A GRAPES-based mesoscale ensemble prediction system for tropical cyclone forecasting: configuration and performance

Xubin Zhang*

CMA/Guangdong Provincial Key Laboratory of Regional Numerical Weather Prediction, Institute of Tropical and Marine Meteorology, CMA, Guangzhou, China

*Correspondence to: X. Zhang, CMA/Guangdong Provincial Key Laboratory of Regional Numerical Weather Prediction, Institute of Tropical and Marine Meteorology, Guangzhou 510640, China.
E-mail: xbzhang@grmc.gov.cn

To improve short-range tropical cyclone (TC) forecasting in China from both deterministic and probabilistic standpoints, a mesoscale ensemble prediction system (TREPS) based on the Global/Regional Assimilation and Prediction System (GRAPES) is constructed as a planned operational system. TREPS comprises one unperturbed and 30 perturbed forecasts ranging up to 60 h at 0.09° horizontal resolution. Downscaling perturbations from the European Centre for Medium-range Forecasts’ (ECMWF) ensemble prediction system (ENS) are used to generate initial perturbations by blending them with balanced random perturbations implemented around the TC and to generate perturbations for lateral boundary conditions. The surface temperature around the TC is also perturbed, with a combination of multiphysics and stochastically perturbed parametrization tendencies considered to represent uncertainties in the model physics. These perturbation methods contribute a full-area, multi-scale and multi-variable mutual complement to increasing forecast perturbations.

TREPS is evaluated for 19 TCs making landfall in China during 2014–2016. For deterministic guidance, TREPS significantly improves TC forecasting for intensity, rainfall and wind compared to ECMWF deterministic and ensemble forecasts. TREPS demonstrates significant advantages over its deterministic forecasts in predicting track, intensity, (extremely) heavy rainfall and wind. Furthermore, TREPS effectively reduces the degree of underdispersion in intensity forecasting compared to ENS. For probabilistic forecasting, TREPS outperforms ENS in accurately predicting extremely heavy rainfall and (extremely) strong wind owing to its better discriminating ability, but performs worse in predicting modest wind, light and heavy rainfall. Two post-processing techniques – optimal member and optimal probability – to improve, respectively, deterministic and probabilistic guidance of TREPS in rainfall and wind forecasting are proposed. The former is shown to be effective only for light rainfall, while the latter demonstrates consistent (partial) improvements in rainfall (wind). The superiorities and inferiorities of TREPS, as well as possible reasons for these, are intuitively illustrated in a case-study for Rammasun.

Key Words: ensemble prediction system; mesoscale; tropical cyclone; rainfall; wind

Received 16 July 2017; Revised 17 October 2017; Accepted 1 December 2017; Published online in Wiley Online Library 25 January 2018

1. Introduction

Improvements in numerical weather prediction (NWP) models, observing systems and data assimilation techniques have significantly improved the forecasting of tropical cyclone (TC) tracks over the last two decades (Rappaport et al., 2009; Goldenberg et al., 2015; Gopalakrishnan et al., 2016). However, the forecasting of TC intensity has shown slow improvement and remains a challenge (Yu et al., 2013; DeMaria et al., 2014; Cangialosi and Franklin, 2016). Because hazardous conditions associated with TCs, such as strong wind, heavy rainfall and storm surge, are usually related to TC intensity (Nishijima et al., 2012; Chen et al., 2013), improving the skill of TC intensity and related wind and rainfall forecasting is of particular importance and urgency.

In recent years, much research has gone into developing operational NWP models at fine resolution and new techniques for assimilating inner-core observations, and optimizing ensemble...
forecasts of TC intensity (Rogers et al., 2013). As TC intensity is primarily determined by small-scale dynamics and moist processes, improvements in intensity prediction and structural realism as a result of increased horizontal resolution have been demonstrated (Davis et al., 2010; Gopalakrishnan et al., 2012; Goldenberg et al., 2015).

The chaotic nature of TCs means that their intrinsic predictability – and therefore, the accuracy of their forecasting – is limited. Additionally, deficiencies in the current data assimilation methods and observation systems, as well as imperfections in the present NWP model dynamics and physics configuration, lead to inaccurate initial conditions (ICs) and imperfect representations of TC evolution for NWP models. These factors underlie the inevitable uncertainties in TC forecasting (Lorenz, 1963; Tribbia and Baumhefner, 1988) and result in limited information in single (deterministic) forecasts (Wilks, 2006). Ensemble forecasting is an important and useful approach for coping with such forecast uncertainties (Epstein, 1969; Leith, 1974; Wilks, 2006). The mean forecasts of a properly designed ensemble prediction system (EPS) are often superior to the deterministic forecasts from any individual member of the EPS (Leith, 1974; Toth and Kalnay, 1997); furthermore, EPS can provide useful quantitative information (e.g. ensemble spread and probability) for estimating forecast uncertainties (confidence) and the probability of weather system evolution (Murphy, 1990; Tracton and Kalnay, 1993).

To date, the application of EPS has produced encouraging improvements in TC forecasting. Ensemble mean track forecasts of TCs significantly outperform deterministic forecasts or forecasts from the best individual member of the EPS (Goess et al., 2004; Rappaport et al., 2009; Yamaguchi et al., 2009). The advantages of ensemble mean forecasts of TC intensity over those produced by individual members have also been demonstrated (Weber, 2005; Sampson et al., 2008; Krishnamurti et al., 2011; Yu et al., 2013), although this advantage is not as evident in the prediction of TC tracks. For the promising benefits of ensemble forecasting technique on TC forecasting, the developments of EPS, especially at high resolution, become increasingly popular and important (Hamill et al., 2012; Gall et al., 2013).

TCs are one of the most serious types of natural hazards and cause severe damage in China each year (Xu et al., 2015). However, TC forecasting in China using NWP models lags behind other advanced operational centres, e.g. the European Centre for Medium-range Forecasts (ECMWF) (Wang, 2016). In particular, research on ensemble forecasting of TCs in China is still in its inception (Tan and Liang, 2012; Tu et al., 2014; Wang, 2016) and operational ensemble forecasting of TCs has not been fully implemented (Tan and Liang, 2012). However, the Numerical Weather Prediction Center of the China Meteorological Administration (CMA) implemented a regional EPS in 2014 (Zhang et al., 2015). This EPS, named GRAPES-REPS, is based on a regional version of the Global/Regional Assimilation and Prediction System (GRAPES_Meso: Xue and Liu, 2007; Chen et al., 2008) and primarily covers land areas in China with a horizontal resolution of 0.15°. Nevertheless, applications of this EPS still cannot meet the requirements of operational refined weather services for TC prediction, necessitating the development of a high-resolution regional EPS for TCs.

In 2006, the Institute of Tropical and Marine Meteorology (ITMM) of CMA established a new tropical regional atmosphere model for the South China Sea (TRAMS) based on the GRAPES framework. During the past few years, TRAMS has shown TC forecasting results comparable to some advanced operational global NWP models (Lei et al., 2016), especially with respect to TC track prediction. At present, TRAMS has become one of the main operational TC forecasting models used in China (Yu et al., 2013; Chen et al., 2014). To further improve the short-range forecasting of TC track and intensity, particularly of the strong wind and heavy rainfall caused by TC landfall, a high-resolution version of TRAMS (TRAMS9KM) and a corresponding mesoscale EPS (TREPS), in which the horizontal resolution has been updated from 0.36° to 0.09°, have recently been developed. Based on its performance during its experimental phase (2014–2016) and the need to implement it as an operational weather service, ITMM plans to commence TREPS operation by the end of 2017.

To the best of our knowledge, issues regarding the development of high-resolution (~10 km) mesoscale regional EPS for use in operational TC forecasting in China have not been well investigated. This study attempts to assess the performance of TREPS in TC forecasting in China over the past 3 years, with a particular eye to evaluating whether its performance is competitive with or superior to that of ECMWF, whose products are used most frequently in operational TC forecasting for forecasters. To achieve this objective, a batch experiment covering 2014–2016 and a super-typhoon case-study are carried out. The TREPS configuration and an overview of the 19 TCs studied here are described in section 2. Section 3 provides some characteristics of TREPS perturbations. Results of the batch experiment and case-study are presented in sections 4 and 5, respectively. Section 6 concludes the article and provides further discussion.

2. Methodology

2.1. Forecast model

The NWP model used in TREPS is TRAMS9KM (Chen et al., 2014), a non-hydrostatic regional model that adopts a semi-implicit, semi-Lagrangian scheme for integration over time. It employs a horizontal grid designed on a longitude–latitude mesh with Arakawa C-grid staggering, including 385 x 305 horizontal grid points with horizontal resolutions of 0.09° x 0.09°. The centre of the model domain is located at the position of the TC centre, which is determined in the official real-time warning information from CMA (WARNING, hereafter). The vertical coordinate is terrain-following based on Charney–Phillips vertical layer skipping (Charney and Phillips, 1953); 55 vertical layers up to 35 km are used. The main forecast variables include zonal and meridional velocity (U, V), Exner pressure (Pi), potential temperature (θ), and specific humidity (Q). The physics parametrization schemes used include the Weather Research and Forecasting (WRF) Single-Moment 6-class (WSM6) microphysics scheme (Hong and Lim, 2006), the five-layer thermal diffusion (SLAB) land surface model (Dudhia, 1996), the rapid radiative transfer model (RRTM) long-wave radiation scheme (Mlawer et al., 1997), and the Dudhia short-wave radiation scheme (Dudhia, 1989).

2.2. TREPS configurations – basic set-up

Because the objective of TREPS is to improve the skill of short-range TC forecasts during landfall, and computing resources are limited, TREPS operation will not be initiated until a TC approaches mainland China, and is suspended after the TC makes landfall. TREPS issues 60 h forecasts twice per day at 0000/1200 and 0600/1800 UTC can both be used to construct forecasts up to 36 h. The main forecast variables include wind speed, viz. $V_{\text{max}}$ (maximum sustained wind speed); and central pressure, viz. $P_{\text{min}}$. The former becomes less (greater) than 1000 hPa (10 m s$^{-1}$). For a TC forming over the western North Pacific (WNP), TREPS initiates if the following two conditions are met:

- The forecasted minimum central pressure of the TC is predicted to become less than 1000 hPa within the next 12 h.
- The forecasted maximum wind speed of the TC is predicted to reach or exceed 10 m s$^{-1}$ within the next 12 h.

The 60 h forecasts are used for constructing the initial conditions of the 120 h ensemble forecasts, which are used for ensemble forecasting. The ensemble size is 10 members, and the ensemble spread is determined by perturbing the initial conditions and the model parameters. The ensemble spread is used to estimate the forecast uncertainties, which are then used to assess the performance of TREPS. The ensemble forecasts are used to construct the ensemble mean forecasts, which are used as the initial conditions of the 120 h forecasts.

The 120 h forecasts are used to construct the final forecasts, which are used for operational TC forecasting. The final forecasts are based on the ensemble mean forecasts and the ensemble spread. The ensemble mean forecasts are used to construct the median forecasts, which are used as the initial conditions of the final forecasts. The ensemble spread is used to estimate the forecast uncertainties, which are then used to assess the performance of TREPS. The final forecasts are used for operational TC forecasting.

In summary, the methodology of TREPS involves using the TRAMS9KM model to produce 60 h forecasts, which are used to construct the initial conditions of the 120 h ensemble forecasts. The ensemble forecasts are used to construct the ensemble mean forecasts, which are used as the initial conditions of the 120 h forecasts. The 120 h forecasts are used to construct the final forecasts, which are used for operational TC forecasting. The methodology of TREPS is designed to improve the skill of short-range TC forecasts during landfall, and the ensemble spread is used to estimate the forecast uncertainties, which are then used to assess the performance of TREPS.
satisfied simultaneously: (i) when the TC is reported to pass over the 48 h warning line prior to landfall at China (0°N, 105°E; 0°N, 120°E; 15°N, 132°E; 34°N, 132°E) and is predicted to make landfall on mainland China within 60 h by CMA, and (ii) when the TC $P_{\text{min}}$ ($V_{\text{max}}$) reported in WARNING is less (greater) than 1000 hPa (10 m s$^{-1}$). Here, only one of the conditions of $P_{\text{min}}$ and $V_{\text{max}}$ is satisfied $\delta$.

The control (deterministic or unperturbed) forecasts of TREPSS (DETER, hereafter) are cold started, with the 0.125° × 0.125° HRES analyses and forecasts used as ICs and lateral boundary conditions (LBCs), respectively. The simplified Arakawa–Schubert (SAS) cumulus parametrization (CU) scheme (Pan and Wu, 1995; Han and Pan, 2011) and the medium-range forecast model (MRF) planetary boundary layer (PBL) scheme (Hong and Pan, 1996) are used in DETER.

2.3. TREPSS configurations – perturbation generation

The uncertainties in ICs, LBCs and model physics processes are all considered in generating the perturbed ensemble members of TREPSS. Downscaling perturbations are calculated by subtracting the ensemble mean of the first 30 perturbed ensemble members of the 0.5° × 0.5° ENS forecasts from each perturbed ensemble member and then interpolated to the TRAMSMKDM domain.

The initial perturbations are generated by a linear combination of downscaling perturbations (ICDp; with a weighting factor $\alpha$) and balanced random perturbations (BRp; with a weighting factor $\beta$), representing, respectively, the large-scale and mesoscale uncertainties in the TC. These perturbations are added to DETER ICs to construct perturbed ICs. $U$, $V$, $\Pi$, $\theta$ and $Q$ of ICs are all perturbed in this manner. $\theta$ and $Q$ are left unperturbed if the perturbed $\theta$ reaches an unperturbed value or exceeds a critical supersaturation value. BRp (Barker, 2005; Zhang, 2005; Schwartz et al., 2014) are produced by taking Gaussian random draws within zero mean and covariances from the regional background error covariances (B) provided by the GRAPES three-dimensional variational (3D-Var) system (Xue et al., 2008). The variances, horizontal correlation scales, and parameters of the vertical correlation function of B are all tuned to meet the needs of proper representation of the mesoscale uncertainties in TC. ICDp are implemented over the entire domain and at all vertical levels, while BRp are used only around the TC centre and in the troposphere. Specifically, BRp are switched on at the grid points at which the following two conditions are satisfied simultaneously: (i) the distance between the grid point and the TC centre reported in WARNING is smaller than the length $\gamma$, and (ii) the difference of equivalent potential temperature between 1000 and 500 hPa, which is calculated using the unperturbed ICs and used as the representative of convective instability, is larger than 10 K. Meanwhile, BRp are switched off near the model top (<50 hPa) and multiplied by a factor that is set equal to 0 near 50 hPa and linearly increases until it approaches 1 near 100 hPa. BRp are used to initialize the mesoscale perturbations in the significant convectively unstable region around the TC, and will evolve to become flow-dependent under the constraints of atmospheric dynamics in the NWP model after several hours of integration (Zhang, 2005). Specifically, in TREPSS, BRp will generally grow to be coherent and organized after about 6 h forecasts, and reflect the large uncertainty of wind where background wind has large horizontal gradient (not shown), representing the mesoscale uncertainty around the TC inner core to some extent.

Downscaling perturbations (multiplied by a tuning factor $\varepsilon$) are added to DETER LBCs in intervals of 6 h to generate the LBCs of the perturbed forecasts (LBCDp). The surface temperature is also perturbed to take into account the uncertainties from the sea-surface temperature, which is used as the surface boundary condition. Specifically, a perturbed surface temperature is produced by adding Gaussian random numbers with the mean of 0 and standard deviation of $\delta$ to the DETER surface temperature. The surface temperature perturbations (TSp) are switched off unless the distance between the grid point and TC centre reported in the WARNING valid at the initial time is smaller than the length $\gamma$. Several further steps are included in TSp implementation: (i) the perturbations are multiplied by a factor that is equal to 0 at distance $\gamma$ and increases linearly to 1 at the wind radius of 17.2 m s$^{-1}$ (gale radius, hereafter) reported in WARNING; (ii) the amplitude of TSp is restricted to the range $[-\lambda\delta, \lambda\delta]$ to limit excessive perturbations; and (iii) the preceding steps are only implemented at the initial time and keep perturbations unchanged during the model integration, resulting in constant forcing from the surface boundary to the atmosphere.

A combination of multi-physics (MP: Houtekamer et al., 1996; Stensrud et al., 2000; Hacker et al., 2011) and stochastically perturbed parametrization tendency (SPPT: Buizza et al., 1999; Palmer et al., 2009; Boustier et al., 2012; Berner et al., 2015) schemes are used to simulate model physics uncertainties arising from imperfections in the parametrization of subgrid-scale processes.

Based on previous tests of both operational performance and numerical integration stability, CU parametrization schemes, including Kain–Fritsch (KF: Kain and Fritsch, 1990) and SAS, and PBL parametrization schemes, including MRF and Yonsei University (YSU: Hong et al., 2006), are used to construct four combinations of physics packages for parametrizing the CU and PBL processes in the MP schemes. Among the 30 perturbed ensemble members, 8 members are selected at random to use SAS/MRF scheme Ds, as is the case in DETER; of the remaining members, 8, 7 and 7 are selected randomly to use SAS/YSU, KF/MRF and KF/YSU schemes, respectively.

In the SPPT schemes used here, perturbations with a random mesoscale pattern are added to the total parametrized tendency of physical processes for variables including $U$, $V$, $\theta$ and $Q$. In this process, a random field $r$ drawn from a Gaussian distribution with mean zero, standard deviation 0.5, spatial correlation scale $\kappa$, and temporal correlation scale $\tau$ is produced during model integration. Here, $r$ is bounded within the range of ±2 standard deviations to avoid unrealistic perturbations causing numerical instability. Next, the physical tendency is multiplied at each time step by $f = 1 + r$ to form perturbed tendency. To avoid the problems mentioned in Palmer et al. (2009), no tendency perturbations are applied near the surface (below about 100 m above ground) and near the model top (<30 hPa), and the perturbations in the transition layers (100–1200 m and 100–500 hPa) are smoothly ramped up to full amplitude. Additionally, perturbations of $Q$ are checked in a manner similar to the initial perturbations to determine whether to implement tendency perturbations to both $\theta$ and $Q$.

2.4. TREPSS configurations – parameter settings

Following the application of the process above, several key parameters used in the generation of perturbations remain to be determined. However, there is currently no clear theoretical or universally applicable guidance for setting such parameters. For example, the three critical parameters (i.e. standard deviation, $\kappa$ and $\tau$) with primary impact on the performance of SPPT show apparent differences in settings between NWP models with different dynamical cores and resolutions (Palmer et al., 2009; Boustier et al., 2012; Lang et al., 2012; Romine et al., 2014; Berner et al., 2015). Furthermore, previous sensitivity tests have revealed considerable impact of parameter settings for nearly all parameters mentioned above (i.e. $\alpha, \beta, \gamma, \varepsilon, \delta, \lambda, \kappa$ and $\tau$) on TC forecasting (appendix). As a preliminary solution to parameter settings, a pragmatic approach based on WARNING in conjunction with tuning experimentation based on performance assessment of several case-studies (appendix) is proposed (Figure 1). WARNING is used to provide parameter settings with the TC information at the initial time and in the first 6 h of forecasts. Specifically, WARNING reports the number of TCs over SCS.

© 2017 The Authors. Quarterly Journal of the Royal Meteorological Society published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society. Q. J. R. Meteorol. Soc. 144: 478–498 (2018)
and WNP and their centre positions in terms of longitude and latitude, intensities in terms of $P_{\text{min}}$ and $V_{\text{max}}$, and scales in terms of $\text{gale}_\text{radius}$. The averaged initial $\text{gale}_\text{radius}$ over the 48 forecasts (about 300 km) was used to classify the TC scale, with the $\text{gale}_\text{radius}$ above (below) 300 km for a large (small) TC; the value of 980 hPa was selected to classify the TC intensity by reference to the national standard of China (see table 2 of Yu et al. (2013)), with $P_{\text{min}}$ above (below) 980 hPa for a strong (weak) TC; while the selection of 40 km to classify the TC extension was based on empirical estimation, with the change in $\text{gale}_\text{radius}$ from the initial time to 6 h later over (below) 40 km for an extended (unextended) TC.

If there is a single TC reported to influence the target area, the weighting factors for ICDp and BRp ($\alpha$ and $\beta$) are set to 0.8 and 0.6, respectively, the distance of double $\text{gale}_\text{radius}$ is used as the value of the maximum range of BRp ($\gamma$), and there is no tuning ($\epsilon = 1$) exerted on LBCDp. For a strong (weak) TC, the standard deviation of TSp ($\delta$) is set to 2.0 (4.0). The clipping parameter for TSp ($\lambda$) is set to 0.75 (1.0) to restrict the probably excessive perturbations for the strong extended (unextended) TCs; in the case of weak extended (unextended) TCs, the value of $\lambda$ is set to 1.5 (2.0). For large (small) TCs, the spatial correlation scale ($\alpha$) and temporal correlation scale ($\tau$) are set to 500 (300) km and 5 (3) h, respectively.

If there are multiple TCs reported in WARNING, the TC predicted by CMA to make landfall at mainland China in the shortest time is selected as the ‘target’ TC and its centre position is used as the centre of the model domain. The distances between the ‘target’ TC centre and those of the other TCs are calculated and compared with the ‘impacting’ length, which is defined as the sum of the other TCs’ $\text{gale}_\text{radius}$ and the half width of the model domain. If this distance does not exceed the ‘impacting’ length, the corresponding TC is identified as an ‘impact’ TC. In the case of more than one ‘impact’ TC, only the strongest is selected. Among the parameters that must be determined in the case of multiple TCs, $\alpha$, $\beta$, $\delta$, $\lambda$, $\epsilon$, and $\gamma$ are independent of the characteristics (e.g. intensity) of the TC and are set to 2.0, 0.8 and 2.0, respectively. In this situation, including more large-scale uncertainties from ENS in TREPS leads to better forecasting performance, which is the result of previous sensitivity tests (appendix). BRp and TSp are implemented for both ‘target’ and ‘impact’ TCs, with $\gamma$ set to the distance of triple $\text{gale}_\text{radius}$ for the corresponding TCs, $\delta$ set to 0.5, and $\lambda$ set to 2.0 (1.0) for extended (unextended) TCs. The two key parameters in SPPT schemes, i.e. $\kappa$ and $\tau$, are set based on the ‘target’ TC in the same way as in the case of a single TC.

2.5. Experimental design and cases overview

To evaluate the performance of TREPS, retrospective forecasts of TREPS were conducted during 2014-2016, when 19 TC cases (Figure 2) with 48 TREPS forecasts were collected. Most of these TCs formed in WNP and moved northwestward to make landfall on the southeast coast of China; several formed in SCS and mainly made landfall on the Guangxi (GX), Guangdong (GD) and Hainan (HN) provinces. There were some simultaneous TCs in the forecasts, including Linfa/Chan-hom, Chan-hom/Nangka, Meranti/Malakas and Haima/Sarika. As the strongest TC making landfall on mainland China during the past 3 years, super-typhoon Rammunus (in 2014) caused great damage in HN, GD and GX (Zhang et al., 2017). Following its genesis, Rammunus moved from WNP to SCS and arrived at SCS at 0000 UTC 16 July as a typhoon. Owing to favourable environmental conditions over SCS, Rammunus strengthened rapidly to become a super-typhoon with a peak maximum sustained wind speed of 72 m s$^{-1}$ at 0600 UTC 18 July. In view of the serious nature of the resulting disaster and its rapid intensification (RI), which is difficult to skilfully predict at present (Rogers et al., 2013), Rammunus was selected as the case-study.

2.6. Forecast verification

This study focuses on the performance of both deterministic and probabilistic guidance of TREPS. Specifically, the deterministic guidance provided by TREPS includes the ensemble-mean (EM) track (TC centre position) and intensity ($P_{\text{min}}$ and $V_{\text{max}}$) and the probability-matched mean (PM: Ebert, 2001; Schwartz et al., 2014) rainfall and 10 m wind; the probabilistic guidance includes the probability of rainfall and 10 m wind at different thresholds. To investigate the advantages of TREPS in TC forecasting, its deterministic guidance is compared with that of ENS, HRES and DETER, while its probabilistic guidance is
compared with that of ENS. To make an equitable comparison, only the first 30 perturbed ensemble members of ENS are used. All comparisons are based on the skill scores illustrated as follows.

In this study, the TC centre is defined as the minimum of the geopotential height at 850 hPa, while the lowest (highest) sea-level pressure (10 m wind speed) within 100 (250) km of the TC centre is used to define $P_{\min}$ ($V_{\max}$) (Torn, 2010). For each forecast time, the EM track and intensity are computed as, respectively, the average of the TC centre positions and $P_{\min}$ ($V_{\max}$) from the perturbed forecasts; these factors are only computed if at least two-thirds of the perturbed ensemble members predict $V_{\max}$ larger than 10 m s$^{-1}$ (Majumdar and Finocchio, 2010). The CMA best-track dataset (Ying et al., 2014; http://tcdata.typhoon.org.cn) was used to calculate the average absolute track error (Hamill et al., 2011) and intensity error over the set of 48 forecasts.

3. Perturbation characteristics

Knowledge of the perturbation characteristics of different perturbation methods for different perturbation sources is shown in Figure 2, and VP for the forecasts initialized at 0000 UTC 17 July is from 30 to 60 h. Both PM rainfall and 10 m wind were calculated within the VA.

To verify the rainfall and 10 m wind for EPS deterministic guidance as well as for the unperturbed forecasts, both the threat score (TS: Gilbert, 1884) and fraction skill score (FSS: Roberts and Lean, 2008) were used. The TS (FSS) was computed for the 6 h (1 h) accumulated rainfall in consideration of the more significant ‘double penalties’ due to the displacement error in the verification of high-resolution forecasts for heavy rainfall accumulated over a shorter period (Ebert, 2008). In calculating FSS, ‘neighbourhood length’ was defined as the length of area over which the fractions were computed. Probabilistic guidance verification was achieved by computing Brier score (BS) and its reliability aspect (Wilks, 2006), and the area under the relative operating characteristic curve (AROC: Mason and Graham, 2002).

For the comparison of skill scores, including the absolute track/intensity error, TS, FSS, BS, reliability, and AROC, statistical significances of the differences between different forecast sources were assessed using a bootstrap resampling procedure. Specifically, random samples of skill scores were generated with replacement, and then the differences of skill scores were calculated. This procedure was repeated 1000 times and bootstrapping was performed on the score differences between the paired samples. The rank at which the resampled score differences crossed zero was used as the significance level (Davis et al., 2010) to represent the probability that two skill scores were distinct. Here, a 90% significance level indicates a 90% probability that two skill scores differed.
important for the design of EPS. Recently, the time evolution of the spatial perturbation structure for different perturbation methods was investigated, revealing significant differences among these methods for TC cases (Yamaguchi and Majumdar, 2010; Lang et al., 2012). Moreover, Lang et al. (2012) indicated the importance of deciding how much variance should be introduced into EPS by each perturbation method in the design of EPS.

3.1. Design of parallel comparison experiments

As a preliminary attempt at establishing an operational mesoscale EPS based on TRAMS9KM, the ideas for the ensemble design of TREPS are simply incorporating perturbations from all uncertainty sources using some leading perturbation methods. Therefore, to confirm their suitability and improve future design of TREPS, the impacts of various perturbation methods on the total forecast perturbations around the TC were investigated. In addition to the retrospective forecasts (ALL, hereafter) described in section 2.5, six additional parallel comparison experiments, called noICD, noBR, noLBCD, noTS, noSPPT and noMP, with ICDp, BRp, LBCDp, TSp, SPPT and MP removed from ALL respectively, were carried out.

We used the same analysis method as that used in Lang et al. (2012). Specifically, the perturbations were defined as the differences between perturbed and unperturbed forecasts for $U$, $V$, temperature ($T$) and $Q$. The kinetic energy ($KE$), internal energy ($IE$) and latent heat energy ($LHE$) for perturbations were then expressed as follows:

$$ KE = \frac{1}{2} \int_0^{P1} \int_S (U'^2 + V'^2) \, dpds, $$

$$ IE = \int_0^{P1} \int_S \frac{C_p T'}{T_r} \, dpds, $$

$$ LHE = \int_0^{P1} \int_S \frac{L^2 Q'^2}{C_p T_r} \, dpds, $$

where $U'$, $V'$, $T'$ and $Q'$ denote the perturbations for $U$, $V$, $T$ and $Q$, respectively; $C_p = 1006.0 \text{ J kg}^{-1} \text{ K}^{-1}$ is the specific heat capacity at constant pressure; $T_r = 287 K$ is the reference temperature; $L = 2.5 \times 10^8 \text{ J kg}^{-1}$ is the latent heat of vaporization; and $\int_S$ and $\int_0^{P1}$ denote horizontal and vertical integrations respectively. These perturbation energies were all calculated in the ‘target’ domain within the distance of the double initial gale_radius of the TC centre in DETER. The differences of perturbation energies between ALL and the additional six experiments, i.e. noICD, noBR, noLBCD, noTS, noSPPT and noMP, were used to represent the relative contributions of ICDp, BRp, LBCDp, TSp, SPPT and MP to the total forecast perturbations, respectively.

3.2. Contribution to total forecast perturbations

The differences of perturbations, averaged over all ensemble members, all 48 forecasts, the domain within the distance of the double initial gale_radius, and the vertical levels, are shown in Figure 3. Overall, ICDp contributed the largest portion of the total forecast perturbation around the TC, followed by MP and LBCDp. However, the contribution of MP to perturbation $IE$ increased quickly and became the largest during model integration. For BRp, the contribution to perturbation $KE$ was significantly higher than that to $IE$ or $LHE$ at the initial time (Figure 3(b)), revealing some deficiencies in BRp, likely caused by the unsatisfactory balance constraints described in B used for generating perturbations; however, this contribution appeared to decrease significantly during model integration, because some random perturbations without dynamically unstable mode dissipate. Furthermore, TSp contributed generally the least to all three perturbation energies, especially perturbation $LHE$ (Figure 3(f)), indicating the probably undesirable implementation of TSp as a constant forcing around the initial TC centre. The contribution of SPPT was not as large as that in Lang et al. (2012), which might be attributable in part to the less than satisfactory parameter settings, e.g. standard deviation.

Composites of the azimuthally integrated perturbations energies were calculated and averaged over all ensemble members and all 48 forecasts. An azimuthal Fourier analysis of these perturbation energies was then calculated to investigate the scale and distribution characteristics of the relative contribution of different perturbation methods.

At the initial time, the contribution of ICDp was greater than that of BRp by and large. Generally, ICDp produced a positive contribution in most of the regions around TC and at nearly all scales, with its contribution most evident at large scales (wave-number $<3$) (Figure 3(a)). Compared with ICDp, BRp produced a substantially higher positive contribution to perturbation $KE$ outside the TC core (at radii $>40\%$ of gale_radius) at small scales (wave-number $\geq 3$) (Figure 3(b)). For perturbation $IE$ and $LHE$, BRp exhibited its primary contribution at large scales, especially near the TC core (at radii $<40\%$ of gale_radius). Thus, BRp provided the initial total perturbations with additional information, particularly at small scales, in addition to partial information similar to that of ICDp.

For the 36 h forecast total perturbations, the main positive contributions from both ICDp and BRp were located outside the TC core at most scales and near the TC core at small scales (Figures 3(c) and (d)). Around the TC eyewall (at 20–40\% of gale_radius), BRp contributed a more substantial small-scale perturbation $IE$ than ICDp. LBCDp chieflly contributed to the large-scale perturbations outside the TC core (Figure 3(e)); TSp and SPPT both led to significant contribution to the small-scale perturbation $LHE$ near the TC core (Figures 3(f) and (g)), which was missing in the contribution of LBCDp. SPPT also produced an evident contribution to the large-scale perturbation $KE$ near the TC core (Figure 3(g)). MP exhibited a broader distribution of positive contributions than the other perturbation methods, with the most noteworthy contributions to large-scale perturbations for $KE$ near the TC core, $IE$ around the TC eyewall, and $LHE$ outside the TC core (Figure 3(h)). Overall, different perturbation methods exhibited different contributions to the total forecast perturbations in different regions and at different scales, with the discrepant contributions complementing each other in some cases.

4. Results of batch experiment

The performances of both deterministic and probabilistic guidance of TREPS in the retrospective forecasts for 2014–2016 TC cases were verified and compared with those of ENS, HRES and DETER. Note that, in operational use, ECMWF forecast data are received at intervals of 6 h; therefore, the comparison between TREPS and ECMWF was conducted only for the verification of variables at 6 h intervals, while the verification comparison between TREPS unperturbed and perturbed forecasts was also carried out at 1 h intervals to investigate the impact of TREPS on high-resolution TC forecasting.

4.1. Deterministic guidance

The TREPS EM track forecasts were slightly better than those of HRES during the first 48 h of forecasts, and seemed to be comparable to those of ENS EM, with which there were no statistically significant differences during the overall forecast period (Figure 4(a)). For the insignificant degeneration of EM track forecasts of TREPS relative to ENS, the more serious bias...
in track forecasts for TRAMS9KM than for ECMWF is likely responsible (not shown). Furthermore, track forecasting was more skilful in TREPS EM than in DETER at nearly all valid times. In this study, the lead time at which TC made landfall often lay between 24 and 54 h according to the conditions that must be satisfied to issue TREPS. Neither TRAMS9KM nor ECMWF could well reproduce the intensities of intense TCs, because of their insufficient resolutions (Gentry and Lackmann, 2010). Thus, during the initial 36 h of forecasts when the TC was making landfall, large intensity errors of both ENS EM and TREPS EM were dominated by serious weak biases (Figures 5(b) and (c)), leading to the reduction of intensity errors with increasing forecast lead times (Figures 4(b) and (c)). For \( P_{\min} \) forecasts, TREPS EM showed smaller (larger) errors than the others in the first 36 h (beyond 36 h) of forecasts, which statistically was significant (not significant) (Figure 4(b)). These larger errors may be partially explained by the more serious bias in \( P_{\min} \) forecasts for TRAMS9KM than for ECMWF (Figure 5(b)), and partially related to the different characteristics of bias between different physics schemes used in MP. The most noticeable differences among the various forecasts were in the \( V_{\max} \) results, with TREPS EM exhibiting the smallest errors in the whole forecasts (Figure 4(c)).

Considering the bias in TC intensity forecasting, the calibrated forecasts were briefly calculated by subtracting the bias from the raw forecasts at each lead time. After calibrating, intensity errors before TC landfall decreased, with larger reduction in ENS than in TREPS (Figures 5(b) and (c)). However, calibrating degraded ENS EM \( P_{\min} \) forecasts after TC landfall when biases did not dominate intensity errors (Figure 5(b)). In general, intensity errors after calibrating were smaller in TREPS than in ENS.

Ensemble spreads of both ENS and TREPS were calculated according to the approach of Hamill et al. (2011) and averaged over the 48 forecasts. It is seen that both ENS and TREPS were overdispersive for track forecasting, with the former generally producing more overdispersions (Figure 5(a)). In intensity forecasting, TREPS exhibited greater spreads and smaller underdispersions than ENS, particularly in the forecasts of \( V_{\max} \) (Figures 5(b) and (c)). Although underdispersions of ENS decreased after calibrating, they were still more serious than those of TREPS. Thus, TREPS provided better estimation of uncertainties in TC intensity forecasting, despite biases in the forecasts.

Figure 3. Differences of the decomposition for perturbation \( KE \) (shaded; unit: \( m^2 s^{-2} \)) and \( LHE \) (thin black contour, solid for positive values, dashed for negative, and dotted for zero values; intervals are 0.5 and 2 for (b) and others, respectively; unit: \( J kg^{-1} \)) at 850 hPa, and for perturbation \( IE \) (thick green contour, solid for positive values, dashed for negative, and dotted for zero values; intervals are 1 and 2 for (b) and others, respectively; unit: \( J kg^{-1} \)) at 500 hPa between ALL and noICD, noBR, noLBCD, noTS, noSPPT and noMP, denoted by (a, c)‘ICDp’, (b, d)‘BRp’, (e)‘LBCDp’, (f)‘TSp’, (g)‘SPPT’ and (h)‘MP’ respectively. ‘00’ (a–b) and ‘36’ (c–h) indicate the differences at the initial time and the 36 h forecast time, respectively. The differences of the decomposition amplitudes are depicted as a function of wave number and radius, which is measured by the factor of \( g\text{a}_\text{le}_\text{r}a\text{d}u\text{s} \); the differences are scaled by factors of 10^5, 10^2 and 10^3 for perturbation \( KE \), \( IE \) and \( LHE \) respectively. Italic numbers in parentheses following the title represent the differences of perturbation \( KE \), \( IE \) and \( LHE \) averaged over the domain within the distance of the double initial \( g\text{a}_\text{le}_\text{r}a\text{d}u\text{s} \) over the levels from 1000 to 100 hPa.
Generally, EM underestimates the peak of heavy rainfall and overestimates the area of light rainfall, due to its simple averaging over all ensemble members (Du et al., 1997); the same is true for TC wind forecasting. PM was proposed by Ebert (2001) to improve the rainfall estimation from EM by blending the spatial pattern of EM with the frequency distribution of the whole EPS. Accordingly, PM rainfall and wind forecasts are also important products of TREPS. The PM field at a given time is produced in three steps. First, the frequency distribution is formed by pooling the rainfall (wind) amounts from all 30 ensemble members at each grid point in VA, ranking them from largest to smallest, and keeping every 30th value. Second, the EM rainfall (wind) amounts in VA are also ranked from largest to smallest. Third, the grid point corresponding to the largest EM rainfall (wind) amount is assigned the largest value in the frequency distribution, and so on.

Obviously, the accuracies of both the EM spatial pattern and EPS frequency distribution determine the accuracy of PM. However, EM has been shown to become physically unrealistic during model integration as a result of nonlinearity at both synoptic and convective scales (Ancell, 2013; Hollan and Ancell, 2015). Thus, a technique called ‘best member’ was proposed to cope with these issues in EM and PM. Ancell (2013) defined two best-member techniques: one uses the single member closest to EM over the whole forecasts, and the other uses a forecast patched together from the members closest to EM at each forecast time. It was determined that the latter method produced better forecasts of sea-level pressure; the latter technique was also used in Schwartz et al. (2014) for convective precipitation forecasting, but did not
yield significant improvements relative to PM and sometimes produced unrealistic convective evolution. It is therefore still not clear how to effectively design a best-member technique for precipitation forecasting. Recently, the short lead time error based on latest observations such as warning position was considered as a means for selecting ensemble members to determine the track consensus at long lead times (Goerss et al., 2004; Sampson et al., 2006; Qi et al., 2014). This attempt had proven effective in improving TC track forecasting. Based on the previous studies mentioned above, an ‘optimal member’ method for both rainfall and wind forecasting was verified in this study to explore how to best produce deterministic guidance from EPS.

Specifically, a single member was selected to be the optimal member based on the closeness of the perturbed ensemble members’ forecasts to both EM forecasts (for both track and intensity) and the latest observations of WARNING. Following Torn (2010), the optimal member was determined by minimizing the cost function:

$$J_i = \frac{1}{\mu_i} \left[ \frac{1}{t} \sum_{j=0}^{6} \mu_j \left\{ \frac{2}{T_j^2} \frac{T_j^i}{\sigma_T^i} + \left( \frac{P_j^i}{\sigma_P} \right)^2 + \left( \frac{V_j^i}{\sigma_V} \right)^2 \right\} \right]^2$$

$$+ \left\{ \frac{1}{t} \sum_{j=1}^{2} \frac{2}{T_j^2} \frac{T_j^i}{\sigma_T^i} + \frac{1}{p} \sum_{j=p+1}^{2} \left( \frac{P_j^i}{\sigma_P} \right)^2 + \frac{1}{w} \sum_{j=w+1}^{2} \left( \frac{W_j^i}{\sigma_W} \right)^2 \right\}$$

(4)

The first term on the right-hand side of Eq. (4) measures the closeness to the latest observations, where $T_j^i$, $P_j^i$, and $W_j^i$ denote absolute errors of the track, $P_{\text{min}}$ and $V_{\text{max}}$ respectively, against WARNING observations for member $i$ at lead time $j$. Here, only the forecasts at the initial and 6 h forecast times are measured, because of the 6 h delay time in operational use. Moreover, the term is tuned by the factor $\mu_i$ in light of larger uncertainties in the WARNING observations for weaker TCs. The second term on the right-hand side measures the closeness to EM forecasts, where $T_j^i$, $P_j^i$, and $W_j^i$ denote absolute differences of the track, $P_{\text{min}}$ and $V_{\text{max}}$ respectively compared with EM forecasts for member $i$ at lead time $j$. The forecasts of track, $P_{\text{min}}$ and $V_{\text{max}}$ were respectively measured from lead time $t_1$, $p_1$ and $w_1$ to $t_2$, $p_2$ and $w_2$ in intervals of 6 h, with the lead times chosen according to the evaluation results illustrated above. Clearly, the performances of both TREPS and ENS EM forecasts for TC track and intensity forecasting changed with lead time (Figure 4). It was naturally assumed that better performance indicates a higher similarity between EM forecasts and observations; therefore, only the lead times at which EM forecasts behaved better were chosen. For TREPS, $t_1$, $t_2$, $p_1$, $p_2$, $w_1$ and $w_2$ were set to 6, 54, 6, 36, 6 and 60, respectively. The closeness to the EM $V_{\text{max}}$ forecasts was not measured in the cost function for ENS, considering the poor performance (Figure 4(c)); and $t_1$, $t_2$, $p_1$ and $p_2$ were set to 6, 60, 54 and 60, respectively, with $t$, $p$ and $w$ used to represent the total number of lead times used in the calculation of the cost function. $\sigma_T$, $\sigma_P$ and $\sigma_W$ denote the standard deviations of the TC track (15 km), $P_{\text{min}}$ (4 hPa) and $V_{\text{max}}$ (1 m s$^{-1}$), respectively, which were determined empirically and used to normalize the various variables to maintain their mutual balance. If more than one member had a minimum cost function, optimal-member (OPT, hereafter) forecasts were produced by averaging over these members.

Figure 6 shows TS for 6 h accumulated rainfall and 10 m wind, where larger TS values indicate higher skill. It was not surprising that TS was nearly 0, or the differences among various forecasts were not very significant, during the initial 24 h of forecasts for both rainfall and wind with larger thresholds, as there were few observations or forecast samples for this period of TC making landfall.

For light rainfall (>0.1 mm per 6 h), TREPS PM had limited superiority, which was significant only in comparison with ENS PM (Figure 6(a)). Additionally, ENS OPT yielded the highest TS, followed by TREPS OPT. The superiority of TREPS OPT to TREPS PM was most evident beyond 24 h of forecasts, while its significant superiority to HRES, ENS PM and DETER was nearly ubiquitous for the whole forecasts. In view of heavy rainfall (>25 mm per 6 h), the highest TS was produced by TREPS.
PM, followed by TREPS OPT, although there was no significant difference between their results (Figure 6(b)). Both of them significantly outperformed the other forecasts (except ENS OPT) during most of the forecasts. The most noticeable improvements of TREPS PM over ENS OPT were found beyond 36 h of forecasts. Again, ENS OPT exhibited the highest TS among the ECMWF forecasts, indicating the effectiveness of the optimal-member method proposed here. For extremely heavy rainfall (>100 mm per 6 h), the TREPS forecasts behaved significantly better than the ECMWF forecasts (Figure 6(c)), probably due to the former’s higher resolution; the performance of TREPS PM was comparable to (better than) that of TREPS OPT (DETER).

The forecasts of 10 m wind exceeding 17.2 m s$^{-1}$ (modest wind, hereafter) were also verified (Figure 6(d)). Although TREPS PM behaved worse than ENS OPT from 30 to 60 h lead time, its performance was comparable to that of TREPS OPT. However,
Finally, if the ratio of heavy rainfall (exhibited no skill in forecasting extremely strong wind, while wind (improvement over TREPS OPT was slightly significant. For strong TREPS PM still displayed significant superiority to DETER, HRES (Figures 6(e)). Similar to the difficulties in forecasting extremely heavy rainfall among the various forecasts (Figures 9(b) and (c)). However, their superiorities over ENS shown). For this reason, TREPS yielded less reliable forecasts specifically, underforecasting was present for light rainfall (overforecasting) existed before (after) TC landfall for both ENS and TREPS (not shown). Overall, TREPS forecasts with (without) the optimal-probability technique were then assessed. BS measures the forecast skills of the probabilistic forecasts for both TREPS and ENS with (OPT, hereafter) and without the optimal-probability technique were then assessed. BS measures the accuracy of EPS for predicting probability, with smaller values corresponding to better probabilistic forecasts. BS can be decomposed into three terms: reliability, resolution, and uncertainty (Candille and Talagrand, 2005). Reliability measures the agreement between predicted probability and mean observed frequency, while resolution measures the ability of a forecast to discern situations with different frequencies of occurrence; smaller reliability indicates more reliable probabilistic forecasts. Considering that AROC measures the ability of a forecast to discriminate between two alternative outcomes, and thus measures the resolution, verification based on the resolution was not considered here. For AROC, values exceeding 0.5 indicate successful discriminating ability, while higher values correspond to better discrimination.

For light rainfall, BS for ENS was significantly smaller than that for TREPS (Figure 8(a)). The optimal-probability technique greatly improved the accuracy for TREPS during nearly the entire forecasting period; but the opposite was true for ENS. Originally, TREPS exhibited higher BS than ENS for heavy rainfall, especially beyond 30 h of forecasts (Figure 8(b)); excitingly, this inferiority was significantly eliminated by using the optimal-probability technique. Beyond 24 h of forecasts, TREPS OPT produced a higher accuracy than ENS. However, the optimal-probability technique slightly degraded ENS. Generally, there were few differences in extremely heavy rainfall forecasts between TREPS and ENS (Figure 10(c)), with the only significant difference being that TREPS was more skilful than ENS for 60 h forecasts; and the optimal-probability technique did not change the forecast performance for either TREPS or ENS.

ENS was significantly superior to TREPS in probabilistic forecasts for modest wind (Figure 8(d)); however, TREPS behaved remarkably better than ECMWF during the initial 18 h of forecasts for strong wind (Figure 8(e)). The optimal-probability technique degraded both ENS and TREPS, particularly the former (Figures 8(d) and (e)). For extremely strong wind forecasting, the superiority of TREPS over ENS was evident in half of the forecasts (Figure 8(f)). The benefits of the optimal-probability technique were found for ENS, despite only being significant for partial forecasts.

There were some biases found in both ENS and TREPS; specifically, underforecasting was present for light rainfall forecasting and was more serious in TREPS than in ENS (not shown). For this reason, TREPS yielded less reliable forecasts than ENS (Figure 9(a)). For heavier rainfall, underforecasting (overforecasting) existed before (after) TC landfall for both ENS and TREPS (not shown). Overall, TREPS forecasts with (without) the optimal-probability technique had the best reliability for heavy (extremely heavy) rainfall among the various forecasts (Figures 9(b) and (c)). However, their superiorities over ENS were only significant at limited lead times. Regarding modest wind, there was overforecasting for both ENS and TREPS (not shown). For (extremely) strong wind, underestimation of ENS was serious and thereby resulted in few samples used in calculating reliability. Therefore, ‘jumpiness’ of reliability for ENS was observed (Figure 9(f)). The differences of reliability between different forecasts were generally only statistically significant for modest wind (Figure 9(d)), and overall, ENS pronouncedly outperformed TREPS for reliability, with the optimal-probability technique generating some deteriorations in both of them.

The AROC comparison indicated that ENS provided a significant improvement in discrimination over TREPS for

\[
\text{BS} = \frac{\sum_{i=1}^{n} \text{BS}_i \times \text{TH}_i}{\sum_{i=1}^{n} \text{TH}_i},
\]

where \( \text{BS}_i \) is the reliability for the forecast of probability. As a preliminary attempt, \( \text{TH}_i \) was set to 0.045. However, there was no member with a ratio below 0.045 in some forecasts, in which case \( \text{TH}_i \) was increased gradually in increments of 0.005 until at least ten members were included.

The forecast skills of the probabilistic forecasts for both TREPS and ENS with (OPT, hereafter) and without the optimal-probability technique were then assessed. BS measures the accuracy of EPS for predicting probability, with smaller values corresponding to better probabilistic forecasts. BS can be decomposed into three terms: reliability, resolution, and uncertainty (Candille and Talagrand, 2005). Reliability measures the agreement between predicted probability and mean observed frequency, while resolution measures the ability of a forecast to discern situations with different frequencies of occurrence; smaller reliability indicates more reliable probabilistic forecasts. Considering that AROC measures the ability of a forecast to discriminate between two alternative outcomes, and thus measures the resolution, verification based on the resolution was not considered here. For AROC, values exceeding 0.5 indicate successful discriminating ability, while higher values correspond to better discrimination.

For light rainfall, BS for ENS was significantly smaller than that for TREPS (Figure 8(a)). The optimal-probability technique greatly improved the accuracy for TREPS during nearly the entire forecasting period; but the opposite was true for ENS. Originally, TREPS exhibited higher BS than ENS for heavy rainfall, especially beyond 30 h of forecasts (Figure 8(b)); excitingly, this inferiority was significantly eliminated by using the optimal-probability technique. Beyond 24 h of forecasts, TREPS OPT produced a higher accuracy than ENS. However, the optimal-probability technique slightly degraded ENS. Generally, there were few differences in extremely heavy rainfall forecasts between TREPS and ENS (Figure 10(c)), with the only significant difference being that TREPS was more skilful than ENS for 60 h forecasts; and the optimal-probability technique did not change the forecast performance for either TREPS or ENS.

ENS was significantly superior to TREPS in probabilistic forecasts for modest wind (Figure 8(d)); however, TREPS behaved remarkably better than ECMWF during the initial 18 h of forecasts for strong wind (Figure 8(e)). The optimal-probability technique degraded both ENS and TREPS, particularly the former (Figures 8(d) and (e)). For extremely strong wind forecasting, the superiority of TREPS over ENS was evident in half of the forecasts (Figure 8(f)). The benefits of the optimal-probability technique were found for ENS, despite only being significant for partial forecasts.

There were some biases found in both ENS and TREPS; specifically, underforecasting was present for light rainfall forecasting and was more serious in TREPS than in ENS (not shown). For this reason, TREPS yielded less reliable forecasts than ENS (Figure 9(a)). For heavier rainfall, underforecasting (overforecasting) existed before (after) TC landfall for both ENS and TREPS (not shown). Overall, TREPS forecasts with (without) the optimal-probability technique had the best reliability for heavy (extremely heavy) rainfall among the various forecasts (Figures 9(b) and (c)). However, their superiorities over ENS were only significant at limited lead times. Regarding modest wind, there was overforecasting for both ENS and TREPS (not shown). For (extremely) strong wind, underestimation of ENS was serious and thereby resulted in few samples used in calculating reliability. Therefore, ‘jumpiness’ of reliability for ENS was observed (Figure 9(f)). The differences of reliability between different forecasts were generally only statistically significant for modest wind (Figure 9(d)), and overall, ENS pronouncedly outperformed TREPS for reliability, with the optimal-probability technique generating some deteriorations in both of them.

The AROC comparison indicated that ENS provided a significant improvement in discrimination over TREPS for
Figure 8. BS for the forecasts of 6 h accumulated precipitation with thresholds of (a) 0.1, (b) 25, and (c) 100 mm, as well as forecasts of 10 m wind with thresholds of (d) 17.2, (e) 24.5, and (f) 32.7 m s$^{-1}$ as a function of lead time from various forecast sources. Grey (black) diamonds and asterisks in the bottom (top) axis indicate the lead times for which the significance level of the BS differences is larger than 90% for the comparison of TREPS (TREPS OPT) with ENS and ENS OPT respectively. Crosses (pluses) in the bottom (top) axis indicate the lead times for which the significance level of the BS differences between TREPS (ENS) and TREPS OPT (ENS OPT) is larger than 90%.

light-rainfall forecasting (Figure 10(a)). However, the optimal-probability technique consistently improved the discrimination of TREPS, with no significant differences between TREPS OPT and ENS seen at later forecast times. The impact of this technique on ENS was neutral. For heavy rainfall, the optimal-probability technique did not produce significant changes in AROC for ENS but evidently raised the rank of AROC for TREPS from last to first, especially during later forecast periods (Figure 10(b)). The ENS discrimination was poor for extremely heavy rainfall forecasting; by stark contrast, TREPS displayed good discriminating ability (Figure 10(c)). It is clear that the optimal-probability technique functioned well for the two EPSs, especially for TREPS.

The optimal-probability technique significantly reversed the significant inferiority of TREPS compared to ENS in terms

© 2017 The Authors. Quarterly Journal of the Royal Meteorological Society published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.
of discrimination for modest wind (Figure 10(d)). However, the degradation in reliability dominated the improvement in discrimination, resulting in partial deterioration in BS (Figure 8(d)). Moreover, the technique did not produce any significant impact on the ENS discrimination. For strong wind, the TREPS discrimination was comparable to or slightly better than that of ENS (Figure 10(e)). The optimal-probability technique played positive roles for both TREPS and ENS, producing more significant improvements in the former. Again, the unsatisfactory performance for reliability associated with the technique contributed more to the poor performance for accuracy than did the exciting performance for discrimination (Figure 8(e)). It was difficult for ENS to yield probabilistic forecasts for extremely strong wind with sufficient discrimination (Figure 10(f)); this poor performance was significantly improved in the TREPS forecasts, especially for forecasts using the optimal-probability technique.

DETER were compared with the probabilistic forecasts of the perturbed members to further examine the advantage of TREPS in probabilistic forecasts of short-time heavy rainfall. For this purpose, DETER was transformed into binary fields by setting grid points exceeding a certain threshold to a value of 1, while all other points were given a value of 0. DETER, with larger BS, was less accurate than TREPS.
perturbed forecasts (Figure 11(a)). Furthermore, the optimal-probability technique did not significantly change the accuracy performance. A comparison based on reliability and AROC revealed the advantages of TREPS probabilistic forecasts over DETER, with the most significant improvement in reliability during the period of TC landfall (Figures 11(b) and (c)). The impact of the optimal-probability technique on reliability was ambiguous, with slight improvements (degradations) in the late (middle) period (Figure 11(b)); however, the impact was consistently significant and positive on the discrimination (Figure 11(c)).

5. Results of case-study

Forecasts of Rammasun initialised at 0000 UTC 17 July 2014 are presented in this section to further illustrate the performance of both deterministic and probabilistic guidance of TREPS.

5.1. Deterministic guidance

Among the various forecasts, the southwestward bias in TC track forecasting was almost the least for TREPS EM; the landfall locations at both GD and GX were accurately predicted by TREPS EM. Although there was a significant bias for TREPS EM after
landfall at GX, some perturbed members did predict tracks similar to the best track. Thus, TREPS could well capture the possibility of the best track by properly predicting the potential spread of the track. In addition, the track spread increased as the track error of TREPS EM grew after landfall at GX, suggesting the success of TREPS in predicting the forecast uncertainty. Although evident RI was observed, nearly all forecasts failed to capture this scenario (Figures 12(b) and (c)). However, TREPS EM exhibited the most accurate prediction for both intensification rate and peak intensity. Furthermore, the rapid weakening after landfall was predicted accurately by TREPS EM. It was noteworthy that some perturbed members predicted peak intensity close to the best track despite their low biases. Thus, TREPS could capture the potential threat and thereby displayed the value in disaster prevention. As well, the intensity spread increased during the RI period and reached a maximum at the time of peak intensity, confirming once again the capability of predicting forecast uncertainty.

Overall, TREPS predicted the rainfall amount better than ECMWF (Figure 13). For heavy rainfall in HN and southwestern GD, TREPS PM exhibited the highest skill (Figure 13(b)), followed by TREPS OPT (Figure 13(c)). Note that the differences in the spread of light rainfall between TRAMS9KM and ECMWF were clear, with more underestimations in the former. Short-time heavy rainfall observed in both the strait and HN, with several spots exceeding 50 mm h\(^{-1}\) (Figure 13(h)), was predicted well by TREPS PM for both location and amount but significantly underestimated by both DETER and TREPS OPT. Generally, strong-wind forecasting did not show remarkable differences among TREPS forecasts in this case (not shown).

In this case-study, TREPS OPT did not display any improvement and even caused slight degradation relative to TREPS PM, which was also the miniature of the optimal-member technique for TREPS. To investigate the cause of the incomplete effectiveness of this technique, the four members used to calculate the forecasts of TREPS OPT for this case were compared with TREPS EM for track and intensity (Figure 12), which showed that there was no member which can match TREPS EM consistently for both track and intensity over the entire period. In fact, for the four selected members, one often matched TREPS EM much better at partial forecast times for track or intensity but then deviated from

Figure 11. (a) BS, (b) reliability, and (c) AROC, for forecasts of 1 h accumulated precipitation with thresholds of 15 mm as a function of lead time from various forecast sources. Pluses on the markers of TREPS (TREPS OPT) indicate the lead times for which the significance level of the score differences between TREPS (TREPS OPT) and DETER is smaller than 90%. Crosses in the top axis indicate the lead times for which the significance level of the score differences between TREPS and TREPS OPT is smaller than 90%.
Figure 12. Forecasts of (a) track, (b) $P_{\min}$, and (c) $V_{\max}$ for Rammasun initialized at 0000 UTC 17 July 2014. ‘OBS’ represents the observation from CMA best-track dataset; ‘TREPS$_m$’ indicates the ensemble members used in calculating the ‘optimal member’; the yellow lines indicate all perturbed members of TREPS.

TREPS EM more seriously at other forecast times than the others. Thus, simple averaging over these selected members at a given forecast time would inevitably lead to unreasonable smoothing, and thereby unrealistic results. This smoothing was more serious for rainfall (wind) with larger amount and smaller scale, resulting in worse performance of the optimal-member technique in heavier rainfall (stronger wind). Moreover, (extremely) strong wind seemed to be more localized than (extremely) heavy rainfall (Figures 13(g) and (i)), since there were different characteristics of scale and distribution between wind and rainfall. Therefore, the optimal-member technique worked better in forecasts of rainfall than those of wind (Figure 6).

The principles used to select the optimal member introduced in section 4.1 might be responsible for its deficiency, because the excessive smoothing of averaging often occurred when there were too many members with the same minimum cost function. Nevertheless, this excessive smoothing was relatively rare for ECMWF ENS owing to large differences in the cost functions among the various members, which was likely attributable to calculating the cost function with reduced forecast times.

5.2. Probabilistic guidance

TREPS provided higher probability of extremely heavy rainfall covering areas that better matched the observations than ENS (Figure 14). This increased the probability of detection, which was consistent with the higher AROC results shown in Figure 10(c). The optimal-probability technique enhanced the magnitude of the high probability and lessened the area of the low probability (Figures 14(b) and (d)) as the common result of reducing ensemble size. Consistent with its goal of selecting ensemble members closer to the true state in calculating probability, the effectiveness of the optimal-probability approach was confirmed based on its results of increasing realistic high probability while reducing false low probability.

For short-time heavy rainfall, DETER missed its occurrence around the peninsula of GD (Figure 13(h)) that was highlighted in the TREPS probabilistic forecasts with probabilities above 40% (Figure 14(e)). Although the optimal-probability technique increased the forecast probability at northeastern HN from 80 to 85% (Figure 14(f)), indicating an improvement in guidance for forecasters, the areas with non-zero probability were reduced in other locations, resulting in mixed impacts. In southern HN and the coastal area of southwestern GD, this reduction in area indicated some deterioration in identifying the potential threat; in other areas, false alarms due to overestimations could be alleviated in this situation. Heavy rainfall exceeding 50 mm h$^{-1}$ in the strait and northern HN was well detected by TREPS; however, the enhanced high probability owing to the optimal-probability technique was apt to produce misleading guidance if biases were present in the forecasts of rainfall location (Figure 14(f)).
Compared with ENS, TREPS indicated the possibility of strong wind occurring around the peninsula of GD with larger forecast probability (Figure 14(a)); TREPS also predicted larger areas with non-zero probability, especially at southeastern HN, where strong wind was observed (Figure 13(i)). In this manner, TREPS showed considerable advantage over ENS in identifying areas with severe weather. Nevertheless, false alarms in probabilistic forecasts were still more notable in TREPS compared with ENS, suggesting the necessity of improving the reliability for TC wind forecasting. The impact of the optimal-probability technique for strong wind was overall similar to the above finding for heavy rainfall.

The optimal-probability technique brought more improvement than degradation in application to TREPS but produced slight impacts for ENS. The risk of reducing ensemble size may be responsible for the degradation, as the excluded ensemble members could still provide useful information in the description of probability. Moreover, because the ENS ensemble members varied more significantly in terms of cost function, the threshold $TH_j$ was usually set larger to include more members; as a result, the optimal-probability technique excluded fewer members overall, resulting in a small influence on ENS.

6. Conclusions and discussion

With the goal of filling the current void in the development and application of high-resolution regional EPS for operational TC forecasting in China, TREPS, a mesoscale EPS based on GRAPES, was established for a planned operational run by the end of 2017. In this article, TREPS was fully described and its performance in forecasting landfall TCs in China was examined for the period 2014–2016.

TREPS is composed of one unperturbed and 30 perturbed members with horizontal resolutions of 0.09°. It issues 60 h forecasts twice per day and focuses solely on TCs in WNP and SCS. Initial perturbations, generated by blending ICDp from ENS with BRp implemented around TCs from GRAPES 3D-Var are added to the ICs of unperturbed members to produce the ICs of perturbed members. Perturbations for the LBCs are downscaled from ENS, with additional perturbations in surface temperature around the TC also considered. The uncertainties in the model physics are considered by combining MP and SPPT. The parameters exerting key impacts are determined using a pragmatic approach based on WARNING using tuning.

The contributions of different perturbation methods for different perturbation sources to the total forecast perturbation around a TC were investigated firstly to determine the rationality of the TREPS design. Although there were different contributions among various perturbation methods to the total forecast perturbations in different regions, at different scales, and for different variables, these methods mutually complemented in increasing the total forecast perturbation in the whole region, at every scale, and for all variables. Note that some benefit-limiting deficiencies were present in some of the perturbation methods. Further improvements are required for the balance constraints used in generating BRp, the range of implementing TSp, and the setting of key parameters determined by only several case-studies now.

Objective verification for both deterministic and probabilistic guidance was carried out to evaluate the performance of TREPS
GRAPES-based Mesoscale EPS for TC

Figure 14. (a–d) Probability (%) of the 6 h accumulated rainfall exceeding 100 mm (shaded) and 10 m wind exceeding 24.5 m s$^{-1}$ (contour with interval of 20, black solid for values larger than 50, and grey dashed for values smaller than 50), as well as (e, f) probability (%) of the 1 h accumulated rainfall exceeding 15 mm (shaded) and exceeding 50 mm (contour; green, purple and grey for values of 0, 10 and 20, respectively) for 36 h forecasts at 1200 UTC 18 July 2014. The range is the same as VA shown in Figure 2.

and to investigate its advantages over ECMWF during the past 3 years. Additionally, the case of Rammasun initialized at 0000 UTC 17 July 2014, was studied to investigate the performance of TREPS more intuitively and to further understand the underlying physical causes of its performance.

TREPS EM provided better track forecasts than either HRES or DETER but was comparable to ENS EM. In terms of intensity forecasting, TREPS EM significantly outperformed the other forecasts in the first 36 h of forecasts for $V_{\text{min}}$ and yielded the highest skill for $V_{\text{max}}$. TREPS produced the ensemble spread more consistent with the EM error than ENS by effectively reducing the underdispersions in intensity forecasting. The case-study showed that TREPS EM could indeed provide more accurate prediction of track, intensification rate and peak intensity than the other forecasts. Furthermore, the ensemble spread of TREPS could not only properly estimate the forecast uncertainty but also well capture the potential threat from high-impact weather, for both track and intensity. Since the track (intensity) forecasting was overdispersive (underdispersive), the perturbations used to describe the uncertainties in large-scale circulation controlling the TC motion seem to be adequate, while the perturbations used to represent the mesoscale uncertainties in the TC structure related to its intensity might still be insufficient. Accordingly, the introduction of more small-scale perturbations to TREPS may be responsible for its more skilful and reliable performance for intensity than ENS. However, the TREPS intensity forecasting was still underdispersive, indicating the necessity of increasing the ensemble spread. Likely, further improvements in the perturbation methods mentioned above could partially address this issue.

For 6 h accumulated rainfall and 10 m wind forecasting, TREPS PM was considerably superior to both HRES and ENS PM for almost all thresholds. However, TREPS PM significantly outperformed DETER only for heavy and extremely heavy rainfall, but underperformed for light rainfall; in terms of wind forecasting, TREPS PM was generally better than DETER. The advantage of TREPS PM over DETER for strong wind was not as significant as that for heavy rainfall. Besides, TREPS PM displayed obvious improvements over DETER in short-time heavy rainfall forecasting. The advantages of TREPS PM over ECMWF and DETER for heavy rainfall were well demonstrated in the case-study.

In order to improve deterministic guidance for both rainfall and wind forecasting from EPS, an ‘optimal member’ technique to select members closest to both EM and the latest WARNING observations was proposed and will be used in the operational version. The effectiveness of this technique was well confirmed in ECMWF forecasts for both rainfall and wind; however, the technique produced mixed results in TREPS forecasts, with significant improvements over PM in light rainfall but degradations in strong and extremely strong wind. As was illustrated in the case-study, simply averaging over the ensemble members with nearly the same minimum cost function resulted in excessive smoothing, which may partially explain the unsatisfactory performance of this technique. Thus, the principles used to select the optimal member require improvement to cope with this issue. Applying this technique to TREPS and ENS led to differing impacts, which probably arise from differences between the two EPSs in principles used in selecting the optimal member. Therefore, further investigation of the differing impacts is expected, to further improve this technique. Considering variations in the performance of ensemble members at different forecast times, the patched-together forecasting proposed by Ancell (2013) may be useful and will be tested in the future.

ENS yielded significantly more accurate probabilistic forecasts than TREPS for light rainfall, heavy rainfall and modest wind. At root, the poor performance of TREPS in both reliability and discrimination for these variables was responsible for its poor accuracy. The advantage of TREPS in accuracy over ECMWF was apparent in the forecasts of extremely heavy rainfall, strong wind, and extremely strong wind, but with limited significance in all three cases. This advantage of TREPS stemmed primarily from its better discrimination, as its superiority in reliability was not highly significant. More serious underforecasting was present in TREPS than in ENS for light rainfall forecasting, causing TREPS to yield less reliable forecasts than ENS for this variable. One suspected cause for this is the difference in representing stratiform precipitation between TRAMS9KM and ECMWF, which may mainly influence light-rainfall forecasting.
Similarly, the poor reliability of TREPS in predicting both heavy rainfall and modest wind resulted from its serious overforecasting. From a high-temporal-resolution probabilistic standpoint, the TREPS probabilistic forecasts showed outstanding improvements in accuracy, which was mainly attributed to its improved reliability over DETER during TC landfall. This reveals the evident advantage of ensemble forecasting over single-deterministic forecasting in predicting the threat of short-time heavy rainfall, as the important signal of threat missing in the latter is covered well by some ensemble members. The capacity of TREPS to highlight severe weather was confirmed intuitively in the case-study.

An ‘optimal probability’ technique based on the optimal-member technique was developed to improve probabilistic guidance. In this technique, the ensemble members close to the optimal member within a certain threshold were selected to construct a new ensemble that produced a new forecast probability. This technique resulted in consistent improvements in the discrimination for both rainfall and wind forecasting in TREPS and substantial improvements in the TREPS reliability for rainfall forecasting. Generally, accuracy for rainfall forecasting was improved by this technique. By contrast, the impact of this technique was mixed on ENS, with major deterioration in reliability and partial improvement in discrimination. For short-time heavy rainfall, this technique contributed the most to improve the optimal-probability technique. Note that the bias was still considerable in the case of simultaneous TCs. Thus, values of the key parameters were finally determined based on the better performance. Moreover, parameter settings according to the TC scale or intensity showed better performance than static or simplified settings; increasing the proportion of ICDp and LBCDp also improved forecasts of simultaneous TCs. Thus, values of the key parameters were finally determined based on the better performance. Moreover, improved performance of ALL over CTL due to including multi-source perturbations confirmed the rationality of the TREPS design.

Acknowledgements

This research was sponsored by the National Natural Science Foundation of China through Grant 41405104, the National Basic Research and Development Project (973 program) of China under Grant 2015CB452802, the Natural Science Foundation of Guangdong Province through Grant 2017A030313225, and the Meteorology Public Welfare Scientific Research Projects under Grant GYHY201506007. The author thanks three anonymous reviewers for their helpful comments, which improved the quality of this paper; and also thanks Qilin Wan (ITMM) for helpful discussions.

Appendix

Sensitivity tests to determine parameter settings

Sensitivity tests based on several cases were conducted to investigate the impacts of parameter settings on TC forecasting. The forecasts for three TCs (i.e. Soudelor, Mujigae and Linfa) in 2015, initialized at 1200 UTC 6 August, 1200 UTC 2 October and 0000 UTC 7 July respectively, were selected in sensitivity tests. The first two TCs were good representations of TCs with various scales and intensities (Figure A1(a)), while the third TC was representative of the situation with simultaneous TCs (section 2.5).

The experiment of CTL with ICDp, LBCDp and MP was designed as a benchmark, to which additional perturbations were added in sensitivity tests. Figure A1 showed partial results of sensitivity tests. For SPPT, the sensitivity of forecasts to $\kappa$ was tested by modifying its value around the averaged initial gale_radius (i.e. 300 km); while for TSp, variation of $\delta$ around the value of 1.5 was tested based on some previous work (Lavaysse et al., 2013; Bouttier et al., 2016). Besides, sensitivity to both $\alpha$ and $s$ in the case of simultaneous TCs was also tested. Among these tests, not only absolute errors of track and $P_{\text{max}}$ but also rainfall distributions (not shown) were compared. Changing parameter settings according to the TC scale or intensity showed better performance than static or simplified settings; increasing the proportion of ICDp and LBCDp also improved forecasts of simultaneous TCs. Thus, values of the key parameters were finally determined based on the better performance. Moreover, improved performance of ALL over CTL due to including multi-source perturbations confirmed the rationality of the TREPS design.

Figure A1. Absolute errors of the TC (a) track and (b) $P_{\text{max}}$, averaged over the whole 60 h forecasts for three TC cases (Soudelor, Mujigae and Linfa). Italic numbers in parentheses following the TC names in the orange box in (a) represent the initial $P_{\text{max}}$ (red; unit: hPa), gale_radius (green; unit: km) and change in gale_radius (blue; unit: km) during the first 6 h for the corresponding TCs. The detailed designs of sensitivity tests are showed in the orange box in (b). The legend of ‘ALL’ represent the retrospective forecasts as illustrated in section 3.1.

© 2017 The Authors. Quarterly Journal of the Royal Meteorological Society published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.
Rogers R, Aberson S, Aksoy A, Annane B, Black M, Cione J, Dorst N, Dunion J, Gamache J, Goldenberg S, Gopalakrishnan S, Kaplan J, Klotz B, Lorsolo S, Marks F, Murillo S, Powell M, Reasor P, Sellwood K, Uhlhorn E, Vukicevic T, Zhang J, Zhang X. 2013. NOAA's hurricane intensity forecasting experiment: a progress report. Bull. Am. Meteorol. Soc. 94: 859–882.

Romine GS, Schwartz CS, Berner J, Smith KR, Snyder C, Anderson JL, Weisman ML. 2014. Representing forecast error in a convection-permitting ensemble system. Mon. Weather Rev. 142: 4519–4541. https://doi.org/10.1175/MWR-D-14-00100.1.

Sampson CR, Goerss JS, Weber HC. 2006. Operational performance of a new barotropic model (WBAR) in the Western North Pacific Basin. Weather and Forecasting 21: 656–662.

Sampson CR, Franklin JL, Knaff JA, DeMaria M. 2008. Experiments with a simple tropical cyclone intensity consensus. Weather and Forecasting 23: 304–312.

Schwartz CS, Romine GS, Smith KR, Weisman ML. 2014. Characterizing and optimizing precipitation forecasts from a convection-permitting ensemble initialized by a mesoscale ensemble Kalman filter. Weather and Forecasting 29: 1295–1318. https://doi.org/10.1175/WAF-D-13-00145.1.

Stensrud DJ, Bao J, Warner TT. 2000. Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. Mon. Weather Rev. 128: 2077–2107.

Tan Y, Liang XD. 2012. An ensemble forecast experiment of a landing typhoon. J. Trop. Meteorol. 18: 314–321.

Torn RD. 2010. Performance of a mesoscale Ensemble Kalman Filter (EnKF) during the NOAA high-resolution hurricane test. Mon. Weather Rev. 134: 4375–4392.

Toth Z, Kalnay E. 1993. Operational ensemble prediction at the National Meteorological Center: practical aspects. Mon. Weather Rev. 121: 3297–3319.

Tracton MS, Kalnay E. 1997. Ensemble forecasting at NCEP and the breeding method. Mon. Weather Rev. 125: 2397–2399.

Tracton MS, Kalnay E. 1997. Operational ensemble prediction at the National Meteorological Center: practical aspects. Weather and Forecasting 8: 378–398.

Tribbia JJ, Baumhefner DP. 1988. The reliability of improvements in deterministic short-range forecasts in the presence of initial state and modeling deficiencies. Mon. Weather Rev. 116: 2276–2288.

Tu XP, Yao RS, Zhang CH, Chen YL. 2014. Operational ensemble forecasting and analysis of tropical cyclones over the western North Pacific (including the South China Sea). J. Trop. Meteorol. 20: 87–92.

Wang CX. 2016. Ensemble prediction experiments of typhoon track based on the stochastic total tendency perturbation. J. Trop. Meteorol. 22: 305–317.

Wang C, Huang S, Chen S, Chang C, Tsuboki K. 2016. Cloud-resolving typhoon rainfall ensemble forecasts for Taiwan with large domain and extended range through time-lagged approach. Weather and Forecasting 31: 151–172. https://doi.org/10.1175/WAF-D-15-0045.1.

Weber HC. 2005. Probabilistic prediction of tropical cyclones. Part II: Intensity. Mon. Weather Rev. 133: 1853–1864.

Wilks DS. 2006. Statistical Methods in the Atmospheric Sciences. Academic Press: New York, NY.

Xu XL, Sun DW, Guo TJ. 2015. A systemic analysis of typhoon risk across China. Nat. Hazards 77: 461–478. https://doi.org/10.1007/s11069-015-1386-0.

Xue JS, Liu Y. 2007. Numerical weather prediction in China in the new century: progress, problems and prospects. Adv. Atmos. Sci. 24: 1099–1108. https://doi.org/10.1007/s00376-007-1099-1.

Xue JS, Zhuang SY, Zhu GF, Liu ZQ, Liu Y, Zhuang ZR. 2008. Scientific design and preliminary results of three-dimensional variational data assimilation system of GRAFES. Chin. Sci. Bull. 53: 3446–3457. https://doi.org/10.1007/s11434-008-0416-0.

Yamaguchi M, Majumdar SJ. 2010. Using TIGGE data to diagnose initial perturbations and their growth for tropical cyclone ensemble forecasts. Mon. Weather Rev. 138: 3634–3655.

Yamaguchi M, Sakai R, Kyoda M, Komori T, Kadowaki T. 2009. Typhoon ensemble prediction system developed at the Japan Meteorological Agency. Mon. Weather Rev. 137: 2592–2604.

Ying M, Zhang W, Yu H, Lu X, Feng J, Fan Y, Zhu Y, Chen D. 2014. An overview of the China Meteorological Administration tropical cyclone database. J. Atmos. Oceanic Technol. 31: 287–301. https://doi.org/10.1175/JTECH-D-12-00119.1.

Yu H, Chen P, Li Q, Tang B. 2013. Current capability of operational numerical models in predicting tropical cyclone intensity in the western North Pacific. Weather and Forecasting 28: 353–367. https://doi.org/10.1175/WAF-D-11-00100.1.

Zhang F. 2005. Dynamics and structure of mesoscale error covariance of a winter cyclone estimated through short-range ensemble forecasts. Mon. Weather Rev. 133: 2876–2893.

Zhang HB, Chen J, Zhi XF, Wang Y, Wang YN. 2015. Study on multi-scale blending initial condition perturbations for a regional ensemble prediction system. Adv. Atmos. Sci. 32: 1143–1155. https://doi.org/10.1007/s00376-015-4232-6.

Zhang XL, Duan YH, Wang YQ, Wei N, Hu H. 2017. A high-resolution simulation of Supertyphoon Rammasun (2014) – Part I: Model verification and surface energetics analysis. Adv. Atmos. Sci. 34: 757–770. https://doi.org/10.1007/s00376-017-6255-7.