Typhoon Ensemble Prediction System Developed at the Japan Meteorological Agency

MUNEHIKO YAMAGUCHI, RYOTA SAKAI, MASAYUKI KYODA, TAKUYA KOMORI, AND TAKASHI KADOWAKI

Japan Meteorological Agency, Tokyo, Japan

(Manuscript received 24 June 2008, in final form 6 February 2009)

ABSTRACT

The Japan Meteorological Agency (JMA) Typhoon Ensemble Prediction System (TEPS) and its performance are described. In February 2008, JMA started an operation of TEPS that was designed for providing skillful tropical cyclone (TC) track predictions in both deterministic and probabilistic ways. TEPS consists of 1 nonperturbed prediction and 10 perturbed predictions based on the lower-resolution version (TL319L60) of the JMA Global Spectral Model (GSM; TL959L60) and a global analysis for JMA/GSM. A singular vector method is employed to create initial perturbations. Focusing on TCs in the western North Pacific Ocean and the South China Sea (0°–60°N, 100°E–180°), TEPS runs 4 times a day, initiated at 0000, 0600, 1200, and 1800 UTC with a prediction range of 132 h. The verifications of TEPS during the quasi-operational period from May to December 2007 indicate that the ensemble mean track predictions statistically have better performance as compared with the control (nonperturbed) predictions: the error reduction in the 5-day predictions is 40 km on average. Moreover, it is found that the ensemble spread of tracks is an indicator of position error, indicating that TEPS will be useful in presenting confidence information on TC track predictions. For 2008 when TEPS was in operational use, however, it was also found that the ensemble mean was significantly worse than the deterministic model (JMA/GSM) out to 84 h.

1. Introduction

In 1997, the Japan Meteorological Agency (JMA) started to provide public users with 3-day track forecasts of tropical cyclones (TCs) in the western North Pacific Ocean and South China Sea based on numerical weather prediction [NWP; the (Regional Specialized Meteorological Center) RSMC Tokyo-Typhoon Center 1997]. Since then, the remarkable progress of the JMA NWP system has brought considerable improvement of the track forecasts. According to the verification of the JMA global NWP system, the 3-yr running mean of position errors of 5-day predictions in 2007 (451 km, the average of 2005, 2006, and 2007) is smaller than that of 3-day predictions in 1997 (472 km, the average of 1995, 1996, and 1997), indicating that we have succeeded in obtaining a 2-day lead time of deterministic TC track predictions over the past 11 yr (see Fig. 1).

While the accuracy of TC track forecasts has improved significantly, ensemble techniques have been attracting much attention because they are expected to improve deterministic TC track forecasts and also provide the uncertainty information (WMO 2008a), based on ensemble mean and ensemble spread, respectively (e.g., Jeffries and Fukada 2002; Vijaya Kumar and Krishnamurti 2003; Sampson et al. 2006; Goerss 2007). For the western North Pacific basin, Goerss et al. (2004) have shown that the consensus of three models, the Navy Operational Global Atmospheric Prediction System (NOGAPS; Hogan and Rosmond 1991; Goerss and Jeffries 1994), the Met Office (UKMO) global model (Cullen 1993; Heming et al. 1995), and the JMA typhoon model (TYM; Kuma 1996), has reduced forecast errors with respect to the best of the individual models. For 2002, for example, the improvement rate of the consensus with respect to the best of the individual models is 17%, 14%, and 11% at 24, 48, and 72 h, respectively. Goerss (2000) has also shown the effectiveness of consensus forecasts for both the western North Pacific and the Atlantic basin: using the JMA global model (GSM; Kuma 1996), NOGAPS, and UKMO for

Corresponding author address: Munehiko Yamaguchi, 1-3-4 Otemachi, Chiyoda-ku, Tokyo, 100-8122 Japan.
E-mail: munehiko.yamaguchi@gmail.com

DOI: 10.1175/2009MWR2697.1

© 2009 American Meteorological Society
the western North Pacific basin; and the Geophysical Fluid Dynamics Laboratory Hurricane Prediction System (GFDL; Kurihara et al. 1993, 1995, 1998), NOGAPS, and UKMO for the Atlantic basin. For the western North Pacific during 1997, the improvement rate with respect to the best of the individual models is 16%, 13%, and 12% at 24, 48, and 72 h, and for the 1995–96 Atlantic hurricane seasons, the improvement rate is 16%, 20%, and 23% at 24, 48, and 72 h, respectively. Furthermore, Komori et al. (2007) have verified the ensemble mean track of three global models from operational NWP centers, the European Centre for Medium-Range Weather Forecasts (ECMWF), JMA, and UKMO. In the verification of 48- and 96-h predictions for the western North Pacific basin from 1991 to 2005, and for the Atlantic basin from 1999 to 2005, the ensemble mean has the best performance with respect to the best of the individual models with few exceptional years. For a study on evaluating the ensemble spread of TC track predictions, Goerss (2000) and Elsberry and Carr (2000) have demonstrated the benefit of the ensemble spread as a measure of confidence in ensemble TC predictions: a small spread of tracks is often an indication of a small error of ensemble mean track prediction. These arguments form the rationale for the development of the ensemble prediction system (EPS).

Recently, JMA has developed a new Typhoon EPS (TEPS). Different from the multimodel EPS mentioned above, TEPS is based on a sole NWP model with perturbed initial conditions. To evaluate whether the ensemble mean and ensemble spread of TEPS have similar characteristics as compared with the multimodel EPS used in previous studies, we conducted a series of quasi-operational runs of TEPS from May to December 2007. This paper describes the verification results of TEPS during the quasi-operational period as well as its specifications: first, the skill of ensemble mean track predictions is verified; and second, the effectiveness of confidence information on track predictions using the ensemble spread is verified. Moreover, the verification of confidence information is performed by dividing the ensemble spread into along-track and cross-track directions since the prediction of uncertainty in timing and location of TC impacts is essential (Aberson 2001). Section 2 describes the specifications of TEPS: the NWP system and the method to create initial perturbations. Section 3 describes the performance of TEPS, and section 4 discusses the performance. The summary and conclusions are presented in section 5.

2. Specifications of TEPS

a. Overview of TEPS

In February 2008, JMA started an operation of TEPS (Fig. 2). TEPS uses a lower-resolution version of the JMA/GSM. The model resolution of TEPS is TL319L60, the spectral triangular truncation TL319 (the “L” indicating the linear grid option) and 60 vertical levels, while that of JMA/GSM is TL959L60 (Iwamura and Kitagawa 2008). As TEPS does not have its own data assimilation system, a global analysis for JMA/GSM, which is based on a four-dimensional variational data system (4DVAR; Kadowaki 2005; JMA 2007), is used for an initial condition of the nonperturbed run of TEPS after interpolating it from TL959L60 to TL319L60.

Forecasters on duty at JMA initiate the operation of TEPS when the following two conditions are satisfied:
1) TCs are located in the RSMC Tokyo-Typhoon Center’s responsibility area of 0°–60°N, 100°E–180°W, or are expected to move into the area within 24 h.
2) The TC’s maximum sustained wind speed exceeds 34 kt or is expected to reach the threshold within 24 h (in that sense, the name of TEPS might not be appropriate because we call TCs with a maximum sustained wind of 64 kt or more “typhoon”).

TEPS runs 4 times a day, initiated at 0000, 0600, 1200, and 1800 UTC with a prediction range of 132 h. The ensemble size is 11. A singular vector (SV) method (Buizza 1994; Molteni et al. 1996; Puri et al. 2001) is employed to create initial perturbations (see section 2b for more details). A method dealing with the uncertainty of an NWP model itself such as a stochastic physics method (Buizza et al. 1999) is yet to be considered.

b. Singular vector method

Under the assumption that a perturbation grows linearly, a SV with a large singular value represents a
fast-growing perturbation (Lorenz 1965). Consider a growth rate of a perturbation $\mathbf{x}$ as shown in Eq. (1):

$$\frac{\|\mathbf{x}(t = t_a)\|}{\|\mathbf{x}(t = t_0)\|} \mathbf{x} \in \mathbb{R}^n, \quad (1)$$

where $\mathbf{x}(t = t_0)$ is a perturbation at a base time $t_0$, $\mathbf{x}(t = t_a)$ is one at an optimization time $t_a$ ($t_a > t_0$), and $\| \cdot \|$ denotes the norm associated with the Euclidean inner product. The growth rate of a perturbation given by Eq. (1) changes into Eq. (2), using a tangent forward propagator $M$:

$$\frac{\|\mathbf{x}(t = t_a)\|}{\|\mathbf{x}(t = t_0)\|} = \sqrt{\left(\mathbf{E}^f \mathbf{T} \mathbf{M} \mathbf{x}(t = t_0)\right)^T \left(\mathbf{E}^f \mathbf{T} \mathbf{M} \mathbf{x}(t = t_0)\right) / \left(\mathbf{x}(t = t_0), \mathbf{E}^f \mathbf{T} \mathbf{M} \mathbf{x}(t = t_0)\right)} \mathbf{x} \in \mathbb{R}^n, \quad (2)$$

where $\mathbf{E}_0$ and $\mathbf{E}_f$ are norm operators at $t_0$ and $t_a$, respectively. The local projection operator $\mathbf{T}$ sets all elements of a vector to 0 outside a prescribed domain, which enables the calculation of perturbations with maximum amplitude at $t_a$ over the targeted area. Here $[\cdot]$ denotes the Eulerian inner product. The growth rate equation still changes to Eqs. (3) and (4) from Eq. (2), using

$$\dot{x} = \mathbf{E}^f_0 \mathbf{x}(t = t_0), \quad \mathbf{A} = \mathbf{E}^f_0 \mathbf{T} \mathbf{M} \mathbf{E}_0^{-1/2}$$

and adjoint matrix represented by the superscript asterisk:

$$\frac{\|\mathbf{x}(t = t_a)\|}{\|\mathbf{x}(t = t_0)\|} = \sqrt{\left(\mathbf{E}^f \mathbf{T} \mathbf{M} \mathbf{x}(t = t_0)\right)^T \left(\mathbf{E}^f \mathbf{T} \mathbf{M} \mathbf{x}(t = t_0)\right) / \left(\mathbf{x}(t = t_0), \mathbf{E}^f \mathbf{T} \mathbf{M} \mathbf{x}(t = t_0)\right)} \mathbf{x} \in \mathbb{R}^n, \quad (3)$$

$$\frac{\|\mathbf{x}(t = t_a)\|}{\|\mathbf{x}(t = t_0)\|} = \sqrt{\left(\mathbf{A}^* \mathbf{A} \mathbf{x}\right)^T \left(\mathbf{A}^* \mathbf{A} \mathbf{x}\right) / \left(\mathbf{x}, \mathbf{x}\right)} \mathbf{x} \in \mathbb{R}^n, \quad (4)$$

Equation (4) represents that SVs, or forward SVs of matrix $\mathbf{A}$, grow linearly about a given trajectory with their growth rates of the corresponding singular values. Therefore the first SV, which has the largest singular value, maximizes the ratio in Eq. (1), and the second SV gives the fastest-growing perturbation following the first SV, and so on. These SVs are the solutions of the eigenvalue problem shown in Eq. (5):

$$\mathbf{A}^* \mathbf{A} \mathbf{x} = \lambda \mathbf{x} \quad (\lambda: \text{eigenvalue}). \quad (5)$$

The SVs are computed applying an iterative Lanczos procedure (e.g., Strang 1986) to the linear propagator $\mathbf{M}$ instead of solving Eq. (5) directly.

In TEPS, $\mathbf{M}$ and $\mathbf{M}^*$ are the tangent-linear and adjoint models used for the 4DVAR, which has been in operation since February 2005 (Kadowaki 2005; JMA 2007). While their resolutions are T159L40 for the 4DVAR as of February 2009, TEPS uses the lower-resolution version, T63L40. They consist of dynamics based on Eulerian integrations and physical processes containing

![Figure 2: Time line of the upgrade of JMA NWP systems: JMA/GSM, TYM, WEPS, and TEPS.](image)
representations of vertical diffusion, gravity wave drag, large-scale condensation, longwave radiation, and deep cumulus convection. The SVs based on the tangent-linear and adjoint models including the full physical processes (the simplified physical processes without moist processes) are called moist (dry) SVs. In TEPS, dry SVs are calculated targeting for a midlatitude area in the RSMC Tokyo-Typhoon Center’s responsibility area, aiming to identify the dynamically most unstable mode of the atmosphere like the baroclinic mode (Buizza and Palmer 1995). Moist SVs are calculated targeting for TC surroundings where moist processes are crucial (Barkmeijer et al. 2001).

TEPS can target up to three TCs at the same time. If more than three TCs exist, three TCs are selected in the order of concerns of the RSMC Tokyo-Typhoon Center. The targeted area of dry SV calculations is fixed, 20°–60°N, 100°E–180°W, and that of moist SV calculations is a rectangle, 10° in latitude and 20° in longitude with its center at a TC central position of the 24-h forecast. The evaluation time interval is 24 h for both dry and moist SVs.

A norm to evaluate the growth rate of dry and moist SVs is based on a total energy norm including a specific humidity term (Barkmeijer et al. 2001):

$$\mathbf{v} = \frac{1}{2} \int_0^1 \left[ \nabla \Delta^{-1} \xi_s \cdot \nabla \Delta^{-1} \xi_s + \nabla \Delta^{-1} D_z \cdot \nabla \Delta^{-1} D_z \right. $$

$$+ g \left( \Gamma - \Delta^{-1} T_z T_r \right) + w \left( \frac{L^2}{c_p T_r} q_s q_s \right) + \left. \int \frac{R_T T_r}{P_s} P_s dS \right] d\eta \tag{6}$$

with $\xi_s$, $D_z$, $T_z$, $q_s$, and $P_s$ being the vorticity, divergence, temperature, specific humidity, and surface pressure components, respectively, of a vector $\mathbf{x}$. The temperature lapse rate $\Gamma$ is taken into consideration in an available potential energy term (Lorenz 1955). Here $g$ is the acceleration of gravity, $c_p$ is the specific heat of dry air at constant pressure, $L_e$ is the latent heat of condensation, and $R_g$ is the gas constant for dry air. Here $p(\eta)$ is the pressure at eta levels. $T_r = 300$ K is a reference temperature, $P_r = 800$ hPa is a reference pressure, and $w_q$ is a constant ($w_q = 1$ in this study). For $\Gamma$, its representative value, $2/3\Gamma_0$ is used, and then an available potential energy term is written as $3c_p T_z T_r / \Gamma_r$.

In Eq. (6), the vertical integration of the kinetic energy term and the available potential energy term is performed with an initial and final norm set to zero above 100 hPa (model level 15). Otherwise, as is the case with Barkmeijer et al. (2001), SVs have a shallow vertical structure in the upper troposphere or have a large specific humidity amplitude in the upper troposphere where the amount of specific humidity is relatively small. Through some numerical experiments, it is found that those SVs have little impact on an ensemble of TC track predictions, and therefore, we have adopted the above conditions in the vertical integration.

Figures 3a,b show the vertical distributions of the total energy components of moist SVs at an initial and evaluation time, respectively. These are the composites of all moist SVs computed during the quasi-operational period (the total number of samples is 3077). In both figures, the energy is normalized by the peak value of a certain component at a certain model level. As Fig. 3a shows, the initial SVs are mostly explained by the kinetic energy component centered at about 650 hPa (model level 11). This kinetic energy component is largely explained by the temperature lapse rate term in Eq. (6). In the experimental SV calculations not considering the lapse rate, where the available potential energy term is written as $c_p T_z T_r / \Gamma_r$, the available potential energy is dominant in the calculated initial SVs. Through some numerical experiments, it is found that those SVs need larger initial amplitudes of total energy in order to get the same amount of ensemble spread in terms of TC tracks with respect to SVs as shown in Fig. 3a. This result would indicate that wind perturbations cause a larger ensemble spread of tracks than temperature perturbations. Final SVs are explained by the specific humidity energy at the lower troposphere and kinetic energy from the lower to upper troposphere. These distributions have a good agreement with those of total energy of forecast errors by JMA/GSM in tropical regions and TC surroundings (figures not shown).

The composite of the vertically accumulated total energy of the initial moist SVs calculated during the quasi-operational period is shown in Fig. 4a for the leading moist SVs, and Fig. 4b for all moist SVs. Just like Figs. 3a,b, the energy is normalized by the peak value of the energy in the 4000 km × 4000 km domain centered on analyzed TC position. As Fig. 4a shows, the distribution is asymmetric, and a sensitive region spreads from north–west to east related to the storm center. The peak point is approximately 500 km away from the center in both Figs. 4a,b, which is consistent with the results of Reynolds et al. (2009) and Peng and Reynolds (2006) that suggest that the high sensitivity appears in the region where the potential vorticity gradient of the vortex first changes sign.
c. Generation of initial perturbations

The initial perturbations are obtained by linearly combining the SVs:

\[ p_i = \mathbf{E}_i^{1/2} \mathbf{\hat{p}}_i = \mathbf{E}_i^{1/2} \sum_{j=1}^{n} r_{ij} \mathbf{x}_j, \]  

(7)

where \( p_i \) is the \( i \)th initial perturbation, \( \mathbf{\hat{p}}_i = \mathbf{E}_i^{1/2} \mathbf{p}_i \), \( r \) are the coefficients of the linear combination, \( n \) is the number of computed SVs, and \( \mathbf{x}_j \) is a SV (1 \( \leq j \leq n \)). The linear combination gives spatially spreading initial perturbations because each SV tends to be confined to a local area. In each SV calculation, up to 10 SVs can be computed depending on an operationally allocated calculation time, which means we have up to 40 SVs (\( n \approx 40 \)), that is 10 dry SVs and 30 moist SVs, at one prediction event. Note that there is a possibility that the calculated SVs are similar to each other. For example, if two TCs exist at the same time, and the targeted areas for the two TCs are overlapped, a SV for one TC can be similar to the SV for the other TC. To avoid duplication, if the value of the inner product of any two SVs is 0.5 or more \([\langle \mathbf{x}_i, \mathbf{x}_j \rangle \geq 0.5]\), one of the two SVs is eliminated from the SVs used in Eq. (7).

The method to decide the coefficients is based on a variance minimum rotation. Consider an \( m \times n \) matrix \( \mathbf{A} \) whose columns are composed of each SV:

\[ \mathbf{A} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n), \]  

(8)

where \( m \) is the dimension of a SV and \( n \) is the number of columns of \( \mathbf{A} \). Then, we compute an \( n \times n \) orthogonal matrix \( \mathbf{R} \), which determines the coefficients:

\[ \mathbf{B} = \mathbf{A} \mathbf{R}, \]  

(9)

\[ \mathbf{B} = (\mathbf{\hat{p}}_1, \mathbf{\hat{p}}_2, \ldots, \mathbf{\hat{p}}_n). \]  

(10)

and

\[ \mathbf{R} = \begin{pmatrix} r_{11} & \cdots & r_{1n} \\ \vdots & \ddots & \vdots \\ r_{n1} & \cdots & r_{nn} \end{pmatrix}. \]  

(11)

Here \( \mathbf{R} \) is obtained through an iterative calculation by minimizing the value of \( V \) defined in Eq. (12):

\[ V = \sum_{i=1}^{n} V_i = \sum_{i=1}^{n} \left[ \frac{1}{m} \sum_{j=1}^{m} \frac{b_{ji}^2}{m} \sum_{k=1}^{m} b_{kj}^2 \right]^2, \]  

(12)

where \( b \) is a factor of \( \mathbf{B} \) and \( V_i \) represents the deviation of the square of factors of \( \mathbf{\hat{p}}_i \). A small value of \( V_i \) means small deviation of factors of \( \mathbf{\hat{p}}_i \), indicating \( \mathbf{\hat{p}}_i \) can be an initial perturbation with a large spatial spread. Therefore, we can obtain a set of initial perturbations with a spatial spread by minimizing the sum of \( V_i \). For the 10 perturbed runs, the first to fifth column vectors of \( \mathbf{B} \) are
selected, and $E_i^{-1/2}$ are operated on them to obtain initial perturbations. Finally, by adding positive and negative perturbations to an initial condition of the non-perturbed run, 10 initial fields are produced. When adding the initial perturbations, the initial amplitudes are adjusted so that the maximum zonal or meridional wind speed is 6.0 m s$^{-1}$.

As the kinetic energy is dominant in the initial SVs, and the wind perturbations are found to be sensitive to the ensemble spread of tracks, we selected wind speed as a variable to control the amplitudes of the initial perturbations. The value of 6.0 m s$^{-1}$ was obtained through the preliminary tests of TEPS, where the ensemble spread of tracks was statistically verified. However, as it is arbitrary and totally dependent on the NWP system, we will need to adjust the value again when the system is changed.

Table 1 summarizes the specifications of TEPS. Note that JMA operates one-week EPS (WEPS; WMO 2008b). While WEPS is designed for medium-range forecasts, both TEPS and WEPS have similar characteristics: they use the same NWP model, the same initial condition for the nonperturbed run, and the similar procedure to make initial perturbations based on a SV method. This is for the efficient development and maintenance of the operational NWP systems. In principle, we could merge those two systems, considering that the other NWP centers like ECMWF are successful in both medium-range forecasts and TC track forecasts with a single EPS, and that TEPS and WEPS follow the ECMWF’s EPS (e.g., Buizza 1994; Buizza and Palmer 1995; Palmer et al. 1998; Barkmeijer et al. 2001; Puri et al. 2001) in many aspects, especially in the calculation of initial perturbations based on a SV method. The merger of the systems might yield further benefits through saving computer resources.

3. Performance of the TEPS

a. Case studies

Figure 5 shows the examples of TEPS. The top figures are for Typhoon Maria, initiated at 1200 UTC 6 August 2006, and the bottom figures are for Typhoon Chaba, initiated at 1200 UTC 28 August 2004. The left panels show the track prediction by JMA/GSM (solid) with the best track (dash), and the right panels show all tracks by TEPS (these are the results of the preliminary test of TEPS before the quasi-operational TEPS). In Maria’s case, a large ensemble spread exists: some of the ensemble members predict the same scenario as JMA/GSM that says Maria heads toward the west of Japan. On the contrary, the other members recurve and head for the Kanto region. In reality, as the best track shows, Maria
recurved and hit the south of Kanto area. It is noteworthy that TEPS had captured the possibility of the best track. From a perspective of disaster prevention or mitigation, it is of great importance to figure out all possible scenarios in advance and take measures as needed. TEPS is expected to capture such a potential spread of tracks. As opposed to Maria’s case, Chaba’s case illustrates that the ensemble spread is quite small, which means the confidence of the prediction is relatively high. In fact, the deterministic prediction by JMA/GSM was almost perfect. Just like those two cases, TEPS is expected to provide the confidence information of track prediction based on the ensemble spread that could vary from TC to TC and initial time to initial time.

b. Quasi operation

To evaluate the performance of TEPS, we conducted quasi-operational runs of TEPS from May to December 2007 and verified the ensemble mean track predictions and a relationship between the ensemble spread of tracks and the position error of ensemble mean. Note that the specifications of the quasi-operational TEPS are different from those of the operational TEPS in several points. For example, as Fig. 2 shows, the analysis field for the quasi-operational TEPS comes from a global analysis for JMA/GSM with a resolution of TL319L40 because TL959L60 JMA/GSM was implemented on 21 November 2007 (note that TL319L40 JMA/GSM had been operated even after 21 November 2007 in order to provide the analysis field for the quasi-operational TEPS, which means the quasi-operational TEPS uses the global analysis of TL319L40 JMA/GSM throughout the quasi-operational period).

c. Ensemble mean track predictions

Figure 6a shows the position error of ensemble mean track predictions for TCs of tropical-storm-or-greater intensity. Figure 6b includes the extratropical-transition stages of the verifying TCs. The verifications are based on the best-track data produced by RSMC Tokyo-Typhoon Center. The position error of the ensemble mean (MEAN) is smaller than those of the nonperturbed runs (CTL) in 4- and 5-day predictions. For predictions up to 3 days, they have almost the same performance. The error reduction in 5-day predictions shown in Fig. 6a is 40 km. A t test performed with a confidence level of 95% shows that differences between MEAN and CTL is significant for 5-day predictions (Figs. 6c,d). JMA plans to provide public users with 5-day track forecasts. TEPS is expected to improve the deterministic TC track forecasts over the extended forecast range (note that JMA/GSM is operated 4 times a day, 0000, 0600, 1200, and 1800 UTC, but the forecast range is 84 h except for 1200 UTC, where the forecast range is 216 h).

d. Confidence information

Any variable that is correlated with the position error of ensemble mean can be used as an indicator of the position error. Considering the sum of the spread of every 6 h as a variable, we verified a relationship between the 5-day prediction error of ensemble mean and the spread of tracks of 5-day prediction (Fig. 7). Here, the ensemble spread is defined by the sum of the spread of every 6 h from initial time to 5 days so that the spread represents the uncertainty of tracks, not positions. The verifying TCs are exactly same as those for Fig. 6b, and each dot...
gives the verification result of each prediction event. When the ensemble spread is relatively small, the position error of the corresponding prediction event is also small and, more importantly, there is no case that has a large position error. When the ensemble spread is relatively large, there is a possibility that the position error becomes large. Note that the large ensemble spread does not guarantee large position error, but a possibility of it.

While Fig. 7 is the verification only for 5-day predictions, the same relationship can be seen in verifications for other prediction times. Based on this relationship, we classify the confidence of TC track prediction and assign a confidence index, A, B, or C every 6 h from the initial time to 5 days, where A, B, and C represent categories of the highest, the middle level, and the lowest confidence, respectively, and the frequency of each category is set to 40%, 40%, and 20%, respectively. As shown in Fig. 8, when collecting all predictions judged as A (ensemble spread is relatively small) and then verifying the position errors of the ensemble mean, it proves that the average position errors are considerably small compared with those of B. On the contrary, when collecting predictions judged as C (ensemble spread is relatively large) and then verifying the position errors of
the ensemble mean, the average position errors are larger than those of B. The position error of A is better than B by 14%, 25%, 34%, 40%, and 39% at 24, 48, 72, 96, and 120 h, respectively. On the other hand, the position error of C is worse than B by 96%, 98%, 93%, 55%, and 28% at 24, 48, 72, 96, and 120 h, respectively.

Another way to explain Fig. 8 is that the solid line in Fig. 6b showing the position errors of ensemble mean is divided into three lines only based on the information of ensemble spread. All cases at a certain prediction time in Fig. 6b are divided into three categories based on the ensemble spread, and the position error of the ensemble mean is verified at each category. Therefore, the three lines that are clearly separated in Fig. 8 imply that the ensemble spread of TEPS can be an indicator of position error.

Figure 9 gives an example of a possible application using the above confidence information. JMA plans to use the ensemble spread of TEPS to optimize the size of probability circle, which has been used in JMA’s operational track forecasts to express forecast uncertainty, and whose size is currently based on the statistics of the recent year’s verifications of JMA/GSM.

The reason why each category is set to 40%, 40%, and 20%, not 33%, 33%, and 33%, is to split the three lines in Fig. 8 as much as possible. Figure 10 shows the position error of each 3-day prediction in 2007 by JMA/GSM, and the error is sorted in an ascending order. As Fig. 10 shows, the distribution is not uniform, and the number of cases that have relatively large position errors is about 10%–20% of the total number of cases. Therefore, we make the rate of category C smaller than that of categories A and B.

A position error can also be broken into along-track (AT) and cross-track (CT) components (e.g., Goerss 2000; Aberson 2001). Accordingly, an ensemble can be used to predict the spread in each of these components around the ensemble mean. Here, we verify whether a relationship exists between the ensemble spread in the
AT and CT directions and their associated errors. First, the AT direction is defined at a given time $T$ as the vector difference between ensemble mean forecasts at $T$ and $T-1$ day. Then, the CT direction is set to the orthogonal direction against the AT direction. Figures 11a,b show the AT and CT direction verifications. Like Fig. 8, we prepare three confidence indices: A, B, and C with the frequency being 40%, 40%, and 20%, respectively. As both figures show, three lines are clearly split, indicating the fact that confidence information would be applicable to the uncertainty forecasts in the AT and CT directions.

4. Discussion

Through the quasi-operational TEPS, we found two things: one is that the position error of the ensemble mean is better than that of the nonperturbed run for predictions longer than three days; the other is that the ensemble spread can be used as an indicator of position error. As we mentioned in section 3d, optimizing the size of the probability circle based on TEPS is now under development. Quantitative verification with
respect to the current system will be future work. However, considering the verification results shown in
Figs. 7, 8, and 11, the ensemble spread of TEPS shows promise in providing the uncertainty forecasts. For ex-
ample, consider a probability circle, which represents that a TC is expected to move into the circle with a
probability of 70% at a certain forecast time. Based on
the ensemble mean of TEPS in 2007, the radius of the
circle is 335 km for 3-day predictions and 489 km for
5-day predictions on average. In the case of A, where
ensemble spread is relatively small, however, the radius
is 207 and 347 km for 3- and 5-day predictions, which
are smaller than the average radius by 38% and 29%,
respectively. On the other hand, in the case of C, where
ensemble spread is relatively large, the radius is 527 and
675 km for 3- and 5-day predictions, which are larger than
the average radius by 57% and 38%, respectively, while
in the case of B, the radius is almost the same as the av-
erage radius, 348 and 499 km for 3- and 5-day predictions.

As for deterministic predictions, on the other hand,
the following two questions still remain. First, how does
the skill of the ensemble mean of TEPS compare to that
of JMA/GSM, and second, how does the skill of the ensemble mean of TEPS compare with respect to that of
the multimodel EPS? Figure 12a shows the preliminary
verification result of the ensemble mean of the opera-
tional TEPS as compared with JMA/GSM. The verifi-
cation is based on the 1st (NEOGURI) to 18th (BAVI)
TC in 2008 analyzed by the RSMC Tokyo-Typhoon
Center, and like Fig. 6b, the extratropical-transition
stages are included in the verification. Note that the
number of cases decreases after 90 h because JMA/
GSM predicts over 90 h only at 1200 UTC. A \textit{t} test
shows that the ensemble mean of TEPS is significantly
worse than JMA/GSM up to 84 h (Fig. 12b). Considering
the differences of specifications between the quasi-
operational and operational TEPS, the difference of
analysis field might cause the decline of the performance.
Though the horizontal resolutions of the model and data
assimilation are the same in the quasi-operational TEPS,
that is TL319, we interpolate the analysis field from
TL959 to TL319 in the operational TEPS. As we have
not seen such decline of the performance in the quasi-
operational TEPS, and it would be ideal that the model
and data assimilation scheme have the same resolution,
we believe that there is room for improvement in the
technique of the interpolation. When the ensemble
mean of TEPS is compared with that of multimodel
EPS, the improvement is limited after 3-day predictions.
In the multimodel EPS presented by Goerss (2000),
Goers et al. (2004), and Komori et al. (2007), the en-
semble mean of the multimodel EPS is better than the
best of the individual models from the early prediction
time, say 24 h. This result may imply that dealing with
the uncertainty of NWP model and accounting for
model errors in ensemble techniques would result in the
improvement of ensemble mean tracks in TEPS, especially for the early prediction time.

5. Summary and conclusions

JMA started an operation of a new EPS, the Typhoon EPS, in February 2008, aiming to improve both deterministic and probabilistic TC track forecasts. TEPS is composed of 11 integrations with a TL319L60 global model and operated for TCs in the western North Pacific and South China Sea. The method to create initial perturbations is based on a singular vector (SV) method, but a method to deal with the uncertainty of the NWP model itself is yet to be considered.

The SVs are calculated targeting both TCs (up to three TCs at one prediction event) and a midlatitude area in the RSMC Tokyo-Typhoon Center’s responsibility area. For the 10 perturbed runs, 5 initial perturbations, which are added to and subtracted from an analysis field, are created by linearly combining the SVs. Coefficients are determined so that the spatial distributions of the perturbations are sufficiently spread. Moist SVs are calculated for TCs, and dry SVs are calculated for the midlatitude. The growth rate of both moist and dry SVs is based on a moist total energy norm. The evaluation interval time is 24 h for both types of SVs. The composites of vertically accumulated total energy of moist initial SVs have an asymmetric pattern and the most sensitive area is approximately 500 km away from the center of TCs. The total energy of moist initial SVs is mainly explained by the kinetic energy component, which leads to large ensemble spreads in TC track predictions.

Through the verifications of the quasi-operational TEPS, we found two benefits. First, the position error of ensemble mean is better than that of nonperturbed run for predictions beyond 3 days; the other is that the ensemble spread can be used as an indicator of position error. For 2008 when TEPS was in operational use, however, it was also found that the ensemble mean was significantly worse than the deterministic model (JMA/GSM) out to 84 h. Beyond 84 h, JMA/GSM was still better than the ensemble mean of TEPS though the sample size was small.

Future work includes the development of the application for uncertainty forecasts, using the ensemble spread of TEPS. The interpolation technique of the analysis field from TL959L60 to TL319L60 and the introduction of model errors into the ensemble perturbation technique would also be issues to be addressed. In addition, the prediction range of TEPS, 132 h, allows us to make up the lagged average forecast (LAF); hence, verifying the impact of LAF technique would also be a challenge for the future.

Acknowledgments. Special thanks are given to Sharan Majumdar of the University of Miami for a thorough and constructive review of the manuscript, and to Edward Fukada of the Joint Typhoon Warning Center and Buck Sampson of the Naval Research Laboratory for their helpful comments on multimodel EPS. We also thank Hirokatsu Onoda, Toshiyuki Ishibashi, Ko Kiozumi, and Masashi Nagata of JMA for their helpful comments and review of the manuscript. This work was supported by Grant-in-Aid for Scientific Research (A) 19201037. The manuscript was revised and completed under the support of the Office of Naval Research Grant N000140810250.

REFERENCES
