Tracking and Verification of East Atlantic Tropical Cyclone Genesis in the NCEP Global Ensemble: Case Studies during the 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, Maryland

(Manuscript received 15 July 2009, in final form 2 May 2010)

ABSTRACT

This study evaluates the performance of the NCEP global ensemble forecast system in predicting the genesis and evolution of five named tropical cyclones and two unnamed nondeveloping tropical systems during the NASA African Monsoon Multidisciplinary Analyses (NAMMA) between August and September 2006. The overall probabilities of the ensemble forecasts of tropical cyclone genesis 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 are also made with high-resolution deterministic forecasts from the NCEP Global Forecast System (GFS). It is found that the ensemble forecasts have high probabilities of genesis for the three strong storms that formed from African easterly waves, but failed to accurately predict the pregenesis phase of two weaker storms that formed farther west in the Atlantic Ocean. The overall accuracy for the genesis forecasts is above 50% for the ensemble forecasts initialized in the pregenesis phase. The forecast uncertainty decreases with the reduction of the forecast lead time. The probability of tropical cyclone genesis reaches nearly 90% and 100% for the ensemble forecasts initialized near and in the postgenesis phase, respectively. Significant improvements in the track forecasts are found in the ensemble forecasts initialized in the postgenesis phase, possibly because of the implementation of the NCEP storm relocation scheme, which provides an accurate initial storm position for all ensemble members. Even with coarser resolution (T126L28 for the ensemble versus T384L64 for the GFS), the overall performance of the ensemble in predicting tropical cyclone genesis is compatible with the high-resolution deterministic GFS. In addition, false alarm rates for nondeveloping waves were low in both the GFS and ensemble forecasts.

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 the model initial conditions and imperfect physical parameterization schemes. Specifically, forecasts of the genesis and evolution of tropical cyclones remain a great challenge for NWP, partially due to a lack of in situ observations over vast ocean areas. Advances in computer science and computer power have, however, made it possible to use ensemble forecasts to account for uncertainties 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 the hurricane season, tropical cyclone forecasting is a high priority in many of these operational centers. Owing to the serious economic and social impacts tropical cyclones can cause, it is important to predict their genesis and evolution with enough lead time and accuracy. In the last two decades, advancements in numerical modeling and data assimilation have lead to significant improvement in track forecasts for mature tropical storms. However, forecasting tropical cyclone genesis and intensity changes remains a challenging problem (Aberson 2001; Rogers et al. 2006).

Corresponding author address: Dr. Zhaoxia Pu, Dept. of Atmospheric Sciences, University of Utah, Rm. 819, 135 S 1460 E, Salt Lake City, UT 84112.
E-mail: zhaoxia.pu@utah.edu

DOI: 10.1175/2010WAF2222332.1

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 gain predictive skill in the tropics (Rennick 1999; Pasch et al. 2002). However, genesis prediction and differentiation between developing and nondeveloping 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 potentially helping to estimate forecast uncertainty. So far, tropical forecasters have largely taken a consensus (multimodel) approach when utilizing ensemble forecasting. Commonly, all (or some) models available (e.g., those from different operational centers) are utilized. Another approach statistically combines different models that have a history of performing well in the tropics, such as those developed by Goerss (2000), Krishnamurti et al. (2000), and Weber (2003).

At many operational centers, such as the National Centers for Environmental Prediction (NCEP) and the European Centre for Medium-Range Weather Forecasts (ECMWF), ensemble forecasts are produced from a single model, which is usually the same as the deterministic prediction but run at a coarser resolution in consideration of computational expense. With a single model, the ensemble is produced by perturbing initial conditions. However, most perturbation methods used for current operational models (e.g., breeding vectors, singular vectors, etc.) are based on midlatitude variability and may not be suitable for tropical ensemble forecasting (Zhang and Krishnamurti 1997; Cheung and Chan 1999; Mackey and Krishnamurti 2001). Despite the importance of forecasting tropical systems, so far there has been little work done to evaluate the skill of operational ensembles in the tropics. Recent studies by Aberson et al. (1998), Aberson (1999), and Marchok (2002) showed that the NCEP's global forecasting model and ensemble forecasting system were skillful for tropical cyclone track prediction, but little attention was focused on the skill of the ensemble forecasts for tropical cyclone genesis.

Theoretically, since ensemble forecasts provide probabilistic information on both the vortex spinup 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 forecast predicts these systems. Forecast probabilities will be assessed in both a pregenesis and a postgenesis environment. Track errors will also be evaluated. Detailed case studies will be performed for five developed tropical systems and two nondeveloped tropical waves as predicted by the NCEP global ensemble forecast system during the National Aeronautics and Space Administration (NASA) African Monsoon Multidisciplinary Analyses (NAMMA) field experiment between August and September 2006 (Zipser et al. 2009).

This paper is organized as follows. Section 2 briefly describes the NCEP ensemble system and the data used in this study. Our 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 contains the overall evaluation with a discussion of a few important issues. Section 6 presents the concluding remarks.

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

The ensemble forecasts used in this study were obtained from the NCEP archive. There were 14 ensemble members produced from the Global Forecast System (GFS) forecasts at T126L28 resolution with perturbed initial conditions generated using the ensemble transform breeding method (Wei et al. 2008). A hurricane vortex relocation method was implemented in the model (Liu et al. 2006) to locate the storm in the observed position. In addition, the storm intensity was also perturbed within 5% of the magnitude of the wind speed. Tropical systems that occurred during the NAMMA field experiment between August and September 2006 were considered for this study. They include Tropical Storm Debby, and Hurricanes Ernesto, Florence, Gordon, and Helene in the Atlantic basin. In addition, two nondeveloping 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 were briefly investigated for comparative purposes. Figure 1 shows the tracks of the five named tropical systems studied in this paper.

For brevity, the evaluation concentrates on the forecasts starting at 0000 UTC each day and proceeds to 120 h. The deterministic GFS forecasts, produced at T384L64 resolution but available at $1° \times 1°$ latitude–longitude grid, were also utilized for comparison.

3. The tracking methods

The primary focus of this study is to track the tropical systems in both the pregenesis and postgenesis phases.

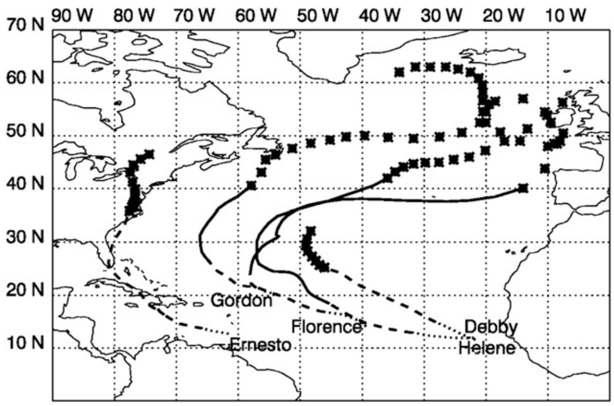

FIG. 1. NHC best tracks for the five name storms investigated. Dotted lines represent a tropical depression, dashed lines represent a tropical storm, solid lines represent a hurricane, and crosses represent a remnant low.

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 system 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 tropical storm intensity to be the genesis time.

Systems are tracked from approximately 3 days before they were designated a tropical depression by NHC to 2 days after they were designated. All track comparisons are made to the NHC's best-track data.

Identifying the genesis of tropical cyclones often proved to be challenging when tracking a system since the center of the system is not always obvious, especially in the pregenesis phase. Therefore, a tracking method has been developed to accurately represent the track of the cyclone.

a. Cyclone-tracking method

Three types of tracking methods for identifying 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 high frequency of the cloud data; however, cloud clusters do not always correspond to the center of the tropical disturbance. Another popular manual tracking method employs a large-scale analysis or reanalysis and uses the low-level vorticity and the meridional wind component (to identify passage of the wave trough). Hovmöller 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 that can make the wave track ambiguous. Kerns et al. (2008) reiterated these difficulties along with complications from spurious small-scale vorticity ''bull's-eyes'' and vorticity maxima 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 maxima appeared in their analysis.

Thorncroft and Hodges (2001) developed an automated tracking system that 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 those used by NCEP (Marchok 2002). Several studies (e.g., Burpee 1972; Albignat and Reed 1980; Pytharoulis and Thorncroft 1999) have utilized statistical approaches including power spectra, composite charts, and the kinetic energy of horizontal winds. However, such analyses suffer from excessive smoothing of the finescale structure of the tropical disturbances. Both automated and statistical methods can also be limited by the need to manually define criteria and thresholds.

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

b. Identifying the vortex center

Except for the mature hurricanes, defining the center of the weaker tropical waves in the pregenesis phase is often very challenging. To evaluate the different parameters for their effectiveness in tracking the center of the system, geopotential height minima, vorticity maxima, and a circulation appearing in the wind field were tested for their importance in depicting system structure. Each storm was tracked independently with each

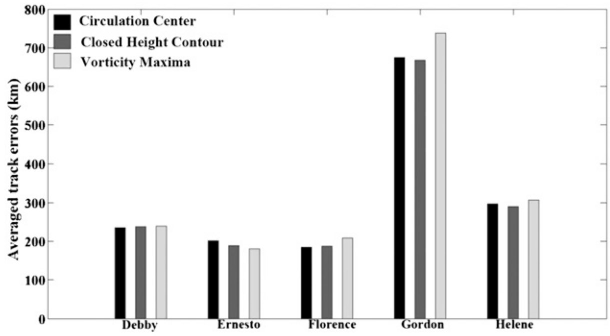

FIG. 2. Distance error (km) for each tracking variable, averaged over all five tropical storm cases for both pre- and postgenesis phases for 1–5-days ensemble forecasts over all members from all different lead times (1–3 days before genesis and 1–2 days after genesis).

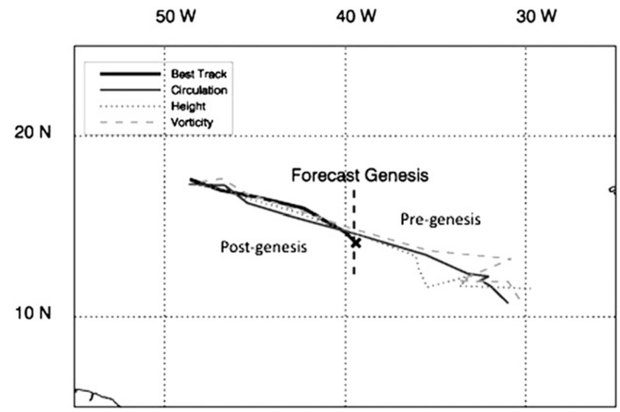

FIG. 3. An example of the track differences among the three variables used. The ensemble mean 120-h forecast tracks are plotted from 66 h before Florence was designated a tropical depression (0000 UTC 1 Sep 2006).

variable, and the tracks were then compared with the available NHC best-track data.

In a majority of the cases, there was little difference in the tracks using the three different variables. Figure 2 shows the overall averaged track errors, compared to the best track, with different variables, over all cases for both pregenesis and postgenesis phases for 1–5-day ensemble forecasts of all members. Overall, the averaged track errors for each case are similar regardless of which variable is used for the tracking. However, the greatest differences occur for very weak systems (e.g., in the wave phase) or for more complex systems with multiple competing waves (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 system tracking starts at 0000 UTC 1 September 2006, 66 h before the system is designated a tropical depression. It is apparent that the discrepancies among the tracks using different variables are larger when a storm is in its weak phase. However, after the system was forecast 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 that demonstrate the need for multiple tracking variables. For instance, in the pregenesis phase (Fig. 4a), the center of the vorticity maximum was generally most useful since the system was often broad and erratic. After genesis (Fig. 4b), the center of the closed height contours best represented the storm. In some cases, the system had both a broad area of vorticity and closed height contours (Fig. 4c), in which case wind vectors helped to identify the center of the system. However, 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 lead times, it was concluded that all three variables were necessary to best find and track the center of the system.

Similarly, no standard height level has been used to track systems in previous studies. To examine which level is most efficient for tracking the systems, the tracking experiments were conducted at different pressure levels. Because the primary focus of this study is to determine if the model predicts genesis, we mainly consider lower pressure levels (e.g., 925, 850, and 700 hPa).

System tracking was performed at each of these pressure levels (at 925, 850, and 700 hPa) and compared to the best-track data. For stronger storms, the tracks were virtually identical at the different pressure levels. In weaker systems, particularly before genesis, there were variations of the tracks among the different levels, but no systematic pattern was identified. Figure 5 shows two opposing examples of systems approximately 3 days before they were designated tropical depressions. Figure 5a shows a very weak wave (pre-Gordon on 8 September 2006) that had widely varying signatures at different vertical levels. Meanwhile, the pregenesis stage of Helene (on 10 September 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. Cyclones also tended to deteriorate from the top down, which confirms that stronger storms tend to be identifiable through a deeper layer in the

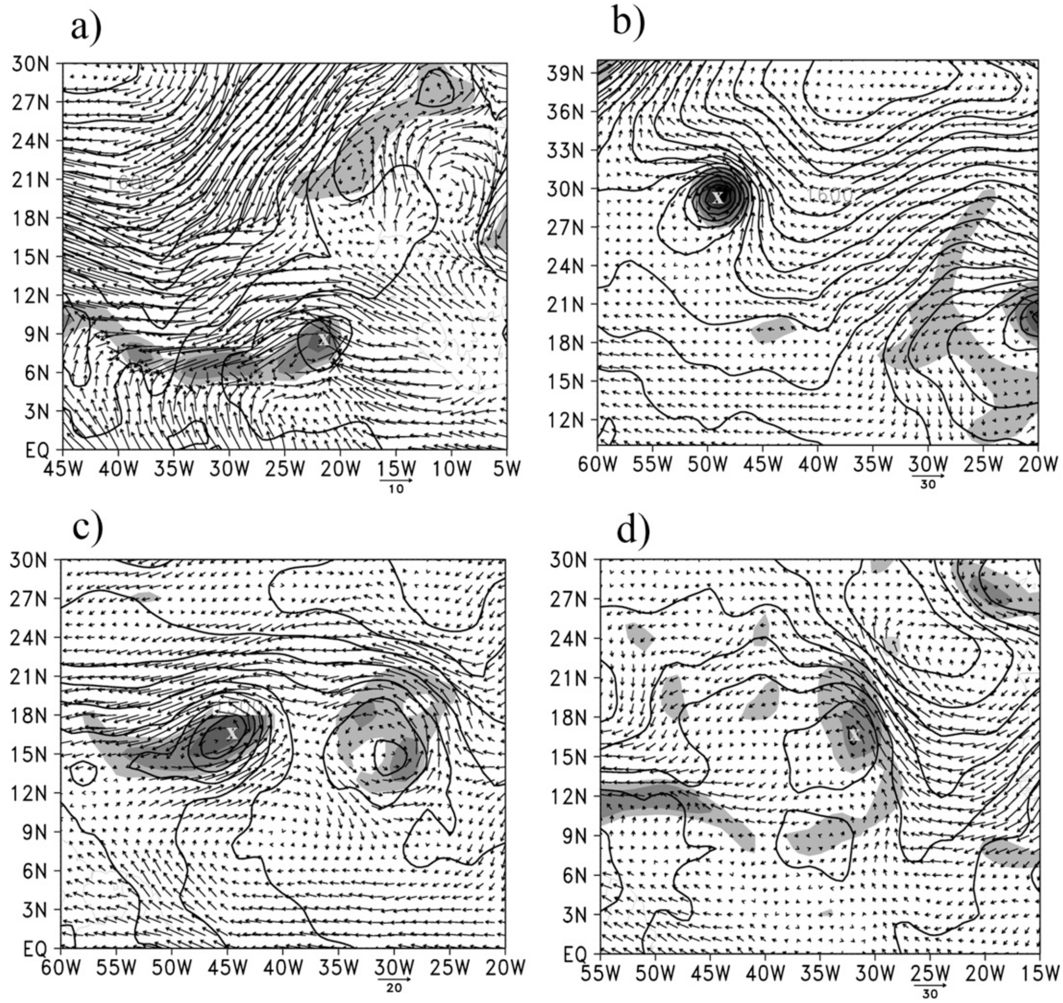

FIG. 4. The usefulness of different variables in tracking systems ("×" denotes the center of the system) in different situations. Shown are the vorticity (10^{-5} s^{-1}, shaded contours), vector wind, and geopotential heights (line contours) at 850 hPa for arbitrary cases. (a) Vorticity is useful in cases with broad waves. (b) For developed 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 useful.

atmosphere. These observations were also seen when comparing the manual tracks to the best-track data.

The magnitudes of the track errors for all cases with different forecast lead times at different pressure levels are compared (table omitted). After assessing and comparing the benefits of all levels, the tracks at 850 hPa were deemed to minimize the track errors and were used 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 the effects 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 (Kerns et al. 2008).

c. *Definition of tropical cyclone genesis*

Tropical waves, depressions, and storms are often characterized by geopotential height minima, vorticity maxima, a developed warm core, and cyclonic circulation. How well these characteristics are captured by numerical models strongly depends on the model resolution. To make a consistent definition of tropical cyclone genesis in the ensemble resolution, an analysis was conducted using the NCEP GFS final analysis (FNL) at the same grid spacing. By analyzing the magnitude and structure of the height, vorticity, and circulation fields in the GFS analysis at lower pressure levels (e.g., 700, 850, and 925 hPa) at the time closest to the actual genesis of the

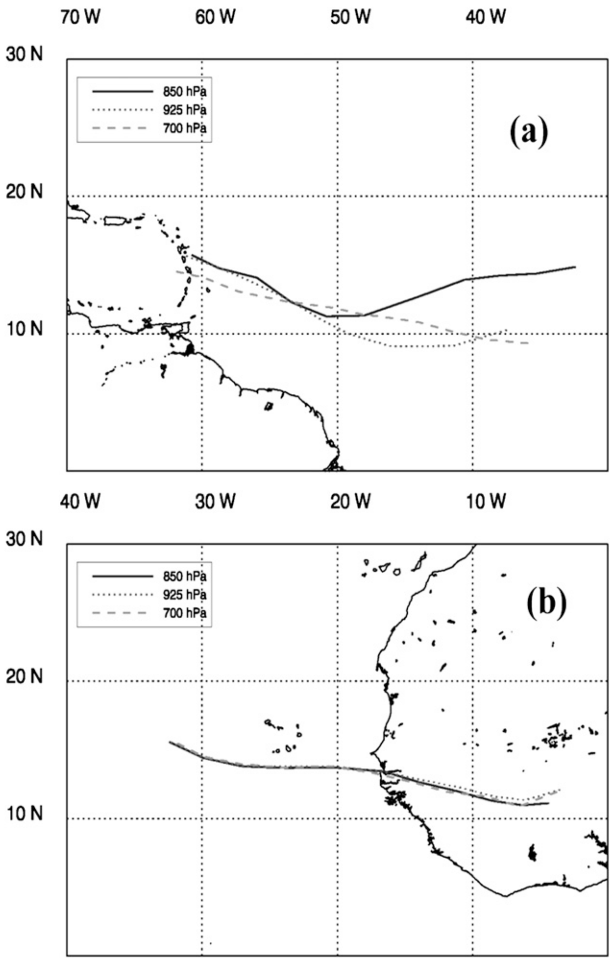

FIG. 5. Example of the variance in tracks at different atmospheric levels. (a) Ensemble mean 0–120-h forecast tracks from 66 h before Gordon was designated a tropical depression (0000 UTC 8 Sep 2006). As a weak system, the tracks widely varied at different vertical levels. (b) Ensemble mean 0–120-h forecast tracks from 60 h before Helene was designated a tropical depression (0000 UTC 10 Sep 2006). Stronger systems usually had similar tracks at different vertical levels.

five cases being studied, it is apparent that most cyclones were represented consistently at the 850-hPa pressure level by multiple closed height contours (at a 5-m interval) within 5° of the low pressure center, cyclonic circulation in the wind field, and a relative vorticity maxima, as well as a local maximum in the temperature between 200 and 500 hPa.

Therefore, a TC vortex is defined to have formed in the NCEP GFS deterministic and ensemble forecasts if the following conditions are satisfied:

1) At the 850-hPa pressure level,
   - a local maximum exists in the relative vorticity and the value of this relative vorticity at a grid point is larger than all points within 4° latitude, and
   - there are multiple closed height contours (at a 5-m interval) within 5° of the vortex center.
2) For the thermal field,
   - a local maximum is found in the average temperature between 200 and 500 hPa. This local maximum, which is considered to be the warm core center, must be within 2° latitude from the vortex center.

The successful genesis forecast is defined only if all of the above criteria are satisfied within 12 h of the actual genesis event. Most of these criteria are consistent with those that were used in Cheung and Elsberry (2002), where tropical cyclone formation over the western North Pacific was tracked by the Navy Operational Global Atmospheric Prediction System (NOGAPS).

Therefore, each ensemble forecast initialized in the pregenesis phase was given one of four designations. When all of the above criteria were met in a forecast, it was noted that the forecast predicted the cyclone's "genesis." If one or two of those conditions were depicted in the ensemble forecasts, then we label the forecasted system as "vortexlike." "Dissipation" was defined to mean that the parameters for genesis exist but do not persist for longer than 48 h from a given time. All forecasts for each member of the ensemble forecasts are tagged with one of these labels.

4. Tracking results

Based on the methods and criteria set up in section 3, we performed the system tracking for five named systems and the two nondeveloping waves. These named systems were tracked in ensemble forecasts from 3 days prior until 2 days after being designated a tropical depression by the NHC. The tracking results for all five named storms are summarized in Table 1. A detailed description for each case is given as follows.

a. Debby

The African easterly wave (AEW) that developed into Debby was disorganized and somewhat ambiguous over land and was, thus, not well resolved in the model. After moving offshore, though, circulation developed quickly. 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. Due to dry and stable air, along with marginal sea surface temperatures (SSTs), Debby never intensified into a hurricane. Shear associated with an approaching upper-level trough eventually caused the cyclone to dissipate.

In a 5-day forecast started from 0000 UTC 19 August, eight ensemble members predicted the development of

TABLE 1. Predictability of each storm for different lead times (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) and nondevelopment (N), respectively. Tracking results from the deterministic GFS forecast are also shown (single member).

Case	Debby		Ernesto		Florence		Gordon		Helene	
Forecast lead time (days)	No. of ensemble members in G/N	GFS	No. of ensemble members in G/N	GFS	No. of ensemble members in G/N	GFS	No. of ensemble members in G/N	GFS	No. of ensemble members in G/N	GFS
−3	8/6	N	1/13	N	12/2	G	0/14	N	14/0	G
−2	14/0	G	0/14	G	12/2	G	0/14	N	14/0	G
−1	14/0	G	4/10	G	14/0	G	0/14	N	14/0	G
0	14/0	G	14/0	G	14/0	G	5/9	N	14/0	G
1	14/0	G	14/0	G	14/0	G	14/0	G	14/0	G
2	N/A	N/A	14/0	G	14/0	G	14/0	G	14/0	G

Debby (Table 1). A majority of the members (12 out of 14) predicted a track that was well south of the best track (Fig. 6a). Most of the members that initially predicted genesis failed to ultimately strengthen the cyclone. In the forecast from 20 August 2006, all ensemble members were in better agreement about developing a storm (Fig. 6b). As the forecast time approached genesis, most members followed a similar track. Consistency among different ensemble members improved as the initialized disturbance became better organized closer to genesis. Tracks were aligned and the spread was smallest for the forecasts starting from 21 and 22 August (Figs. 6c and 6d). In particular, the forecast tracks from 22 August started from the same location due to the NCEP implementation of the vortex relocation after the system became a classified tropical depression (Fig. 6d). The ensemble mean is also close to the best track in forecasts from 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. 6e). Overall, the skill of the track predictions for Debby showed tremendous improvement as the system became more organized, especially after developing into a tropical depression (Figs. 6d and 6e).

Compared with the ensemble forecasts, the deterministic GFS forecast did not show any development from the predominant vorticity maximum, but rather showed a brief rapid deepening of an unrelated vorticity maximum to the southwest of the actual storm's location in the forecast from 0000 UTC 19 August (Fig. 6a). For the remainder of the forecasts, the GFS tended to depict a strong system. The forecasts started from 20 and 21 August were slower and more northerly than the ensemble mean (Figs. 6b and 6c), while the forecasts from 22 and 23 August aligned more closely to the ensemble members (Figs. 6d and 6e).

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. Most ensemble members did not develop a depression in the forecasts initialized in the pregenesis phase, although the number of vortexlike systems increased with a short forecast lead time (e.g., a day before the observed genesis 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. 7a). While more members produced weak vortexlike circulations in the forecasts that started on 23 and 24 August, predictions of a well-developed tropical system remained nonexistent (Figs. 7b and 7c). The track of the wave/vorticity maximum remained much farther south than the course the actual system tracked, and was much faster than the actual system. In forecasts initialized from 0000 UTC 25 August, almost all ensemble members continued to predict a westward track, even after the system developed into a tropical storm (Fig. 7d). The ensemble mean showed a track into the Gulf of Mexico, while the actual system made landfall in southern Florida. This pattern continued in forecasts from 26 August and no members produced a well-developed storm, and many completely dissipated it. Those members that held the system together showed deepening over the Gulf Stream. 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 pregenesis cases, the forecasts only produced vortexlike tropical waves. The disagreement among the members implies

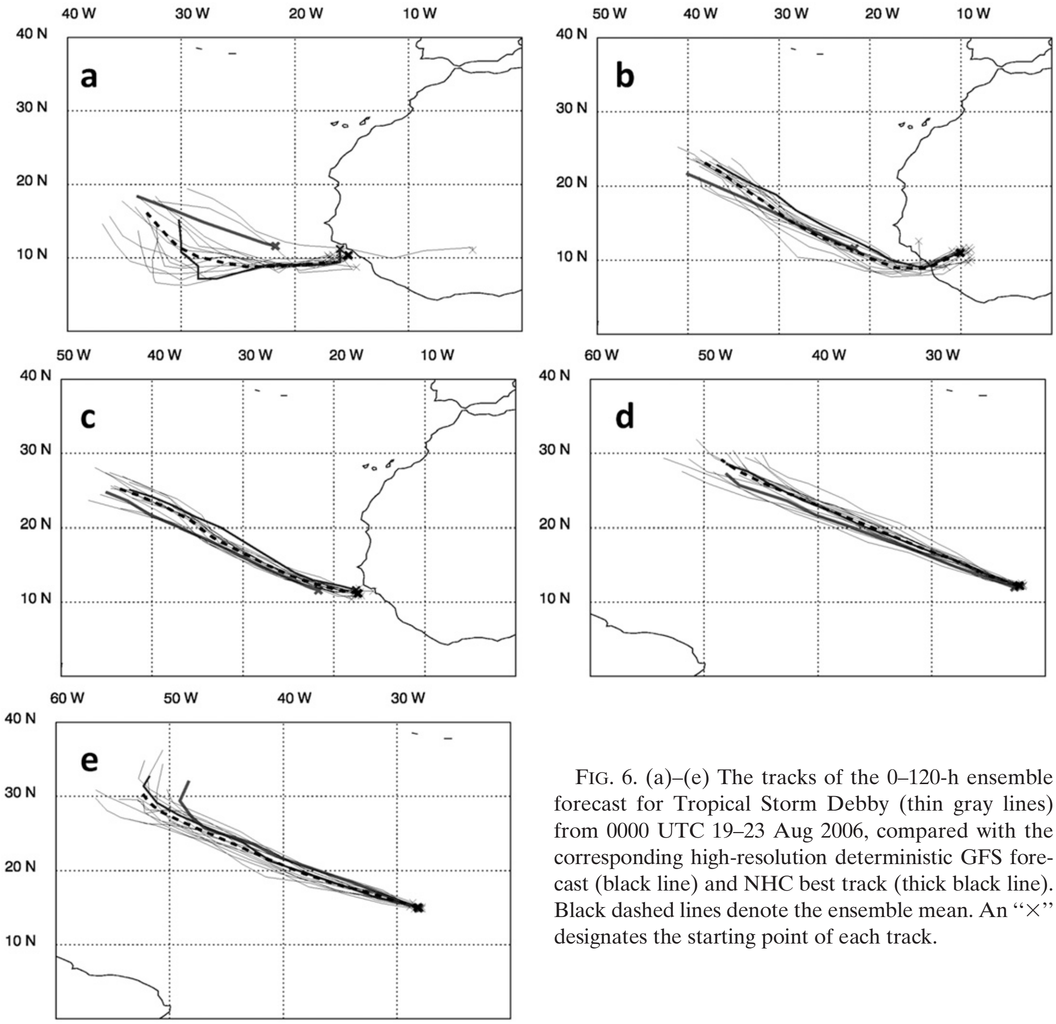

FIG. 6. (a)–(e) The tracks of the 0–120-h ensemble forecast for Tropical Storm Debby (thin gray lines) from 0000 UTC 19–23 Aug 2006, compared with the corresponding high-resolution deterministic GFS forecast (black line) and NHC best track (thick black line). Black dashed lines denote the ensemble mean. An "×" designates the starting point of each track.

a difficult forecast. After its genesis the track forecasts showed some improvement (Figs. 7d, 7e, and 7f), while the intensity forecasts remained too weak.

The GFS forecast starting from 0000 UTC 22 August was on the north side of the ensemble envelope. In the forecasts initialized on 23 August the GFS forecasted development, but it occurred from a vorticity maximum that split off from the wave and developed northeast of the ensemble mean and the actual track. The GFS forecasts from 24 to 26 August were generally south of the ensemble mean. However, the GFS forecast starting on 0000 UTC 27 August (Fig. 7f) produced a system with tropical storm intensity and had a track comparable to the best 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 3 September 2006 and gradually strengthened to a tropical storm by 0600 UTC 5 September 2006. The system presented a complex situation in the pregenesis environment. Specifically, there were three vorticity maxima over 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 developed the storm from the first wave, while other members developed it from the second wave or from the merger of the first and second waves. On a few occasions, there would be a second merger with the third wave, or the depression would form from the third wave itself. Some ensemble members accurately predicted the genesis, but also generated an equal or even stronger system to the northeast of Florence. In general, the continuity was inconsistent and the tracking was difficult in the pregenesis environment of these ensemble forecasts.

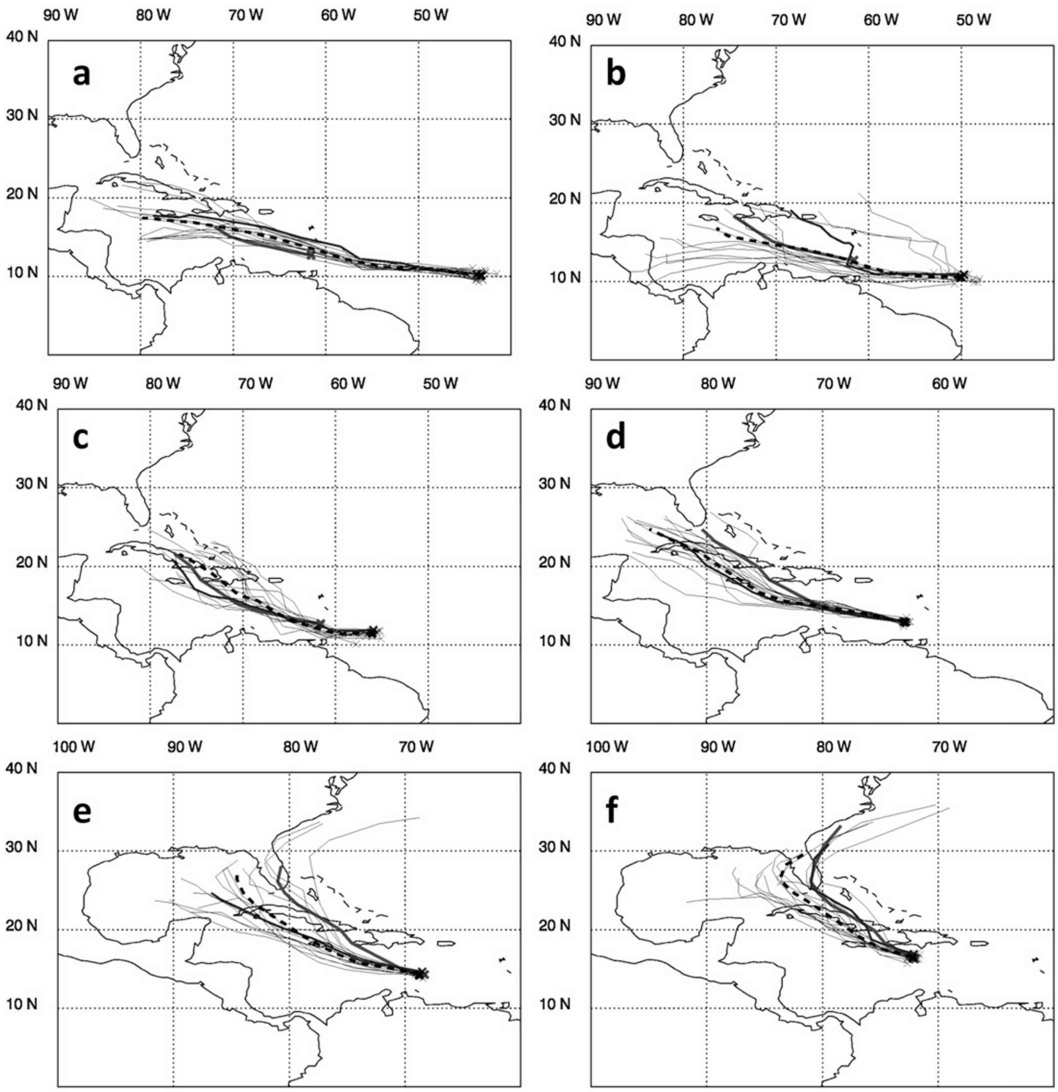

FIG. 7. (a)–(f) As in Fig. 6, but for Hurricane Ernesto and forecast from 0000 UTC 22–27 Aug 2006.

This ambiguity was present in the forecasts from 1–3 September, although the forecasts became more consistent as time progressed. For example, Figs. 8a–c shows a large ensemble spread present in the forecast from 1 to 3 September, with few members representing the best track. After the genesis stage (forecasts from 4 to 6 September; see Figs. 8d–f), the system tracks were well predicted by the ensemble forecasts, with small spread and a mean track close to the best track. Compared with Figs. 8a–c, there is significant improvement in the ensemble forecast after the system strengthened to a tropical depression. The improved postgenesis track forecasts may be attributed to the NCEP storm relocation scheme.

In the GFS deterministic forecast, the third wave was responsible for forming the strongest storm in the forecasts, starting from 1 and 2 September. For the forecast starting on 3 September, the GFS presented a complex and discontinuous solution that was similar to a subset of the ensemble members (Figs. 8a–c). In the postgenesis phase, the GFS was 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 forecasts started at 0000 UTC 8 and 9 September (Figs. 9a and 9b), none of the ensemble members predicted the actual position of the depression. Even when the ensemble showed a strengthening of the disturbance, it was too far south or

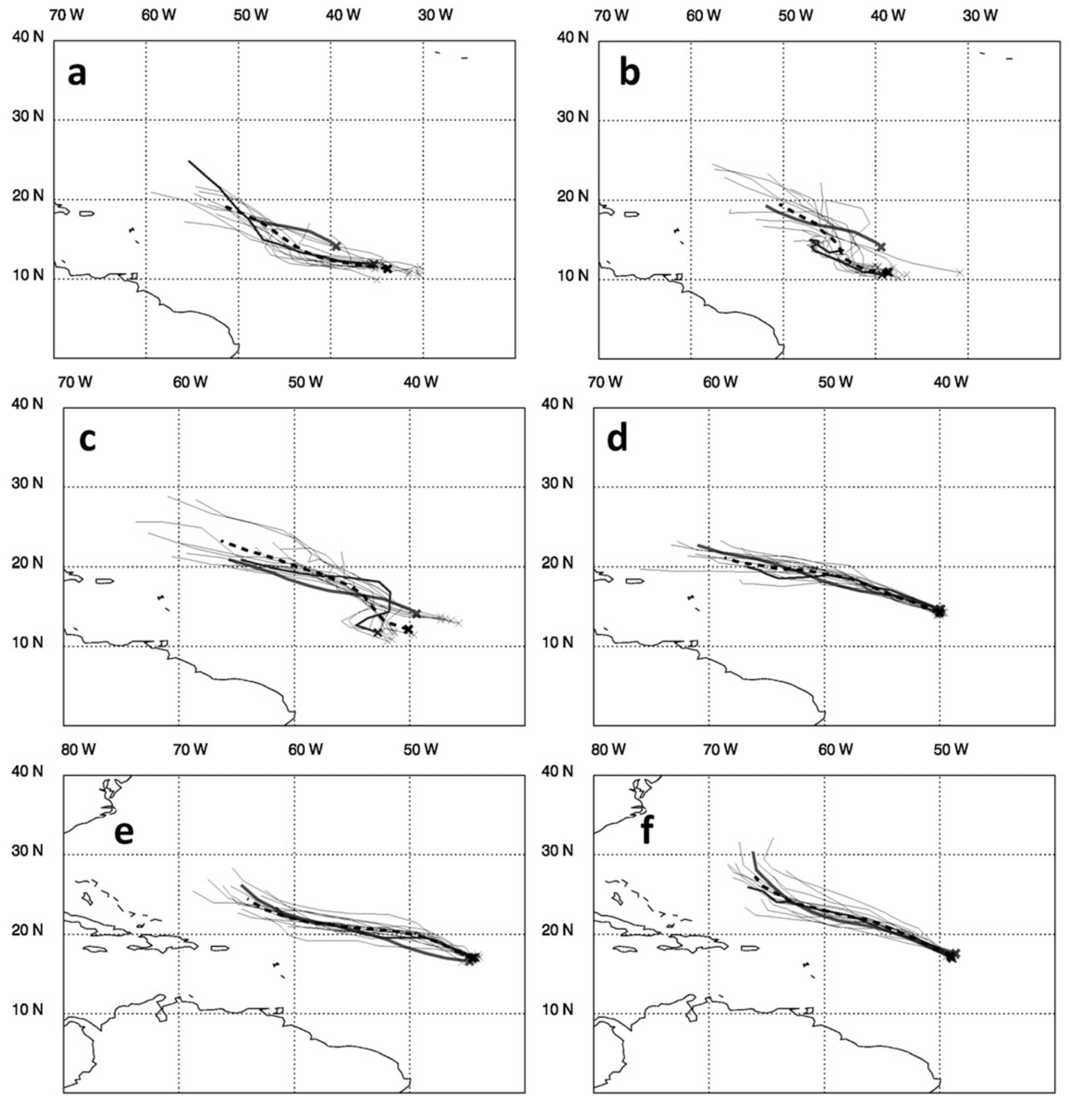

FIG. 8. (a)–(f) As in Fig. 6, but for Hurricane Florence and forecast from 0000 UTC 1–6 Sep 2006.

west of the actual track. Figure 9a (forecasts from 8 September) shows how almost all of the tracks were well south of the actual genesis location. In most of these cases, the initial vorticity maximum in most of these members tracked well south and merged into a strip of high vorticity. The forecast from 10 September (Fig. 9c) produced a wide array of tracks, with only one member correctly showing the system curving northward. By 11 September (after the tropical depression had developed), the model forecast agreed with the best track (Fig. 9d). However, none of the ensemble members suggested that a strong hurricane would eventually develop. In addition, ensemble forecasts were consistently too fast with movement, leading to erroneous forecasts of dissipation or absorption by midlatitude troughs. Once again, the NCEP ensemble forecast showed significant improvement in track forecasts after the system became a tropical depression (Fig. 9e).

The GFS forecasts from 8 and 9 September also produced a weak system that tracked too far south (Figs. 9a and 9b). In the forecasts starting from 10 September, the original vorticity maximum dissipated, while another wave formed within the vorticity strip. The GFS forecasts from 12 September proved to be very close to the actual track, while maintaining strength and circulation (Fig. 9e). Initializing on 13 September, the GFS track mirrored both the best track and the ensemble mean, but was displaced to the west (Fig. 9f).

e. Helene

This case was the easiest to track, as a strong AEW developed quickly as it moved westward off the coast.

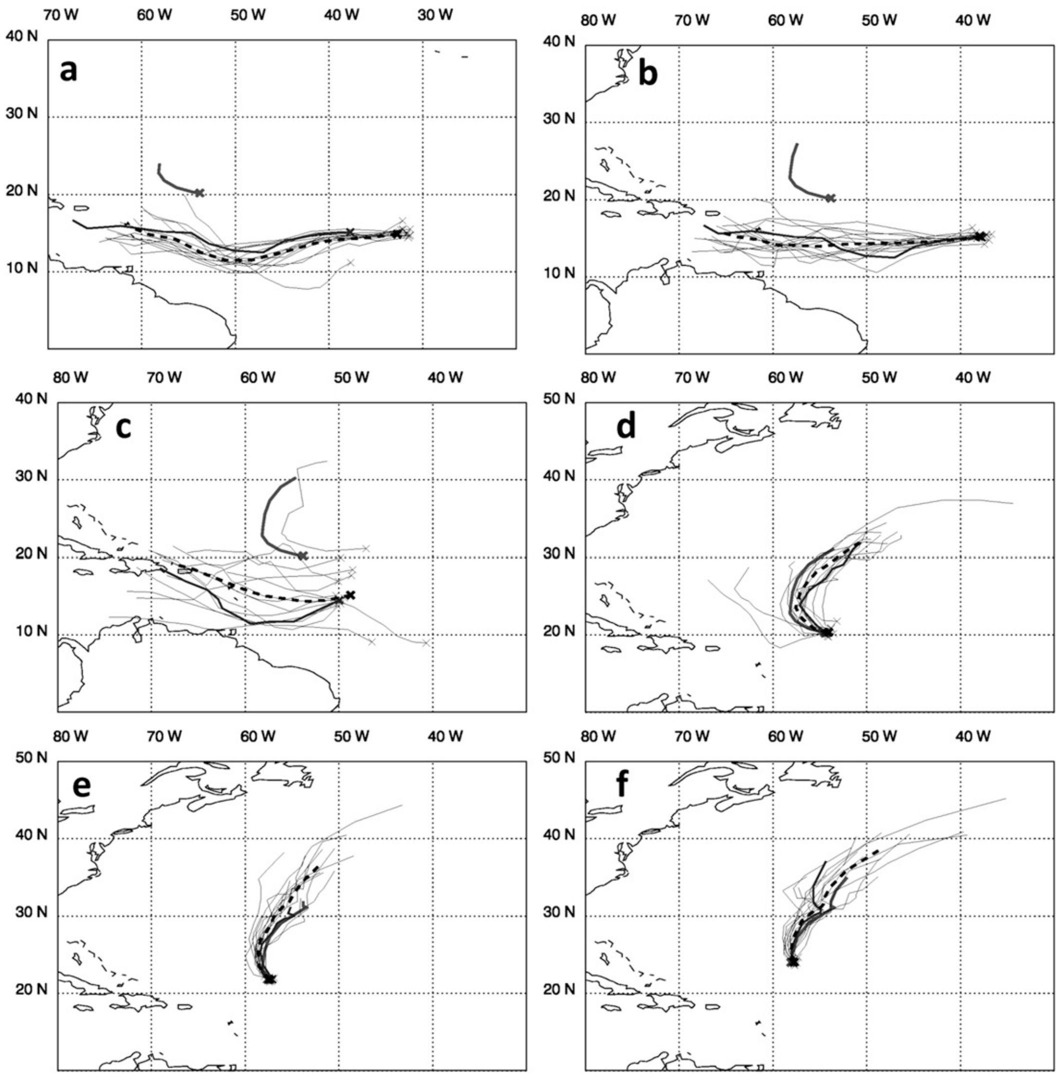

FIG. 9. (a)–(f) As in Fig. 6, but for Hurricane Gordon and forecast from 0000 UTC 8–13 Sep 2006.

It was classified as a tropical depression at 1200 UTC 12 September 2006 and was upgraded to a tropical storm at 0000 UTC 14 September. Even in the pregenesis phase over Africa, the vorticity maxima and wind circulation were well defined in the model initialization. The forecast from 0000 UTC 10 September showed a strong developing system as soon as the disturbance exited the coast, with all 14 members predicting genesis. In Fig. 10a, the track forecasts are close to the actual track with little spread, although eight members predicted slower movement. The forecasts from 11 September (Fig. 10b) suggested a weaker cyclone, and one member dissipated the cyclone. The cyclones 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. 10d. The ensemble mean is close to the best track with a classical conelike ensemble spread. Ensemble members handled the timing of the deepening with relative accuracy. The forecast from 15 September did not 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. 10b), where the system movement was much slower than the ensemble members and showed a northerly turn. Otherwise, the GFS forecasts were similar to the ensemble.

f. Nondeveloping cases

Although the aforementioned five tropical storms all developed from AEWs, not all easterly waves develop

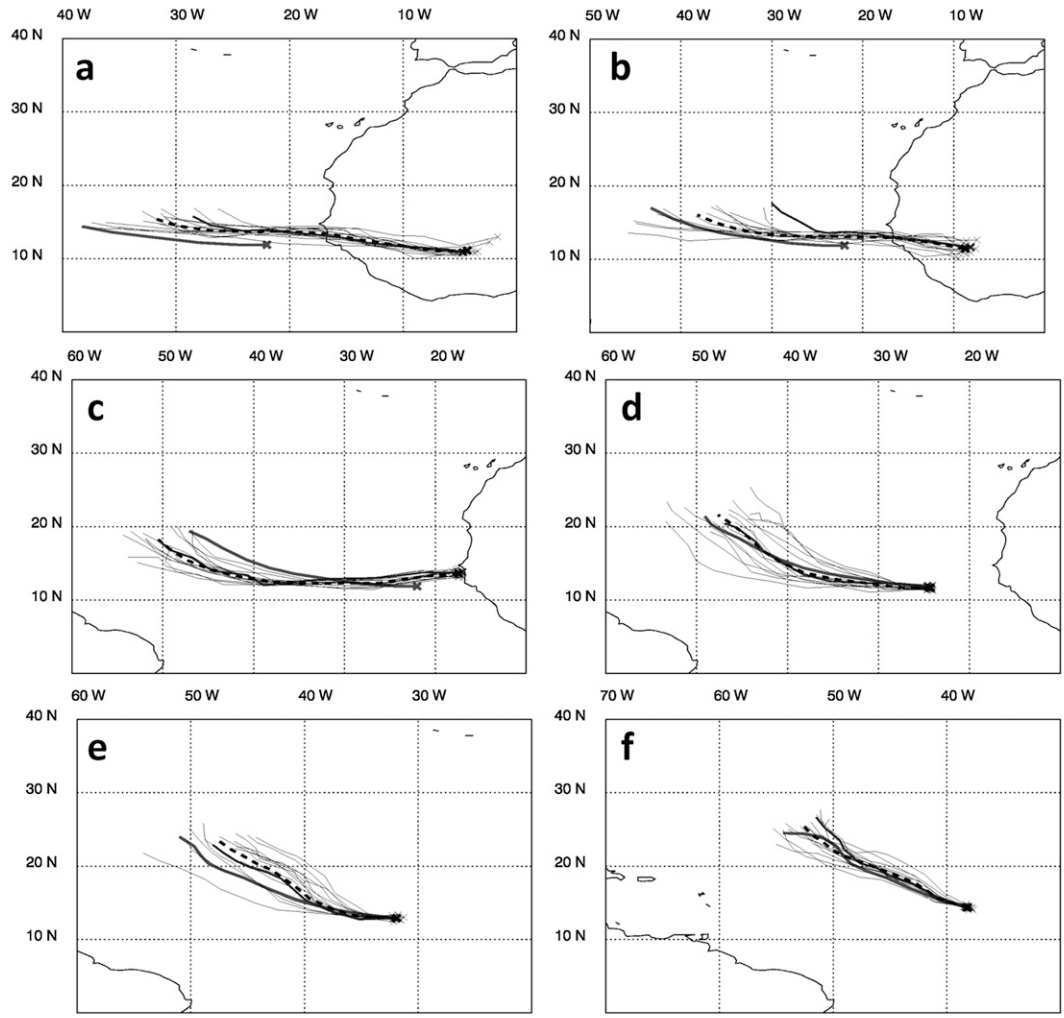

FIG. 10. (a)–(f) As in Fig. 6, but for Hurricane Helene and forecast from 0000 UTC 10–15 Sep 2006.

into tropical storms. Some studies, such as Thorncroft and Hodges (2001), have sought to determine a relationship between AEWs and tropical cyclone genesis. While they found a correlation between the number of these tropical waves and the subsequent number of tropical cyclones, not all the waves strengthen into storms. When an AEW (commonly called a tropical disturbance) does not reach tropical depression status as determined by the NHC, it is considered a nondeveloping system. Forecasts that predict tropical cyclone genesis within these nondeveloping tropical waves are referred to as false alarms. There were only two nondeveloping systems during NAMMA. To evaluate whether the ensemble produces such false alarms, two nondeveloping systems during this period are examined. The same criteria as mentioned in section 3b are applied for the cases.

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. Afterward, it slowly weakened, never developing into a tropical storm. The ensemble handled the tracks of wave 3 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 an interaction between one vorticity maximum moving westward, 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 were diffuse, especially in forecasts initialized on 24 and 25 August (850- and 700-hPa tracks diverge in opposite directions). A broad, weak circulation was evident initially in most of the forecasts, and by our criteria constituted vortexlike structures. However, there were no forecasts of genesis or maintained strength (Table 2), meaning the model performance for

TABLE 2. Predictability of wave 3 in 5-day forecasts started from different lead times. Values represent the number of members (out of 14) predicting genesis, vortexlike development, premature dissipation, and nondevelopment, respectively.

Forecast date	23 Aug	24 Aug	25 Aug	26 Aug	27 Aug
Genesis	0	0	0	0	0
Vortex like	10	8	14	10	8
Nondevelopment	3	6	0	3	6
Dissipation	1	0	0	1	0

TABLE 3. As in Table 3, but for the predictions of the genesis for wave 6.

Forecast date	7 Sep	8 Sep	9 Sep	10 Sep
Genesis	0	0	0	0
Vortex like	5	0	1	1
Nondevelopment	9	14	13	13
Dissipation	0	0	0	0

intensity was reasonable. While the deterministic GFS occasionally spun up small vorticity centers within the wave, none of them met genesis criteria (except one vortexlike case on 26 August) and most quickly dissipated.

Wave 6 was another AEW that tracked due west from Africa. Its strength was much weaker and more diffuse than wave 3 but was better defined at 700 hPa. Wave 6 could only be consistently tracked in the 5-day forecasts from 7 through 10 September. Over time, this wave was generally forecast to dissipate and merge with a southern strip of vorticity. The track forecasts had more spread than wave 3, although its 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 vortexlike structure were very brief in nature. Tables 2 and 3 show the high rate of nondevelopment forecasts, indicating the high accuracy of the ensemble in predicting this nondeveloped system. The GFS also predicted a weak wave in all of the evaluated forecasts.

5. Overall evaluation

Table 1 shows the fraction of tropical cyclone genesis in ensemble forecasts that started before and after genesis. The predictive skill of tropical cyclone genesis varies from case to case as described in section 4. Specifically, the skill in pregenesis forecasts is highly case dependent. To make an overall evaluation, Table 4 shows the overall predicted probability of genesis in ensemble forecasts over all five cases by summarizing the results from Table 1. It is apparent that the accuracy for the forecast genesis was equal to or above 50% for ensemble forecasts initialized in the pregenesis phase. The forecast uncertainty generally decreases with the reduction of the forecast lead time. The predicted probabilities of genesis were above 87% and 100% for ensemble forecasts initialized in the near- and postgenesis phases, respectively. The predicted probability for development was over 57% for the forecasts in the pregenesis phase. The probability of nondevelopment was less than 43% for the five named systems. The probability for development was over 87% for the forecasts near the actual genesis time (e.g., 0-day lead time). The tropical cyclone vortices were well represented in all of the ensemble members in the postgenesis phase. These results show that the ensemble forecast offered a reasonable indication of the possibility of tropical cyclone genesis in these cases.

The track forecasts were significantly improved in the postgenesis phase for all the cases (Figs. 6–10), possibly because of the implementation of the NCEP storm relocation scheme in the operational ensemble forecasting system. Further studies are needed to evaluate the impacts of the storm relocation scheme on the ensemble track forecasting. In addition, both the GFS and ensemble forecasts handled the two nondeveloped tropical waves very well. None of the forecasts predicted a false alarm of tropical cyclone genesis in any of the different forecast lead times (Tables 2 and 3).

The performance of the higher-resolution deterministic GFS control forecast also varied between the cases. Including all five cases, a total of 29 forecast periods were evaluated. In 14 of the forecasts, the GFS track forecasts nearly overlapped with the track of the ensemble mean. For the forecasts in which the ensemble mean and GFS differed, the GFS was closer to the best-track data in two cases, while the ensemble mean was closer to the best track six times. For another eight instances, GFS forecasts did not compare closely to the ensemble mean or actual track. In five forecasts, 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 did not change the overall ensemble mean significantly, even when the GFS had a large deviation from the main ensemble members. This fact mainly occurred because the ensemble size is 14, which is much larger than the sample of a single control forecast. However, even with coarser resolution (T126L28 for ensemble versus T384L64 of GFS), the overall level of performance of the ensemble

TABLE 4. Probability of system development in the ensemble for each lead time (forecast from the number of days), combined for all five named storms.

Lead time (days)	Genesis	Nondevelopment
−3	35/70 (50%)	35/70 (50%)
−2	40/70 (57%)	30/70 (43%)
−1	46/70 (66%)	24/70 (34%)
0	61/70 (87%)	9/70 (13%)
1	70/70 (100%)	0/70 (0%)
2	56/56 (100%)	0/56 (0%)

in predicting tropical cyclone genesis is compatible to the high-resolution deterministic GFS.

6. Concluding remarks

NCEP global ensemble forecasts for five developed and two nondeveloped tropical systems from the 2006 North Atlantic hurricane season were evaluated. The primary focus of the study was to determine how skillfully the ensemble performed in predicting the genesis and evolution of the tropical systems. For this study, each system was tracked using a manual method that utilized vorticity, geopotential height at 850 hPa, and the average temperature between 200 and 500 hPa.

Overall, the ensemble forecasts predicted high probabilities of genesis for the three strong storms (Debby's mature phase, Florence, and Helene) that formed from AEWs, but failed to accurately predict the pregenesis phase of two weaker storms that formed farther west in the Atlantic Ocean (Ernesto and Gordon). The differences in the pregenesis environment may play an important role in the forecast accuracy. Further investigation needs to be done in future work. In addition, although ensemble forecasts performed poorly for Ernesto and Gordon, disagreement among the ensemble members implies a difficult forecast for these cases.

Statistically, the overall accuracy for the genesis forecasts is above 50% for the NCEP ensemble forecasts initialized in the pregenesis phase. The forecast uncertainty generally decreases with the reduction of the forecast lead time. The probabilities of the ensemble forecasts predicting and maintaining tropical cyclone strength reach, respectively, 87% and 100% for the forecasts initialized near and in the postgenesis phase (we are not predicting genesis after genesis has occurred). The skill of the ensemble track forecasts was significantly improved in forecasts in the postgenesis phase, possibly because of the implementation of the NCEP storm relocation scheme, which makes an accurate initial storm location for all ensemble members. Further studies are needed to evaluate the impacts of the storm relocation scheme on the ensemble track forecasting.

Even with the coarser resolution, the overall performance of the ensemble in predicting tropical cyclone genesis is compatible to the high-resolution deterministic GFS. False-alarm rates for nondeveloping waves were low in both the GFS and ensemble for two cases presented in this paper.

The cases studied in this paper demonstrate that an operational ensemble forecast system may prove valuable in tropical prediction. However, with only five developed storms and two nondeveloped systems, it must be noted that the sample size of the evaluation is small in this study. More case studies need to be completed in the future to fully evaluate the skill of the NCEP ensemble in predicting tropical cyclone development. In addition, in order to enhance the skill of the ensemble forecasts, it is necessary to evaluate the impacts 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.

Acknowledgments. This study is supported by ONR Grant N000140810308 through financial support to the first two authors. The 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 the University of Utah are also appreciated. The review comments from two anonymous reviewers, Dr. Joshua Hacker, and Prof. Da-Lin Zhang were very helpful in improving the manuscript.

REFERENCES

Aberson, S. D., 1999: Ensemble-based products to improve tropical cyclone forecasting. Preprints, *23rd Conf. on Hurricanes and Tropical Meteorology,* Dallas, TX, Amer. Meteor. Soc., 843–844.

——, 2001: The ensemble of tropical cyclone track forecasting models in the North Atlantic Basin (1976–2000). *Bull. Amer. Meteor. Soc.,* **82,** 1895–1904.

——, M. A. Bender, and R. E. Tuleya, 1998: Ensemble forecasting of tropical cyclone tracks. Preprints, *12th Conf. on Numerical Weather Prediction,* Phoenix, AZ, Amer. Meteor. Soc., 290–292.

Albignat, J. P., and R. J. Reed, 1980: The origin of African wave disturbances during phase III of GATE. *Mon. Wea. Rev.,* **108,** 1827–1839.

Beven, J. L., 1999: The boguscane—A serious problem with the NCEP medium range forecast model in the Tropics. Preprints, *23rd Conf. on Hurricanes and Tropical Meteorology,* Dallas, TX, Amer. Meteor. Soc., 845–848.

Buizza, R., P. L. Houtekamer, Z. Toth, G. Pellerin, M. Wei, and Y. Zhu, 2005: A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Mon. Wea. Rev.,* **133,** 1076–1097.

Burpee, R., 1972: The origin and structure of easterly waves in the lower troposphere of North Africa. *J. Atmos. Sci.,* **29,** 77–90.

Chen, T. C., 2006: Characteristics of African easterly waves depicted by ECMWF reanalyses for 1991–2000. *Mon. Wea. Rev.,* **134,** 3539–3566.

Cheung, K. W., and J. C. L. Chan, 1999: Ensemble forecasting of tropical cyclone motion using a barotropic model. Part II: Perturbations of the vortex. *Mon. Wea. Rev.,* **127,** 2617–2640.

——, and R. L. Elsberry, 2002: Tropical cyclone formation over the western North Pacific in the Navy Operational Global Atmospheric Prediction System forecasts. *Wea. Forecasting,* **17,** 800–820.

Fink, A. H., D. G. Vincent, P. M. Reiner, and P. Speth, 2004: Mean state and wave disturbances during phases I, II, and III of GATE based on ERA-40. *Mon. Wea. Rev.,* **132,** 1661–1683.

Goerss, J. S., 2000: Tropical cyclone track forecasts using an ensemble of dynamical models. *Mon. Wea. Rev.,* **128,** 1187–1193.

Hennon, C. C., and J. S. Hobgood, 2003: Forecasting tropical cyclogenesis over the Atlantic basin using large-scale data. *Mon. Wea. Rev.,* **131,** 2927–2940.

Kerns, B., K. Greene, and E. Zipser, 2008: Four years of tropical ERA-40 vorticity maxima tracks. Part I: Climatology and vertical vorticity structure. *Mon. Wea. Rev.,* **136,** 4301–4319.

Krishnamurti, T. N., C. M. Kishtawal, Z. Zhang, T. LaRow, D. Bachiochi, E. Williford, S. Gadgil, and S. Surendran, 2000: Multimodel ensemble forecasts for weather and seasonal climate. *J. Climate,* **13,** 4196–4216.

Liu, Q., S. Lord, N. Surgi, Y. Zhu, R. Wobus, Z. Toth, and T. Marchok, 2006: Hurricane relocation in global ensemble forecast system. Preprints, *27th Conf. on Hurricanes and Tropical Meteorology,* Monterey, CA, Amer. Meteor. Soc., P5.13. [Available online at http://ams.confex.com/ams/pdfpapers/108503.pdf.]

Mackey, B. P., and T. N. Krishnamurti, 2001: Ensemble forecast of a typhoon flood event. *Wea. Forecasting,* **16,** 399–415.

Marchok, T. P., 2002: How the NCEP Tropical Cyclone Tracker works. Preprints, *25th Conf. on Hurricanes and Tropical Meteorology,* San Diego, CA, Amer. Meteor. Soc., P1.13. [Available online at http://ams.confex.com/ams/pdfpapers/37628.pdf.]

McBride, J. L., and R. Zehr, 1981: Observational analysis of tropical cyclone formation. Part II: Comparison of non-developing versus developing systems. *J. Atmos. Sci.,* **38,** 1132–1151.

Pasch, R. J., J.-G. Jiing, F. M. Horsfall, H.-L. Pan, and N. Surgi, 2002: Forecasting tropical cyclogenesis in the NCEP global model. Preprints, *25th Conf. on Hurricanes and Tropical Meteorology,* San Diego, CA, Amer. Meteor. Soc., 178–179.

Perrone, T. J., and P. R. Lowe, 1986: A statistically derived prediction procedure for tropical storm formation. *Mon. Wea. Rev.,* **114,** 165–177.

Pytharoulis, I., and C. Thorncroft, 1999: The low-level structure of African easterly waves in 1995. *Mon. Wea. Rev.,* **127,** 2266–2280.

Reed, R., A. Hollingsworth, W. Heckley, and F. Delsol, 1988: An evaluation of the performance of the ECMWF operational system in analyzing and forecasting easterly wave disturbances over Africa and the tropical Atlantic. *Mon. Wea. Rev.,* **116,** 824–865.

Rennick, M. A., 1999: Performance of the navy's tropical cyclone prediction model in the western North Pacific basin during 1996. *Wea. Forecasting,* **14,** 297–305.

Reynolds, C. A., J. Teixeira, and J. G. McLay, 2008: Impact of stochastic convection on the ensemble transform. *Mon. Wea. Rev.,* **136,** 4517–4526.

Rogers, R., and Coauthors, 2006: The Intensity Forecasting Experiment: A NOAA multiyear field program for improving tropical cyclone intensity forecasts. *Bull. Amer. Meteor. Soc.,* **87,** 1523–1537.

Thorncroft, C., and K. Hodges, 2001: African easterly wave variability and its relationship to Atlantic tropical cyclone activity. *J. Climate,* **14,** 1166–1179.

Toth, Z., and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.,* **125,** 3297–3319.

Weber, H. C., 2003: Hurricane track prediction using a statistical ensemble of numerical models. *Mon. Wea. Rev.,* **131,** 749–770.

Wei, M., Z. Toth, R. Wobus, and Y. Zhu, 2008: Initial perturbations based on the ensemble transform (ET) technique in the NCEP global operational forecast system. *Tellus,* **60,** 62–79.

Zawislak, J. A., and E. J. Zipser, 2009: Observations of seven African easterly waves in the east Atlantic during 2006. *J. Atmos. Sci.,* **67,** 26–43.

Zhang, Z., and T. N. Krishnamurti, 1997: Ensemble forecasting of hurricane tracks. *Bull. Amer. Meteor. Soc.,* **78,** 2785–2795.

Zipser, E. J., and Coauthors, 2009: The Saharan air layer and the fate of African easterly waves—NASA's AMMA 2006 field program to study tropical cyclogenesis. *Bull. Amer. Meteor. Soc.,* **90,** 1137–1156.