Two-Dimensional Retrieval of Typhoon Tracks from an Ensemble of Multimodel Outputs

TRAN TAN TIEN, CONG THANH, HOANG THANH VAN, AND KIEU QUOC CHANH

Laboratory of Weather and Climate Research, Hanoi College of Science, Vietnam National University, Hanoi, Vietnam

(Manuscript received 5 June 2011, in final form 28 September 2011)

ABSTRACT

In this study a method of retrieving optimum information of typhoon tracks in a multimodel ensemble of forecasts is explored. By treating the latitudes and longitudes of typhoon centers as components of two-dimensional track vectors and using the full ensemble mean as a first guess, it is shown that such a two-dimensional approach for the typhoon track forecast can be formulated as a multivariate optimization problem. Experiments with five nonhydrostatic primitive equation models during the 2004–08 typhoon seasons in the western North Pacific basin show some noticeable improvements in the forecasts of typhoon tracks in terms of the forecast errors and track smoothness with this multivariate approach. The advantages of the multivariate optimization approach are its portability and simplicity, which could make it easily adaptable to any operational typhoon forecast center that synthesizes typhoon track forecast products from different sources.

1. Introduction

Accurate typhoon (TY) track forecasting is a challenging problem due to the existence of unpredictable components associated with large-scale flows and multiscale interactions of the TYs with the background environment (e.g., Lander and Holland 1993; Wu and Emanuel 1993; Simpson et al. 1997; Ritchie and Holland 1997). While there has been some steady improvement in hurricane track forecasts in the Atlantic basin during recent decades (e.g., Brown et al. 2010), the TY track forecasts in the western North Pacific (WPAC) have some issues to resolve (Goerss et al. 2004; Kehoe et al. 2007; Payne et al. 2007). Complicated multiscale interactions of the environmental flows with nearby topography and frequent direct vortex–vortex interaction in the WPAC often result in irregular TY paths (Cheung and Chan 1999; Payne et al. 2007; Liu and Chan 2008; Yang et al. 2008). An example of Typhoon Nari (2008) with a three-loop track to the south of Taiwan together with its multiple intensification phases demonstrates typically unusual patterns of TY behavior in the WPAC (e.g., Huang et al. 2005; Yang et al. 2008). Another example is Supertyphoon Parma (2009), whose track crossed the northern tip of the Philippine several times before making a southern turn and had its final landfall in northern Vietnam. (More detailed reports on Typhoon Parma can be found online: http://www.nasa.gov/mission_pages/hurricanes/archives/2009/h2009_Parma.html.)

These examples highlight the complex TY movement in the WPAC, where a single deterministic model forecast may not capture the most likely TY track. Various modeling factors contribute to uncertainties in the numerical TY track forecasts, including model initialization, boundary conditions, model resolution, or parameterization schemes. As a result, the ensemble approach with either multiple models or a set of optimized initial conditions for a model becomes a growing trend in operational TY track forecasts. Indeed, the ensemble technique has long been employed in the hurricane track forecasts in the past because the spread of the ensemble could allow for an estimation of the forecast reliability in addition to the track forecast (e.g., Neumann and Pelissier 1981; Goerss 2000).

Formally, one can divide the production of ensemble TY track forecasts into two stages. The first is to create an ensemble of forecasts, and the second is to draw from the ensemble the most information about TY movement.

Corresponding author address: Dr. Chanh Kieu, Laboratory of Weather and Climate Research, Hanoi College of Science, Vietnam National University, 334 Nguyen Trai, Thanh Xuan, Hanoi 10000, Vietnam.
E-mail: chanhkq@vnu.edu.vn

DOI: 10.1175/WAF-D-11-00068.1

In the first stage, two approaches are generally used for creating the ensemble. The first approach is based on a set of initial conditions generated by an ensemble system such as the ensemble assimilation Kalman filter or the breeding method (see, e.g., Toth and Kalnay 1997; Evensen 1994; Krishnamurti et al. 1997). The second approach relies on an ensemble of different models, sometimes referred to as the superensemble or multimodel approach. In the second stage, the most common method is to simply take a full ensemble mean. The advantage of this ensemble mean approach is not only its simplicity but also that, statistically, the ensemble mean has a standard deviation that decreases as a square root of the number of the ensemble members (Daley 1992). Goerss (2000) showed that the simple ensemble average (or consensus forecast) derived from a combination of several global and regional operational models could indeed help improve the quality of TY track forecasts at all lead times. As will be shown in section 4, the simple ensemble mean, nevertheless, quite often results in crisscross movements due to the independent averaging of the latitudes and longitudes of TY centers, especially when the number of ensemble members is small.

Other TY track forecast treatments that have been proposed specifically for TY track forecasts include the selective mean (Payne et al. 2007), the cluster mean (Zhang and Krishnamurti 1997), and the ensemble linear regression (Leslie and Fraedrich 1990). Although each has its own advantages, the recent study by Payne et al. (2007) appeared to show that the selective mean tends to have the best performance. In a different manner, Weber (2003) presented an effective "downhill" method for estimating hurricane centers based on information related to hurricane structure, location, and motion in the season preceding the forecast period. Using the downhill technique, the hurricane centers can be estimated as the locations with the maximum probability. While this probability approach is systematic, it requires substantial amounts of information, such as maximum surface wind, radius of maximum wind, radius of the outermost closed isobar, or the radii of 35- and 50-kt winds, that are not always available operationally in the WPAC, especially at the warning centers for developing countries within the WPAC region [winds of 35 and 50 knots (kt) are equivalent to 18.0 and 25.7 m s^{-1}, respectively, where 1 kt = 0.514 m s^{-1}].

Given the current uncertainties in our understanding of TY dynamics and a wide range of TY models, a question of interest is how to retrieve the maximum information of TY movement out of the ensemble outputs. Unlike the ensemble forecast of scalar quantities such as temperature or humidity for which a single regressive equation could give the best estimation, the ensemble TY track forecast requires a simultaneous prediction of both the latitude and longitude of a TY center. Because forecasting models tend to have a systematic bias, some cross correlation between the latitude and longitude errors of the forecasted TY centers always exists. Thus, the simple ensemble mean of either latitude or longitude separately appears to discard a significant piece of information. In addition, each model has its own accuracy for a certain forecast range and region. In this study, an ensemble method will be presented for the TY track forecast that can retrieve the maximum information of TY movement from an ensemble of model outputs. The main objective is to take into account the cross correlation of the latitudinal and longitudinal forecast errors such that the ensemble track forecast will have the fewest errors at each forecast range.

In the next section, detailed formulation for the application of multivariate optimization to the TY track forecast will be presented. Sections 3 and 4 describe the model configuration as well as the dataset used in our ensemble experiments. In section 5, comparisons between different ensemble mean approaches and the multivariate optimization will be discussed. Some discussion and our conclusions are given in the final section.

2. Methodology

As mentioned above, the forecasting of TY tracks differs from that of scalar variables since both the latitudes and longitudes of TY centers are required at the same time instead of a single value. If the latitudinal and longitudinal forecast errors of the TY centers are statistically independent, it would be reasonable to optimize the latitudes and longitudes separately using the linear regression approach (e.g., Leslie and Fraedrich 1990). In reality, the latitudes and longitudes always have some cross correlation since TYs move with the environmental steering flow. Therefore, a two-dimensional track vector needs to be simultaneously forecasted. Given a set of ensemble track forecasts from different models, predicting the track thus becomes a multivariate optimization problem. Note that some models have more forecast skill than others at short forecast lead times but may have lower skill at the longer ranges. As a result, it would be desirable to take into account such time-dependent accuracies of different models in the ensemble forecast. Assume a set of M models has provided the center positions of a TY at forecast times of 12, 24, 36, 48, 60, and 72 h. A TY center position vector of the ith model is defined as $\mathbf{v}_i \equiv (x_i, y_i)$, where x_i and y_i are the longitude and latitude, respectively. A simple ensemble mean at each forecast time T is first calculated as follows:

$$\overline{\mathbf{v}}(T) = \frac{\sum_{i=1}^{M} \mathbf{v}_i(T)}{M}, \qquad (1)$$

where M is the number of models. This ensemble mean is taken to be the first guess of the center position at time T. The advantage of using such an ensemble mean as the first guess of the TY location is that the mean value is the first-moment approximation that tends to converge to an expected value when the number of ensemble members is large (Goerss 2000). The objective of this study is to improve this first guess by adding some further corrections, based on the accuracy of each individual model at each lead time T, as follows:

$$\mathbf{v}(T) = \overline{\mathbf{v}}(T) + \sum_{i=1}^{M} \mathbf{W}_i(T)[\mathbf{v}_i(T) - \overline{\mathbf{v}}(T)], \qquad (2)$$

where $\mathbf{W}_i(T)$ is the 2×2 matrix specifying the weight of the each model that varies with forecast lead time T. The time dependence of these weight matrices is expected as each model has its own skill at different forecast ranges. This approach differs from the Leslie and Fraedrich (1990) regression equations in that the regressive linear combinations of the latitudes and longitudes are not calculated directly, as this will occasionally cause the track to have discontinuous jumps. Furthermore, the multivariate optimization ensures that the corrected track will not greatly deviate from the ensemble mean value, which is desired when the number of ensemble models is large enough. The aim now is to find the corrections to the ensemble mean in (2) such that information from the best models can be taken into account as much as possible.

Finding the weight matrix $\mathbf{W}_i(T)$ is essentially the multivariate linear regression problem (Daley 1992), in which the weight matrix $\mathbf{W}_i$ can be obtained by minimizing the residual error vector $\mathbf{e}(t)$ defined as

$$\mathbf{e}(T) = \mathbf{v}_t(T) - \overline{\mathbf{v}}(T) - \sum_{i=1}^{M} \mathbf{W}_i(T)[\mathbf{v}_i(T) - \overline{\mathbf{v}}(T)], \qquad (3)$$

where $\mathbf{v}_t(T)$ denotes the TY best-track vectors from previous observations. Standard multivariate minimization calculations with the residual error vectors $\mathbf{e}(T)$ given by (3) lead to

$$\mathbf{W}_i(T) = \mathbf{B}(T)[\mathbf{R}_i(T) + \mathbf{B}(T)]^{-1}, \qquad (4)$$

where $\mathbf{B}(T)$ is the error covariance matrix for the ensemble mean and $\mathbf{R}_i(T)$ is for each member i of the ensemble, which are, respectively, given by

$$\mathbf{B}(\mathrm{T}) = E\{[\overline{\mathbf{v}}(T) - \mathbf{v}_t][\overline{\mathbf{v}}(T) - \mathbf{v}_t]^{\mathrm{T}}\} \quad \text{and} \qquad (5)$$

$$\mathbf{R}_i(T) = E\{[\mathbf{v}_i(T) - \mathbf{v}_t(T)][\mathbf{v}_i(T) - \mathbf{v}_t(T)]^{\mathrm{T}}\}. \qquad (6)$$

Here, $E\{\cdot\}$ denotes the statistical expectation operator. Given a set of the track forecasts generated by all models and the corresponding best-track data, the error covariance matrices $\mathbf{B}(T)$ and $\mathbf{R}_i(T)$ can be calculated during the training period. Since the performance of each model varies with forecast seasons and basins, one should in principle construct these error covariance matrices for different months and different basins. Due to the shortage of data and computational resources in this study, the error covariance matrices $\mathbf{B}(T)$ and $\mathbf{R}_i(T)$ are assumed to be constant for the entire forecasting season. In addition, the application here is limited to TYs in the South China Sea, which is the domain of interest due to their direct impacts on Vietnamese coastal areas.

For comparisons of the impact of the different ensemble means, three ensemble mean approaches mentioned in section 1 are also examined: 1) the full ensemble mean, in which the simple mean of all members is computed; 2) the cluster ensemble mean, in which a cluster analysis is performed and the cluster with the largest number of members is selected; and 3) a selective mean, in which the outliers of the ensemble track are excluded in calculating the mean track, which is similar to the selective consensus method in Payne et al. (2007). Note that in the third approach, only model track vectors in the 12-h forecast that have the largest deviation from the ensemble mean are omitted.

3. Model

In this study five nonhydrostatic primitive equation models are used to generate a set of track forecasts for TYs in the past. The models include the Weather Research and Forecasting Model (WRF, version 2.1.2), the Eta Model [International Centre for Theoretical Physics (ICTP) version 2005], the fifth-generation Pennsylvania State University–National Center for Atmospheric Research (Penn State–NCAR) Mesoscale Model (MM5), the Regional Atmospheric Modeling System Model (RAMS), and another version of RAMS in which a bogussed vortex initialization is activated when a depression can be identified (RAMB).

All models are initialized with Global Forecast System (GFS) initial conditions with a resolution of 1° × 1°, available from the National Centers of Environmental Prediction (NCEP), and with lateral boundary conditions updated every 6 h. The models are configured with a single domain of 161 × 161 grid points, 26 vertical levels, and resolution of 28 km that is centered at (15°N, 110°E).

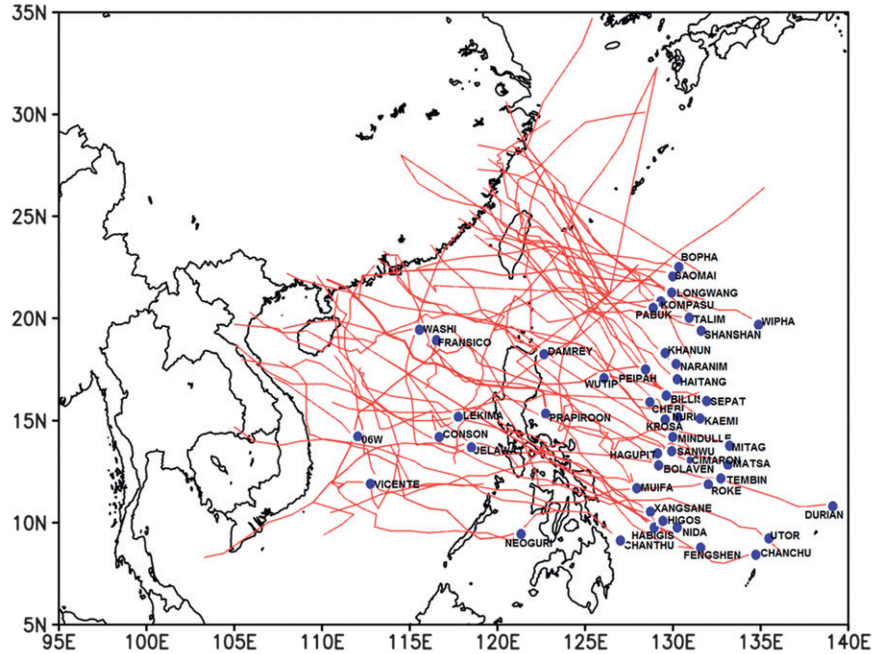

FIG. 1. Model domain for the ensemble experiments during the 2004–08 western North Pacific typhoon seasons together with the best tracks of the typhoons. Blue dots indicate the first GFS initial position of each typhoon.

This horizontal domain covers the entire Vietnam and South China Sea area with northern and southern boundaries at $-5°$S and 35°N, which is large enough to include part of the equatorial circulation, the Siberian high, and the western North Pacific subtropical high (Fig. 1). This domain has been selected based on numerous observational and modeling studies to include those synoptic-scale systems that influence the TY movement in the desired area (Fig. 1).

The WRF, MM5, RAMS, and RAMB's physical schemes used in this study are (i) the Kain and Fritsch (1990) cumulus parameterization scheme in which deep convection and a broad range of shallow convection patterns are both parameterized, (ii) the Yonsei University planetary boundary layer (PBL) parameterization with the Monin–Obukhov surface layer scheme, (iii) the Rapid Radiative Transfer Model (RRTM) scheme for both longwave and shortwave radiation with six molecular species (Mlawer et al. 1997), and (iv) the Lin et al. (1983) cloud microphysics scheme with six classes of hydrometeors, namely, water vapor, cloud water, rain, snow, graupel, and cloud ice. For the Eta Model, the default longwave scheme (Lacis and Hansen 1974), the shortwave scheme (Fels and Schwarzkopf 1975), and the Ferrier et al. (2002) microphysics scheme are chosen. See Table 1 for a summary of each model's configuration.

Although it would be desirable to use as diverse a set of different physical parameterizations as possible such that the ensemble would include a larger spread of TY movement, the forecast track difference due to various treatments of the dynamical core, finite-difference grids, surface layers, or boundaries among ensemble models is in practice sufficiently large that no attempt at using multiple physics is employed. Despite using the same physical parameterization schemes, the spectrum of the TY intensities and tracks from different models is believed to be wide enough to ensure the statistics of the TY movement, as will be seen in section 5. It should be

TABLE 1. Summary of the models and their physical parameterizations used in the ensemble forecasts.

Model	Microphysics	Surface	PBL	Cumulus	Radiation
WRF	Lin et al.	Noah	YSU	Kain–Fritsch	RRTM
MM5	Lin et al.	Noah	YSU	Kain–Fritsch	RRTM
Eta	Ferrier et al.	Noah	Mellor–Yamada 2.5	Kain–Fritsch	Lacis and Hansen 1974
RAMS	Lin et al.	Noah	YSU	Kain–Fritsch	RRTM
RAMB	Lin et al.	Noah	YSU	Kain–Fritsch	RRTM

mentioned also that the use of the above five regional models is not based on their previous qualified performances in TY track forecasts. Instead, these models are needed simply to create a dataset of TY track forecasts that can be used to examine the performance of different ensemble mean methods. As a result, no attempt has been made to optimize the model physical parameterization schemes, tuning parameters, or configuration in these models for the track forecasts.

4. Data

To obtain the error covariance matrices $\mathbf{B}(T)$ for the first guess (i.e., ensemble mean) and $\mathbf{R}_i(T)$ for each member at each forecast time T, experiments were conducted with all five models for a set of TYs for which the best-track data were available. To generate a number of history cases that is as large as possible within our computational capability and the data available, each model was integrated for $K = 52$ typhoons during the 2004–08 WPAC typhoon seasons (see Table 2 and Fig. 1 for the initial locations of the TYs). Note that these typhoons are only a subset of the total number of TYs in the WPAC during this period. The reason for choosing these TYs is simply because they had a history of significant impacts on Vietnam, and their best-track data for the model integrations were available.

For each typhoon i and model m, $L_{im}(T)$ T-hourly (T-h) forecasts are made. Each T-h forecast is defined as a model integration with an updated GFS initial condition for T hours. For example, a 72-h forecast for Typhoon Damrey (2005) starting at 0000 UTC 18 September 2005, and a 72-h forecast for the same typhoon starting at 0000 UTC 19 September 2005, are considered as two different forecasts (hereafter referred to as cases). To avoid a potential serial correlation between two consecutive GFS initializations for one specific typhoon (Aberson and DeMaria 1994), only initializations separated at least 12 h apart are counted as different cases (due to different environmental conditions and timing, forecasts for different typhoons are always considered to be different cases). So, the total number of cases generated by the model m at the lead time T is $N_m(T) = \sum_{i=1}^{K} L_{im}(T)$. Note further that each typhoon has its own lifetime that may or may not be as long as T-h for the forecast lead time T. In addition, not all models could make a successful T-h forecast as a TY might have dissipated before T hours due to the model's own dynamics and physics. Thus, it is not always possible to make an entire T-h forecast for all of the TYs, and the number of the cases will vary not only with T but also with each of the models. To simplify the statistics and notation, only the T-h forecasts for which the T-h tracking is successful

TABLE 2. List of 52 typhoons during the 2004–08 western North Pacific typhoon seasons used for the training set and independent cases (boldface) with the starting and ending dates of the typhoons, maximum surface wind speed (kt), and the category of each typhoon. These typhoons were selected because they had some influence on the Vietnam coastal areas.

Typhoon	Date	Vmax (kt)	Category
Chanthu	**9–13 Jun 2004**	**75**	**1**
Kompasu	13–16 Jul 2004	45	TS
Muifa	14–26 Nov 2004	115	4
Conson	4–11 Jun 2004	95	2
Mindulle	23 Jun–4 Jul 2004	125	4
Nida	13– 21 May 2004	140	5
Rananim	7–13 Aug 2004	90	2
Washi	28–31 Jul 2005	<35	TS
Vicente	15–18 Sep 2005	40	TS
Damrey	21–27 Sep 2005	90	2
Roke	13–17 Mar 2005	65	1
Haitang	11–19 Jul 2005	140	5
Matsa	31 Jul–6 Aug 2005	90	2
Sanwu	10–13 Aug 2005	65	1
Talim	26 Aug–1 Sep 2005	125	4
Khanun	5–11 Sep 2005	115	4
Kai_Tak	**28 Oct–2 Nov 2005**	**90**	**2**
Tembin	7–11 Nov 2005	45	TS
Bolaven	13–20 Nov 2005	75	1
Longwang	25 Sep–2 Oct 2005	130	4
Chanchu	**8–18 May 2006**	**135**	**4**
Jelawat	26–29 Jun 2006	45	TS
Prapiroon	31 Jul–3 Aug 2006	70	1
17W	22–25 Sep 2006	<35	TD
Xangsage	**25–30 Sep 2006**	**125**	**4**
Cimaron	27 Oct–4 Nov 2006	140	5
Saomai	4–10 Aug 2006	140	5
Kaemi	21–25 Jul 2006	85	2
Bilis	9–14 Jul 2006	55	TS
Bopha	6–10 Aug 2006	50	TS
Shanshan	10–17 Sep 2006	120	4
Chebi	9–14 Nov 2006	115	4
Durian	**26 Nov–5 Dec 2006**	**135**	**4**
Utor	**7–14 Dec 2006**	**100**	**3**
06W	**2–6 Aug 2007**	**<35**	**TD**
Habigis	18–27 Nov 2007	85	2
Wutip	7–9 Aug 2007	40	TS
Sepat	12–19 Aug 2007	140	5
Wipha	15–19 Sep 2007	135	4
Francisco	23–25 Sep 2007	45	TS
Krosa	1–8 Oct 2007	130	4
Mitag	20–27 Nov 2007	95	2
Pabuk	5–9 Aug 2007	65	1
Lekima	**30 Sep–3 Oct 2007**	**70**	**1**
Peipah	3–9 Nov 2007	75	1
Neguri	14–19 Apr 2008	95	2
Jangmi	26–28 Sep 2008	70	1
Fengshen	17–25 Jun 2008	110	3
Nuri	16–23 Aug 2008	100	3
Hagupit	17–25 Sep 2008	125	4
Higos	23 Sep–1 Oct 2008	140	5

TABLE 3. Total number of cases at each lead time in the training set for all models. A case is defined as a model integration initialized from one GFS update cycle in which the TY center can be detected with a corresponding forecast lead time. Each of the typhoons listed in Table 2 could generate many cases depending on the number of initializations (see text for more details).

Forecast lead time (T, h)	No. of cases
6	320
12	311
18	302
24	290
30	216
36	203
42	234
48	220
54	201
60	187
66	165
72	147

for all models are chosen so that $N_m(T) = N(T)$ for all models. Table 3 lists the total number of cases at each forecast lead time T.

Thus, a set of $N(T)$ T-h forecasts for all TYs has been generated for each of five models in which each forecast is a set of two-dimensional vectors, $\mathbf{v}_i(T) = [x_i(T), y_i(T)]$, where $x_i(T)$ and $y_i(T)$ are the longitude and latitude of the TY center of the ith forecast at time T, respectively. Note that the index i from 1 to $M = 5$ indicates the forecast made by the ith model. Because the best-track data are available for all TYs during the training period, the error matrices $\mathbf{B}(T)$ and $\mathbf{R}_i(T)$ can be computed quickly using Eqs. (5) and (6) in which the expectation operator $E\{\cdot\}$ is simply an average from $i = 1, \ldots, K$. An independent dataset of eight typhoons including Chanthu (2004), Kaitak (2005), Chanchu (2006) Xangsage (2006), Utor (2006), Durrian (2006), Depression 06W (2007), and Lekima (2007) are selected because their tracks had either sharp changes in direction or had sudden deceleration and/or accelerating movements that most of the models could not capture individually. Note that since the main objective of this study is to examine the relative performance of the ensemble postprocessing techniques, the size of the independent dataset is not essential. Numerous experiments with independent datasets of different sizes have been conducted and all showed consistent results. So we hereinafter present results with the independent dataset containing eight typhoons as listed above.

5. Results

a. Individual model performance

The total, the cross-track, and the along-track forecast errors of each of the individual models for the training

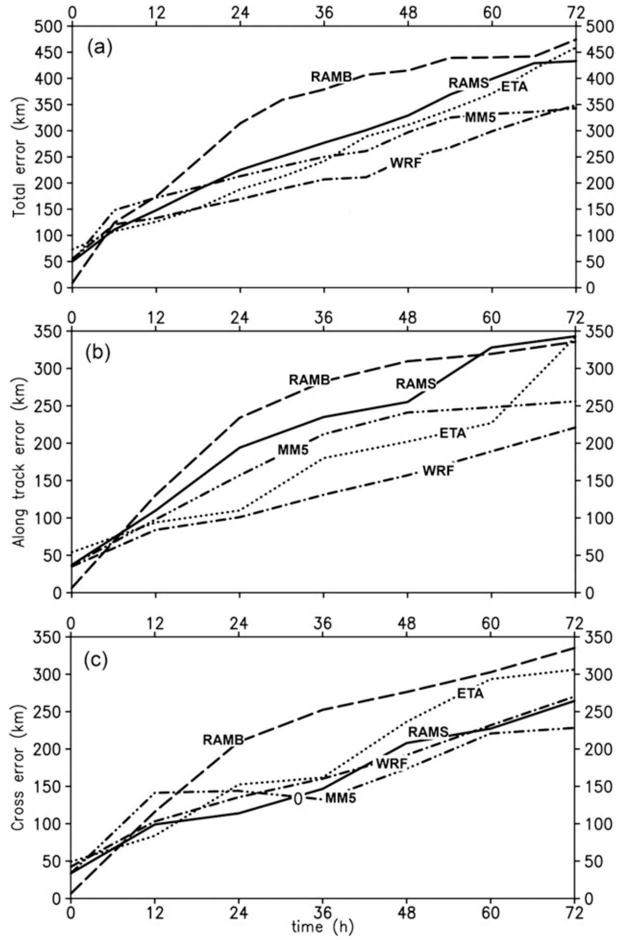

FIG. 2. Mean track errors (km) for RAMS (solid), RAMB (dashed), Eta (dotted), WRF (dotted–dashed), and MM5 (dotted–dotted–dashed) for the (a) total, (b) along-track, and (c) cross-track errors.

dataset are first shown in Fig. 2. Overall, all models have fairly consistent total mean errors, which are roughly ~198 km at 24-h, 287 km at 48-h, and 395 km at 72-h forecast lead times (Fig. 2a). These mean errors are noticeably higher than those documented in the Atlantic Ocean basin from the NCEP/National Hurricane Center (Pike and Neumann 1987; Brown et al. 2010), which may be attributed partly to the present experiments that are not fully optimized. However, the lower skill of the track forecasts in the WPAC basin appears to be consistent with the consensus forecasting in the Systematic Approach Forecasting Aid (Carr et al. 2001), which is based on a consensus of three global and two regional models and thus indicates to some extent the irregular patterns in track behavior for TYs in the WPAC basin. Of the all models, the WRF appears to be the most accurate at long lead time, with 72-h mean error of ~349 km, as compared to the other models. For the shorter forecast intervals, the Eta and RAMS consistently perform better, with mean errors ~110 km at 12-h and ~170 km at 24-h lead time.

It is worth mentioning that except for RAMB, for which a bogussed vortex is inserted into the initial conditions according to the best-track location, the initial positions of the TYs in all other models contain some initial error (~50 km; Fig. 2a). This initial error is mostly caused by the coarse resolution of the GFS data used to initialize the forecasts in this study. In addition, such incorrect initial positioning could be related to some other factors such as the center-tracking algorithm, or some difference between the initial positions in the GFS and the best-track analysis. Except for experiments with RAMB, no attempt has been made to relocate the TY centers for other models in this study, as the main objective is focused more on the relative effectiveness of different ensemble mean methods.

In terms of the along-track errors (Fig. 2c), the RAMS and RAMB have the worst performance at long lead times, with 3-day errors of roughly 350 km. Examination of the large-error cases created by the RAMS and RAMB shows that these models tend to produce abnormally strong subtropical ridges over central China, which seem to slow down the westward movement of TCs and lead to larger eastward biases as compared to the other models. Previous studies have shown that the along-track errors tend to be larger than the cross-track errors (see, e.g., Buckingham et al. 2010). However, comparison of the along-track and cross-track mean errors in Figs. 2c and 2d shows that the WRF and Eta do not exhibit such statistics, especially the WRF during the first 36 h of the model forecasts. This appears to indicate that the WRF could capture fairly well the translational speed but has some systematical bias across the track. As pointed out in the recent studies by Kehoe et al. (2007), the source of such large cross-track errors could be attributed to either the physical processes that are poorly represented by the models or to strong interactions with nearby cyclones or midlatitude influences.

While the performance of the models in all of the experiments is obtained for the configuration at 28-km resolution and the specific parameterizations discussed in the section 4, several experiments with higher resolution did not show significant track error differences. The exception was the WRF, which had a slight improvement in its along-track error forecast (not shown). Due to the limited computational resources, higher-resolution forecasts for all 52 TYs could not be carried out, and the mean errors shown in Fig. 2 will be considered hereafter to be a benchmark for later verification.

The cross correlations between the latitudinal and longitudinal error positions, which are the basis for the multivariate optimization presented in Section 2, are shown in Fig. 3. These cross correlations of the latitudinal and longitudinal errors are the off-diagonal components of the matrix $\mathbf{R}_i(T)$ in Eq. (6). Although these correlations are small for the first 12 h, they increase with time and reach magnitudes of ~0.3–0.4 after 54 h. This significant correlation indicates that independent ensemble averages of the latitudes and longitudes of TY centers would become inaccurate at longer lead times. In general, correlation exists between the total mean error and the correlation between the latitudinal and longitudinal errors; the larger the total mean error is, the larger the correlation is. The negative correlations between latitudinal and longitudinal errors in the RAMS and MM5 imply that the negative latitudinal errors due to a slow translational speed of a TY would correspond to positive longitudinal errors. As a result, the TY centers in these models tend to stay in the southeastern quadrant with respect to the best-track center for westward-moving TYs, consistent with the along-track errors (Fig. 2b). The high correlation between the latitudinal and longitudinal track error suggests that any good ensemble mean should take this information into account so that the retrieval of ensemble information is maximized.

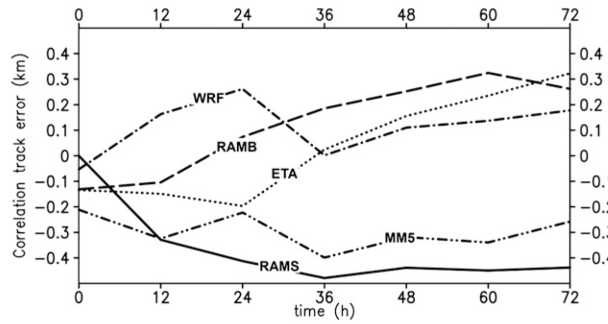

FIG. 3. As in Fig. 2, but for the cross correlation between latitude and longitude track errors computed for the training dataset.

While the focus here is on the TY tracks, it is noted that all models have fairly low skill in intensity forecasting, especially during the rapid intensification. Of all of the models, only MM5 and RAMB seemed to be capable of occasionally capturing some phases of the intensity change. Reasons for such low skill in simulating intensity changes could be related to the coarse horizontal model resolution used in this study and insufficient in situ observation data for TYs in the WPAC basin. Inclusion of radar and satellite observations in the models will be presented in a future study.

b. Ensemble track forecasts

In this section four different ensemble mean methods for processing multimodel outputs are compared. The objective of the comparisons is to examine the performance

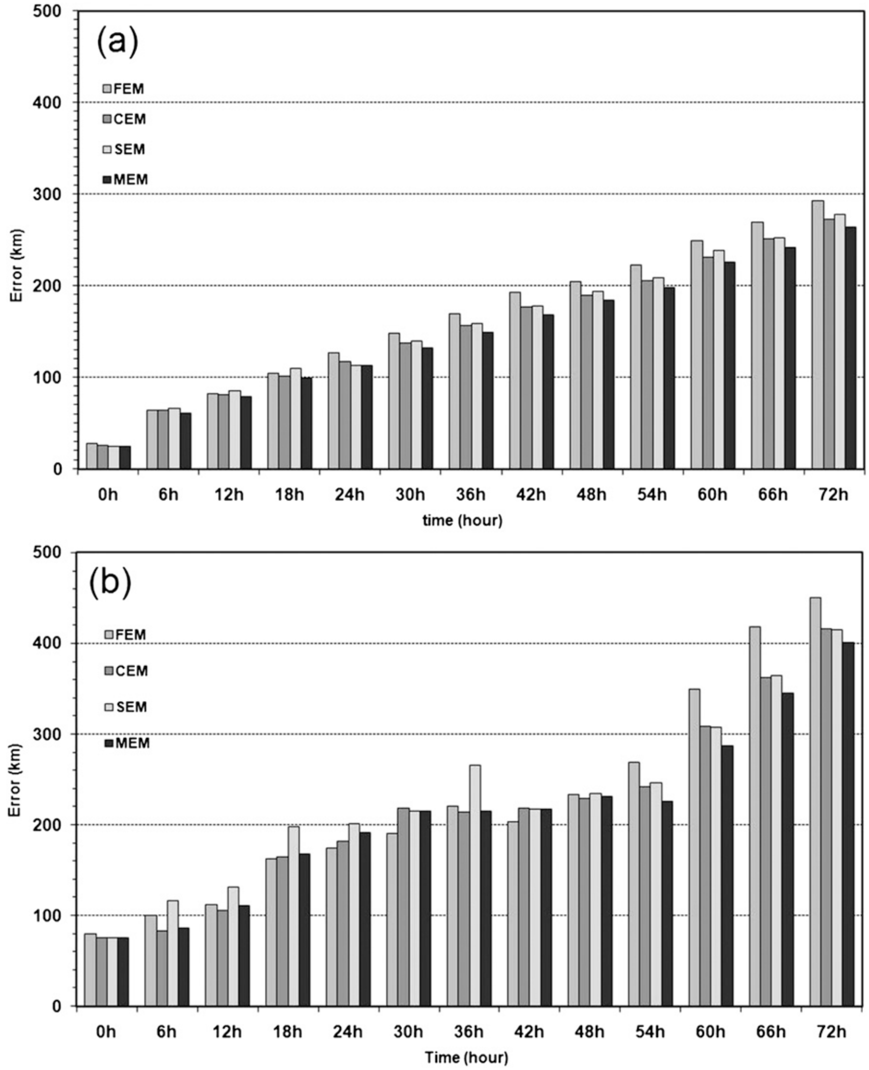

FIG. 4. Total track errors (km) for four different ensemble mean methods including FEM (medium gray), CEM (white), SEM (light gray), and MEM (dark gray) with the (a) training and (b) independent datasets.

of the multivariate ensemble mean (MEM) *relative to* the simple full ensemble mean (FEM), the cluster ensemble mean (CEM), and the selected ensemble mean (SEM) methods discussed in section 2. Due to the small number of models, it is not always possible to identify different clusters in the CEM approach. So only cases where two distinct clusters can be identified are counted in the CEM mean. Similarly for the SEM approach, only cases in which outliers can be definitely detected are included in this mean calculation.[1]

[1] For the sake of automatic computation, an outlier is detected when its 1-day error relative to the ensemble mean is larger than 250 km.

Comparisons of the various ensemble means for eight independent typhoons and for the dependent cases are given in Fig. 4. Table 4 offers a list of the dependent cases corresponding to each forecast lead time. In general, all methods give a similar result for the first 36 h with mean errors of ~150 km for the 24-h forecast. At the later 36 h, MEM has the smallest errors with 72-h track errors (~325 km) as compared to FEM (355 km), CEM (381 km), and SEM (383 km) at the 95% confident level. The better performance of either CEM or SEM as compared to the FEM approach at the later time may be understood if one recalls that the RAMB tends to produce increasingly large cross-track errors with time. So, eliminating the RAMB forecasts in the CEM and SEM approaches could indeed reduce the total track

TABLE 4. As in Table 3, but for the number of independent cases for eight typhoons given in Table 2.

Forecast lead time (T, h)	No. of cases
6	135
12	115
18	104
24	95
30	88
36	83
42	77
48	76
54	65
60	58
66	49
72	36

error. The better performance of the MEM approach in reducing the total track mean error is most evident after 36 h when the correlation between the latitudinal and longitudinal track errors becomes significant (Fig. 3). During this later period, the high correlation of the track errors results in more constrained correction to the first guess since the MEM method uses the full ensemble mean only as the first guess and then corrects the mean value according to the accuracy of each model (cf. Fig. 3).

Although the actual percentage of the track error reduction may vary with the numbers of the ensemble models or the size of the training/independent dataset, it should be noted that the relative improvement of the MEM approach with respect to other ensemble mean methods is still noticeable for all sizes of the independent dataset. Experiments with the size of the independent dataset varying from 2 to 12 TYs all confirm the improvement found in the MEM approach. The larger size of the independent dataset (>12 TYs) will nevertheless result in insufficient statistics for the weight matrices, thus causing the ensemble mean tracks to be nearly indistinguishable among different methods.

In addition to reduced track mean errors, the MEM approach produces TY tracks that are smoother when the number of ensemble models is small. This improvement is a result of the latitudes and longitudes of the TY centers not being independently corrected due to their cross correlation, thus preventing sharp jumps in the averaged latitudes and longitudes. To illustrate this point, the 72-h forecasts for Utor initialized at 1200 UTC 11 December 2006, Chanthu initialized at 1200 UTC 12 June 2004, Durian initialized at 0000 UTC 1 December 2006, and 06W initialized at 0000 UTC 3 August 2007 are shown in Fig. 5. Note that for the above examples, initializations were selected such that one of the model track forecasts is clearly deviating from the rest of the forecasts. This ensures that both the CEM and the SEM classifications can be applied. In these examples, the CEM and SEM tracks are identical, as only one model track (the WRF in the forecast of Utor and the RAMB in the forecast of 06W, Chanthu, and Durian) diverges far to the west compared to the rest of the tracks. The smoothness of the track forecast in the MEM approach can be seen evidently in Fig. 5 with most of sudden jumps in Utor's forecast locations during the first 18 h removed. The irregularities of the simulated track during this period are caused by the interaction of Utor with the Philippines archipelagic area. Similarly, the sharp zigzags in the 06W, Durian, and Chanthu tracks during the first 12 h are also removed in the MEM approach. Note that 06W was only a marginal tropical storm at its peak intensity, which presented some challenges to predicting its movement and intensity. Apparently, the predictability of 06W is quite low as all models fail to capture its sharp turn as it changed its direction suddenly near 0000 UTC 5 August (Fig. 5b). Since the WRF model had shown the most consistent performance in the training set (Fig. 4b), it is given more weight in the MEM track forecast for Utor. The inclusion of the WRF forecast, which is eliminated in the CEM and SEM approaches, could indeed pull the track farther north (not shown). In this case, the track would have been closer to the observations that neither FEM nor CEM/SEM could capture.

6. Conclusions

In this study a multivariate optimization approach to extracting the most information about typhoon tracks from a set of multimodel forecasts has been presented. Unlike the ensemble forecast of a single variable for which a simple ensemble mean can provide statistically meaningful information about the predicted value, the typhoon track forecast should be considered to be a multivariate optimization problem in which the latitudes and longitudes of forecast positions are treated as two-dimensional vectors. The basis for such a two-dimensional approach is based on the fact that the components of these track vectors possess considerable correlation due to model systematic biases. Such correlation increases with time and becomes significant after 1 day into the forecast, which suggests that separate ensemble averages of the latitude and longitude of typhoon centers would lose substantial information about the track constraint.

In the two-dimensional approach, the simple ensemble mean was used as a first guess and the multivariate optimization corrected this first guess with information from each model, based on each model's skill from a training set. Experiments with five mesoscale models have

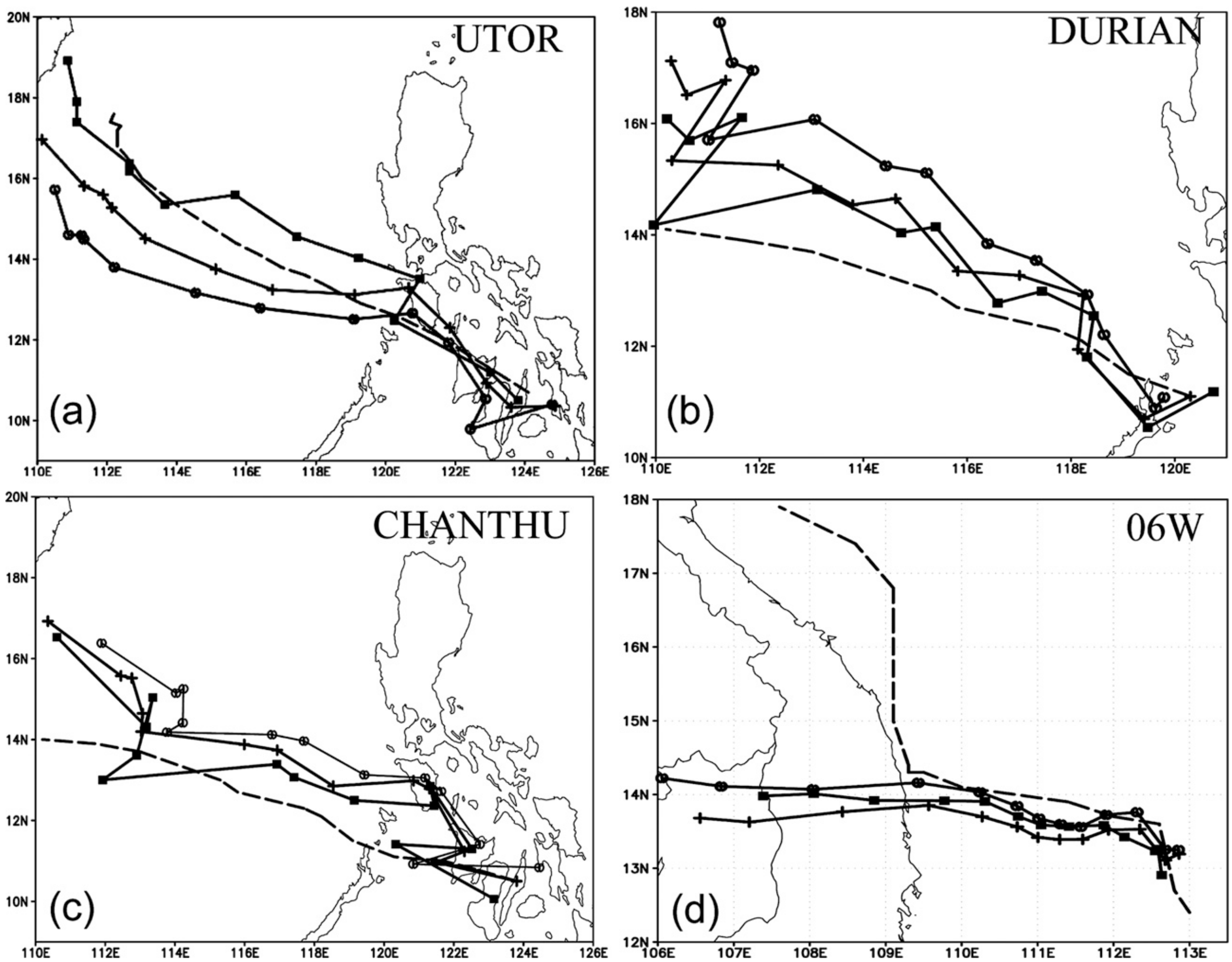

FIG. 5. Comparison of the 72-h track forecasts between FEM (boldface plus sign), MEM (boldface square), CEM and SEM (boldface circle), and best track (boldface dashed) for four independent cases: (a) Typhoon Utor initialized at 1200 UTC 11 Dec 2005, (b) Durian initialized at 0000 UTC 1 Dec 2006, (c) Chanthu initialized at 1200 UTC 12 Jun 2004, and (d) Depression 06W initialized at 0000 UTC 3 Aug 2007.

shown that the multivariate optimization overall provides better performance both in terms of total errors and the smoothness of the typhoon track forecast. For an independent set of eight typhoons, the total track error was reduced by about 5% as compared to the full ensemble mean, selective mean, or cluster mean after 36 h of integration. The advantage of the multivariate approach is its simplicity as it does not require overly detailed information. While we have presented only 72-h typhoon forecast statistics due to computational constraints, the multivariate approach in this study can be extended readily to longer forecast lead times to extract the most information from an ensemble of outputs. Provided that a history of previous ensemble forecasts is made available, our approach will always allow for maximizing the information of the TY tracks from the ensemble at all forecast lead times.

Acknowledgments. This research was supported by the Vietnam Ministry of Science and Technology Foundation (KC.08.05/06-10 and NCCB-DHUD.2011-G/10). The authors thank two anonymous reviewers for their invaluable comments and suggestions, which have improved substantially the quality of this work. Credit is also given to Dr. Nguyen Minh Truong for his discussions and to the Vietnam National Hydro-Meteorological Forecasting Center and the Vietnam Institute of Meteorology, Hydrology and Environment for their help in synthesizing model outputs and collecting datasets.

REFERENCES

Aberson, S. D., and M. DeMaria, 1994: Verification of a nested barotropic hurricane track forecast model (VICBAR). *Mon. Wea. Rev.,* **122,** 2804–2815.

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

——, and ——, 2003: Targeted observations to improve operational tropical cyclone track forecast guidance. *Mon. Wea. Rev.,* **131,** 1613–1628.

Brown, D. P., J. L. Beven, J. L. Franklin, and E. S. Blake, 2010: Atlantic hurricane season of 2008. *Mon. Wea. Rev.,* **138,** 1975–2001.

Buckingham, C., T. Marchok, I. Ginis, L. Rothstein, and D. Rowe, 2010: Short- and medium-range prediction of tropical and transitioning cyclone tracks within the NCEP Global Ensemble Forecasting System. *Wea. Forecasting,* **25,** 1736–1754.

Carr, L. E., R. L. Elsberry, and J. E. Peak, 2001: Beta test of the systematic approach expert system prototype as a tropical cyclone track forecasting aid. *Wea. Forecasting,* **16,** 355–368.

Cheung, K. K. W., and J. C. L. Chan, 1999: Ensemble forecasting of tropical cyclone motion using a barotropic model. Part I: Perturbations of the environment. *Mon. Wea. Rev.,* **127,** 1229–1243.

Daley, R., 1992: *Atmospheric Data Analysis*. Cambridge University Press, 472 pp.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.,* **99,** 10 143–10 162.

Fels, S. B., and M. D. Schwarzkopf, 1975: The simplified exchange approximation: A new method for radiative transfer calculations. *J. Atmos. Sci.,* **32,** 1475–1488.

Ferrier, B. S., Y. Jin, Y. Lin, T. Black, E. Rogers, and G. DiMego, 2002: Implementation of a new grid-scale cloud and precipitation scheme in the NCEP Eta model. Preprints, *19th Conf. on Weather Analysis and Forecasting/15th Conf. on Numerical Weather Prediction,* San Antonio, TX, Amer. Meteor. Soc., 280–283.

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

——, C. R. Sampson, and J. M. Gross, 2004: A history of western North Pacific tropical cyclone track forecast skill. *Wea. Forecasting,* **19,** 633–638.

Huang, C.-Y., Y.-H. Kuo, S.-H. Chen, and F. Vandenberghe, 2005: Improvements in typhoon forecasts with assimilated GPS occultation refractivity. *Wea. Forecasting,* **20,** 931–953.

Kain, J. S., and J. M. Fritsch, 1990: A one-dimensional entraining/detraining plume model and its application in convective parameterization. *J. Atmos. Sci.,* **47,** 2784–2802.

Kehoe, R. M., M. A. Boothe, and R. L. Elsberry, 2007: Dynamical tropical cyclone 96- and 120-h track forecast errors in the western North Pacific. *Wea. Forecasting,* **22,** 520–538.

Krishnamurti, T. N., R. Correa-Torres, G. Rohaly, and D. Oosterhof, 1997: Physical initialization and hurricane ensemble forecasts. *Wea. Forecasting,* **12,** 503–514.

Lacis, A. A., and J. E. Hansen, 1974: A parameterization of the absorption of solar radiation in the earth's atmosphere. *J. Atmos. Sci.,* **31,** 118–133.

Lander, M., and G. J. Holland, 1993: On the interaction of tropical-cyclone scale vortices. I: Observations. *Quart. J. Roy. Meteor. Soc.,* **119,** 1347–1361.

Leslie, L. M., and K. Fraedrich, 1990: Reduction of tropical cyclone position errors using an optimal combination of independent forecasts. *Wea. Forecasting,* **5,** 158–161.

Lin, Y.-L., R. D. Farley, and H. D. Orville, 1983: Bulk parameterization of the snow field in a cloud model. *J. Climate Appl. Meteor.,* **22,** 1065–1092.

Liu, K. S., and J. C. L. Chan, 2008: Interdecadal variability of western North Pacific tropical cyclone tracks. *J. Climate,* **21,** 4464–4476.

Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmosphere: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.,* **102,** 16 663–16 682.

Neumann, C. J., and J. M. Pelissier, 1981: Models for the prediction of tropical cyclone motion over the North Atlantic: An operational evaluation. *Mon. Wea. Rev.,* **109,** 522–538.

Payne, K. A., R. L. Elsberry, and M. A. Boothe, 2007: Assessment of western North Pacific 96- and 120-h track guidance and present forecast ability. *Wea. Forecasting,* **22,** 1003–1015.

Pike, A. C., and C. J. Neumann, 1987: The variation of track forecast difficulty among tropical cyclone basins. *Wea. Forecasting,* **2,** 237–241.

Ritchie, E. A., and G. J. Holland, 1997: Scale interactions during the formation of Typhoon Irving. *Mon. Wea. Rev.,* **125,** 1377–1396.

Simpson, J., E. Ritchie, G. J. Holland, J. Halverson, and S. Stewart, 1997: Mesoscale interactions in tropical cyclone genesis. *Mon. Wea. Rev.,* **125,** 2643–2661.

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

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

Wu, C.-C., and K. A. Emanuel, 1993: Interaction of a baroclinic vortex with background shear: Application to hurricane movement. *J. Atmos. Sci.,* **50,** 62–76.

Yang, M.-J., D.-L. Zhang, and H.-L. Huang, 2008: A modeling study of Typhoon Nari (2001) at landfall. Part I: Topographic effects. *J. Atmos. Sci.,* **65,** 3095–3115.

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