Tropical Cyclone–like Vortices Detection in the NCEP 16-Day Ensemble System over the Western North Pacific in 2008: Application and Forecast Evaluation

HSIAO-CHUNG TSAI and KUO-CHEN LU
Central Weather Bureau, Taipei, Taiwan

RUSSELL L. ELSBERRY
Graduate School of Engineering and Applied Sciences, Department of Meteorology, Naval Postgraduate School, Monterey, California

MONG-MING LU
Central Weather Bureau, Taipei, Taiwan

CHUNG-HSIUNG SUI
Institute of Hydrological and Oceanic Sciences, and Department of Atmospheric Sciences, National Central University, Jhongli City, Taiwan

(Manuscript received 1 March 2010, in final form 6 October 2010)

ABSTRACT

An automated technique has been developed for the detection and tracking of tropical cyclone–like vortices (TCLVs) in numerical weather prediction models, and especially for ensemble-based models. A TCLV is detected in the model grid when selected dynamic and thermodynamic fields meet specified criteria. A backward-and-forward extension from the mature stage of the track is utilized to complete the track. In addition, a fuzzy logic approach is utilized to calculate the TCLV fuzzy combined-likelihood value (TFCV) for representing the TCLV characteristics in the ensemble forecast outputs. The primary objective of the TCLV tracking and TFCV maps is for use as an evaluation tool for the operational forecasters. It is demonstrated that this algorithm efficiently extracts western North Pacific TCLV information from the vast amount of ensemble data from the NCEP Global Ensemble Forecast System (GEFS). The predictability of typhoon formation and activity during June–December 2008 is also evaluated. The TCLV track numbers and TFCV averages around the formation locations during the 0–96-h period are more skillful than for the 102–384-h forecasts. Compared to weak tropical cyclones (TCs; maximum intensity ≤ 50 kt), the storms that eventually become stronger TCs do have larger TFCVs. Depending on the specified domain size and the ensemble track numbers to define a forecast event, some skill is indicated in predicting the named TC activity. Although this evaluation with the 2008 typhoon season indicates some potential, an evaluation with a larger sample is necessary to statistically verify the reliability of the GEFS forecasts.

1. Introduction

Many studies have endeavored to improve the predictive skill in forecasting tropical cyclones (TC) due to their high hazardous impacts. For the 5-day forecasts, the outputs from high-resolution numerical weather prediction (NWP) models have been demonstrated to be useful track forecast guidance. The track forecast skill within 5 days has been steadily improved due to improved numerical model guidance and the utilization of consensus techniques that combine the predicted tracks from multiple skillful NWP models (Goerss et al. 2004; Elsberry 2007; Goerss 2007). The level of skill will also depend on the initial state via the observations and data assimilation in the ensemble prediction system and the perturbation method (Froude 2010). However, the uncertainty of physical-process parameterizations of
subgrid-scale motions in NWP models leads to limited skill beyond this time scale, such as week-2 forecasts and seasonal climate predictions (Shapiro and Thorpe 2004). In addition, TCs may form at longer lead times; thus, the TC forecast approach at these time scales generally focuses on the TC activity (i.e., numbers or occurrences).

Several TC activity forecasts have been successfully developed. One approach to providing TC activity forecast guidance is to establish specified indices or statistical relationships by utilizing the predicted or observed environmental conditions (Gray 1975; Gray 1979; Gray et al. 1992; Watterson et al. 1995; DeMaria et al. 2001; Hennon and Hobgood 2003; Emanuel and Nolan 2004; Leroy and Wheeler 2008). The selected factors in the numerical model output or remotely sensed data employed by these methods may include sea surface temperature, vertical wind shear, cyclonic (anticyclonic) vorticity in the lower (upper) troposphere, instability, brightness temperature, water vapor, etc.

The alternate strategy is to directly detect and track the tropical cyclone–like vortices (TCLVs) in the numerical model outputs. The basis for directly utilizing NWP outputs is that the dynamical model is explicitly predicting the favorable environmental conditions for TC development and intensification, and also predicting the vortices that will develop in response to those conditions.

Two types of TC tracking algorithms have been utilized. The first type locates the TCLV center in the model by searching for a minimum mean sea level pressure (MSLP) or a maximum 850-hPa relative vorticity around the observed TC center position (i.e., the first-guess position). Once the initial center position is located in the chosen field, the automated program will then search for the TCLV centers in the succeeding forecast time steps. This type of TC tracking is often applied at short forecast lead times and may be regarded as a “semi-automated” algorithm. That is, the TC is only tracked once it has been designated as a tropical cyclone or tropical storm by the forecast center. Similar algorithms are developed and implemented in the operational applications on both deterministic models and the ensemble system (Heming 1994; Sampson and Schrader 2000; Marchok 2002; ECMWF 2004; Lu et al. 2007).

The second type of TC tracking algorithm attempts to detect and track TCLV centers without specification of first-guess positions. A TCLV center is objectively detected in the model grid output when selected dynamic and thermodynamic fields meet specified criteria. This approach has been effective in several global and climate model studies (Vitart et al. 1997; Nguyen and Walsh 2001; Camargo and Zebiak 2002; Oouchi et al. 2006; Yoshimura et al. 2006; Bengtsson et al. 2007; Froude et al. 2007; Tsai et al. 2007; Walsh et al. 2007). This type of TC tracking could be regarded as an “automated” algorithm since it provides not only the track, but also the TCLV formation information if the tracks can be obtained in the early stage.

An automated tracking strategy is developed and applied in this study, and a fuzzy logic approach is also proposed to obtain the combined TCLV information from the model output. The potential operational forecast usefulness of the automated algorithm will be demonstrated with the output from the National Centers for Environmental Prediction Global Ensemble Forecast System (NCEP GEFS). Two advantages of the NCEP GEFS are its high temporal resolution (each 6 h) and long duration (16 days) forecasts for identifying the TCLV. For purposes such as hydrological or disaster preparedness operations, it would be useful to have such longer-term outlooks for the possible TC impacts. Moreover, the NCEP GEFS has 21 ensemble forecast members. These ensemble members may provide some uncertainty information (e.g., TC track spread) to assist in the forecast decision procedure (Froude 2010; Majumdar and Finocchio 2010; Yamaguchi and Majumdar 2010). For operational forecast purposes, once the target information (e.g., TC center) is objectively and efficiently extracted from the ensemble model outputs, the forecasters may utilize this information as another real-time forecast guidance product, especially when multiple models or ensemble forecast outputs are used.

The primary objective of this study is to demonstrate the detection and tracking algorithm as a tool for forecast guidance in the operational environment. The NCEP GEFS data are described in section 2, and the automated detection and tracking method is illustrated in section 3. Section 4 provides some evaluations of the application with the NCEP GEFS. The evaluation results of the predictability of named TC activity and also the forecast skill are described in section 5. Finally, a summary and discussion are given in section 6.

2. Data

This study uses the NCEP GEFS forecasts from June to December 2008. The NCEP GEFS output is available four times a day (0000, 0600, 1200, and 1800 UTC) at T126 resolution (1° × 1° latitude–longitude grid) with 28 vertical levels. The GEFS employs an ensemble transform method to create the initial perturbations (Wei et al. 2008). It has 21 ensemble members (i.e., 1 control run and 20 perturbed runs), and each ensemble member provides a forecast out to 16 days with 6-hourly output. More information on the operational configurations, documentation, and future implementations can be found.
on the NCEP ensemble Web site (http://www.emc.ncep.noaa.gov/gmb/ens/index.html). Although the domain in this study is the western North Pacific (WNP) basin from 0°–50°N to 100°E–180°, in principle the technique is applicable to other TC basins.

The 6-hourly Joint Typhoon Warning Center (JTWC) best-track data in 2008 were used for forecast evaluation. Following Ritchie and Holland (1999), Cheung (2004), and Lee et al. (2008), the formation time of each typhoon is defined when the maximum sustained surface wind speed reaches 25 kt (about 13 m s\(^{-1}\)). During June–December 2008, 21 named TCs formed in the WNP basin. Admittedly, this number of TCs is not adequate to establish the statistical reliability of the GEFS forecasts of TC formation. Rather, the objective here is a demonstration of the potential usefulness of the approach for operational forecasting of tropical cyclone formation with ensemble model output.

3. Methodology

Whereas the TCLV detection methodology in this section will be applied on the individual ensemble member forecast outputs, it could also be applied for multiple deterministic NWP models, which generally have higher horizontal resolution.

a. TCLV tracking

The TCLV tracking method used to obtain the TCLV tracks in the model forecast fields is patterned after previous studies (Vitart et al. 1997; Camargo and Zebiak 2002; Tsai et al. 2007). This tracking method is the type without a specification of first-guess positions. As described by Tsai et al. (2007), this algorithm is divided into two stages: (i) TCLV center detection and (ii) TCLV center tracking.

1) TCLV CENTER DETECTION

The TCLV center detection criteria used in this study are (i) minimum MSLP in a centered 7° × 7° latitude–longitude grid-box domain, which is taken as the storm center; (ii) 850-hPa relative vorticity (\(\zeta_{850}\)) \(\geq 3 \times 10^{-5} \text{ s}^{-1}\); (iii) maximum surface wind speed (10-m height wind; \(V_{10m}\)) in the centered 7° × 7° grid box \(\geq 10 \text{ m s}^{-1}\); (iv) 300-hPa air temperature (\(T_{300}\)) anomaly \(\geq 0\), where the anomaly is defined as the difference between the average over a 3° × 3° grid box and the average over a 7° × 7° grid box excluding the inner 3° × 3° grid box; (v) 300-hPa temperature anomaly \(\geq -850\)-hPa temperature (\(T_{300}\)) anomaly \(\geq 0\); and (vi) 850-hPa EKE average (\(\text{EKE}_{850}\)) \(-300\) hPa EKE average (\(\text{EKE}_{300}\)) \(\geq 0\), where the eddy kinetic energy (EKE; Shenoi et al. 1999) difference is over a 7° × 7° grid box. Instead of computing wind speed differences, the main purpose for using EKE in criterion vi is to reduce the influence of the background flow.

The TCLV center is defined at the position of minimum MSLP. Criteria ii – vi are used to determine whether it is a tropical cyclone or not. The TCLV center detections based on the first four criteria are denoted as the results from the “basic criteria” as they compose a simple criterion for warm-core feature detection, and those detections based on all six criteria are denoted as “all criteria” detections. The low centers found using the above detection criteria are considered to represent a single event by using the “nearest neighbor method.”

The distance between two detected low-MSLP centers must be less than two grid boxes over two consecutive 6-h time steps, which implicitly assumes that the TCLV is in the tropics where the translation speed would not exceed 0.47° h\(^{-1}\) (\(\approx 50 \text{ km h}^{-1}\)). Finally, a low-MSLP center detection is only considered to be a TCLV if its life span is at least 24 h (i.e., it exists for five consecutive forecast periods). The TCLV center detection procedure is only performed on NCEP GEFS grid points. The main purpose is not focused on model track accuracy, but on whether TCLVs are detected in the model output or not.

2) EXTENSION OF TCLV TRACKS

The TCLV tracks obtained from the ensemble model output as in section 3a(1) may be relatively short compared to the actual model forecast TC track because all of the detection criteria may not be met throughout the life cycle (Camargo and Zebiak 2002). In addition, global numerical models may have a limited capability to represent the TCLV structure due to low horizontal resolution. To overcome this deficiency during the early stages, the TCLV center tracking algorithm is extended as in Camargo and Zebiak (2002) with some modifications: (i) the time at which the MSLP is a minimum during the period of the track obtained by the first TCLV center detection procedure is denoted as the “mature stage” and becomes the starting time for the track extension and (ii) the TCLV centers are tracked backward and forward in time from the mature stage if the distance between two low-MSLP centers is less than two grid intervals over two consecutive 6-h time steps. The TCLV backward track extension procedure is stopped if the 850-hPa vorticity < \(2.0 \times 10^{-3} \text{ s}^{-1}\). Finally, the TCLVs with the same tracks are counted as a single one.

The TCLV center-tracking logic has a simple criterion to avoid misinterpreting the circulation as a tropical cyclone during extratropical transitions. When the standard deviation of \(T_{300}\) hPa over a 7° × 7° grid box (\(T_{\text{std,300}}\) hPa) is > 1.5 K during two consecutive time steps, the TCLV is
considered to have undergone extratropical transition. The justification for the extratropical transition criterion is that the TCLV is being tracked as a local minimum in MSLP. However, extratropical storms may reintensify during the extratropical transition (Klein et al. 2000; Ritchie and Elsberry 2007). Tsai et al. (2007) found that using the simple index (\(T_{\text{std,300 hPa}}\)) provides a useful distinction between TCs and extratropical storms, and may be used as an effective alternative to schemes such as those employed by Hart (2003), Evans and Hart (2003), and Guishard et al. (2009).

b. Fuzzy logic TCLV center detection

A fuzzy logic approach is applied to combine the TCLV center detection criteria at each grid point in the study domain. The distinguishing feature of fuzzy logic is a modification of the yes–no decision by using mathematical functions (so-called membership functions) to provide the likelihood of membership. User-specified membership functions and weights are applied to the detection criteria in section 3a(1) to produce a TCLV fuzzy combined-likelihood value (TFCV) for each grid box in each ensemble member output:

\[
\text{TFCV} = \sum_{i=1}^{n} \omega_i \cdot f_i(x_i),
\]

where \(\omega_i\) are the user-specified weights, \(x_i\) are the five criteria for TCLV center detection in section 3a(1), and \(f_i\) are the membership functions for \(x_i\). The weights and the membership functions used in this study are shown in Table 1 and Fig. 1, respectively. The weighting factors used here put more emphasis on the \(\xi_{850}\) and \(V_{sfc}\) recognitions, and the warm-core feature and TCLV structure identifications are regarded as auxiliary criteria. Similar applications utilizing fuzzy logic to combine selected information for meteorological forecasts can be found in the National Center for Atmospheric Research (NCAR) Auto-Nowcast System (Mueller et al. 2003).

The display of the TFCVs in each grid box is denoted as the TCLV fuzzy combined-likelihood map (TFCM).

<table>
<thead>
<tr>
<th>Detection criterion</th>
<th>Weighting factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\xi_{850})</td>
<td>Fuzzy logic all criteria: 0.30</td>
</tr>
<tr>
<td>(V_{sfc})</td>
<td>0.25</td>
</tr>
<tr>
<td>(T_{300}) anomaly</td>
<td>0.15</td>
</tr>
<tr>
<td>(T_{300}) anomaly (-) (T_{850}) anomaly</td>
<td>0.15</td>
</tr>
<tr>
<td>(EKE_{850}) (-) (EKE_{300})</td>
<td>0.15</td>
</tr>
</tbody>
</table>

Table 1. Weighting factors for fuzzy-based TCLV center detection algorithms.

FIG. 1. Membership functions applied to fuzzy logic TCLV center detection criteria for the (a) 850-hPa vorticity, (b) 10-m wind speed, and (c) \(T_{300}\) anomaly, \(T_{300}\) anomaly \(-\) \(T_{850}\) anomaly, and \(EKE_{850}\) \(-\) \(EKE_{300}\).
The TFCM represents the likelihood that a TCLV will form or pass over a specific region. Since the TCLVs represent moving weather systems with translation speeds of about 10–30 km h\(^{-1}\), the TFCMs can be jointly compared over several ensemble integrations. For example, several 6-hourly TFCMs can be combined to represent the “maximum TFCV” over each grid among all ensemble members during the forecast period (denoted as TFCM-max hereafter). If the ensemble members have significant TCLV signals and noticeable spread during the forecast period, the region with larger likelihood values in TFCM-max might indicate the TCLV activity and likely path. Other statistical values of the TFCV such as the average (TFCM-avg) and minimum (TFCM-min) values could also be used to produce similar TFCMs for operational forecasting purposes. It may be possible then to relax the original TCLV tracking restriction that a TCLV must continuously pass all of the detection criteria for at least 24 h. If the model output is too limited (e.g., output at 24-h intervals or longer) to implement the tracking procedure in section 3, these TFCMs may still have the potential to indicate the locations of TCLV centers in the model.

c. Fuzzy-based TCLV tracking

After applying the membership functions and weights as in Eq. (1) to produce combined-likelihood fields, a threshold criterion of 0.9 was determined to result in more reliable regions of TCLV activity. When TFCVs of the grid boxes pass this threshold value, and if the distance between two centers is less than two grid boxes in two consecutive 6-h time steps, these TFCVs are considered to be one event. If such an event continuously passes the threshold value for over 24 h, the TCLV tracking procedure in section 3a(2) can be applied to obtain the whole track, for which the results are denoted as “fuzzy logic basic criteria” or “fuzzy logic all criteria” tracking products.

4. Applications to NCEP GEFS

A TCLV monitoring system was established to provide three main products from the NCEP GEFS: (i) original

**FIG. 2.** Examples of TCLV tracks from the 0000 UTC 19 Aug 2008 NCEP GEFS output using the (a) fuzzy logic basic criteria, (b) basic criteria, (c) fuzzy logic all criteria, and (d) all criteria. The thick lines are the tracks from the control ensemble member, and the thin lines are from 20 perturbed runs. The filled circles denote the starting positions of the TCLVs, while the unfilled circles are the ending positions. The unfilled diamonds are considered to represent the extratropical cyclone stage. The TCLV starting and ending times (h) are also shown. The best track of TS14 is overlaid in Fig. 2d (dotted gray line).
TCLV tracking based on basic criteria and all criteria, (ii) fuzzy-logic TCLV tracking based on basic criteria and all criteria, and (iii) TFCM or TFCM-max.

After considerable testing during the early 2008 typhoon season, four tracking products seemed to provide the most useful guidance to the forecasters: (i) original TCLV tracking—all criteria, (ii) original TCLV tracking—basic criteria, (iii) fuzzy-based TCLV tracking—all criteria, and (iv) fuzzy-based TCLV tracking—basic criteria. A combination of these four tracking products provided a strategy for evaluating likely TCLV tracks. For example, Figs. 2a and 2b are the TCLV tracking products from the 0000 UTC 19 August 2008 NCEP GEFS output based on the fuzzy logic basic criteria and basic criteria. Notice that many extratropical cyclones are also being tracked due to some loose constraints (i.e., simple warm-core feature detection). The corresponding TCLV tracking products based on fuzzy logic all criteria and all criteria are shown in Figs. 2c and 2d, respectively. If all of the detection criteria are used (Fig. 2d), only the most dominant track (i.e., Typhoon Nuri) is strongly represented in the tracking product, and no other model tracks were present.

The TFCM-max for the WNP basin from the NCEP GEFS forecast on 0000 UTC 19 August 2008 is shown in Fig. 3. The 16-day forecast outputs are grouped into 4-day intervals. The areas with large values of TFCV indicate regions with enhanced potential for TCLV occurrences. Notice that the TFCV values are greater than 0.9 around the Philippines during the 0–96-h period in Fig. 3 and the TCLV track resembles the path of Typhoon Nuri. The best track of Typhoon Nuri (top left; gray line) and TS14 (top right and bottom left; black lines) are overlaid.

FIG. 3. The TFCM-max during four forecast intervals (labeled to the top left) from the 0000 UTC 19 Aug 2008 NCEP GEFS output. The dark fuzzy TFCM grid boxes around southern Taiwan with TFCVs > 0.9 represent the path of Typhoon Nuri during the 0–96-h period. The best track of Typhoon Nuri (top left; gray line) and TS14 (top right and bottom left; black lines) are overlaid.
pattern of movement toward the south of Taiwan, the best track still lies within the range of ensemble spread at most of the times.

In the NCEP GEFS forecast on 0000 UTC 19 August 2008 (Figs. 2c, 2d, and 3), no significant development was expected during 102–384 h (i.e., 23 August–4 September). However, a weak tropical storm (TS) formed over the east of the Philippines on 0600 UTC 26 August (denoted as TS14; shown in Fig. 5a). It was maintained for just 2.5 days, and its maximum intensity was 35 kt (about 18 m s\(^{-1}\)). The ensemble forecasts during 22–24 August show that only some lows and disturbances are located around the east of the Philippines. However, no significant development can be found until 25 August. These

![Diagram](http://example.com/diagram.png)

**Fig. 4.** (top) As in Fig. 3, but for the Taiwan area from the NCEP GEFS for 0000 UTC 19 Aug 2008. (bottom) Time series of the maximum TFCVs in the Taiwan area for each ensemble member, where values >0.9 indicate higher likelihood of a TCLV.

![Diagram](http://example.com/diagram.png)

**Fig. 5.** (a) Best-track data for the named TCs during June–December 2008 with formation locations and (b) track count (see inset) within a 1° × 1° grid domain.
FIG. 6. The NCEP GEFS TCLV track count within 1° × 1° grid boxes during four forecast intervals (labeled to the top left) obtained by the (a) fuzzy logic basic criteria, (b) basic criteria, (c) fuzzy logic all criteria, and (d) all criteria. The NCEP GEFS forecast data used in this study are from June–December 2008.
results suggest that weak and short-lived storms may be hard to predict at long lead times in the GEFS. More detailed evaluations of forecast skill are discussed in the next section.

Similar TFCM displays were established for specific regions (e.g., Taiwan and Guam). An example of such a regional TFCM for the Taiwan area is given in Fig. 4. The higher fuzzy TFCM values along the southern
boundary in the first 96 h (Fig. 4, top left) correspond with the occurrence of Typhoon Nuri. Later in the forecast period, the greater threat is to the north and northeast of Taiwan. The temporal variations (maximum value within the domain) among the 21 ensemble members are also displayed as a time series in Fig. 4 (bottom). For instance, the confidence in the ensemble forecasts of a TCLV in the Taiwan area is enhanced if large values are forecast by most of the ensemble members. Conversely, one might expect that no activity, or at least a reduced level of activity, will occur in the 72–96-h period, during which none of the 21 members have forecast a TCLV (Fig. 4, bottom).

5. Predictability of western North Pacific typhoons in NCEP GEFS during 2008

a. TCLV in the NCEP GEFS

The predictability of named TC occurrences (i.e., whole track; shown in Fig. 5a) during June–December 2008 in the NCEP GEFS is evaluated in this section. The occurrences are summarized as a best-track count within a $1^\circ \times 1^\circ$ grid domain (Fig. 5b). The corresponding tracks count plots based on the four tracking algorithms during the four forecast intervals in the 2008 season are given in Figs. 6a–d. Except for the track spread produced by the ensemble members, the similarity between the track count patterns obtained from the model output (Fig. 6) and from the best-track data (Fig. 5b) should generally be higher when the model forecast skill is better.

In this 2008 sample, almost no developments can be found at low latitudes east of 150°E in the best-track data (Fig. 5). The NCEP GEFS forecasts (Figs. 6a–d) can fairly well predict this aspect of the 2008 season. However, the track count patterns in the fuzzy logic basic criteria (Fig. 6a) and basic criteria (Fig. 6b) are too widespread because of the lack of strict warm-core detection criteria, and many cyclones that are not TCs are detected with these criteria. These results also imply that had even simpler TCLV detection criteria such as just using maximum wind or minimum SLP been used, the TCLV track count patterns would be even more widespread than in Figs. 6a and 6b. When the warm-core detection criteria are also required (Figs. 6c and 6d), more significant tropical cases are identified, and mid-latitude cyclone activity is reduced. The distinctions among these track patterns may indicate that NCEP GEFS does simulate the differences between the structural and thermodynamic characteristics of tropical cyclones and extratropical cyclones.

Notice that the track count patterns in Figs. 6c and 6d are comparable to the best-track pattern (Fig. 5b) only during the 0–96-h period and somewhat less so for the 102–192-h forecasts. The correlation coefficients ($\rho$) between Figs. 5b and 6c during the 0–96- and 102–192-h periods are 0.48 and 0.19, respectively. The $\rho$ values between Figs. 5b and 6d during the 0–96- and 102–192-h periods are slightly increased to 0.50 and 0.20. The absence of agreement with the best-track pattern for the 198–288- and 294–384-h forecast intervals suggests a lack of predictability in TC occurrences by the NCEP GEFS in these forecast intervals. Froude (2009) indicated that the poor predictability might be due to a wider spread in the ensemble model at the longer lead times. It could lead to a general reduction in forecast skill, especially when tropical cyclones move into the midlatitudes.
b. Named TC formation forecast evaluation

The predictability of the TC formation is discussed in this section. The average ensemble track numbers within a $5^\circ \times 5^\circ$ box surrounding the formation locations for all 21 named TCs at the precise JTWC formation times (as defined in section 2; the maximum sustained surface wind speed reaches 25 kt) during four forecast lead-time intervals are shown in Fig. 7a. Relatively large values are predicted only during the first 4-day forecast interval. Notice that the average TCLV track numbers for all of the four criteria are less than 1.0 during the 102–384-h forecast intervals. A similar result is found even if (i) a 4-day tolerance of the formation time error is allowed or (ii) the domain is expanded to $7^\circ \times 7^\circ$ or $9^\circ \times 9^\circ$, except for a slight increase of the TCLV track number average (figure not shown). The track count patterns in Fig. 6 are widespread, but the average track numbers near the actual TC formation locations for all of the four criteria are small (Fig. 7). These results suggest that TCs were generally overpredicted in the GEFS (Fig. 6), but underpredicted near the actual TC formation locations (Fig. 7).

To evaluate the effects of the maximum intensity of the TC during these forecasts, these 21 named TCs are then divided into two categories of maximum intensity ($V_{\text{max}}$): 1) strong TC, with $V_{\text{max}} > 50$ kt ($\sim 26$ m s$^{-1}$), and 2) weak TC, with $V_{\text{max}} \leq 50$ kt. The evaluations for the 10 strong TCs and 11 weak TCs during the study period are shown in Figs. 7b and 7c with the average of ensemble track numbers around the formation locations. On average, the storms that eventually become the stronger TCs (Fig. 7b) have larger ensemble track numbers during the first forecast interval than the weaker TCs (Fig. 7c). At least for this small sample during the 2008 season, no significant signals from any of the four detection criteria exist for longer than 102-h forecast lead times for either category of maximum intensity of TC when the forecast location is required to be within a $5^\circ \times 5^\circ$ grid box at the JTWC formation time.

For the fuzzy detection criterion, the TFCM-max averages within a $5^\circ \times 5^\circ$ box surrounding the formation locations are larger for the stronger TCs than for the weaker TCs for all four forecast intervals (Fig. 8). The difference is significant at the 95% confidence level according to the signed-rank test (Maidment 1993). As one example, consider the number of ensemble tracks within the $5^\circ \times 5^\circ$ grid domain of the formation locations according to JTWC for Jangmi and Haishen at their formation times (Fig. 9).

For the fuzzy detection criterion, the TFCM-max averages within a $5^\circ \times 5^\circ$ box surrounding the formation locations are larger for the stronger TCs than for the weaker TCs for all four forecast intervals (Fig. 8). The difference is significant at the 95% confidence level according to the signed-rank test (Maidment 1993). As one example, consider the number of ensemble tracks within the $5^\circ \times 5^\circ$ grid domain of the formation locations according to JTWC for Jangmi and Haishen at their formation times. The average of ensemble track numbers for the strong TCs is $140$ kt ($\sim 72$ m s$^{-1}$) and $40$ kt ($\sim 21$ m s$^{-1}$), in Figs. 9a and 9b. The average of ensemble track numbers for the strong TCs is generally larger than that for Haishen for all four detection criteria for forecast lead times within 96 h. For lead times longer than 96 h, no significant signals are found even for Supertyphoon Jangmi. However, the number of ensemble tracks during
the 48–72-h forecasts of Haishen (Fig. 9b) is larger than that of Jangmi (Fig. 9a). By comparing with the satellite images, Haishen formed from a gradually developed tropical disturbance. In contrast, GEFS did not predict the rapid development of Jangmi. Thus, GEFS could have more ensemble tracks in the 48–72-h forecasts of Haishen.

The areal averages of the TFCM-max values around the formation location of Typhoon Jangmi range from 0.3 to 0.7 during the 96–384-h forecast intervals (Fig. 10). Thus, some of the ensemble members do produce low pressure centers around the Jangmi formation position. However, these model-produced vortices may be either not sustained long enough (at least 24 h) or be sufficiently well developed to be tracked directly and objectively. By contrast, a strong signal in TFCM-max is present for only about 72 h prior to the formation of Haishen. Notice that there are four 0 TFCM-max values within a $5^\circ \times 5^\circ$ box surrounding the formation location of Typhoon Jangmi during 96–288 h, which implies that no minimum pressure centers are predicted within a $5^\circ \times 5^\circ$ box surrounding the formation location. If the domain is expanded from $5^\circ \times 5^\circ$ to $7^\circ \times 7^\circ$, these values increase to about 0.4. This result represents the limited predictability and also the variability of the ensemble model forecasts.

c. Named TC activity forecast skill test

In the western North Pacific, a TC is assigned a name by the Regional Specialized Meteorological Center-Tokyo Typhoon Center once its intensity reaches the tropical storm stage. The assessment of the intensity is based on various kinds of information, such as satellite observations, aircraft observations, and high-resolution objective analyses. However, numerical model outputs may not detect the precise timing of the formation due to inadequate initial conditions, physical process representations, or the model spatial resolution. Consequently, the forecast performance of the NCEP GEFS is evaluated for the named TC activity (occurrence) in this section.

To assess the NCEP GEFS forecast skill, and to establish an objective TC activity forecast index for the WNP basin (Fig. 11), the 16-day forecasts are grouped into 4-day intervals. An “observed event” is defined such that at least one named TC (based on best-track data; see section 2) exists over the WNP basin during the 4-day forecast intervals. A “forecast event” is defined when the ensemble TCLV track numbers within a predetermined size of domain ($N_d \times N_d$) in the WNP basin exceed a specified threshold number $N_{\text{thres}}$. The contingency table of observed and forecast events based on the above definition is used to obtain the threat score, probability of detection (POD), frequency bias (FB), and false alarm rate (FAR). A perfect forecast would result in the following values: threat score $= 1$, POD $= 1$, FAR $= 0$, and FB $= 1$.

As an example, $N_d = 3^\circ$ (i.e., $3^\circ \times 3^\circ$ domain) is selected and the threshold numbers $N_{\text{thres}}$ for the four tracking algorithms are chosen as (i) at least five TCLV ensemble tracks for the basic criteria and fuzzy logic basic criteria and (ii) at least three tracks for all criteria and fuzzy logic all criteria. For these specifications, the NCEP GEFS forecast performance in predicting named TC activity during June–December 2008 is shown in Fig. 12. As expected, the threat score, POD, and FB values decrease with increasing forecast lead time. The threat score values for the four tracking algorithms range from about 0.35 to 0.6 during the first interval (0–96 h), and decrease to below 0.3 in the 294–384-h interval. Although the POD values for the first two periods are relatively high, the FB values depart from the desired 1.0 value, and the FAR values are also high. For all four tracking
algorithms, the threat score and POD values could be higher with the specification of less stringent criteria. However, the FAR values would also be larger, so the forecaster is faced with a trade-off.

The WNP domain is then divided into five subregions (Fig. 11), although regions 3 and 5 will not be discussed in detail because of the small samples. The performance measures (Fig. 13) are generally similar to the entire domain during the 0–96-h forecast interval. However, the skill is decreased in the longer forecast intervals, especially for region 2. However, the FAR values are also comparatively smaller in region 2, which has the largest fraction of the 21 TCs during 2008. In addition, large variations in the FB and FAR values for the four tracking algorithms are found in region 4 (Fig. 13c) with significant differences when strict warm-core detection criteria are considered. In this region, many extratropical cyclones existed that may have similar features as a TC except for the warm core (i.e., low MSLP, large vorticity, and high 10-m wind speeds). This result was also revealed in Figs. 6a–d since the midlatitude cyclone activity was reduced significantly when warm-core detection criteria were used. The threat scores for TC activity forecasts in region 1 during 0–96 h (about 0.25 to 0.38) is not as high as in other regions (0.4 to 0.6). These results reflect that the GEFS might overintensify some tropical disturbances in region 1.

Sensitivity tests were carried out with other definitions of the “forecast event”: (i) expanding the $N_d \times N_{\text{thres}}$ domain from $3^\circ \times 3^\circ$ to $5^\circ \times 5^\circ$, and to $7^\circ \times 7^\circ$; (ii) varying the threshold number $N_{\text{thres}}$; and (iii) using TCFM-max average (e.g., section 5b) instead of the ensemble track numbers. The first two test results were quite similar to using the $N_d$ and $N_{\text{thres}}$ values as in Figs. 12 and 13. Based on analyses of the ensemble track count patterns (Fig. 6) and the TC formation predictability (Fig. 7), a large $N_{\text{thres}}$ value is not suitable for evaluation. Also, the ensemble spread could affect the predictability analysis since the TC activity will be noticeably reduced if the ensemble spread is large. Thus, the threshold number was varied from 2 to 6. Applying the TFCM-max average as the “forecast event” decision rule (e.g., TFCM-max average within the $3^\circ \times 3^\circ$ domain). $0.85$ resulted in threat score values that increased in the latter three forecast intervals (102–384 h) from below 0.3 to about 0.5. However, the FB values increased from below 1.0 to 1.5–1.7 (i.e., overforecast), and the false alarms were higher. Also, the evaluation result is similar even if the TFCM-avg is used.

Fig. 12. Measures of TC activity forecast skill during June–December 2008 over the entire WNP basin based on four tracking algorithms (see inset) during four forecast intervals: (a) threat score, (b) POD, (c) frequency bias, and (d) FAR.
FIG. 13. As in Fig. 12, but for subregions (a) 1, (b) 2, and (c) 4, as defined in Fig. 11.
6. Summary and discussion

An automated detection and tracking system has been developed for TCLVs in NWP models, and especially for ensemble-based NWP models. This system has been applied to efficiently extract western North Pacific TCLV information based on the operational forecast guidance from NCEP GEFS during June–December 2008. A TCLV is detected in the model grid output when selected dynamic and thermodynamic fields meet specified criteria. A fuzzy-based algorithm has also been developed to combine the detection criteria without having discrete yes–no thresholds.

The TCLV track numbers and TFCM-max average within a $5^\circ \times 5^\circ$ box surrounding the formation locations have higher skill during the 0–96-h forecast interval than for the 102–384-h forecasts. Those TCs that eventually have maximum intensities $>50$ kt have larger TFCV values for the entire 16-day forecast period than for weaker TCs. A forecast skill test of named TC activity in which the ensemble track numbers are used as the objective guidance shows that the threat scores are about 0.35–0.6 for the 0–96-h forecast interval, and then decrease to below 0.3 in the 294–384-h forecast interval. For all four tracking algorithms, the threat score and probability of detection could be higher if the detection algorithm had weaker criteria, but the false alarms would also be larger due to frequency bias increase. The threat score values could be increased to about 0.5 during the 102–384-h forecast interval if the TFCM-max average is used as the objective “forecast event” guidance. Again, larger forecast biases can be found, and the false alarms would be also higher.

This evaluation indicates that it is not easy to objectively identify the precise timing and locations within $5^\circ \times 5^\circ$ of TC formations directly in the NCEP GEFS for longer lead times. Nevertheless, some indication is given that the NCEP GEFS may be simulating the general state of the atmospheric environment at long lead times that is favorable for the stronger typhoon developments. Therefore, statistical approaches that use environmental parameters to predict the TC formation or estimate the formation probability (e.g., DeMaria et al. 2001; Hennon and Hobgood 2003; Schumacher et al. 2009) may be extended by inserting the long-lead NWP model outputs to provide alternate objective guidance for typhoon formation and activity forecasts.

Based on the above results and considerations, the detection and tracking algorithm developed in this study provides an objective tool for identifying and evaluating TC formation and activity forecasts in ensemble prediction models. The objective TCLV tracking and TFCV...
maps could be useful evaluation tools for the operational forecasts. The TC forecast performance still relies on the NWP model. In this study, it is shown that the false alarm of TC activity forecast by NCEP GEFS during the 2008 typhoon season is large. In the future, the ensemble forecast outputs should be accumulated to have an adequate dataset to further assess the usefulness and reliability of the uncertainty information in the GEFS.

Acknowledgments. The authors thank Yen-Ting Lai (National Central University), Chia-Chi Liu (Taiwan Typhoon and Flood Research Institute), Mong-Yang Lee (Central Weather Bureau), Patrick O’Reilly (NCEP), Yuejian Zhu (NCEP), and Doug Schuster (NCAR) for supporting the NCEP GEFS data download processes. The authors would also like to thank Dr. Hua-Lu Pan (NCEP) for the valuable discussions and the encouragement.

REFERENCES


Camargo, S. J., and S. E. Zebiak, 2002: Improving the detection and prediction of the uncertainty information in the GEFS.


——, 2004: Large-scale environmental parameters associated with tropical cyclone formations in the western North Pacific. J. Climate, 17, 466–484.


ECMWF, 2004: Early medium-range forecasts of tropical cyclones. ECMWF Newsletter, No. 102, ECMWF, Reading, United Kingdom, 7–14.


