes. Specifically, forecasts of genesis and the evolution of tropical cyclones remain a great challenge in NWP due to lack of observations over vast ocean areas. Advances in computer science and computer power made it possible to use ensemble forecasts to take into account the errors in model initial conditions. Recently, ensemble forecasting has become operational in many major NWP centers around the world (e.g., Toth and Kalnay 1997; Buizza et al. 2005; Wei et al. 2008; Reynolds et al. 2008).

During hurricane season, tropical cyclone forecasting is a high priority in many of these operational centers. Owing to the serious economic and social impact a tropical cyclone could potentially cause, it is very important to predict the genesis and evolution of tropical cyclones with enough lead time. In last two decades, along with advancements in numerical modeling and data assimilation, track forecasts for mature tropical storms have been improved significantly. However, forecasts of the tropical cyclone genesis and intensity changes remain a challenging problem (Aberson 2001; Rogers et al. 2006). Specifically, operational models have historically struggled with the prediction of genesis and often produce spurious vortices (Beven 1999). Several studies have demonstrated that as operational models become more complex, they do have skill at prediction in the tropics (Rennick 1999; Pasch et al. 2002). However, genesis prediction and differentiation between developing and non-developing systems continues to be a challenging problem, and thus remains a large area of research (e.g., McBride and Zehr 1981; Perrone and Lowe 1986; Hennon and Hobgood 2003; Kerns et al. 2008).

The advent of ensemble forecasting has added a new dimension to tropical cyclone
prediction. Instead of a single deterministic forecast, a suite of forecasts adds a probabilistic dimension to the forecast, thus helping to estimate forecast uncertainty. Model spread for storm tracks as well as atmospheric conditions has led to increased forecast confidence and skill for genesis, storm tracks and landfall effects. So far, tropical forecasters have largely taken a consensus (multi-model) approach when utilizing ensemble forecasting. One common approach is to utilize all (or some) models (e.g., these from different operational centers) available. Another approach is to statistically combine different models that were either designed for or have a history of performing well in the tropics, such as those developed by Goerss (2000), Krishnamurti et al. (2000), and Weber (2003).

In many operational centers, such as U. S. National Centers for Environmental Prediction (NCEP) and the European Centre for Medium-range Weather Forecasting (ECMWF), ensemble forecasts are produced from the single operational model, usually at a coarser resolution in consideration of computational expense. With a single model, the ensemble is achieved by using perturbed initial conditions. However, almost all perturbation methods used for current operational models (e.g., breeding vectors, singular vectors) are based on mid-latitude variability and may not be very suitable for tropical ensemble forecasting (Zhang and Krishnamurti 1997; Mackey and Krishnamurti 2001; Cheung and Chan 2001). So far, there has been little work done to evaluate the skill of operational ensembles in the tropics. Recent studies from Aberson et al. (1998), Aberson (1999), and Marchok (2002) showed that the NCEP’s global forecasting model and the ensemble forecasting system were skillful in the tropical cyclone track prediction, but little attention was focused on the skill of ensemble forecasting on tropic cyclone genesis forecasts.

Theoretically, since ensemble forecasts provide probabilistic information on both vortex
spin-up and ambient environmental conditions, they could prove to be useful in predicting tropical cyclone genesis. Therefore, the purpose of this study is to evaluate the NCEP ensemble forecast system for its ability to predict the genesis of tropical cyclones. The primary focus will be tracking the development of tropical storms and then evaluating how well the ensemble system predicts these systems from their early genesis phase to a tropical depression then as it strengthens to a tropical storm. Predictability will be assessed in both a pre-genesis and post-genesis environment. Track errors, mean, and spread will also be evaluated. Detailed case studies will be performed for five tropical systems and two non-developed tropical waves as predicted by the NCEP Global Ensemble Forecast System during the NASA African Monsoon Multidisciplinary Analyses (NAMMA) field experiment between August and September 2006 (Zipser et al. 2009).

The paper is organized as follows. Section 2 describes briefly the NCEP ensemble system and the data used in this study. Tracking methods and criteria are evaluated and defined in Section 3. In Section 4, the results are summarized from each case study, and Section 5 discusses the overall evaluation on a few important issues. Section 6 makes the concluding remarks.

2. Brief description of NCEP ensemble system and the data used in this study

The ensemble forecasts used in this study were obtained from the NCEP archive and the data were available at one-degree latitude by one-degree longitude horizontal resolution. There were 14 ensemble members at T126L28 resolution available for each set of ensemble forecasts. Initial perturbations were generated using the ensemble transform breeding method as described by Wei et al. (2008). A hurricane vortex relocation method was also implemented in the model (Liu et al. 2006) to locate the storm in the right location. In addition, the storm intensity is also perturbed
with 5% of the magnitude of the wind speed. Tropical systems that fully occurred during the NAMMA field experiment between August and September 2006 were considered for this study. In order to reduce the data flow, we concentrated the evaluation on the forecasts starting at 0000 UTC each day. The operational Global Forecast System (GFS) forecasts, at T382L64 resolution, were also utilized for comparison. Five tropical systems are analyzed in this study. They include Tropical Storm Debby, and Hurricanes Ernesto, Florence, Gordon, and Helene. All five storms fully occurred during the months of August and September 2006 in the Atlantic Basin. Debby, Florence, and Helene were studied and sampled during NAMMA field program. In addition, two non-developing waves (unnamed and numbered as Waves 3 and 6 in sequential order, following Zawislak and Zipser 2009) also traversed the eastern Atlantic during this time period, and these were briefly investigated for comparative purposes. Figure 1 shows the tracks of five named tropical systems studied in this paper.

3. Evaluation of the tracking methods

The primary focus of this study is to track the tropical systems in both the pre-genesis and post-genesis phases. The **genesis time** in this study is defined as the time when the system was designated by the National Hurricane Center (NHC) as a tropical depression in the best track data. We choose this time as a reference point because it serves as a clear dividing line between an open wave and closed circulation. In addition, each storm took a different amount of time to reach its tropical storm status, thus making direct comparison difficult if we were to define the time when a system reaches the tropical storm intensity to be a genesis time.

Storms are tracked from approximately three days before they were named to two days after they have been named. All track comparisons are made to the NHC’s best track data.

Tracking the genesis of tropic cyclones often proved to be challenging since the center of the
system is not always obvious, especially in the pre-genesis phase. Therefore, an efficient tracking method has to be developed in order to accurately represent the track of the storm.

**a. Review of previous tracking methods**

Several tracking methods for tropical disturbances can be found in the literature: manual, automated, and statistical. Manual tracking involves human analysis of data utilizing a number of different variables. Several studies have tracked cloud clusters using satellite data (e.g., McBride and Zehr 1981, Perrone and Lowe 1986, Hennon and Hobgood 2003). One advantage of this method is the frequency of the data; however, cloud clusters do not always correspond to the center of the tropical disturbance. Another popular manual tracking method was one that has been applied in largescale analysis or reanalysis. It uses the low level vorticity and the meridional wind component (to identify passage of the wave trough) for tracking purposes. Hovemoeller diagrams have also been used to aid in tracking (Fink et al. 2004, Chen 2006, Kerns et al. 2008). Reed et al. (1988) found that tracking waves consistently was difficult, so they preferred to use vorticity. However, they also noted that weak systems could have multiple vorticity centers which made the wave track ambiguous. Kerns et al. (2008) noted these difficulties along with spurious small-scale vorticity “bullseyes” and vorticity that formed and dissipated within the same wave. To counteract these problems, they instituted a Lanczos bandpass time filter and Gaussian spatial smoothing so that only consistently strong vorticity appeared in the analysis.

Thorncroft and Hodges (2001) developed an automated tracking system, which used a cost function to match features between time steps in the analysis. They noted that multiple vorticity centers and wind shifts from squall lines caused difficulties with their algorithm. Operational centers have also instituted automatic cyclone track algorithms for named storms, such as used
by NCEP (Marchok 2002). Several studies (e.g., Burpee 1972; Albignat and Reed 1980; Pytharoulis and Thorncroft 1999) have utilized the statistical approaches including power spectra, composite charts, and kinetic energy of horizontal winds. However, such analyses suffer from smoothing of important structures of the tropical disturbances. Both automated and statistical methods can be limited due to the criteria and thresholds that must be chosen.

As discussed above, most of the tracking methodology has been developed for tropical cyclone analysis, and not model forecasts. Considering the limitation from both automated and statistical methods, all storms for this study are manually tracked on a plan-view map of model output. By doing the manual tracking, attention could be focused on the details of each system, which can be important in the pre-genesis stage. Meanwhile, since there is not a standard variable (or parameter) and height level that has been identified to be most effective tracking the tropical cyclone genesis, an evaluation has been done first in order to determine a reasonable way to track the storms.

b. Tracking criteria

Tropical waves, depressions and cyclone systems are often characterized by the patterns of geopotential height, vorticity, and wind vectors. The representation of the intensity and structures of these variables strongly depends on the model resolution. In order to make a consistent definition of the tropical cyclone genesis in the one-degree latitude by one-degree longitude resolution, a statistical analysis has been conducted based on the NCEP global forecast system (GFS) final analysis (FNL) at the same resolution. By analyzing the magnitude and features of the height, vorticity and circulation fields in the GFS analysis at the time closest to the actual genesis of all the five cases being studied, it is apparent that most storms were represented by multiple closed height lines (at a 10 m interval) within 5 degrees, circulation in
the wind field, and relative vorticity greater than $7.5 \times 10^{-5} \text{ s}^{-1}$. Therefore, for the tracking process in the pre-genesis phase, when all three of these criteria were met in a forecast, we note that the forecast predicted the “genesis” of the storm. If one or two of those conditions were depicted in the ensemble forecasts, then we label the forecasted system as “vortex-like.” “Dissipation” will be defined to mean that the parameters for genesis exist but do not persist for longer than 48 hours. All forecasts for each member of the ensemble forecasts are tagged with one of these labels. These criteria are consistent with what have been used in Cheung and Elsberry (2002), where tropical cyclone formation over the Western North Pacific was tracked in the Navy Operational Global Atmospheric Prediction System (NOGAPS).

c. Tracking variables

In order to evaluate the different parameters for their effectiveness in tracking the storm genesis and development, three aforementioned variables were tested based on their importance in depicting storm structure: geopotential height, vorticity, and wind vectors. Thermal variables were not considered for tracking purposes because independent investigation revealed that the warm core structure of the storm was not always evident at this model resolution. Each storm was tracked independently with each variable, and the tracks were then compared with the available NHC best track data.

In a majority of the cases, there was very little difference in the tracks using different variables. Figure 2 shows overall averaged track errors with different variables over all cases for both pre-genesis and post-genesis phases for 1 to 5 day ensemble forecasts of all members. Overall, the averaged track errors for each case are very similar regardless which variable is used for the tracking. However, the most drastic differences occur for very weak systems (e.g., in the wave phase) or for complex systems (e.g., the first stages of Florence, discussed in Section 4).
Figure 3 shows a typical example with the tracks of Florence using different variables. The storm tracking starts at 0000 UTC 01 September 2006, 66 hours before the storm becomes a named tropical depression. It is apparent that the discrepancies among the tracks using different variables are quite larger when storm is in its weak phase. However, after the storm is forecasted to organize, all variables tend to agree in terms of accuracy in storm tracking.

In addition, the relative importance of each variable is case dependent. Figure 4 shows plots of vorticity, wind, and geopotential height from actual tracking sequences. These selections are examples of cases which demonstrate the need for multiple tracking variables. For instance, in the pre-genesis phase (Fig. 4a), the center of the vorticity maximum was generally most useful since the system is often broad and erratic. After genesis (Fig. 4b), the center of the closed height lines best represented the storm. In some cases, the storm had both a broad area of vorticity and height contours (Fig. 4c), in which case wind vectors helped to identify the center of the storm, although the wind field was generally the least useful due to the relatively coarse resolution of the model. Additionally, weak and deteriorating waves often had ambiguous centers (Fig. 4d). In these cases, the manual tracking method based on the judgment of all three variables was advantageous in picking the center of the feature consistently.

From the number of cases with various forecasting leading times, it was concluded that all three variables were important to best find the center of the system. Thus, all three variables are necessary for accurate storm tracking.

d. Tracking levels

Similarly, there is not a standard height level with which to track storms that has been suggested by previous studies. In order to examine which level would be most efficient to be used for tracking the storms, the storm tracking experiments also have been conducted for
different pressure levels. Although it is believed that the upper level atmospheric conditions are important for tropical cyclone genesis, since the primary focus of this study is to determine if the model predicts genesis, we mainly consider lower pressure levels for this evaluation (e.g., 925, 850, and 700 hPa pressure levels).

Storm tracking was performed at each of these pressure levels (at 925, 850 and 700 hPa), and subsequently compared with each other as well as to the best track data. Results showed that the tracks were virtually identical at different pressure levels for stronger storms. For weaker storms and especially those well before their genesis, there were variations of the track amongst the different levels, but no discernable pattern was noticed and was best analyzed on a case by case basis. Figure 5 shows two opposing examples of systems approximately three days before they were designated tropical depressions. Fig. 5a shows a very weak wave (pre-Gordon on 08 Sep 2006) that had widely varying signatures at different vertical levels. Meanwhile, the pre-genesis stage of Helene (on 10 Sep 2006) was well organized, and the tracks are virtually the same at each pressure level (Fig. 5b).

In general, the storm signal was stronger at lower levels, but in some cases weaker waves were better depicted at 700 hPa. Storms also tended to deteriorate from the top down, which confirms that stronger storms tend to be deeper. These observations were also seen when comparing the manual tracks to the best track data.

Table 1 illustrates the conceptual magnitude of the track errors for all cases, with different forecast lead times at different pressure levels. After assessing and comparing the benefits of all levels, the tracks at 850 hPa are deemed to minimize the track errors and are sufficiently accurate to be used for most of the cases. We thus conduct and compare all tracks at 850 hPa for all the cases in this paper. In fact, the use of 850 hPa as the tracking level also agrees with the common
justification: due to effect from the ocean or land surface, diurnal vorticity maxima may be present at 925 hPa. In addition, weaker systems may not be depicted at 700 hPa since a weaker system is usually shallower in terms of its entire depth.

e. Daily genesis potential

Originally developed by McBride and Zehr (1981), and recently used by Kerns et al. (2008), the daily genesis potential (DGP) can be used as a tool to assess and track developing tropical disturbances. By thermal wind balance, the difference in relative vorticity within a layer can be used to approximate the strength of the warm core (if one exists). Therefore, it can indicate the likelihood of tropical storm formation. Commonly, the DGP can be defined by the raw difference of relative vorticity at 925 and 200 hPa: namely, \( \text{DGP} = \zeta_{925} - \zeta_{200} \). By this definition, a high DGP is consistent with a deep warm core, and indicates that genesis is proceeding.

The DGP was plotted for tracking in the same manner as vorticity to determine its usefulness as a tracking variable. Compared with low level vorticity, the DGP produced comparable tracks that are similar to all tracking variables. However, it also added to the diversity of the track when tracking the weak systems. Figure 6 shows a typical track that followed this pattern. As a proxy for strength, the DGP provided inconsistent results. Based on these case studies, a threshold value could not be established to justify the genesis. Therefore, although the DGP is a useful parameter, it does not add significant value for our specific purpose in tracking storm genesis. Thus, we decide not to use this DGP for the storm tracking in this study.

4. Storm Case Studies

Based on the methods and criteria set up in Section 3, we performed the storm tracking for five named storms and the two non-developing waves. The storms were tracked in ensemble forecasts from three days in advance until two days after being designated a tropical depression
The tracking results for all five named storms are summarized in Table 2. A detailed description for each case is given as follows.

a. Debby

Debby was classified as a tropical depression at 1800 UTC 21 August 2006. It evolved into a tropical storm at 0000 UTC 23 August 2006. The African easterly wave (AEW) that developed was disorganized and somewhat ambiguous over land, and was thus not well resolved in the model. In the 5-day model forecast started from 0000 UTC 19 August (Fig. 7a), very few ensemble members accurately predicted the evolution of Debby. A majority of the members predicted a track that was well south of the best track. In addition, most of the members that predicted genesis initially, later failed to develop a deep storm. In the forecast from 20 August 2006, all ensemble members were in better agreement about developing a storm (Fig. 7b). As the forecast time approached genesis, most members followed a similar track. Consistency between different ensemble members improved as the initialized disturbance became better organized closer to genesis. Tracks were aligned and spread was minimized for the forecasts starting from 21 and 22 August (Fig. 7c and d). In particular, the forecast tracks from 22 August start from the same location due to the NCEP implementation of the vortex relocation after the storm became a named tropical depression (Fig. 7d). The ensemble mean is also very close to the best track on this day. For the forecasts started at 0000 UTC 23 August 2006, most of ensemble members did not predict the northward turn of the storm at later forecast hours, but did demonstrate an evenly spread ensemble field that generally encompassed the best track (Fig. 7e). Overall, the ensemble predictability of Debby showed tremendous improvement as the storm became more organized, especially after developing into a tropical depression (Fig. 7a-e).

Compared with the ensemble forecasts, the GFS forecast did not show any development from
the predominant vorticity maximum, but did show a brief rapid deepening of a vorticity maximum to the southwest of the actual storm’s location in the forecast from 0000 UTC 19 August (Fig. 7a). For the remainder of the forecasts, the GFS tended to depict a strong storm. The forecasts started from 20 and 21 August are slower and more northerly than the ensemble mean (Fig. 7b and c), while the forecasts from 22 and 23 August align more closely to the ensemble members (Fig. 7d and e).

b. Ernesto

Ernesto formed from a weak tropical wave that was traversing the south-central North Atlantic Ocean. Tropical depression status was reached at 1800 UTC 24 August 2006, followed by tropical storm status at 1200 UTC 25 August 2006. Surprisingly, all forecasts before the genesis time did not develop a storm at all, although the number of vortex-like systems increased with decreasing the forecast lead time. From the 5-day forecasts started at 22 August, the consensus was for the wave to remain weak. Forecasts also moved the wave too quickly to the west, with small ensemble spread until the later stages of the forecast (Fig. 8a). While more members produced weak vortex-like circulations in the forecasts that started on 23 and 24 August, predictions of a well-developed tropical storm remained nonexistent (Fig. 8b and c). The track of the wave/vorticity maximum remained much farther south than where the actual storm tracked, and was much faster than the actual storm. Many ensemble members continued to predict a westward track, even after the storm developed into a tropical depression on 25 August (Fig. 8d). The ensemble mean showed a track into the Gulf of Mexico, while the actual storm made landfall in southern Florida. This pattern continued on 26 August and no members produced a well-developed storm, and many completely dissipated it. Those members that held the storm together showed deepening over the Gulf Stream offshore from the eastern United
States. None of the ensemble members from the forecasts on 27 August (the last day tracked, when the storm initialized near Hispaniola) predicted the eventual landfall in North Carolina.

Overall, the ensemble forecasts of Ernesto largely failed since in most of the pre-genesis cases, the forecasts only produced vortex-like tropical waves. While after its genesis the track forecasts have been improved significantly (Figs. 8 d, e, and f) due to the vortex relocation, the intensity of the storms never reached the tropical genesis stage in terms of the intensity of the systems.

Meanwhile, the GFS forecast starting from 0000 UTC 22 August was on the north side of the ensemble envelope. The GFS forecast from the next day did produce a storm, but it was from a piece of vorticity that split off from the wave and developed northeast of the ensemble mean and the actual track. It did fairly well maintaining the circulation, but the track was closer to the ensemble mean than the actual track. The GFS forecasts from 24-26 August were generally south of the ensemble mean. However, the GFS forecast starting on 0000 UTC 27 August (Fig. 8f) produced a storm with tropical cyclone intensity and had a perfect track (as opposed to the ensemble members), although it moved slightly slower than the actual storm.

c. Florence

Florence was first classified as a tropical depression at 1800 UTC 03 September 2006 and gradually strengthened to a tropical storm by 0600 UTC 05 September 2006. The storm presented a complex situation in the pre-genesis environment. Specifically, there were three vorticity maxima across the central Atlantic, and the forecasts from different ensemble members handled them differently. In the GFS analysis, the second wave was faster than the first, and they combined to form a large circulation with multiple vorticity maxima that eventually became Florence. However, in the ensemble forecasts, some members would develop the storm from the
first wave, while other members would either develop it from the second wave or from the merger between first and second waves. On a few occasions, there would be a second merger with the third wave, or the storm would form from the third wave itself. Another complexity was that ensemble members accurately predicted the genesis, but at the same time the third wave would generate an equal or even stronger storm to the northeast of Florence in the forecasts. In general, the continuity was inconsistent and the tracking was difficult in the pre-genesis environment of these ensemble forecasts. Often, very broad cyclonic circulations would obviously represent a low pressure, but were hard to classify as a tropical storms. This complexity was present in the forecasts from 01, 02, and 03 September, although the forecasts became more consistent as time progressed. For example, Figs. 9a-c shows a large ensemble spread present in the forecast from 01 to 03 September, with few members representing the best track. After the genesis stage (forecasts from 04-06 September, Figs. 9d-f), the storm tracks were well predicted by the ensemble forecasts, with small spread and a mean track close to the best track. Compared with Fig. 6a-c, there is significant improvement in the ensemble forecast after the system strengthened to a tropical depression. The improved post-genesis track forecasts can be again attributed to the NCEP storm relocation scheme.

In the GFS control forecast, the third wave is preferred to form the largest storm from both forecasts starting from 1 and 2 September. For the forecast starting on 3 September, the GFS presents a complex and discontinuous solution that is similar to a subset of the ensemble members (Figs.9a-c). In the post-genesis phase, the GFS is more similar to the ensemble.

*d. Gordon*

Gordon formed from a weak tropical wave in the central Atlantic that reached tropical depression status at 1800 UTC 10 September and tropical storm status at 1200 UTC 11
September. It recurved and progressed slowly northward. For the 5-day forecast started from 8 and 9 September (Fig. 10 a and b), none of the ensemble members predicted the actual position of the storm. Even in the case when the ensemble showed a strengthening of the disturbance, it was too far south or west of the actual track. Fig. 10a (forecasts from 08 September) shows how the majority of the tracks are well south of the actual genesis location. In most of these cases, the initial vorticity maximum drifted southward and merged into a strip of vorticity. The forecast from 10 September (Fig. 10c) produced a wide array of tracks, with only one member that correctly showed the system curving northward. By 11 September (after the tropical depression had developed), the model forecast was in line with the correct movement (Fig. 10d). However, none of the ensemble members suggested that a strong hurricane would eventually develop. Also, ensemble forecasts were consistently too fast with movement, leading to erroneous forecasts of dissipation or absorption by mid-latitude troughs. Once again, the NCEP ensemble forecast showed significant improvement in track forecasts after the system became a tropical depression (Fig. 10e).

The early GFS forecasts were also weak and southerly from the forecasts starting from 9 and 10 September. In the forecasts starting from 10 September, the original vorticity maximum dissipated, while another wave or circulation formed from the vorticity strip. The GFS forecasts from 12 September proved to be very close to the actual track, while maintaining strength and circulation (Figure 10e).

e. Helene

This case was the easiest to track, as a strong AEW developed quickly as it moved westward off the coast. It was classified as a tropical depression at 1200 UTC 12 September 2006, and upgraded to a tropical storm at 0000 UTC 14 September. Even in the pre-genesis phase over
Africa, the vorticity maxima and wind circulation were very well defined in the model initialization. The forecast from 0000 UTC 10 September showed a strong developing storm as soon as the disturbance exited the coast with all 14 members predicting genesis. In Fig 11a, the track forecasts are shown to be very close to the actual track with little spread, although most members predicted slower movement. The forecasts from 11 September (Fig. 11b) suggested a weaker storm, and some members dissipated the storm. The storms in these forecasts generally moved slower and more northerly than the actual track with small track spread. From 12 September onward, the ensemble forecasts were fairly accurate. The first forecast after tropical depression status was reached (0000 UTC 13 September) is shown in Fig. 11d. The ensemble mean is close to the best track with a classical cone-like ensemble spread. Ensemble members handled the timing of the deepening with relative accuracy. The forecast from 15 September didn’t show the eventual western drift of the storm. Overall, all forecasts for Helene displayed a high degree of accuracy.

GFS tracks were similar to the ensemble mean for this particular case. One exception was the forecast started from 11 September (Fig. 11b), where the storm movement in GFS was much slower than the other members and showed a northerly turn.

f. Non-developing cases

Although aforementioned five tropical storms are all developed from easterly waves, in reality, not all of these easterly waves develop tropical storms. Some studies, such as Thorncroft and Hodges (2001), have sought to determine the relationships between African Easterly Waves and tropical cyclone genesis. While they found a correlation between the number of these mainly meridionally-oriented tropical waves and the subsequent number of tropical cyclones, not all the waves strengthen into storms. When an easterly wave (or commonly named “tropical
disturbance”) does not reach tropical depression status as determined by the NHC, it is considered a non-developing system. Predicting tropical cyclone genesis for these non-developing tropical waves is referred to as a “false alarm”. There were two non-developing cases during NAMMA. To evaluate whether the ensemble is likely to make such false alarms, two non-developing waves are tracked in this study.

“Wave 3” was an AEW that was well-defined as it moved westward off the African coast. Tracking was performed between 23 and 27 August during which the wave travels from east to west. Afterward, it slowly weakened with time and never developed into a tropical storm. The ensemble handled the tracks of the first non-developing wave relatively well (figures not shown). Tracks were tightly clustered, and closely followed the objective analysis of Zawislak and Zipser (2009) except for the forecasts from 25 August, which took a more southerly track. “Wave 3” had complex movement off the coast of Africa, with one wave moving from the east, and another traversing southward down the coast from the north (seen more so at 700 hPa). With the weak structure, the mean tracks at various levels are diffuse, especially 24 and 25 August (850 and 700 hPa diverge in opposite directions). Intensity was initially moderate. Broad, weak circulation was evident initially in most of the forecasts, which by our criteria, constituted vortex-like structure. However, there were no forecasts for genesis or maintained strengthening (Table 2), meaning model performance for intensity was reasonable in this case.

“Wave 6” was another AEW that tracked due west from Africa. Its strength was much weaker and diffuse than “Wave 3” and was better defined at 700 hPa. The five-day forecasts could only be consistently tracked from 7 through 10 September. Over time, this wave was generally forecasted to dissipate and merge with a southern strip of vorticity. The track forecasts had more spread than “Wave 3,” although movement was generally to the west with reasonable
speed. This wave was weak and tracked differently at different levels, especially from the forecast that started on 10 September. With weak initial intensity and a trend toward dissipation, none of the forecasts sampled predicted that the wave would strengthen significantly. In fact, even the few cases of vortex-like structure were very brief in nature. Table 3 shows the high rate of non-development forecasts, indicating the high accuracy of the ensemble in predicting this non-developed system.

In addition, GFS forecasts did not predict the development of these two non-developed systems.

5. Overall evaluation

a. Predictability: pre-genesis vs. post-genesis

Table 2 shows the genesis rates for each storm in terms of the probability of the genesis of tropical cyclones in ensemble forecasts started from both lead times for each case before and after the storm genesis. The predictability of the tropical cyclone genesis varies case by case as described in the Section 4. In order to make an overall evaluation, Table 5 shows the overall probability of genesis in ensemble forecasts over all 5 cases by summarizing the results from Table 2. It shows that the overall probability of the ensemble forecast of the genesis, including both genesis and vortex-like cases are over 50%. The overall probability of the non-development for these five developed cases is less than 45%. Combining all three lead times to produce an overall average among these five cases, the ensemble forecast had a genesis/vortex-like predictability of 73%, with the overall rate of 61 % for the forecast in pre-genesis phase (-3 to -1 day leading time) and 87% for the forecast in the post-genesis (0 to 2 day leading time). These results indicate that the ensemble forecast could offer a reasonable indication of the possibility of tropical cyclone genesis.
The track forecast was significantly improved in the post-genesis phase for all the cases (Figure 7-11), mainly because the implementation of the NCEP storm relocation scheme in the operational ensemble forecasting systems.

In addition, both GFS and ensemble forecasts handle the two non-developed tropical wave systems very well, none of the forecasts predicts a “false alarm” of the tropical cyclone genesis in all different forecast leading times (Table 3 and 4).

b. Accuracy of ensemble track forecast vs. ensemble spread

Goerss (2000) states that the ensemble spread could be a good indication of the accuracy of the ensemble track forecast. Specifically, the ensemble spread is an approximation of the upper bound of the forecast track error. In order to confirm this useful point with the case studies in this paper, Table 6 compares the ensemble mean track errors with the standard deviations of the ensemble spread as calculated by following:

\[
\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}} \tag{1}
\]

The values of track errors show a decreasing trend of the errors with the reduction of the leading time of the forecasts in majority of the cases, with the exception of Florence, in which case the track errors remained fairly constant. Track standard deviations of the spread are also varied with time, with Ernesto producing the highest values. For the forecasts that started at pre-genesis phase, Gordon had higher ensemble mean track error but lower standard deviation of the ensemble spread, proving that all ensemble members were consistently inaccurate.

To examine these relationships between the mean track error and ensemble spread, a linear fit has been applied to the two variables, and the coefficient of determination \(R^2\), or the square of the correlation has been calculated to show how much of the variation in mean error is
explained by the spread. Figure 12 shows the correlations between the mean track error and standard deviation of the spread among all storms. The forecasts were divided into two groups by the forecast lead time: both pre- and post-genesis sets. In the pre-genesis subset, there is only minimal correlation overall, with an $R^2$ value of 0.29. However, by removing the outlier (high mean error, low spread) values (the three values in the most right side of Fig. 12a) that resulted from the pre-genesis forecasts of Gordon, the $R^2$ increases to 0.69 and the correlation between the mean track error and standard deviation becomes more obvious. The regression for the post-genesis forecasts is much more impressive, with 83% of the variance explained. These results from Fig. 12 confirm Goerss’s (2000) conclusion that the ensemble spread could be an indication of the uncertainty in the mean track.

c. GFS vs. Ensemble

The performance of the higher resolution GFS control forecast also is varied in different cases. For all five cases, a total of 29 forecast periods have been evaluated. On 14 forecasts out of the total 29 cases, the track forecasts from GFS were very close to the track of the ensemble mean, which represents about half of the cases. In the forecasts in which the ensemble mean and GFS differed, the GFS resembled the best track data on two cases, while the ensemble mean was close to the best track six times. This difference may indicate additional value in the ensemble for tracking purposes. For another eight instances, GFS forecasts did not compare closely to the ensemble mean or actual track. In the remaining five cases, the GFS forecasted a track that significantly deviated from any ensemble member (the track was not within the ensemble “envelope”).

Further investigation found that including the GFS as an additional ensemble member usually does not change the overall ensemble mean significantly, even when the GFS had a large
deviation from the main ensemble members. This is mainly because the ensemble size is 14, which is much larger than the sample of a single control forecast. However, even with a single forecast, the overall performance of GFS in predicting tropical cyclone genesis is equal or better than the ensemble forecasting. Using the overall genesis plus vortex-like prediction rate as a metric for pre-genesis intensity forecasts, the GFS predicted tropical cyclone development 73% of the time, which is compatible to the overall ensemble accuracy of 73% for all cases and 61% for the ensemble forecasts in the pre-genesis phase. The advantage of the GFS model forecast could be attributed to the fact that the forecasts have been produced in a higher resolution. This advantage can also be seen in the forecasts from pre-genesis phase. There were only three cases of vortex-like structure found in GFS: Debby 30 h after genesis and Ernesto at 6 and 30 h after genesis. Additionally, the forecast from 6 h after the genesis of Gordon showed premature dissipation. In all other cases, the GFS correctly showed the storm structure according to genesis criteria. Except in two examples, the GFS fell into the same intensity forecast category as the majority of the ensemble members, which again proves the value of the ensemble.

d. Comments on storm strength

The intensity of the system, both pre- and post-genesis, was usually a good indicator of forecast accuracy. Helene was one of the strongest and most well-developed storms and was forecast the best in the ensemble forecasts. Florence was also well forecasted in the ensemble. Debby was forecast poorly when it was a weak wave, but as Debby developed into a tropical storm, the forecast accuracy improved in the ensemble forecasts. Ernesto and Gordon were weak storms, at least initially, and were forecast poorly. Another factor for these two storms may have been their size. Even after development, their sizes were both small compared to other storms, so ensemble forecasts may have had trouble resolving them. Gordon became a strong hurricane
later in its life cycle, but none of the ensemble members had predicted it. It should be noted that these two poorly forecasted storms (Ernesto and Gordon) formed in the western North Atlantic, while the other three storms formed directly from African easterly waves. The differences in the pre-genesis environment may also play an important role in forecast accuracy.

6. Concluding remarks

Five developed and two non-developed tropical systems from the 2006 North Atlantic hurricane season were evaluated in the NCEP global ensemble forecasts. The primary focus of the study was to determine how skillfully the ensemble performed in predicting the genesis and evolution of tropical storms. Various track methods from the previous studies have been compared and evaluated based on their effectiveness in tracking the tropical cyclone genesis. For this study, each storm was tracked using a manual method that utilized vorticity, wind, and geopotential height at 850 hPa.

Overall, the ensemble forecasts have high probabilities of genesis for the three strong storms (Debby’s mature phase, Florence and Helene), but failed to predict the two weaker storms that formed farther west in the Atlantic Ocean (Ernesto and Gordon) in the forecasts during the pre-genesis phase.

Statistically, it is found that the overall rate of accuracy for the genesis forecasts is slightly above 70% in the NCEP ensemble forecasts. The forecast uncertainty generally decreases with the reduction of the forecast leading time, since the short-range ensemble forecasts tend to be more accurate. In contrast, the skill of ensemble track forecasts is significantly improved for matured storms, mainly because of the implementation of the NCEP storm relocation scheme, which makes an accurate initial storm location for all ensemble members. Although the overall performance of the high resolution GFS forecasts is equal or better than the ensemble forecasting
in almost half of the cases of tropical cyclone geneses, the ensemble forecasting added value in predicting the probability of the tropical cyclone genesis. In addition, the ensemble spread could be a good indication of the forecasted track errors. Probability of detection for non-developing waves was high in both GFS and ensemble for two cases presented in this paper.

The cases studies in this paper demonstrate that an operational ensemble forecast system is valuable in tropical prediction. However, with only five developed storms and two non-developed systems, it must be noted that the sample size of the evaluation is small in this paper. More case studies need to be done in the future in order to fully evaluate the skill of NCEP ensemble in predicting tropical cyclone development. In addition, in order to enhance the skill of ensemble forecasting, it is necessary to evaluate ensemble forecasting itself, particularly the impact of the size and distribution of the initial perturbations on the ensemble skill of forecasting tropical cyclone genesis. Detailed studies are planned in the future.
Acknowledgements

This study is supported by ONR Grant # N000140810308 through financial support to first two authors. Authors would like to thank Dr. Richard Wobus for his help in transferring the NCEP ensemble data to the University of Utah for this study. Some useful discussions with Dr. Carolyn Reynolds at Naval Research Laboratory and Dr. W. James Steenburgh at University of Utah are also appreciated.
References


2617–2640.


McBride, J.L., and R. Zehr, 1981: Observational Analysis of Tropical Cyclone Formation. Part


Table Captions

Table 1. Comparison of track errors amongst the three pressure levels for each forecast leading time (forecast from number of days) relative to a genesis time defined to be the first designation of the tropical depression from National Hurricane Center (NHC). Negative numbers denote the forecasts start number of day (s) before the genesis time and positive numbers represent the forecasts start number of day (s) after the genesis time. “x” denotes the track error less than 150km. “v” represents the track error greater than 150km but the error is 50km [a half of model grid size] of the minimum track error among the three pressure levels.

Table 2. Predictability of each storm for different lead-time (forecast from number of days) relative to the system being designated a tropical depression by NHC. Values represent number of members (out of 14) predicting genesis (G), vortex-like development (V), premature dissipation (D), and non-development (N), respectively. Tracking results from operational GFS forecast is also shown (single member).

Table 3. Predictability of “Wave 3” in 5 day forecasts start from the different leading time, Values represent number of members (out of 14) predicting genesis, vortex-like development, premature dissipation, and non-development, respectively.

Table 4. Similar to Table 3, except for the genesis predictability for “Wave 6”

Table 5. Probability of genesis in ensemble for each lead time (forecast from number of days), combined for all five named storms.

Table 6. Ensemble mean error and spread (km) for all storms. Lead time denotes the forecast from the number of days before the system was designated a tropical depression.
**Figure Captions**

**Fig.1:** National Hurricane Center (NHC) best track for five name storms investigated. Dotted line represents a tropical depression, dashed line represents a tropical storm, solid line represents a hurricane, and cross represents a remnant low.

**Fig.2:** Distance error (in km) for each tracking variables, averaged over all five tropical storm cases for both pre- and post- genesis phases for 1-5 day ensemble forecasts over all members from all different leading time (1-3 days before genesis and one to two days after genesis).

**Fig.3:** An example of the track differences among three variables used. The ensemble mean 120-h forecast tracks are plotted from 66-h before Florence was designated a tropical depression (0000 UTC 01 September 2006). The solid line denotes the track following the center of the circulation. The black dash line presents the track following the vorticity maxima and the gray dot line links the centers of closed height lines. The figure shows that the tracks differ in the complex pre-genesis environment but are more similar after the storm organized.

**Fig. 4.** The usefulness of different variables in tracking storms in different situations. (a) Vorticity is useful in cases with broad waves. (b) For mature storms, the central closed height line provides a good estimate of where the center is located. (c) Systems with broad vorticity and height signatures can be tracked by the approximate center of circulation denoted by the wind vectors. (d) Sometimes the center of the selected system is ambiguous and no variable is most useful.

**Fig.5.** Example of variance in tracks at different atmospheric levels. (a) Ensemble mean 0-120 h forecast tracks from 66 hours before Gordon was designated a tropical depression (0000 UTC 08 September 2006). As a weak system, the tracks widely varied at different vertical levels. (b)
Ensemble mean 0-120 h forecast tracks from 60 hours before Helene was designated a tropical depression (0000 UTC 10 September 2006). Stronger systems usually had similar tracks at different vertical levels.

**Fig. 6.** Two examples of the DGP images. The figures show the ensemble mean tracks at different pressure levels (dash curve for the track at 700mb, thin solid curve for 850mb and dot for 925mb), compared with ensemble mean of DGP (thick gray solid curve). The 5-day forecast from 21 August for Debby shows the case of a stronger system, and the DGP track closely follows the low level track (a). The forecast from 23 August for Ernesto shows the case of a weaker system, and the DGP track deviates far northward of all of the individual levels (b).

**Fig. 7.** The tracks of 0-120 h ensemble forecast for Tropical Storm Debby (purple thin lines) from 0000 UTC 19-23 August 2006 (corresponding to Fig.7a-e, respectively), compared with the corresponding operational GFS forecast (blue line) and NHC best track (thick black line). Red dashed lines denote the ensemble mean. An “X” designates the forecasted genesis.

**Fig.8:** Same as Fig. 7 but for Hurricane Ernesto and forecast from 0000UTC 22-27 August (Fig.8a-f).

**Fig. 9.** Same as Fig. 7, but for Hurricane Florence and forecast from 0000UTC 01-06 Sep 2008 (Fig.9a-f).

**Fig. 10.** Same as Fig. 7, but for Hurricane Gordon and forecast from 0000UTC 08-13 Sep 2008 (Fig.10a-f).

**Fig. 11.** Same as Fig. 7, but for Hurricane Helene and forecast from 0000UTC 10-15 Sep 2008 (Fig.a-f).

**Fig.12.** Scatter plots of ensemble mean error (km) vs. ensemble spread (represented by standard deviations of the ensemble members relative to the ensemble mean in km) for each forecast.
period (1 to 5 day) for all storms. Forecasts are divided into two groups by the forecast leading
time in (a) pre-genesis, and (b) post-genesis.
**Table 1.** Comparison of track errors amongst the three pressure levels for each forecast leading time (forecast from number of days) relative to a genesis time defined to be the first designation of the tropical depression from National Hurricane Center (NHC). Negative numbers denote the forecasts start number of day(s) before the genesis time and positive numbers represent the forecasts start number of day(s) after the genesis time. “x” denotes the track error less than 150km. “v” represents the track error greater than 150km but the error is 50km [a half of model grid size] of the minimum track error among the three pressure levels.

<table>
<thead>
<tr>
<th>Case</th>
<th>Debby</th>
<th>Ernesto</th>
<th>Florence</th>
<th>Gordon</th>
<th>Helene</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-3</td>
<td>-2</td>
<td>-1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>925 mb</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>850 mb</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>v</td>
</tr>
<tr>
<td>700 mb</td>
<td>v</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>v</td>
</tr>
</tbody>
</table>
Table 2. Predictability of each storm for different lead-time (forecast from number of days) relative to the system being designated a tropical depression by NHC. Values represent number of members (out of 14) predicting genesis (G), vortex-like development (V), premature dissipation (D), and non-development (N), respectively. Tracking results from operational GFS forecast is also shown (single member).

<table>
<thead>
<tr>
<th>Case</th>
<th>Debby</th>
<th>Ernesto</th>
<th>Florence</th>
<th>Gordon</th>
<th>Helene</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of ensemble members in G/V/D/N</td>
<td>GFS</td>
<td>Number of ensemble members in G/V/D/N</td>
<td>GFS</td>
<td>Number of ensemble members in G/V/D/N</td>
</tr>
<tr>
<td>Forecast Lead-time</td>
<td>G/V/D/N</td>
<td>GFS</td>
<td>G/V/D/N</td>
<td>GFS</td>
<td>G/V/D/N</td>
</tr>
<tr>
<td>-3</td>
<td>6/4/2/2</td>
<td>V</td>
<td>0/1/0/13</td>
<td>N</td>
<td>10/3/1/0</td>
</tr>
<tr>
<td>-2</td>
<td>4/7/2/1</td>
<td>G</td>
<td>0/3/0/11</td>
<td>G</td>
<td>7/3/0/4</td>
</tr>
<tr>
<td>-1</td>
<td>6/3/4/1</td>
<td>G</td>
<td>0/9/0/5</td>
<td>V</td>
<td>14/0/0/0</td>
</tr>
<tr>
<td>0</td>
<td>10/0/4/0</td>
<td>G</td>
<td>1/13/0/0</td>
<td>V</td>
<td>12/2/0/0</td>
</tr>
<tr>
<td>1</td>
<td>1/6/0/7</td>
<td>V</td>
<td>1/13/0/0</td>
<td>V</td>
<td>13/0/1/0</td>
</tr>
<tr>
<td>2</td>
<td>n/a</td>
<td>n/a</td>
<td>5/6/3/0</td>
<td>G</td>
<td>13/1/0/0</td>
</tr>
</tbody>
</table>
Table 3. Predictability of “Wave 3” in 5 day forecasts start from the different leading time, Values represent number of members (out of 14) predicting genesis, vortex-like development, premature dissipation, and non-development, respectively.

<table>
<thead>
<tr>
<th>Forecast Date</th>
<th>23-Aug</th>
<th>24-Aug</th>
<th>25-Aug</th>
<th>26-Aug</th>
<th>27-Aug</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genesis</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Vortex-Like</td>
<td>10</td>
<td>8</td>
<td>14</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>Non-Development</td>
<td>3</td>
<td>6</td>
<td>0</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Dissipation</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 4. Similar to Table 3, except for the genesis predictability for “Wave 6”

<table>
<thead>
<tr>
<th>Forecast Date</th>
<th>7-Sep</th>
<th>8-Sep</th>
<th>9-Sep</th>
<th>10-Sep</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genesis</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Vortex-Like</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Non-Development</td>
<td>9</td>
<td>14</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>Dissipation</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 5. Probability of genesis in ensemble for each lead time (forecast from number of days), combined for all five named storms.

<table>
<thead>
<tr>
<th>Lead Time</th>
<th>Genesis (G)</th>
<th>Vortex Like (V)</th>
<th>G+V</th>
<th>Dissipation (D)</th>
<th>Non-development (N)</th>
<th>(N+D)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3</td>
<td>30/70</td>
<td>10/70</td>
<td>40/70</td>
<td>3/70</td>
<td>27/70</td>
<td>30/70</td>
</tr>
<tr>
<td>-2</td>
<td>22/70</td>
<td>15/70</td>
<td>37/70</td>
<td>2/70</td>
<td>31/70</td>
<td>33/70</td>
</tr>
<tr>
<td>-1</td>
<td>32/70</td>
<td>19/70</td>
<td>51/70</td>
<td>6/70</td>
<td>13/70</td>
<td>19/70</td>
</tr>
<tr>
<td>0</td>
<td>36/70</td>
<td>20/70</td>
<td>56/70</td>
<td>9/70</td>
<td>6/70</td>
<td>15/70</td>
</tr>
<tr>
<td>1</td>
<td>32/70</td>
<td>29/70</td>
<td>61/70</td>
<td>2/70</td>
<td>7/70</td>
<td>9/70</td>
</tr>
<tr>
<td>2</td>
<td>46/56</td>
<td>7/56</td>
<td>53/56</td>
<td>3/56</td>
<td>0/56</td>
<td>3/56</td>
</tr>
</tbody>
</table>
Table 6. Ensemble mean error and spread (km) for all storms. Lead time denotes the forecast from the number of days before the system was designated a tropical depression.

<table>
<thead>
<tr>
<th>Lead Time</th>
<th>Debby</th>
<th></th>
<th>Ernesto</th>
<th></th>
<th>Florence</th>
<th></th>
<th>Gordon</th>
<th></th>
<th>Helene</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>-3</td>
<td>89.37</td>
<td>371.13</td>
<td>226.29</td>
<td>431.57</td>
<td>134.97</td>
<td>213.25</td>
<td>67.61</td>
<td>943.94</td>
<td>177.61</td>
<td>334.59</td>
</tr>
<tr>
<td>-2</td>
<td>58.31</td>
<td>127.47</td>
<td>123.93</td>
<td>128.93</td>
<td>80.90</td>
<td>107.82</td>
<td>310.99</td>
<td>1025.05</td>
<td>139.46</td>
<td>339.24</td>
</tr>
<tr>
<td>-1</td>
<td>51.91</td>
<td>87.10</td>
<td>64.53</td>
<td>118.81</td>
<td>54.62</td>
<td>199.97</td>
<td>411.83</td>
<td>1005.82</td>
<td>107.49</td>
<td>267.69</td>
</tr>
<tr>
<td>0</td>
<td>60.78</td>
<td>109.39</td>
<td>149.28</td>
<td>163.70</td>
<td>58.62</td>
<td>93.39</td>
<td>82.88</td>
<td>127.81</td>
<td>37.58</td>
<td>84.15</td>
</tr>
<tr>
<td>1</td>
<td>104.49</td>
<td>139.09</td>
<td>162.56</td>
<td>188.54</td>
<td>60.16</td>
<td>108.68</td>
<td>158.29</td>
<td>207.51</td>
<td>150.47</td>
<td>161.78</td>
</tr>
<tr>
<td>2</td>
<td>n/a</td>
<td>n/a</td>
<td>145.06</td>
<td>198.52</td>
<td>96.67</td>
<td>109.17</td>
<td>141.07</td>
<td>150.03</td>
<td>56.55</td>
<td>66.78</td>
</tr>
</tbody>
</table>
**Fig.1**: National Hurricane Center (NHC) best track for five name storms investigated. Dotted line represents a tropical depression, dashed line represents a tropical storm, solid line represents a hurricane, and cross represents a remnant low.
Fig. 2: Distance error (in km) for each tracking variable, averaged over all five tropical storm cases for both pre- and post-genesis phases for 1-5 day ensemble forecasts over all members from all different leading time (1-3 days before genesis and one to two days after genesis).
Fig.3: An example of the track differences among three variables used. The ensemble mean 120-h forecast tracks are plotted from 66-h before Florence was designated a tropical depression (0000 UTC 01 September 2006). The solid line denotes the track following the center of the circulation. The black dash line presents the track following the vorticity maxima and the gray dot line links the centers of closed height lines. The figure shows that the tracks differ in the complex pre-genesis environment but are more similar after the storm organized.
Fig. 4. The usefulness of different variables in tracking storms in different situations. (a) Vorticity is useful in cases with broad waves. (b) For mature storms, the central closed height line provides a good estimate of where the center is located. (c) Systems with broad vorticity and height signatures can be tracked by the approximate center of circulation denoted by the wind vectors. (d) Sometimes the center of the selected system is ambiguous and no variable is most useful.
Fig.5. Example of variance in tracks at different atmospheric levels. (a) Ensemble mean 0-120 h forecast tracks from 66 hours before Gordon was designated a tropical depression (0000 UTC 08 September 2006). As a weak system, the tracks widely varied at different vertical levels. (b) Ensemble mean 0-120 h forecast tracks from 60 hours before Helene was designated a tropical depression (0000 UTC 10 September 2006). Stronger systems usually had similar tracks at different vertical levels.
Fig. 6. Two examples of the DGP images. The figures show the ensemble mean tracks at different pressure levels (dash curve for the track at 700mb, thin solid curve for 850mb and dot for 925mb), compared with ensemble mean of DGP (thick gray solid curve). The 5-day forecast from 21 August for Debby shows the case of a stronger system, and the DGP track closely follows the low level track (a). The forecast from 23 August for Ernesto shows the case of a weaker system, and the DGP track deviates far northward of all of the individual levels (b).
Fig. 7. The tracks of 0-120 h ensemble forecast for Tropical Storm Debby (purple thin lines) from 0000 UTC 19-23 August 2006 (corresponding to Fig.7a-e, respectively), compared with the corresponding operational GFS forecast (blue line) and NHC best track (thick black line). Red dashed lines denote the ensemble mean. An “X” designates the forecasted genesis.
Fig. 8: Same as Fig. 7 but for Hurricane Ernesto and forecast from 0000UTC 22-27 August (Fig. 8a-f).
**Fig. 9.** Same as Fig. 7, but for Hurricane Florence and forecast from 0000UTC 01-06 Sep 2008 (Fig. 9a-f).
Fig. 10. Same as Fig. 7, but for Hurricane Gordon and forecast from 0000UTC 08-13 Sep 2008 (Fig.10a-f).
Fig. 11, Same as Fig. 7, but for Hurricane Helene and forecast from 0000UTC 10-15 Sep 2008 (Fig.a-f).
Fig. 12. Scatter plots of ensemble mean error (km) vs. ensemble spread (represented by standard deviations of the ensemble members relative to the ensemble mean in km) for each forecast period (1 to 5 day) for all storms. Forecasts are divided into two groups by the forecast leading time in (a) pre-genesis, and (b) post-genesis.