


## RESEARCH ARTICLE

# A new approach to represent model uncertainty in the forecasting of tropical cyclones: The orthogonal nonlinear forcing singular vectors

Yichi Zhang<sup>1,2</sup>  | Wansuo Duan<sup>1,2</sup>  | Stéphane Vannitsem<sup>3</sup>  | Han Zhang<sup>1,2</sup>

<sup>1</sup>LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

<sup>2</sup>University of Chinese Academy of Sciences, Beijing, China

<sup>3</sup>Royal Meteorological Institute of Belgium, Brussels, Belgium

**Correspondence**

Duan Wansuo, LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China.  
Email: [duanws@lasg.iap.ac.cn](mailto:duanws@lasg.iap.ac.cn)

**Funding information**

National Natural Science Foundation of China, Grant/Award Numbers: 41930971, 42288101

**Abstract**

Tropical cyclone (TC) track forecasting has been considerably improved in recent decades, while TC intensity forecasting remain challenging. In this study, orthogonal nonlinear forcing singular vectors (O-NFSVs) for emulating the impact of model uncertainties are used to conduct TC ensemble forecasting experiments with the Weather Research and Forecasting (WRF) model, with a focus on improving TC intensity forecasting skill. The O-NFSVs approach is compared with the traditional stochastic kinetic-energy backscatter (SKEB) and stochastically perturbed parametrization tendency (SPPT) schemes. The results demonstrate that the O-NFSVs ensembles generally provide a better representation of the model uncertainties affecting TC intensification, with much better deterministic and probabilistic skills. These results also extend to the ability to forecast TC track, although the perturbations have not been optimized for that specific purpose. The O-NFSVs are therefore appropriate perturbation structures for describing the uncertainties of the TC intensity and track forecasting and are also favourable for recognizing the rapid intensification process.

**KEYWORDS**

ensemble forecast, model uncertainty, nonlinear forcing singular vector, tropical cyclone

## 1 | INTRODUCTION

Tropical cyclones (TCs) are one type of high-impact weather event that often bring strong wind, heavy rain and storm surges, causing severe natural hazards and threatening the safety of people's lives and property. It is therefore important to forecast TCs to mitigate potential disasters. Over the past few decades, TC track forecasting has been significantly improved due to more observations and more advanced models, but TC intensity forecasting still remains a challenge. DeMaria *et al.* (2014) demonstrated that the improvement in TC intensity forecasting skill was only approximately one-third to one-half of that

of TC track forecasting skill at lead times from 24 to 72 hr. Furthermore, the forecasts can hardly capture the rapid intensification (RI) of TCs. Hence, a current important challenge is to improve the TC intensity forecasting skill.

Forecast errors are caused by initial errors and model errors (Lorenz, 1969, 1982; Toth and Vannitsem, 2002; Vannitsem and Toth, 2002; Kalnay, 2003; Nicolis, 2004; Nicolis *et al.*, 2009). Emanuel and Zhang (2016) demonstrated that TC intensity forecasting errors are dominated by initial errors during the first few days. Puri *et al.* (2001) showed that the initial perturbations generated by the singular vectors (SVs) often provide significant spread in the ensemble forecasts for TC tracks but are smaller than the

stochastic physics perturbations in forecasting TC central pressure. Zhang *et al.* (2014) found that both TC track and intensity ensemble forecasts can be improved by implementing a stochastic convective trigger scheme in the Hurricane Weather Research and Forecast (HWRF) model. All these studies suggest that model errors also play an important role in TC forecasting.

There are many different sources of model errors. These can be roughly classified into four types: (a) inaccuracy of physical parametrizations and physical parameters, (b) numerical approximations, (c) insufficient model resolutions, and (d) missing physical and dynamic processes (Toth and Vannitsem, 2002; Nicolis *et al.*, 2009). Each type of model error may contribute to the uncertainties in TC intensity forecasting. Miglietta *et al.* (2015) found that different selections of physical parametrization schemes have a strong influence on uncertainties in TC intensity forecasting. Torn (2016) also showed that the uncertainties occurring in the drag coefficient and enthalpy coefficient model parameters, which are related to the air–sea interaction effect, markedly affect the TC structure and intensity. Davis (2018) and Qin *et al.* (2020) demonstrated that relatively coarse grid spacing may seriously underestimate hurricane wind speed. Obviously, different sources of model errors represent different impacts on TC intensities. Furthermore, these effects are interactive in the forecasting of TCs. It is therefore difficult to identify every source of model error, and an integrated approach to the problem is desired.

Duan and Zhou (2013) proposed a nonlinear forcing singular vector (NFSV) approach that addresses the combined effects of different sources of model errors. Qin *et al.* (2020) demonstrated that uncertainties in TC intensity are more sensitive to the model errors emulated by NFSV-tendency perturbations when TC intensity forecasts are performed at a lead time of 24 hr, and once the NFSV-type model errors are removed by an assimilation method, the TC intensity forecasting skills are significantly improved. Yao *et al.* (2021) also found that the NFSV-type tendency errors of the sea-surface temperature (SST) along the TC track are more likely to lead to large errors in forecasts of TC intensity, especially during the rapid intensification (RI) period. These studies demonstrate that the model errors featured by the NFSV-tendency perturbations have a strong impact on forecasts, which suggests that the NFSV is an appropriate method for representing model uncertainties (Qin *et al.*, 2020; Xu *et al.*, 2022a, 2022b).

Ensemble forecasts are very popular approaches to provide estimations of forecasting uncertainty, which is an essential component of any numerical weather or climate prediction (Duan and Huo, 2016; Buizza, 2019; Duan *et al.*, 2022). Regarding TCs, there are studies based on

ensemble forecasts, illustrating how ensemble forecasts can improve TC track forecasting (Zhang and Krishnamurti, 1997; Hamill *et al.*, 2011; Yamaguchi *et al.*, 2012; Bhatia and Nolan, 2015; Torn, 2016). These studies paid attention to the estimation of initial uncertainties and their impacts. However, for TC intensity, it is inappropriate to build ensemble forecasts based on initial perturbations only (Krishnamurti *et al.*, 2005; Rogers *et al.*, 2006; Hamill *et al.*, 2011). As model errors have a considerable influence on TC intensity forecasting, a thorough exploration of ensemble forecasting skill in the presence of model uncertainties is crucial.

To emulate model uncertainties, stochastic physics schemes were developed [see the reviews of Demayer and Vannitsem (2018) and Buizza (2019)]. Among the methods proposed, two are widely used: (a) the stochastic kinetic-energy backscatter scheme (SKEB: Shutts, 2005; Berner *et al.*, 2009) and (b) the stochastically perturbed parametrization tendencies scheme (SPPT: Palmer *et al.*, 2009; Berner *et al.*, 2015). Reynolds *et al.* (2011) demonstrated that SKEB increases the ensemble spread of TC track ensemble forecasting but does not reduce the root-mean-square error (RMSE) within the US Navy global atmospheric ensemble forecast system. Judt *et al.* (2016) showed that if a convection-permitting model is used, the large-scale SKEB has a strong influence on the uncertainty of TC intensity. Puri *et al.* (2001) demonstrated that SPPT leads to a smaller spread in the TC tracks but to a much larger spread in the central pressures within the ensemble prediction system (EPS) of the ECMWF. Lang *et al.* (2012) investigated the impact of SPPT and SKEB on TC track and central pressure and showed that the perturbations of SKEB often have much larger amplitudes in the TC outer region, while those of SPPT are more concentrated on the TC core and at upper levels. These studies illustrate that stochastic physics schemes may represent model uncertainties to a certain extent. However, the usual underdispersion of ensembles remains a key issue within current ensemble forecasting systems (Novak *et al.*, 2008; Romine *et al.*, 2014; Torn, 2016; Melhauser *et al.*, 2017).

Ensemble forecasts require a group of growing-type perturbations to ensure that the ensemble members substantially deviate from the control forecast and encompass the true value (Molteni *et al.*, 1996; Toth and Kalnay, 1993, 1997). As mentioned earlier, the NFSV is a tendency perturbation that causes the largest deviation from the reference state and may be an appropriate method for representing model uncertainties. Qin *et al.* (2020) showed that the TC intensity errors in control forecasts can be largely reduced by superimposing NFSV-tendency perturbations. This suggests that the NFSV-tendency perturbation can be considered a good candidate for model perturbation in ensemble forecasts. A natural question

is therefore to ask how to perform NFSV-tendency perturbations to generate ensemble forecasts which mimic the impact of model errors on TC intensity forecasts. Duan *et al.* (2022) developed a novel ensemble forecasting method, that is, the C-NFSVs, which generate a group of mutually independent and rapidly growing combined modes of initial and model tendency perturbations. The usefulness of the C-NFSVs for building an ensemble forecast has been demonstrated in the Lorenz-96 model, suggesting that the C-NFSVs method is able to provide a higher ensemble forecasting skill (Duan *et al.*, 2022). One particular group of modes among the C-NFSVs family is the set of orthogonal NFSVs (O-NFSVs), a generalization of the NFSV-tendency perturbation proposed in Duan and Zhou (2013). In the present study, we examine the usefulness of O-NFSVs in describing the model uncertainties associated with TC intensity forecasting using the Weather Research and Forecasting model (WRF) and we explore whether O-NFSVs may improve the ensemble forecast skill.

The rest of the article is organized as follows. Section 2 introduces the O-NFSVs approach. Section 3 gives a brief description of the WRF model and an overview of the TC cases adopted in the present study. In Section 4, the experimental strategy is described. Section 5 demonstrates the performances of the O-NFSVs in producing appropriate ensemble forecasts for TC intensity and track in the WRF at coarse and high resolutions, respectively. In Section 6, a test of the O-NFSVs in a convection-permitting WRF is provided. Finally, conclusions are drawn in Section 7.

## 2 | THE O-NFSVS APPROACH FOR MEASURING THE MODEL ERROR EFFECT

Suppose a state vector  $U$ ; then, its evolution equations can be described as in Equation (1):

$$\begin{cases} \frac{\partial U}{\partial t} = F(U(x, t)) \\ U|_{t=0} = U_0, \end{cases} \quad (1)$$

where  $U_0$  is the initial state and  $(x, t) \in \Omega \times [0, T]$ , with  $\Omega$  belonging to an  $n$ -dimensional Euclidean space  $R^n$ ,  $F$  being a nonlinear operator, and  $t$  being the time. Assuming that Equation (1) and its initial state are known exactly, then the solution of Equation (1) at a future time  $T$  is given by

$$U(x, T) = M_T(U_0), \quad (2)$$

where  $M_T$  is the nonlinear propagator of Equation (1).

In realistic forecast systems, forecasts are generally contaminated by both initial errors and model errors. If we use  $u_0$  and  $f(x, t)$  to represent the initial and model errors, the forecast model can be written as

$$\begin{cases} \frac{\partial(U+u)}{\partial t} = F(U(x, t)) + f(x, t) \\ U + u|_{t=0} = U_0 + u_0. \end{cases} \quad (3)$$

In this case, we use  $M_T(f)$  to denote the propagator of Equation (3) from the initial time  $t = 0$  to the future time  $t = T$ , and the forecast error (denoted by  $u_T$ ) is written as

$$u_T = M_T(f)(U_0 + u_0) - M_T(U_0), \quad (4)$$

where  $M_T(U_0)$  is as in Equation (2) and is not contaminated by any errors, while  $M_T(f)(U_0 + u_0)$  is the forecast influenced by both initial error  $u_0$  and model error (or tendency error)  $f$ . If the initial fields are further assumed to be perfect (i.e.  $u_0 = 0$ ), the forecast error is only caused by the model error  $f$ . Then, Equation (4) is rewritten as

$$u_T = M_T(f)(U_0) - M_T(U_0). \quad (5)$$

Using Equation (5) and assuming a time-constant  $f$ , Duan and Zhou (2013) defined the NFSV, which is interpreted as the time-independent tendency perturbation that causes the largest departure from the given reference state during the forecast period  $T$  (Barkmeijer *et al.*, 2003; Duan and Zhou, 2013). Based on the NFSV, the O-NFSVs are can be defined through the maximization problem Equation (6) below and represent a group of mutually independent tendency perturbations that cause the largest departure from the reference state  $M_T(U_0)$  in orthogonal subspaces (Duan *et al.*, 2022).

$$J(f_j^*) = \max_{f_j \in \Omega_j} \left\| M_T(f_j)(U_0) - M_T(U_0) \right\|_b \quad (6)$$

where  $J$  is a cost function that measures the maximum departure from the reference state  $M_T(U_0)$  by the norm  $\|\cdot\|_b$  and  $f_j \in \Omega_j$  is the constraint condition that limits the amplitude of the tendency perturbations  $f_j$ :

$$\Omega_j = \begin{cases} \{f_j \in \mathfrak{R}^n \mid \|f_j\|_a \leq \delta\}, j = 1 \\ \{f_j \in \mathfrak{R}^n \mid \|f_j\|_a \leq \delta, f_j \perp \Omega_k, k = 1, \dots, j-1\}, \\ j > 1 \end{cases} \quad (7)$$

where the norm  $\|\cdot\|_a$  defines the amplitude of tendency perturbations with the parameter  $\delta$  constant, and  $\perp$  represents the orthogonality of spaces. Finally,  $f_j^*$  ( $j = 1, 2, 3, \dots$ ) provides the O-NFSVs of the 1st NFSV, 2nd NFSV, 3rd NFSV, ..., and the corresponding values of the cost function are ranked as  $J(f_1^*) > J(f_2^*) > J(f_3^*) > \dots >$

$J(f_n^*)$ . Obviously, when  $j = 1$ , Equation (6) provides the NFSV.

In the present study, we use the O-NFSVs to represent model uncertainties that influence TC forecasts and we conduct ensemble forecasting experiments. The Spectral Projected Gradient 2 solver (SPG2: Birgin *et al.*, 2000) is used to solve the optimization problem Equation (6). The SPG2 solves a minimum-value problem, while the O-NFSVs are obtained through maximum-value problems. Therefore, when we use SPG2 to estimate the O-NFSVs, we reverse the maximum-value problem  $J(f_j^*)$  to a minimum-value problem  $-J(f_j^*)$  to search for the O-NFSVs along the descending direction of the gradient of the cost function with respect to the tendency perturbations. More details on the use of that solver can be found in Birgin *et al.* (2000) and Appendix A of Duan *et al.* (2022).

### 3 | THE WRF MODEL AND TC CASES

The Weather Research and Forecasting model version 3.9.1 (WRFV3.9.1) is used in the present study, with three different model resolutions. The coarse-resolution model is set up on a single domain with a horizontal resolution of 54 km, which contains  $120 \times 90$  grid points. The high-resolution model is built on 3 two-way nested domains with horizontal grid spacings of 54, 18 and 6 km, which contain  $120 \times 90$ ,  $61 \times 61$  and  $97 \times 97$  grid points, respectively, and the latter two domains move with the TC centre. The convection-permitting resolution model is built on 4 two-way nested domains with horizontal grid spacings of 54, 18, 6 and 2 km, which contain  $120 \times 90$ ,  $61 \times 61$ ,  $97 \times 97$  and  $202 \times 202$  grid points, respectively, and the latter three domains move with the TC centre. After several tests and according to the literature, 31 vertical levels with the top level up to 20 hPa are adopted. The physical parametrization schemes used in the model include the Lin microphysics scheme (Lin *et al.*, 1983), the Rapid Radiative Transfer Model (RRTM), the long-wave radiation scheme (Mlawer *et al.*, 1997), the Dudhia short-wave radiation scheme (Dudhia, 1989), the Yonsei University (YSU) planetary boundary-layer scheme (Hong *et al.*, 2006), and the Kain–Frisch cumulus scheme (Kain, 2004). Note that the Kain–Frisch cumulus scheme is not used in the domains with 6 and 2 km resolutions.

Note that when calculating the O-NFSVs, the WRFV3.9.1 adjoint model is used to calculate the relevant gradient (see Section 2). However, only simplified physical parametrization schemes (Iscond scheme, surfdrag scheme and ducu scheme) are available for the adjoint model. Therefore, we must use WRF and its simplified

adjoint to solve the O-NFSVs and conduct ensemble forecasting experiments. Such configurations of WRF and its adjoint have been widely used in studies of data assimilation or ensemble forecasting (Xiao *et al.*, 2008; Zhang *et al.*, 2013). In the integration of the WRF model, the initial and boundary conditions are provided by the operational Global Forecast System (GFS) forecast data with a resolution of  $1 \times 1^\circ$ . To evaluate the ensemble forecasting skill for TCs, the best track data from the China Meteorological Administration (CMA) are used, and six TC cases occurring from 2017 to 2020 are selected. These cases all formed in the western North Pacific and underwent a process of intensification or even RI before landing over China. Table 1 shows detailed information on the six TC cases and the start and final times of their respective forecasting periods, including the whole RI process.

### 4 | EXPERIMENTAL STRATEGY

To compute the O-NFSVs that represent the model uncertainties for TC intensity forecasting, we follow the approach proposed in Qin *et al.* (2020) by considering the departure of the sea-level pressure (SLP) from the control forecast as the basis of the cost function and utilize the moisture and potential temperature that are directly related to TC intensity as components of the tendency perturbation. The optimization problem in Equation (6) can then be rewritten as follows.

$$J(f_j^*) = \max_{E_j \in \Omega_j} \left\{ [\text{SLP}_T(u_0 f_j) - \text{SLP}_T(u_0, 0)]^2 \right\}, \quad (8)$$

where

$$\Omega_j = \begin{cases} \{f_j \in \mathfrak{R}^n | E_j \leq \delta\}, j = 1 \\ \{f_j \in \mathfrak{R}^n | E_j \leq \delta, f_j \perp \Omega_k, k = 1, \dots, j-1\}, j > 1 \end{cases} \quad (9)$$

and

$$E_j = \frac{1}{D} \iint_{D\sigma} \left[ \left( \frac{g}{N\theta} \right)^2 \theta_j'^2 + \frac{L^2}{C_p T_r} qv_j'^2 \right] dDd\sigma. \quad (10)$$

Here,  $\text{SLP}_T(u_0, 0)$  and  $\text{SLP}_T(u_0, f_j)$  represent the control forecast of the SLP of the TC centre and its perturbed forecast generated by the tendency perturbation  $f_j$  at the lead time  $T$ , and  $E_j$  is the energy norm for measuring tendency perturbations  $f_j$  with  $\delta$  being a positive number that constrains the amplitude of the tendency perturbations. In Equation (10),  $\theta_j'$  and  $qv_j'$  are the potential temperature and moisture perturbations,  $D$  and  $\sigma$  are the horizontal domain and the vertical level, and the constants  $L = 2.51 \times 10^6 \text{ J} \cdot \text{kg}^{-1}$ ,  $C_p = 1005.7 \text{ J} \cdot \text{kg}^{-1}$ ,  $T_r = 270 \text{ K}$  and  $\bar{N}$  is the



TABLE 1 Information and forecast periods of six TC cases.

| Name    | No.    | Start time<br>(hr: UTC) | End time<br>(hr: UTC) | Peak intensity<br>(hPa; m·s <sup>-1</sup> ) |
|---------|--------|-------------------------|-----------------------|---|
| Hato    | 201713 | 0000 21 Aug             | 0600 23 Aug           | 935; 52                                     |
| Maria   | 201808 | 0000 05 Jul             | 0000 09 Jul           | 925; 58                                     |
| Lekima  | 201909 | 1200 05 Aug             | 1600 05 Aug           | 915; 62                                     |
| Ling    | 201913 | 1200 03 Sep             | 1200 06 Sep           | 930; 55                                     |
| Hagupit | 202004 | 0000 02 Aug             | 0400 04 Aug           | 965; 42                                     |
| Haishen | 202010 | 0000 02 Sep             | 0000 06 Sep           | 920; 60                                     |

Note: The numbers (No.) (year + ordinal no. that year) and intensities (the minimum sea-level pressure and maximum 10 m wind speed) are from the Best-Track data of China Meteorological Administration (CMA).

Brunt–Vaisala frequency. Note that perturbing the wind field directly could often induce a collapse of the forecasts. This type of perturbation therefore will not be considered in the present work.

The WRF model with different resolutions of 54, 6 and 2 km, as stated in Section 3, is used to conduct ensemble forecasting experiments for predicting TC intensities, and the ensemble members are generated by superimposing the O-NFSVs [see Equations (8)–(10)] on the control forecasts. For convenience, we refer to these experiments as EX-54 km, EX-6 km and EX-2 km, respectively. In view of the massive computational costs of ensemble forecasts, we follow the traditional way to calculate the model perturbations. That is, the O-NFSVs are calculated and superimposed on the whole domain for EX-54 km, while for EX-6 km and EX-2 km, the O-NFSVs are calculated over the parent domain and superimposed on the inner nested domains by interpolation in an attempt to keep the tendency perturbations dynamically consistent between the inner and parent domains. Certainly, the 54 km resolution is coarse for TC intensity forecasts; however, some operational ensemble forecasting systems with model resolutions of 10–100 km, such as the European Centre for Medium-Range Weather Forecasts ECMWF-EPS and the National Centers for Environmental Prediction NCEP-GEFS, provide real-time TC ensemble forecasting not only for TC tracks but also for TC intensities at model resolutions of 18 and 34 km, respectively (Yamaguchi *et al.*, 2009; Magnusson *et al.*, 2019; Ono *et al.*, 2021; Zhou, 2022). Therefore, in line with these operational forecasts, we also provide the results of the 54 km resolution on TC forecasts in the present work despite this too coarse resolution.

The skill of the ensemble forecasts generated by the O-NFSVs depends on the tendency perturbation amplitude [indicated by the constraint radii  $\delta$ ; see Equations (8) and (9)] and the number of ensemble members. When we calculate the O-NFSVs, we find that the values of

the cost function for O-NFSVs reach near saturation at the 26th NFSV. This indicates that the NFSVs, whose orders are larger than 26, exhibit almost the same sensitivity as the 26th NFSV and do not help increase the ensemble spread when they are superimposed on the control forecast. Therefore, a total of 26 O-NFSVs are computed that are superimposed on the control forecast with positive/negative signs, leading to 52 perturbed forecasts. Once a positive/negative pair of NFSV perturbations is superimposed on the control forecast, it could lead to either a low pressure or a high pressure depending on their O-NFSVs signs. As the ensemble members of a high-pressure structure are not physically relevant for TCs, we only select those that have a low-pressure structure. From the 52 perturbation forecasts generated by the 26 positive/negative perturbation pairs, a total of 26 perturbed forecasts remain, which, together with the control forecast, are composed of 27 ensemble forecasting members.

The tendency perturbation amplitudes [i.e.  $\delta$  in Equations (8) and (9)] are fixed according to the mean tendency of the WRF using the control forecasts of the six TCs. Specifically, the WRF model is integrated to generate the control forecasts. The tendency of the potential temperature and moisture components (which are respectively denoted by  $F_T$  and  $F_Q$ ) are then calculated as  $F = \frac{U_{t+\Delta t} - U_t}{\Delta t}$  for each TC, where  $U_t$  is the state variable (i.e. potential temperature or moisture) at time  $t$  and  $\Delta t$  is the time step. We take the mean of the tendency  $F$  for six TCs and determine the amplitude interval  $[0.1\bar{F}, 0.2\bar{F}]$  of the tendency perturbations to guarantee the stability of model integration, where  $\bar{F}$  is the mean of the tendency  $F$  for the six TCs. Then, the tendency perturbation energy [measured by  $\delta$  in Equation (8)] is optimized in the range of  $1 \times 10^{-9}$  to  $7 \times 10^{-9} \text{ J} \cdot \text{kg}^{-1} \cdot \text{s}^{-1}$ . In this range, the skill of ensemble forecasting generated by the O-NFSVs with four different amplitudes of  $\delta = 1 \times 10^{-9}$ ,  $3 \times 10^{-9}$ ,  $5 \times 10^{-9}$  and  $7 \times 10^{-9} \text{ J} \cdot \text{kg}^{-1} \cdot \text{s}^{-1}$  are compared

**TABLE 2** Parameter values of SKEB and SPPT for EX-54 km, EX-6 km and EX-2 km.

| Parameter |                    | EX-54 km | EX-6 km | EX-2 km |
|-----------|--------------------|----------|---------|---------|
| SKEB      | tot_backscat_psi   | 1E-5     |         |         |
|           | tot_backscat_t     | 1E-6     |         |         |
|           | Decorrelation time | 6 hr     |         |         |
| SPPT      | gridpt_stddev_sppt | 0.5      | 0.33    | 0.33    |
|           | lengthscale_sppt   | 150 km   |         |         |
|           | timescale_sppt     | 6 hr     | 3 hr    | 3 hr    |

Abbreviations: SKEB, stochastic kinetic-energy backscatter; SPPT, stochastically perturbed parametrization tendency.

for EX-54 km, EX-6 km and EX-2 km. A total of 1944 perturbed forecasts are conducted for the six TCs. Then, the perturbation amplitude providing the best forecasting skill is finally determined as  $\delta_{54\text{km}} = 7 \times 10^{-9} \text{ J} \cdot \text{kg}^{-1} \cdot \text{s}^{-1}$  for EX-54 km,  $\delta_{6\text{km}} = 5 \times 10^{-9} \text{ J} \cdot \text{kg}^{-1} \cdot \text{s}^{-1}$  for EX-6 km, and  $\delta_{2\text{km}} = 3 \times 10^{-9} \text{ J} \cdot \text{kg}^{-1} \cdot \text{s}^{-1}$  for EX-2 km.

To explore the quality of the O-NFSVs perturbations, the ensemble forecasts are compared with those generated by two popular stochastic physics schemes, the stochastic kinetic-energy backscatter scheme (SKEB) and the stochastically perturbed parametrization tendencies scheme (SPPT). SKEB focuses on the reinjection of excessive energy dissipation near the truncation scale at higher scales in the numerical model. It compensates for the dissipated energy by adding a stochastic rotational zonal wind and meridional wind forcing and potential temperature forcing (Shutts, 2005; Berner *et al.*, 2009). The SPPT emulates the uncertainty of the subgrid-scale parametrization process by multiplying the tendency terms with stochastic perturbations (Palmer *et al.*, 2009; Berner *et al.*, 2015). Leutbecher *et al.* (2017) noted that SPPT sometimes provides a risk of unrealistic extreme situations [see also Lang *et al.* (2021)] and suggested ignoring unrealistic ensemble members in advance in operational forecasting. Therefore, for both SPPT and SKEB, we also select the ensemble members of the low-pressure structure as for the O-NFSVs ensemble. Furthermore, we find that when 26 perturbed forecasts are applied to SPPT and SKEB, the ensemble forecasting skill is not substantially modified when more perturbed forecasts are added. Therefore, we also adopt 26 perturbed forecasts for SPPT and SKEB as for the O-NFSVs. Then, a total of 1,458 individual forecasts are produced to compare the forecasting skills of the different model perturbation schemes at 54, 6 and 2 km resolution models.

The SKEB and SPPT schemes are two popular tendency perturbation schemes. Many studies have conducted a large number of comparative experiments to discuss the sensitivity of the parameters determining the amplitude of the tendency perturbations in the WRF

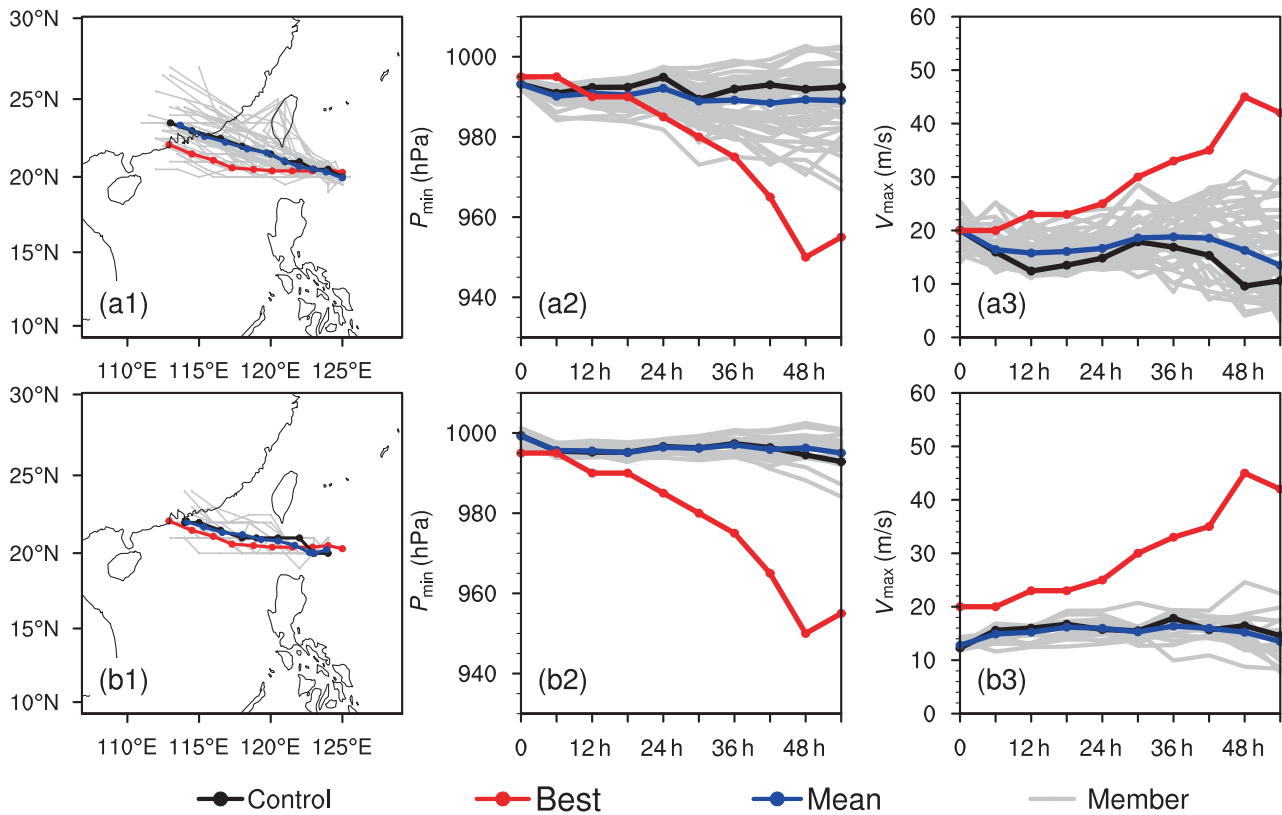
model (Berner *et al.*, 2011; Romine *et al.*, 2014; Ha *et al.*, 2015; Duda *et al.*, 2016; Melhauser *et al.*, 2017). Thanks to these studies, ranges of parameter values for the SKEB and SPPT schemes are available. Taking into account these ranges, we conduct additional numerical experiments to refine the ensemble parameters for the six TC cases to achieve an optimal forecasting skill. The main final parameters of the SKEB and SPPT schemes are provided in Table 2.

The ensemble forecasting skills are measured by the deterministic ensemble mean forecasting error, the probabilistic Brier Score (BS), the Relative Operating Characteristic (ROC) curves, and the Reliability Diagram (RD) for TC track strike probability within a 120 km radius of a given location [see Weber (2003)] for all lead times. The ensemble mean forecasting error and the BS are also used to measure the forecasting skill for TC intensity, graded from tropical depression, tropical storm, typhoon, severe typhoon, to super-typhoon. The BS, the ROC and the RD are described in the Appendix. In addition, the comparison of the ensemble spread to the ensemble mean forecast error is also performed to evaluate the reliability of the ensembles generated by the O-NFSVs (Eckel and Mass, 2005; Buckingham *et al.*, 2010). The details of all these measures are described in the Appendix.

In the following section, we first present the results of EX-54 km and further extend the application of the O-NFSVs to a higher model resolution (i.e. EX-6 km), and then test the approach at a more realistic convection-permitting resolution (i.e. EX-2 km) in Section 6.

## 5 | PERFORMANCE OF THE ENSEMBLE FORECASTING GENERATED BY O-NFSVS FOR TCS

In this section, ensemble forecast experiments for the six selected TC cases (see Table 1) are conducted, and the forecasting skills between the O-NFSVs and the two



**FIGURE 1** The ensemble forecasts conducted by (a) ECMWF and (b) NCEP, for (1) tracks, (2)  $P_{\min}$  and (3)  $V_{\max}$  of TC *Hato* (201713). The black lines are control forecasts, the red lines are the best tracks, the blue lines are the ensemble means, and the grey lines are ensemble members. The corresponding data are from the THORPEX Interactive Grand Global Ensemble (TIGGE) database.

stochastic physics schemes are compared. In the six TC cases, TC *Hato* (201713) experienced an RI and evolved into a super-typhoon within 24 hr before landfall. This typhoon led to serious disasters in Guangdong Province. Operational forecasting, however, failed to produce the RI process. Figure 1 displays the forecasts according to the ensemble forecast products generated by ECWFMF and NCEP, clearly showing that the ensembles generated are both under-dispersive for the TC intensity and miss the RI process of *Hato* (201713). To illustrate the usefulness of the O-NFSVs in representing model uncertainties and improving TC forecasting skill, we first explore the ensemble forecasts of *Hato* (201713). The statistics for the six TC cases investigated are then provided.

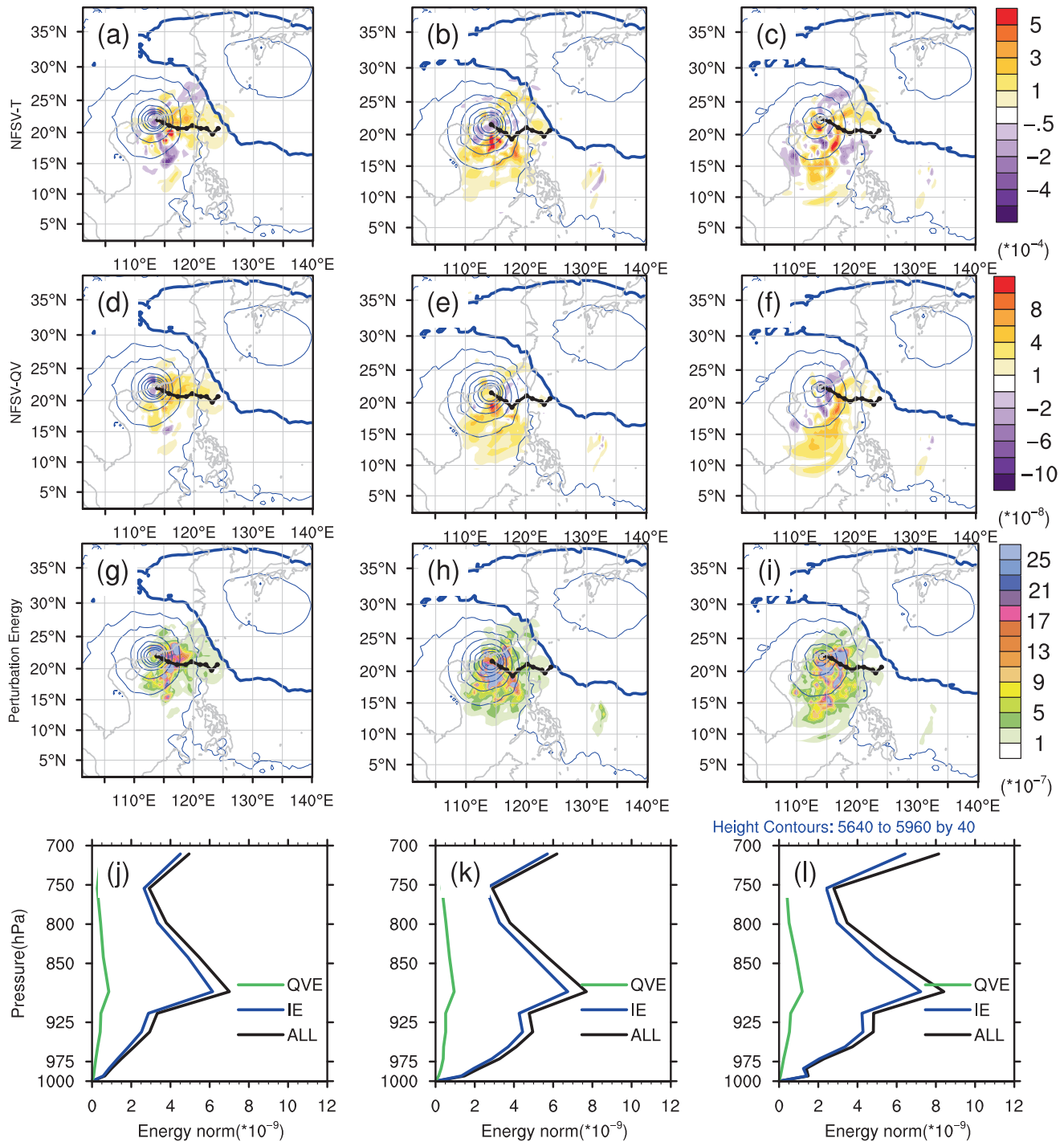
## 5.1 | The ensemble forecasting for TC *Hato* (201713)

### 5.1.1 | Results for the coarse resolution

The ensemble forecasting experiments for TC *Hato* (201713) are conducted in this subsection according to the strategy discussed in Section 4.

We first investigate the structures of the O-NFSVs. Both the NFSVs and their perturbation energies are generally located differently around the TC centre, with their energies mostly spanning the low levels of the atmosphere (i.e. 925–850 hPa). Figure 2 shows the first three NFSVs for TC *Hato* (201713) with their horizontal structures at 900 hPa, vertical integrated energies and vertical structures of the horizontal integrated energies. In fact, for each TC case, we observe that they exhibit similar spatial characteristics. This indicates that the change in TC intensity is more sensitive to the change in the potential temperature and moisture close to the TC centre, and the O-NFSVs tend to mainly describe the uncertainties in the potential temperature and moisture at these locations. Does this sensitivity of O-NFSVs result in much higher ensemble forecasting skill for TC forecasting?

Figure 3 shows the TC intensity measured by the minimum sea-level pressure ( $P_{\min}$ ) and maximum 10 m wind speed ( $V_{\max}$ ) of the control forecast and the ensemble forecasting members for TC *Hato* (201713) using the O-NFSVs, SKEB and SPPT for the coarse-resolution WRF model (i.e. the EX-54 km defined in Section 4). Some of the ensemble members generated by the O-NFSVs display an intensification process measured by both  $P_{\min}$  and  $V_{\max}$  similar to

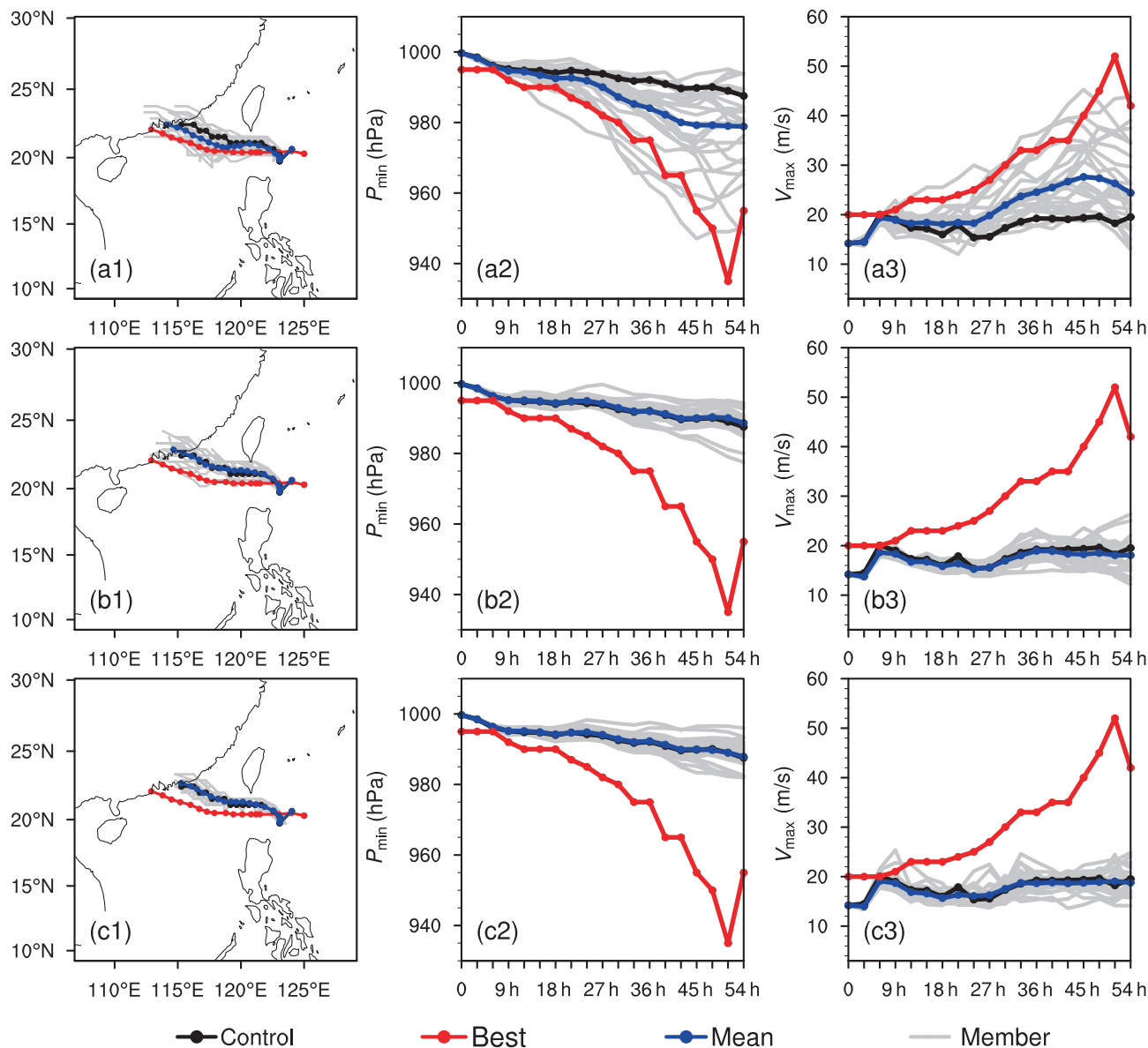


**FIGURE 2** Horizontal structures of the first three NFSV (a–c) moisture and (d–f) potential temperature tendency perturbations (shaded) at 900 hPa and (g–i) horizontal structures of vertically integrated tendency perturbation energies (shaded), and potential height field at 500 hPa (blue lines) for TC *Hato* (201713). The blue bold lines are the 5,880 gpm lines. The black dots and lines represent the TC tracks of perturbed forecasts. The vertical structures of horizontally integrated moisture (green lines), potential temperature (blue lines), and all (black lines) tendency perturbation energies of (j–l) the first three NFSVs, for TC *Hato* (201713).

the one produced in the best track, while the control forecast fails to do so. For SPPT and SKEB, the ensembles have a much smaller spread than those based on the O-NFSVs. Moreover, they all concentrate close to the control forecast, consequently failing to capture the intensification process.

As a result, the ensemble mean of  $P_{\min}$  and  $V_{\max}$  generated by the O-NFSVs is much closer to the observed. To further quantify the improvement achieved by the O-NFSVs, we compute the ensemble mean forecasting errors for all lead times against the control forecast error (Figure 4). It is



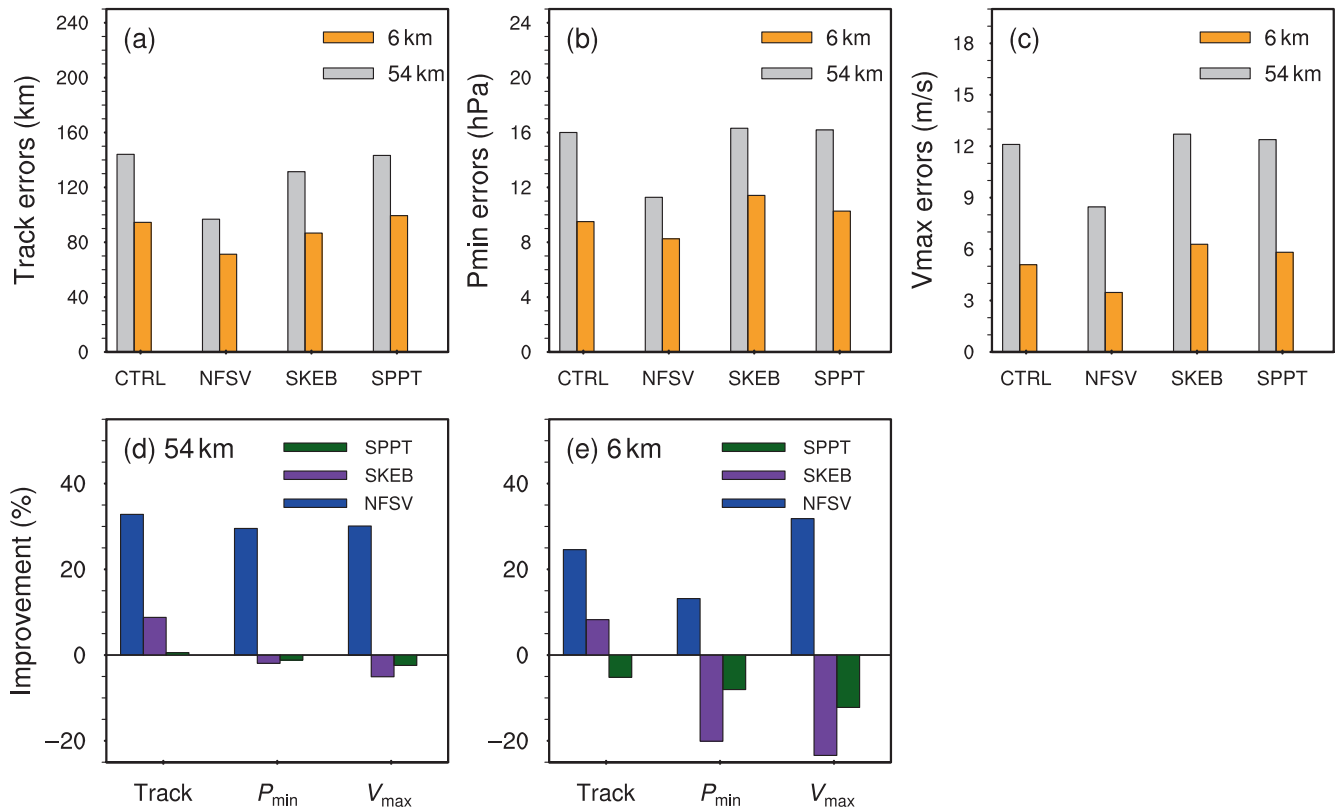


**FIGURE 3** The (a) ensemble forecasts generated by the O-NFSVs, (b) stochastic kinetic-energy backscatter (SKEB) and (c) stochastically perturbed parametrization tendency (SPPT), for (1) track, (2)  $P_{\min}$ , and (3)  $V_{\max}$  of TC *Hato* (201713). The Weather Research and Forecasting (WRF) model version has a coarse resolution. The black lines are the control forecasts, the red lines are the best tracks, the blue lines are the ensemble means, and the grey lines are the ensemble members.

clear that the forecasting errors of  $P_{\min}$  and  $V_{\max}$  are significantly smaller for the ensemble mean of the O-NFSVs than for both SKEB and SPPT. More precisely, the improvements achieved by the O-NFSVs reach 29.5% for  $P_{\min}$  and 30.1% for  $V_{\max}$ , while those obtained with SKEB and SPPT are  $-2.3\%$  and  $-1.2\%$  for  $P_{\min}$  and  $-5.0\%$  and  $-2.4\%$  for  $V_{\max}$ , respectively. In other words, the classical stochastic physics schemes tend to be neutral or even degrade the skill in predicting the intensity of TC *Hato* (201713). A comparison is also made with respect to the reliability of the ensembles generated by the O-NFSVs, SKEB and SPPT by comparing the ensemble spread to the ensemble mean forecast error (Figure 5). The discrepancy between

the ensemble spread and ensemble mean forecast error for the ensemble based on the O-NFSVs is much smaller than those of SKEB and SPPT for both  $P_{\min}$  and  $V_{\max}$  forecasting. The O-NFSVs therefore provide more reliable ensembles than the stochastic physics schemes for TC *Hato* (201713).

Although the focus in this article is on TC intensity, TC track forecasting is also investigated (Figure 3). The ensemble members generated by the O-NFSVs have a spread covering the best track, and the ensemble mean is closer to the best track than the two stochastic physics schemes. Quantitatively, the forecasting error made by the O-NFSVs members is 32.8% smaller than that of the

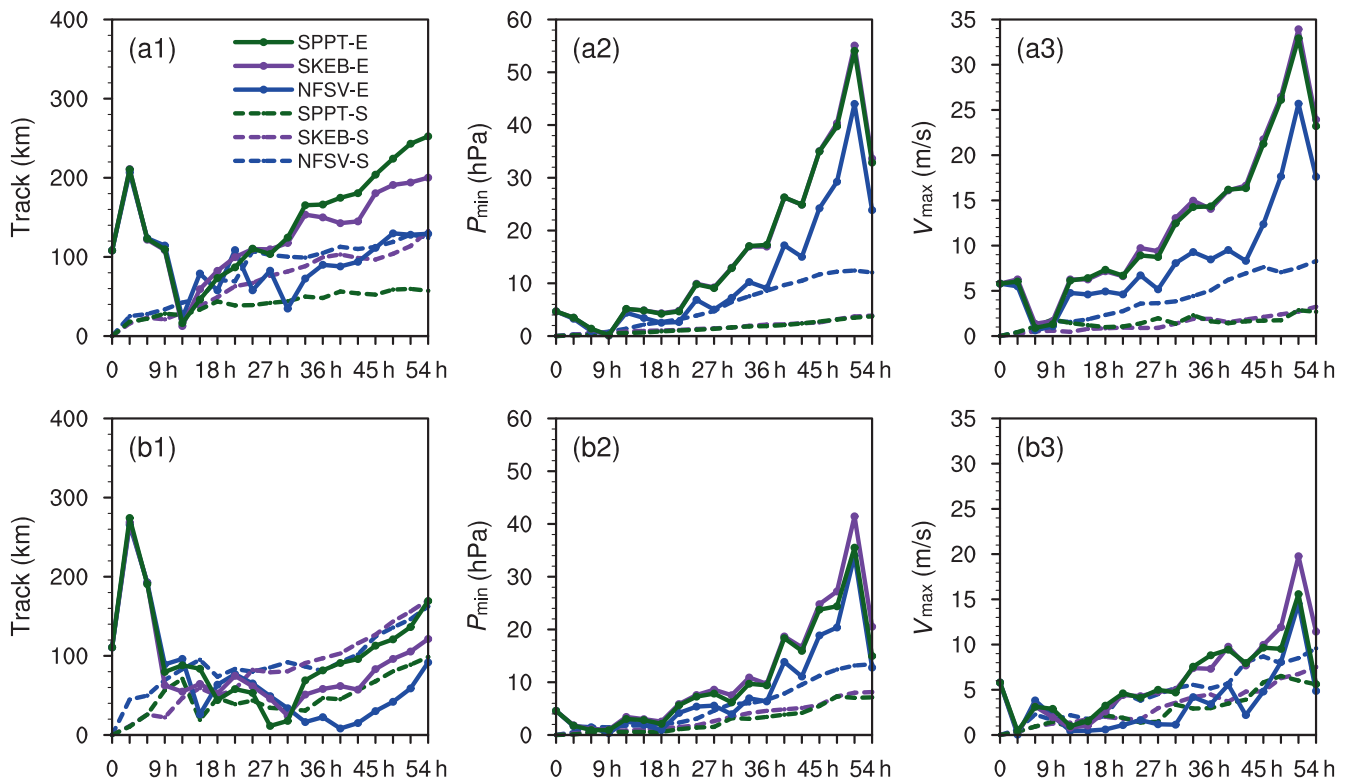


**FIGURE 4** The forecast errors of the control and ensemble mean for (a) track, (b)  $P_{\min}$ , and (c)  $V_{\max}$  for *Hato* (201713) using Weather Research and Forecasting (WRF) at coarse (grey bars) and finer (orange bars) resolutions. The improvements of the ensemble mean forecasts against the control forecasts, where the ensemble means are generated by the orthogonal nonlinear forcing singular vectors (O-NFSVs) (blue bars), the stochastic kinetic-energy backscatter (SKEB) (purple bars) and the stochastically perturbed parametrization tendency (SPPT) (green bars) using the WRF of the (d) coarse and (e) finer resolution, and the improvement is estimated by calculating the percentage of the error reduction averaged over all lead times.

control forecast, while for the SKEB and SPPT members, the forecasting errors are 8.8% and 0.6% smaller than that of the control forecast (Figure 4). In addition, when we examine the reliability of the ensembles for track forecasting, the ensemble generated by the O-NFSVs possesses a smaller discrepancy between the ensemble mean forecasting error and the ensemble spread than those of the SKEB and SPPT schemes (Figure 5). It is therefore clear that the representation of the model uncertainties on the track forecast of TC *Hato* (201713) can considerably be improved by using the O-NFSVs that perturb the potential temperature and the moisture mainly describing the model uncertainties on the TC intensity.

To clarify how the O-NFSVs contribute to optimizing the track of the TC, we plot in Figure 6 the spread of 500 hPa wind speed and the ensemble mean of the 500 hPa geopotential height generated by the O-NFSVs at lead times of 12, 24, 36 and 48 hr for TC *Hato* (201713). When the O-NFSVs are superimposed on the control forecast, the pressure field is perturbed, together with the pressure gradients and the wind fields near the TC. These

fields readjust to reach gradient wind balances. Simultaneously, the pressure gradients and the environmental flow between the TC and its surrounding circulation system (such as the western Pacific subtropical high, depicted by the contour of 5,880 geopotential height in Figure 6) are also progressively modified to reach a quasi-geostrophic balance between pressure and wind. These balances occur for each ensemble member generated by the O-NFSVs, leading to a large wind spread near the TC centre and around the TC track (Figure 6). The spread appropriately captures the uncertainties of the steering flow of the TC, and in turn, an ensemble mean TC track is closer to the best track. To confirm this, we also plot the evolution of the environmental steering flow vectors [i.e. the winds averaged over the radii between 300 and 700 km from the TC centre, from 700 to 500 hPa (Chan and Gray, 1982; Munsell and Zhang, 2014)] of the ensemble mean forecast generated by the O-NFSVs, the SKEB and the SPPT (Figure 7). The environmental steering flow vectors generated by the O-NFSVs are oriented farther to the west than those generated by SKEB and SPPT at lead times of 12 and 24 hr,



**FIGURE 5** The ensemble mean forecast errors (solid lines) and the ensemble spread (dashed lines) generated by the O-NFSVs (blue), stochastic kinetic-energy backscatter (SKEB) (purple) and stochastically perturbed parametrization tendency (SPPT) (green) using Weather Research and Forecasting (WRF) at coarse [(a), as in EX-54 km] and high [(b), as in EX-6 km] resolutions. The panels correspond to the forecasting of the (1) track, (2)  $P_{\min}$ , and (3)  $V_{\max}$  of *Hato* (201713).

which displaces the TC track generated by the O-NFSVs westward, now much closer to the best track than those generated by SKEB and SPPT (Figure 3). This ultimately leads to a higher forecast skill of the track for TC *Hato* (201713).

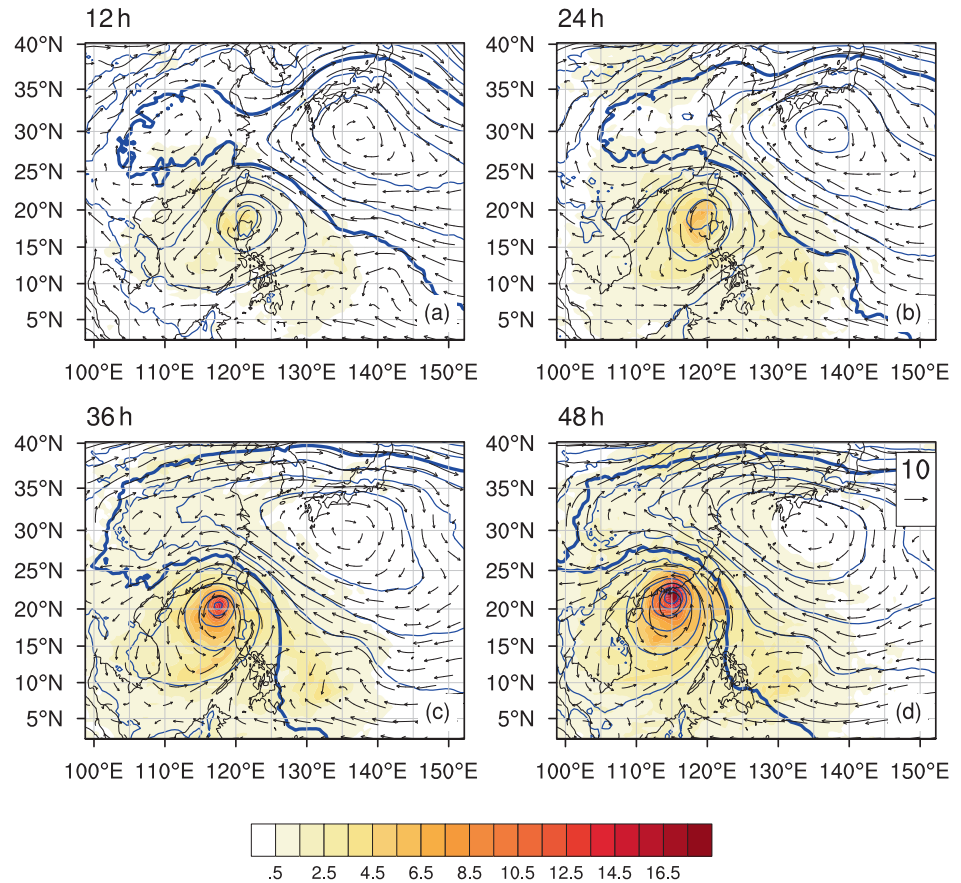
### 5.1.2 | Results with the higher resolution

Although the ensembles generated by the O-NFSVs above show higher forecasting skills for TC *Hato* (201713) compared with those generated based on SPPT and SKEB, the model resolution is relatively coarse. To evaluate the impact of the resolution on the ensemble forecast generated by the O-NFSVs, we extend the analysis to the high-resolution WRF model (i.e. the EX-6 km in Section 3).

The control forecasts and associated ensembles for TC *Hato* (201713) in EX-6 km are plotted in Figure 8. The errors of the control forecast for the TC intensity decrease compared with those in EX-54 km. However, the offshore RI is still not completely captured. It is also important to note that the ensembles generated by the O-NFSVs again present a larger spread for both  $P_{\min}$  and  $V_{\max}$ . Furthermore, there exist a few members ( $\sim 1/6$  of the ensemble

members) that capture the RI processes. This suggests that the O-NFSVs used in the high-resolution model provide a more appropriate representation of the model uncertainty for TC intensity forecasting, particularly for RI forecasting. We also find that the RI timing in the ensemble members generated by O-NFSVs exhibits a wider diversity than those generated by the SKEB and SPPT schemes (Figure 8). Some members display the RI earlier than it occurred in reality, while other members exhibit the RI later, but the ensemble mean forecast is much closer to the reference solution compared with the SKEB and SPPT schemes. Therefore, it seems that the ensemble members generated by the O-NFSVs of the high-resolution model also describe the uncertainty of the RI timing more appropriately than those created by the SKEB and SPPT schemes. In addition, we find that the spread of the ensembles generated by SKEB and SPPT for the TC intensity in EX-6 km increases compared with that in EX-54 km. Romine *et al.* (2014) indicate that small-scale motions exhibit faster instabilities and that stochastic perturbations at this scale may partly project on unstable growing perturbations and then develop faster than those at much larger scales. This may explain why the ensembles generated by SKEB and SPPT with high resolution in EX-6 km have a

**FIGURE 6** The ensemble spread of wind speed at 500 hPa (shaded), and the 500 hPa geopotential height field (lines) and the 500 hPa wind (vectors) of the ensemble mean generated by orthogonal nonlinear forcing singular vectors (O-NFSVs) at lead times of (a) 12, (b) 24, (c) 36 and (d) 48 hr for TC *Hato* (201713) using Weather Research and Forecasting (WRF) at coarse resolution. The blue bold lines are the 5,880 gpm lines.



much larger spread than those with coarse resolution in EX-54 km. Despite this, the O-NFSVs still provide a better approach in describing the nonlinearly growing perturbations and consequently induce an ensemble spread much larger than those generated by SPPT and SKEB, especially for  $P_{\min}$ , with more ensemble members able to encompass the RI process. This also implies that the ensemble mean generated using the O-NFSVs experiences a higher improvement of 13.1% for  $P_{\min}$  and 31.8% for  $V_{\max}$  against the control forecast in EX-6 km (Figure 4), while those for SKEB and SPPT are  $-20.1\%$  and  $-8.1\%$  for  $P_{\min}$  and  $-23.4\%$  and  $-12.2\%$  for  $V_{\max}$ , respectively. In terms of reliability, the ensemble based on the O-NFSVs, as in EX-54 km, also displays a discrepancy between the ensemble mean forecast error and ensemble spread that is much smaller than those of SPPT and SKEB for both  $P_{\min}$  and  $V_{\max}$  (Figure 5). The O-NFSVs, for both the coarse and high resolutions, show a skill superior to those of the SKEB and SPPT schemes.

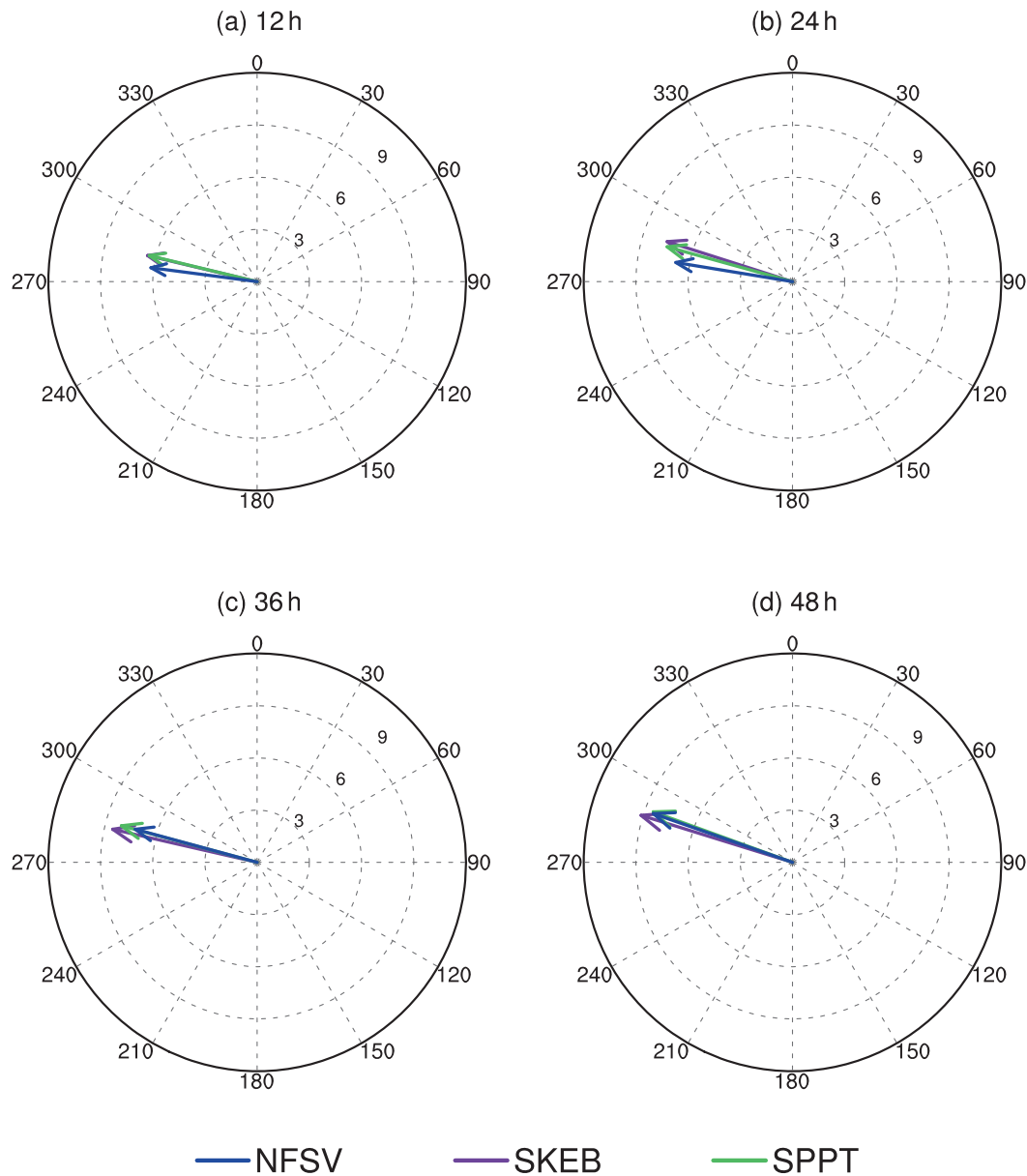
For the track forecast of TC *Hato* (201713), the control forecast at high resolution displays a forecast error smaller than that at coarse resolution. The ensembles generated by the O-NFSVs still provide an ensemble mean whose forecast errors are reduced by 24.0% compared with the control forecast. For the ensemble mean obtained with SKEB and

SPPT, the improvement in the control forecast is equal to 8.3% and  $-5.2\%$ , respectively (Figure 4). The higher performance of the O-NFSVs compared with SKEB and SPPT for TC track forecasting of *Hato* (201713) also holds for the high-resolution model version. Note that although the track errors in the control forecast and corresponding ensemble mean have been reduced due to the use of the high resolution, this is not generally the case; in fact, for the six TC cases investigated in the present study, there are only two TC cases that possess a track forecasting skill at high resolution that is higher than that at coarse resolution, while the other four TC cases have similar forecasting skills at both resolutions. This may suggest that the TC tracks are often controlled by large-scale environmental flow and that finer resolution plays a less important role in improving the track forecasting skill [see also Chan and Gray (1982), Chan (1985) and Munsell and Zhang (2014)]. The details are further discussed in the next section.

## 5.2 | Statistical evaluation of the ensemble forecasting skill for the six TCs

In Section 5.1, it is shown that the O-NFSVs provide remarkable tools to describe the model uncertainties for





**FIGURE 7** The environmental steering flow vectors of the ensemble mean forecasts generated by orthogonal nonlinear forcing singular vectors (O-NFSVs) (blue), stochastic kinetic-energy backscatter (SKEB) (purple), and stochastically perturbed parametrization tendency (SPPT) (green) at lead times of (a) 12, (b) 24, (c) 36 and (d) 48 hr for TC *Hato* (201713) using Weather Research and Forecasting (WRF) at coarse resolution. The numbers around the compass represent the orientation of the environmental steering flow vectors, and the radius represents the magnitudes of the environmental steering flow (unit:  $\text{m}\cdot\text{s}^{-1}$ ).

both track and intensity forecasting of TC *Hato* (201713). To further examine how reliable this conclusion is, we conduct ensemble forecasting experiments for five other TC cases presented in Section 3.

### 5.2.1 | Behaviour of the ensemble mean and spread

Figure 9 displays the ensemble spread and the ensemble mean forecast error averaged for the six TC cases. For

both EX-54 km and EX-6 km, the ensembles generated by the O-NFSVs usually possess a larger spread and have a smaller ensemble mean forecast error than the SKEB and SPPT, particularly during the intensifying period (i.e. 0–72 hr), including during the RI processes (Figure 11). This indicates a higher reliability of the ensembles generated by the O-NFSVs for both  $P_{\min}$  and  $V_{\max}$  forecasting. The improvement of the ensemble mean forecasts generated by the O-NFSVs, compared with the control forecasts averaged over all lead times, is 16.9% for  $P_{\min}$  and 11.4% for  $V_{\max}$  in EX-54 km, while it is 20.7% for  $P_{\min}$  and 14.0%

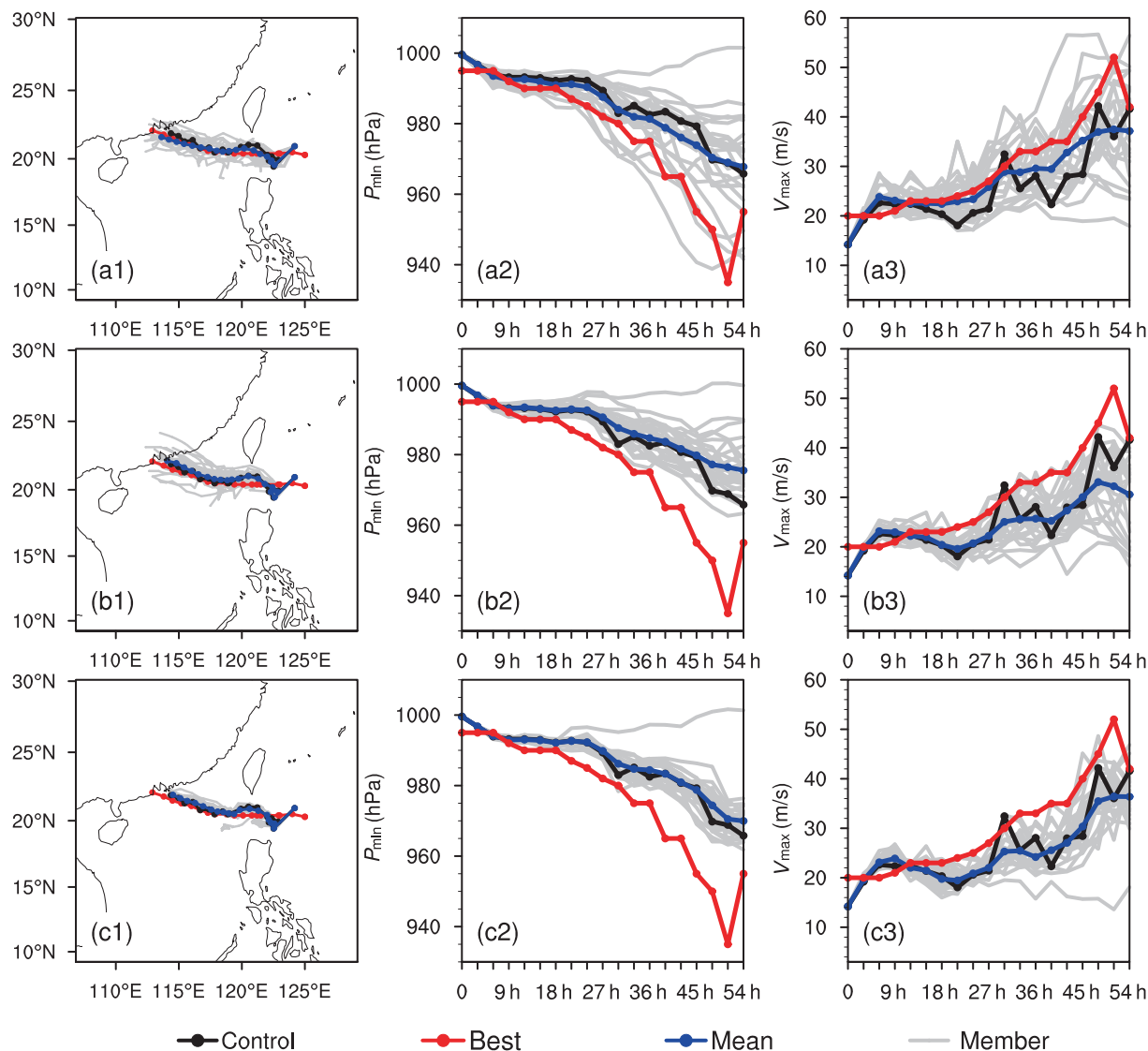


FIGURE 8 (a1–c3) As in Figure 3, but for the high-resolution model (EX-6 km).

for  $V_{\max}$  in EX-6 km. The improvement is also highly significant during the intensifying period (i.e. 0–72 hr) (Figure 11). Those generated by SKEB and SPPT hardly improve the errors, with modifications of  $-5.4\%$  and  $-2.6\%$  for  $P_{\min}$  and  $-3.8\%$  and  $4.5\%$  for  $V_{\max}$  in EX-54 km and of  $-3.1\%$  and  $-0.17\%$  for  $P_{\min}$  and  $-2.5\%$  and  $-3.6\%$  for  $V_{\max}$  in EX-6 km, respectively (Figure 10).

For the track forecasting of the TCs, the ensemble mean forecast error generated by the O-NFSVs is reduced by 32.9% in EX-54 km and by 21.6% in EX-6 km, compared with control forecast error, respectively; these reductions are also more significant than those in the ensemble mean forecasts made by SKEB and SPPT with 14.0% and 9.3% in EX-54 km and 7.3% and 1.9% in EX-6 km, compared with control forecast error, respectively (Figure 10). In addition, the ensemble spread shows the smallest difference from the ensemble mean forecasting error for the

O-NFSVs, and the ensembles generated by the O-NFSVs for track forecasting are more reliable than those of SKEB and SPPT. Therefore, although the ensembles are generated by perturbing the potential temperature and moisture only, the O-NFSVs as defined here represent a considerable improvement over SKEB and SPPT for track forecasting.

When comparing EX-54 km and EX-6 km, we find that the track errors of the control forecasts become much smaller in EX-6 km than in EX-54 km; furthermore, the ensemble mean generated by the O-NFSVs in EX-6 km further reduces the errors by 21.6%. However, the ensemble mean generated by SKEB improves the track forecasting skill to a lesser extent, by exactly 14.0%, while that of SPPT can hardly increase the skill. This result certainly emphasizes the important role of the O-NFSVs in improving the track forecasting skill. For the intensity forecasts, the finer resolution in the EX-6 km greatly enhances the forecasting

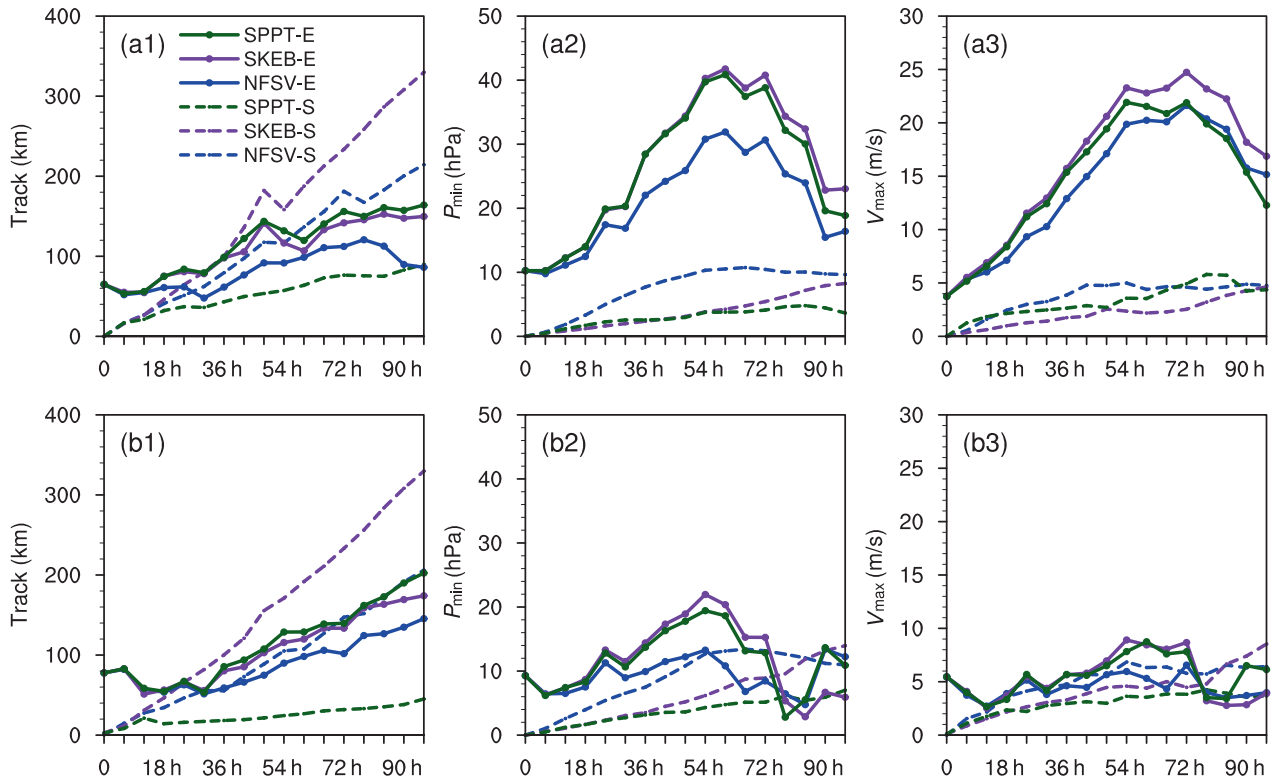


FIGURE 9 (a1–b3) As in Figure 5, but for the results averaged over the six TC cases.

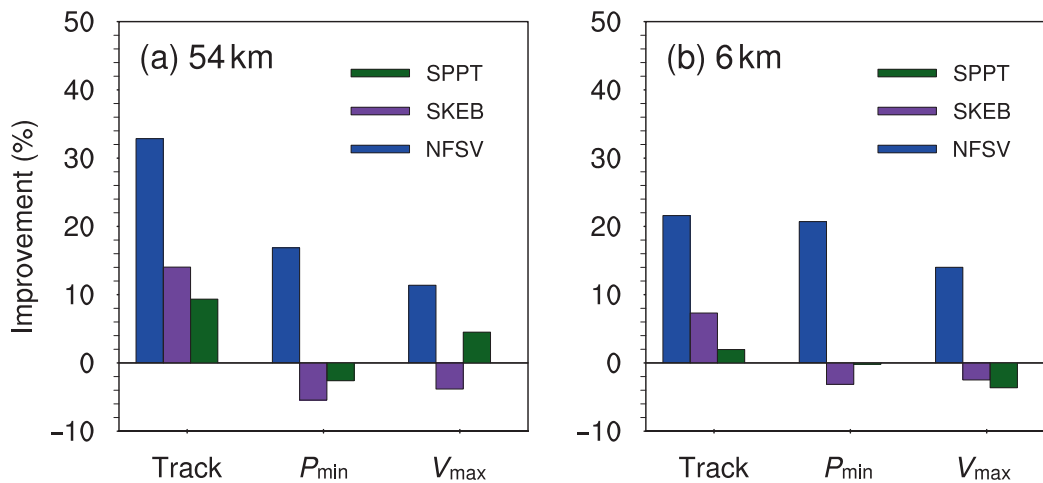
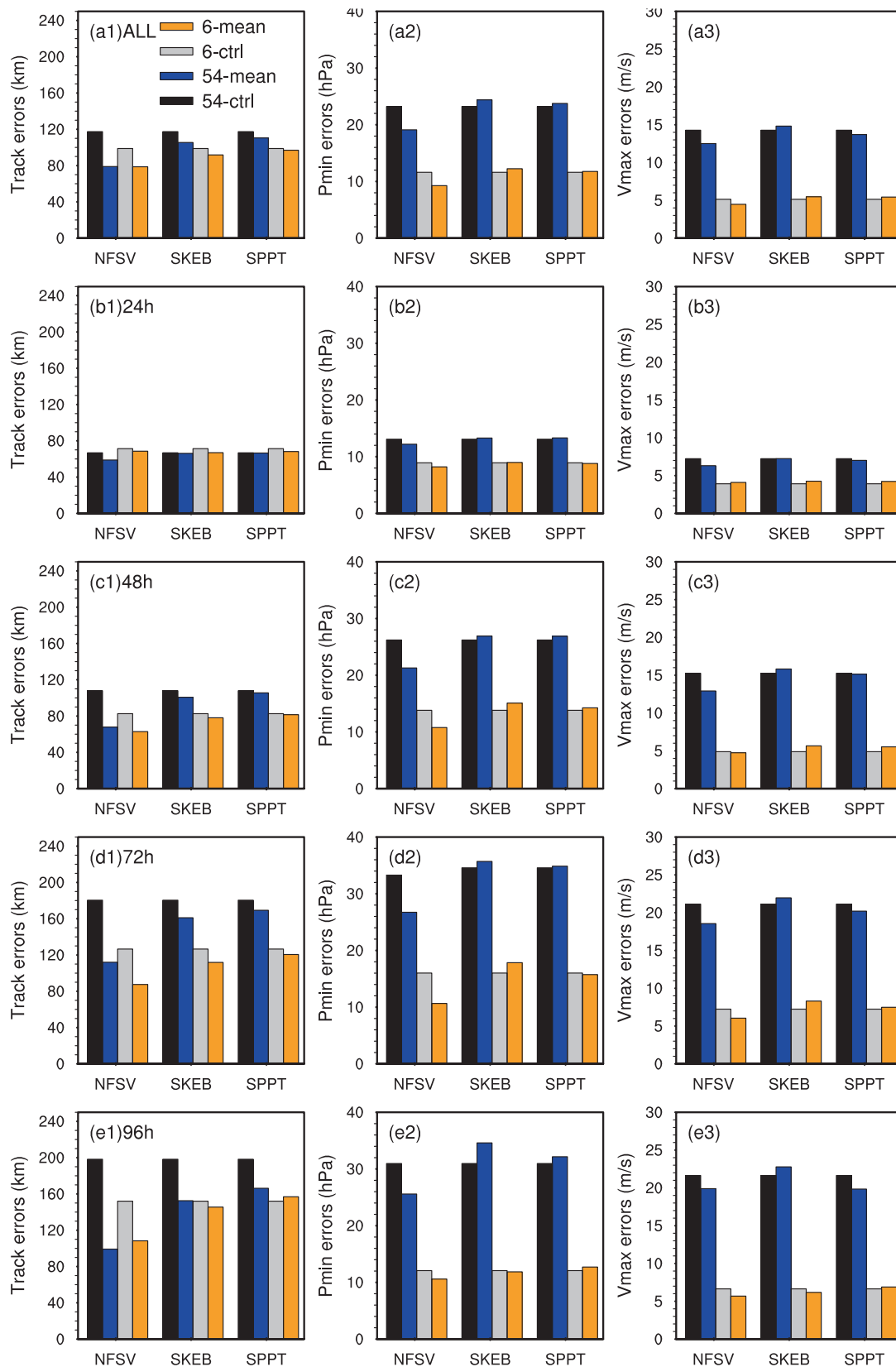


FIGURE 10 The improvements in the ensemble mean forecasts against the control forecasts averaged over all lead times, where the ensemble means are generated by the orthogonal nonlinear forcing singular vectors (O-NFSVs) (blue bars), stochastic kinetic-energy backscatter (SKEB) (purple bars) and stochastically perturbed parametrization tendency (SPPT) (green bars) using Weather Research and Forecasting (WRF) at (a) coarse and (b) finer resolution.

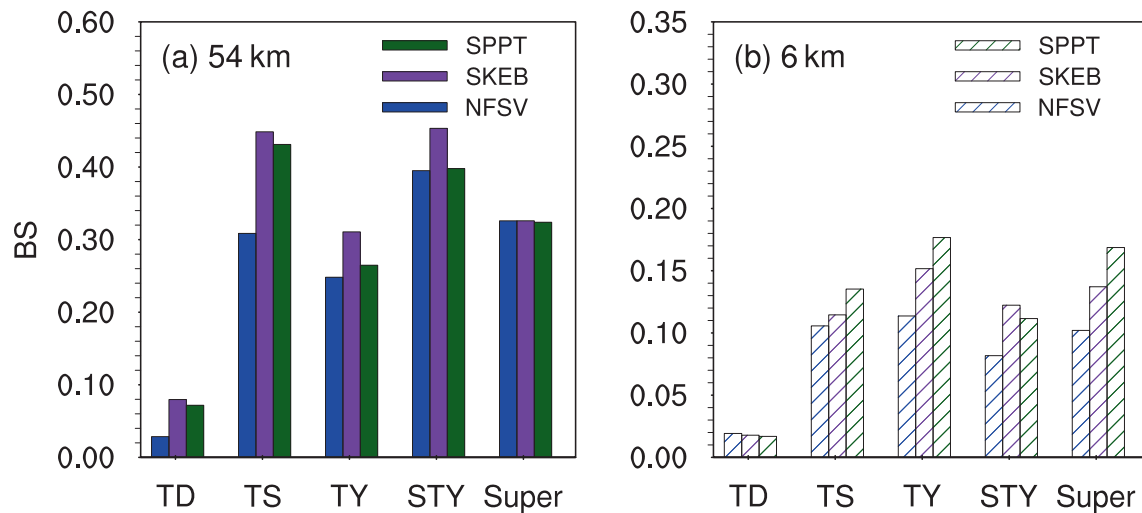
skill of the control forecast compared in the EX-54 km and in this situation, the O-NFSVs still show its great positive effect on the improvement of the forecasting skill for the TC intensity, but both the SKEB and the SPPT often present a negative effect. Therefore, the O-NFSVs show an obvious advantage over SKEB and SPPT in improving TC intensity forecasts.

Figure 11 shows the forecast errors of the control forecasts and the ensemble mean generated by the O-NFSVs, SKEB and SPPT at all lead times for TC track and intensity. The ensemble mean produced by the O-NFSVs in EX-54 km substantially reduces the track errors compared with the control forecast at all lead times. However, the O-NFSVs do not reduce the error any further beyond that



**FIGURE 11** The forecast errors for TC (1) track, (2)  $P_{min}$ , and (3)  $V_{max}$  in the control forecasts and the ensemble mean generated by the orthogonal nonlinear forcing singular vectors (O-NFSVs), stochastic kinetic-energy backscatter (SKEB), and stochastically perturbed parametrization tendency (SPPT). The results are obtained by using the Weather Research and Forecasting (WRF) at both coarse (54 km) and high (6 km) resolutions and then averaged over the six TC cases and (a) all lead times, (b) 0–24, (c) 24–48, (d) 48–72, (e) 72–96 h.





**FIGURE 12** The BS for the five categories of TC intensity [i.e. tropical depression (TD), tropical storm (TS), typhoon (TY), severe typhoon (STY), and super-typhoon (Super)] for the ensembles generated by the orthogonal nonlinear forcing singular vectors (O-NFSVs), stochastic kinetic-energy backscatter (SKEB), and stochastically perturbed parametrization tendency (SPPT) using Weather Research and Forecasting (WRF) at coarse [(a); 54 km] and high [(b); 6 km] resolutions. The statistics are computed on the six TC cases and all lead times.

of EX-54 km when the model resolution becomes finer in EX-6 km. If we further compare the control forecast in EX-6 km and the ensemble mean of the ensemble forecasts based on the O-NFSVs in EX-54 km, it is found that the former has a much larger forecast error. This suggests that the O-NFSVs play a more important role in track forecasting than the finer resolution plays. For both SKEB and SPPT, however, the ensemble mean in EX-54 km has a track error larger than that in the control forecast of EX-6 km. Therefore, the improvement of the ensemble mean generated by both SKEB and SPPT in EX-6 km for track forecasting skill is the result of a combined effect of finer resolution and ensemble perturbation approaches.

In summary, both SKEB and SPPT play a minor role in improving the forecast skill of the TC track, while O-NFSVs greatly improve it. In addition, TC tracks are mainly modulated by large-scale flows, such as the steering flow located between the TC and its surrounding circulation systems (e.g. Carr and Elsberry, 2000; Wu *et al.*, 2004; Torn *et al.*, 2018). This is further illustrated here by the impact of the O-NFSVs in EX-6 km, which provides nearly identical track errors to those in EX-54 km (Figure 11). Furthermore, much smaller track errors for the O-NFSVs than for the SKEB and SPPT illustrate that the O-NFSVs are better at capturing the model uncertainty of the large-scale flow.

The comparison of the TC intensity forecasts between EX-54 km and EX-6 km is also displayed in Figure 11. It reveals that the finer resolution allows for a large decrease in the errors of the control forecast, compared with the track forecasts. The ensemble mean generated

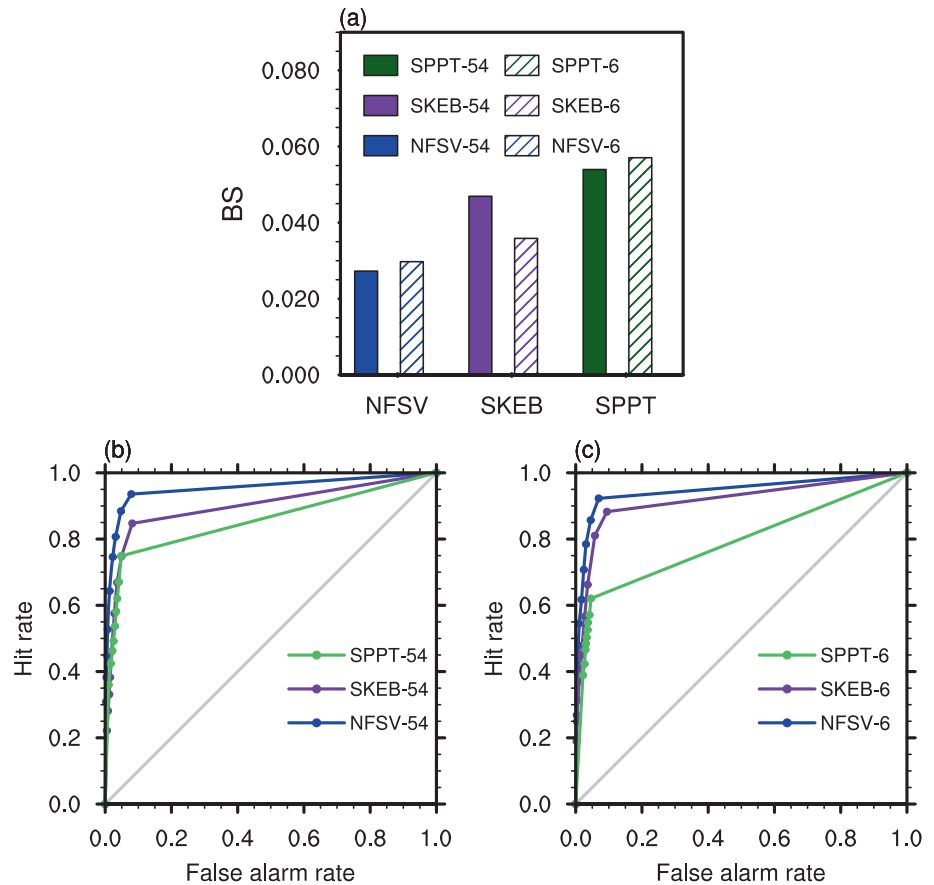
by the O-NFSVs further reduces the errors. Obviously, the reductions achieved by the O-NFSVs result from the combined effect of both finer resolution and the nature of the perturbations. For both SKEB and SPPT, the TC intensity errors of the ensemble mean forecasts are not further reduced from the control forecast in EX-6 km. The decrease in errors should therefore mostly be attributed to the impact of the finer resolution in the EX-6 km and not to the use of the SKEB and SPPT schemes.

### 5.2.2 | Probabilistic forecasts.

For the probability forecasts of TC intensity and track generated by the O-NFSVs, we estimate the Brier Score (BS) based on five categories of TC intensity [i.e. tropical depression (TD), tropical storm (TS), typhoon (TY), severe typhoon (STY), and super-typhoon (Super)] for the six TC cases. For the TC tracks, we also compute the relative operating characteristic (ROC) curve and the reliability diagram (RD) for track strike probability.

The BS averaged over all lead times for all TC cases is shown in Figure 12 for the five categories of TC intensity. The ensembles generated by the O-NFSVs exhibit a lower BS value for almost all categories of TC intensities in both EX-54 km and EX-6 km. This suggests that the probabilistic forecast skill achieved by the O-NFSVs is higher than those of the SKEB and SPPT for almost all categories of TC intensities. Moreover, we can see that when the model resolution becomes much finer in EX-6 km, the BS values obtained by the O-NFSVs, SKEB and SPPT are all

**FIGURE 13** (a) The BS of the track strike probability computed based on the ensembles generated by the orthogonal nonlinear forcing singular vectors (O-NFSVs), stochastic kinetic-energy backscatter (SKEB), and stochastically perturbed parametrization tendency (SPPT) for the six TC cases for the Weather Research and Forecasting (WRF) model at coarse (54 km; solid bars) and high (6 km; slant bars) resolutions. The Relative Operating Characteristic (ROC) curves for the track strike probability forecasts generated by the O-NFSVs, SKEB and SPPT using WRF at (b) coarse and (c) high resolutions are also plotted.



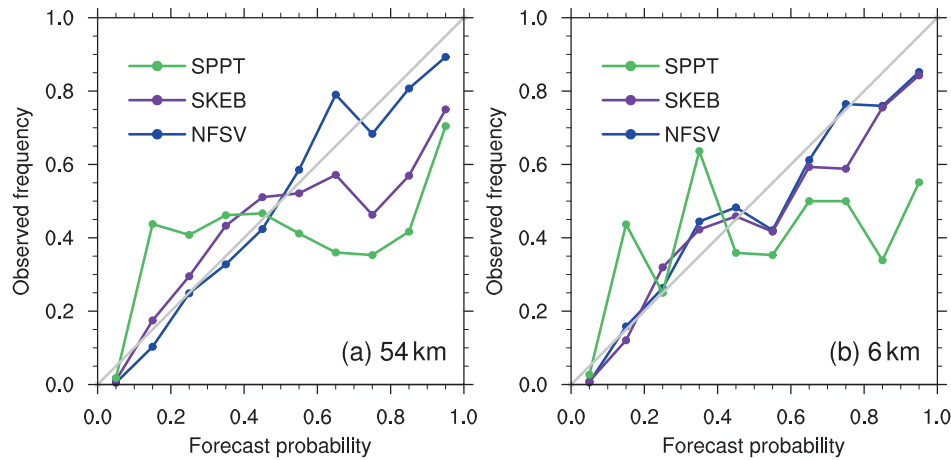
further reduced, which suggests that the finer resolution, as expected, helps the ensembles enhance the TC intensity forecasting skill. For the strike probability of the TC tracks (Figure 13), the ensembles generated by the O-NFSVs have the smallest BS values in both EX-54 km and EX-6 km. Moreover, the corresponding ROC curves display the highest hit rates and the lowest false alarm rates (Figure 13). Figure 14 also suggests that the O-NFSVs provide the most reliable ensembles for estimating the track strike probability with an RD line closest to the diagonal. Comparing the EX-54 km and EX-6 km for track forecasting, the O-NFSVs ensembles show a probabilistic skill as measured by the BS, the ROC curve and the RD line that are similar in both experiments. The SPPT ensembles show a probabilistic skill in EX-6 km with a higher BS, an ROC curve closer to the ROC diagonal line and an RD line farther from the RD diagonal line. While the SKEB ensembles show in the EX-6 km a much better probabilistic skill with a lower BS, an ROC curve far away from the ROC diagonal line and an RD line close to the RD diagonal, both are still less skilful than the O-NFSVs.

In summary, the ensembles generated by the O-NFSVs provide a better representation of the model uncertainty associated with the forecasts of TC intensity and its related track at resolutions of 54 and 6 km.

## 6 | TEST OF THE O-NFSVS IN THE CONVECTION-PERMITTING WRF MODEL

In Section 5, we demonstrated that the O-NFSVs provide a very good forecasting skill for TC track and intensity at both coarse and intermediate resolutions of the WRF model (i.e. EX-54 km and EX-6 km). However, these two versions of models cannot resolve the TC inner core structures and may underestimate the TC intensity and miss the RI process in TC forecasts. A convection-permitting resolution model (<4 km) better resolves the development of convection and would be more appropriate for TC intensity forecasting. Therefore, to test the use of the O-NFSVs in TC forecasting, we conduct ensemble forecast experiments using the WRF model with a 2 km resolution, referred to as EX-2 km, as in Section 4.

A similar experimental strategy as in Section 4 is applied to EX-2 km and, compared with both EX-54 km and EX-6 km, a much smaller amplitude of O-NFSVs, that is  $\delta_{2\text{km}} = 3 \times 10^{-9} \text{ J} \cdot \text{kg}^{-1} \cdot \text{s}^{-1}$  is used (see Section 4). In fact, from EX-54 to EX-2 km, the amplitude of O-NFSVs decreases, as expected since the resolution of the model is improved. From this evidence, it can be concluded that the O-NFSVs appropriately describe errors at the scale of



**FIGURE 14** The reliability diagram for the track strike probability forecasts generated by the orthogonal nonlinear forcing singular vectors (O-NFSVs), stochastic kinetic-energy backscatter (SKEB) and stochastically perturbed parametrization tendency (SPPT) using the Weather Research and Forecasting (WRF) model at (a) coarse and (b) high resolutions.

interest. Again, 26 perturbed forecasts for each TC are generated, which, together with the control forecast, are composed of the 27 ensemble members. For SKEB and SPPT, we also adopt similar experimental strategies as in Section 4 to determine the ensemble parameters for EX-2 km (see Table 2). A comparison is then made among O-NFSVs, SKEB and SPPT, and a total of 486 perturbed forecasts for the six TCs are conducted.

The control forecast and associated ensemble members for TC *Hato* (201713) are plotted in Figure 15. It is found that the errors of the control forecast for TC intensity, compared with those in EX-6 km, further decrease. However, the RI is not yet accurately captured, especially in the  $P_{\min}$  forecast. The O-NFSVs, compared with SKEB and SPPT, still present a much larger spread and better forecast skill for both  $P_{\min}$  and  $V_{\max}$ , which is similar to the results reported for EX-6 km (Figure 8). Furthermore, the O-NFSVs ensemble also describes the RI uncertainty much better than EX-6 km, with approximately 1/3 of the ensemble members capturing the RI process. In addition, the ensemble mean forecast is even closer to the true TC intensity. All these results indicate that even if the control forecast skill is improved due to the use of a convection-permitting model, the O-NFSVs are still able to further increase the TC intensity forecast skill.

The above ensemble forecast experiments are also conducted for the five other TC cases. Figure 16 displays the ensemble spread, the ensemble mean forecast error and the improvement of the ensemble mean forecasts averaged for the six TC cases. The results indicate that the O-NFSVs ensemble provides a large spread and a small ensemble mean forecast error compared with SKEB and SPPT for TC intensity, especially during the intensifying period (0–72 hr). More precisely, the ensemble mean forecasts generated by the O-NFSVs, compared with the control forecasts, decrease the errors by 23.6% for  $P_{\min}$  and 12.7% for  $V_{\max}$ , while those generated by SKEB and SPPT decrease the errors by much less (and sometimes

even increase the errors) by 4.5% and  $-0.1\%$  for  $P_{\min}$  and 7.0% and  $-0.9\%$  for  $V_{\max}$ , respectively. These results further underscore the excellent performance of the O-NFSVs for TC intensity forecasting at EX-2 km, as at EX-54 km and EX-6 km.

Concerning the reliability of the ensemble, the picture is more complicated, as the O-NFSVs ensemble is less reliable in  $V_{\max}$  forecasting in EX-2 km. The ensemble mean forecasting error is very small ( $<5 \text{ m}\cdot\text{s}^{-1}$ ) and not very close to the ensemble spread, suggesting that the spread is too large (Figure 16). This discrepancy for  $V_{\max}$  forecasting is also visible in the probabilistic BS scores for TC intensity categories, now not as good as those for SKEB and SPPT (not shown). However, for  $P_{\min}$  forecasting, the ensemble mean forecast error is closer to the ensemble spread for the O-NFSVs than for both SKEB and SPPT. Furthermore, track forecasting not only possesses an ensemble spread closer to the ensemble mean forecast error but also displays a smaller ensemble mean forecast error (Figure 16). The improvement is further illustrated with a much higher probability forecast skill, including a lower BS score, a better ROC curve and a better RD line (Figure 17). Interestingly, these skills are almost the same as in EX-54 km and EX-6 km.

We can still conclude that the TC track is mainly modulated by large-scale flow and that the O-NFSVs ensemble can much better depict the uncertainties of the large-scale flow, achieving higher track forecasting skills than SKEB and SPPT.

In summary, the superiority of the O-NFSVs over the SKEB and the SPPT is also found in the convection-permitting model for TC track and intensity forecasting. Nevertheless, with the change in model resolution from 6 km to the convection-permitting resolution of 2 km, the impact of SPPT and SKEB improves, and their differences with the O-NFSVs are reduced. In addition, we note in particular that for the six TC cases above, their control forecasts either underestimate (as in TC *Hato*'s

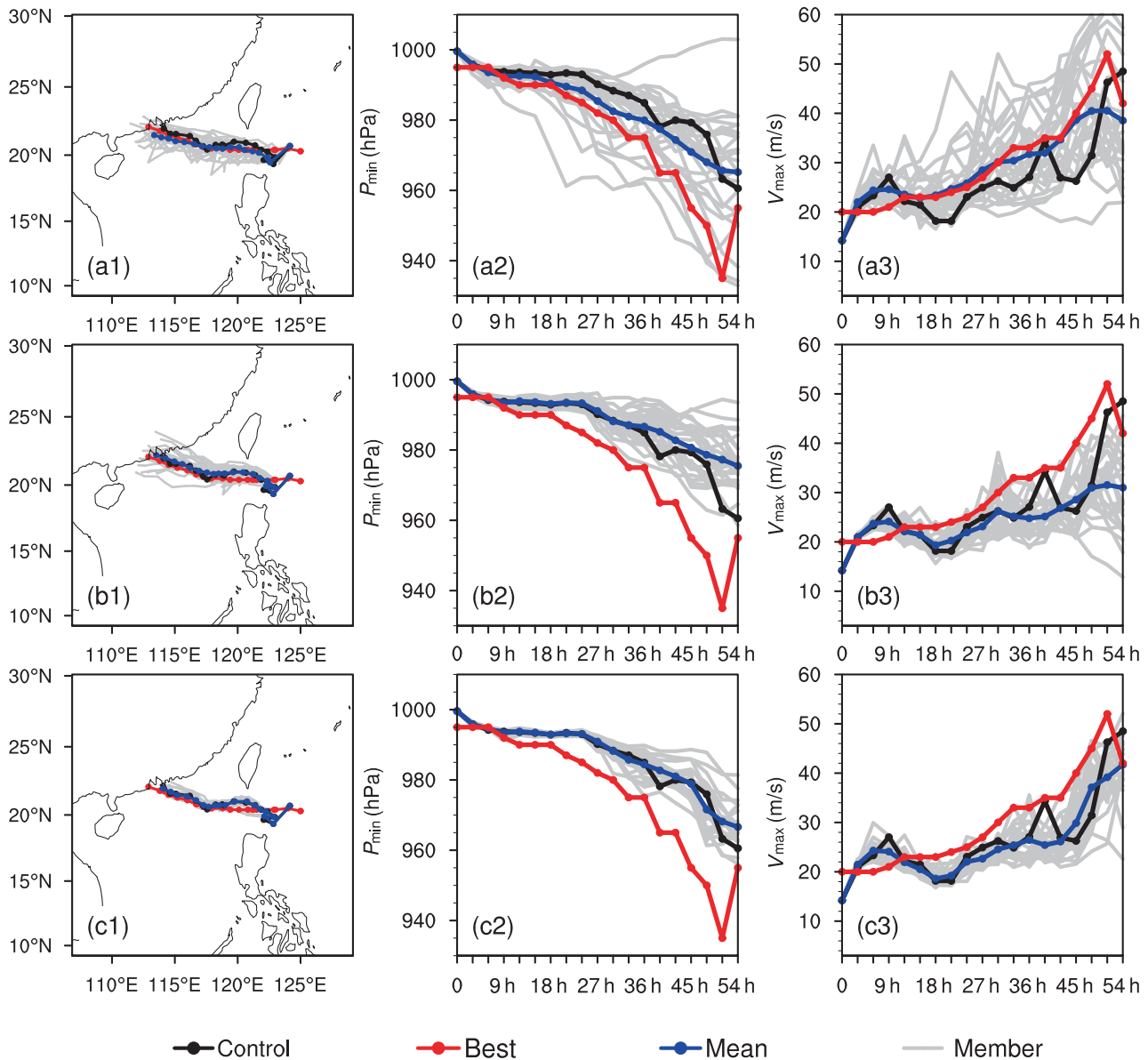


FIGURE 15 (a1–c3) As in Figure 3, but for the convection-permitting model (EX-2 km).

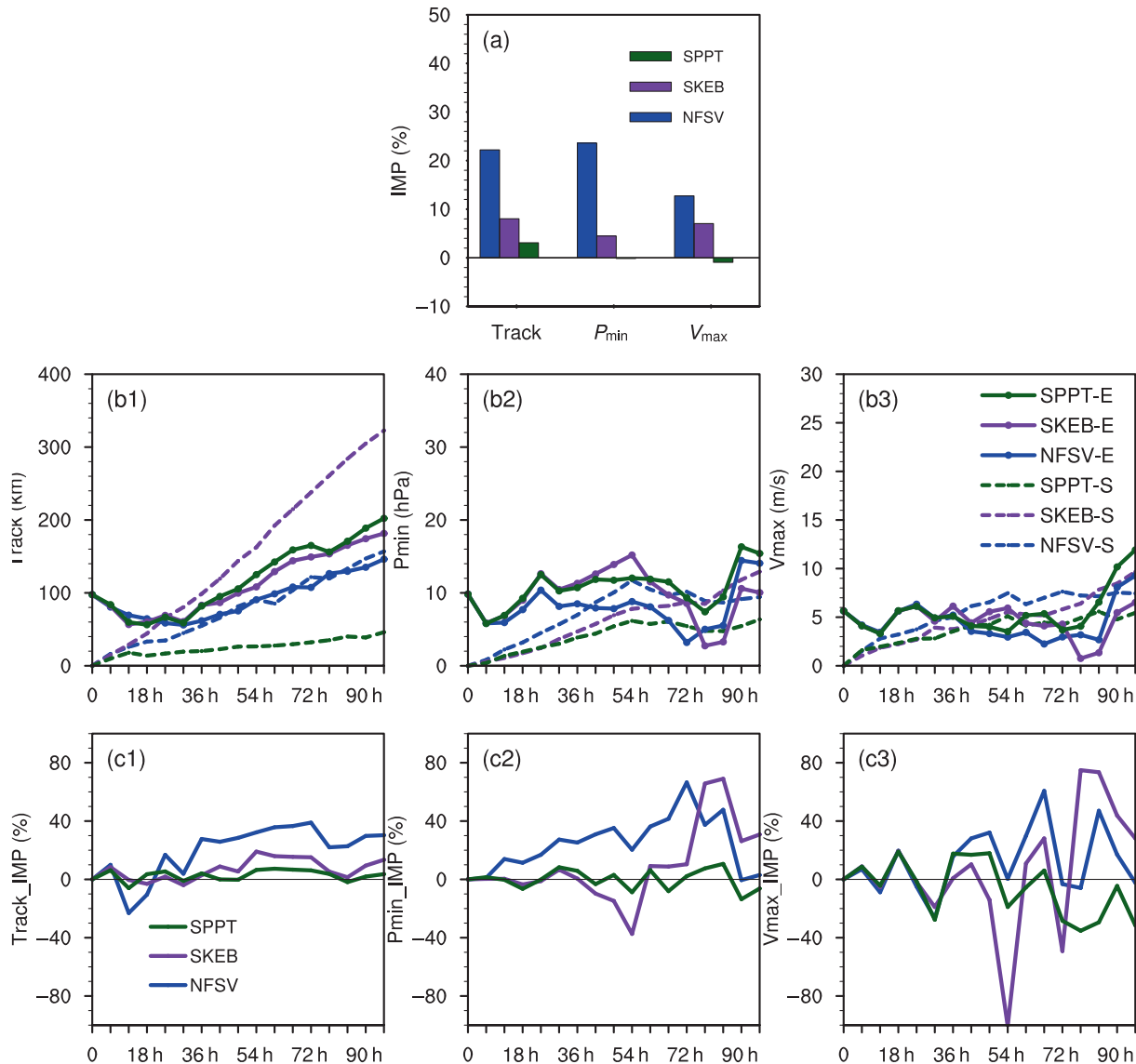
forecasting; see Figure 15) or overestimate the TC intensities (as in TC *Lekima*'s forecasting; the figure is omitted) in the EX-2 km. However, for these two situations, the ensemble forecasts generated by the O-NFSVs still provide larger spreads than the SKEB and the SPPT for intensity forecasting, especially for RI forecasting. Moreover, the ensemble means generated by the O-NFSVs are closer to the observed intensity than the control forecasts and even the ensemble means generated by the SKEB and the SPPT, although a selection of the ensemble members with low-pressure structures is applied. Of course, these results are obtained from a very limited number of TC cases, and to obtain a full picture, more cases should be considered to properly extract the differences between the different schemes and evaluate the genericity of the

superior performance of O-NFSVs as compared to SKEB and SPPT.

## 7 | CONCLUSION AND DISCUSSION

In this study, orthogonal nonlinear forced singular vectors (O-NFSVs) are used to emulate the model uncertainties that limit tropical cyclone (TC) intensity forecasts in the context of the advanced WRF model at three different resolutions: 54, 6 and 2 km. The ensemble forecast experiments for six TC cases are conducted with a focus on the rapid intensification (RI) process. A comparison of the performances is conducted between the O-NFSVs approach





**FIGURE 16** (a) Improvements in the ensemble mean forecasts against the control forecasts averaged over all lead times, where the ensemble means are generated by the orthogonal nonlinear forcing singular vectors (O-NFSVs) (blue bars), stochastic kinetic-energy backscatter (SKEB) (purple bars) and stochastically perturbed parametrization tendency (SPPT) (green bars) using the WRF model with 2 km resolution. (b) The ensemble mean forecast errors (solid lines) and ensemble spread (dashed lines) generated by the O-NFSVs (blue), SKEB (purple) and SPPT (green) using the WRF model with 2 km resolution. (c) The evolution of improvements at each lead time of the ensembles generated the O-NFSVs (blue lines), the SKEB (purple lines) and the SPPT (green lines) using the WRF model with 2 km resolution. The panels correspond to the forecasting of (1) the track, (2)  $P_{min}$ , and (3)  $V_{max}$ .

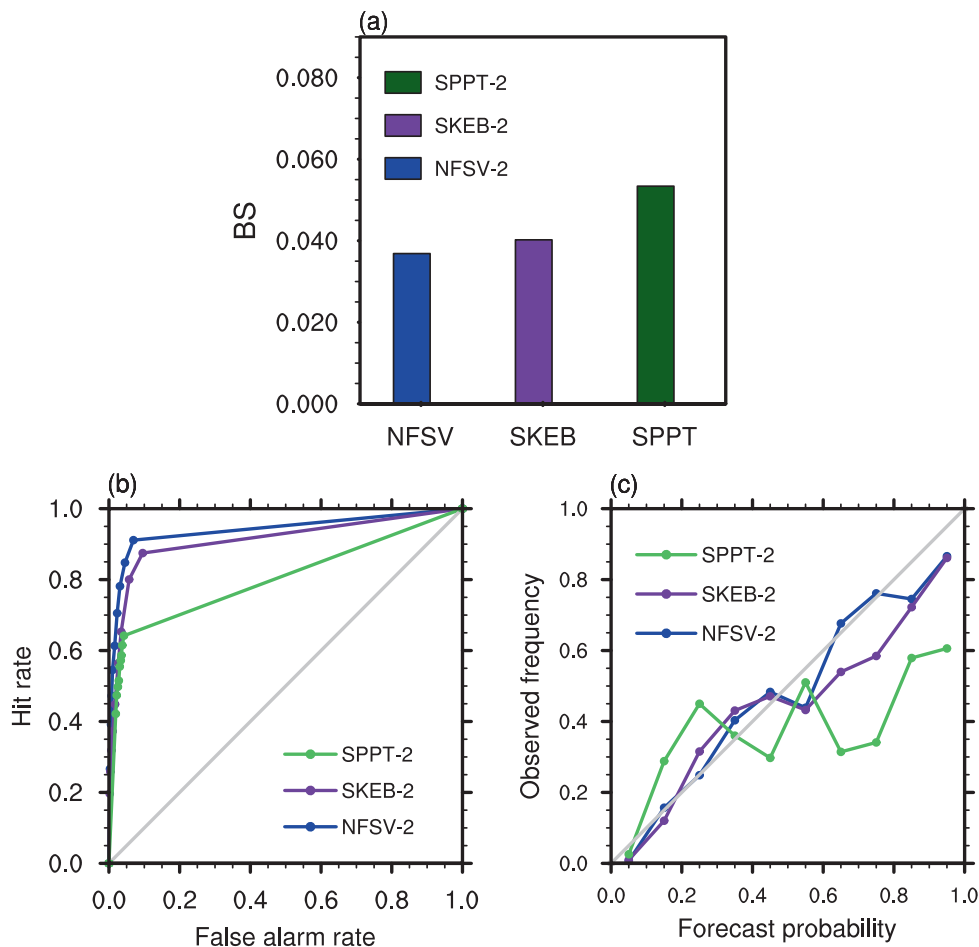
and two traditional stochastic physics schemes (SKEB and SPPT) based on the evaluation of the forecast skill and the reliability of the ensembles.

We first investigate the forecast of TC *Hato* (201713), which experienced an RI within 24 hr before landfall, at resolutions of 54 and 6 km. Our results demonstrate that the O-NFSVs provide a few ensemble members that capture the RI, particularly at the finer model resolution, while both the SKEB and SPPT do not produce any. Furthermore, the O-NFSVs allow for obtaining an ensemble mean whose forecasting errors are often significantly

smaller than those of SKEB and SPPT. It is also demonstrated that the O-NFSVs, although they are optimized based on potential temperature and moisture for TC intensity, produce a distribution of TC track members close to the best track at both model resolutions. SKEB and SPPT generate ensembles that mainly spread close to the control forecast, leading to an ensemble mean far away from the best track.

When statistics are gathered for the six TC cases selected, similar results are obtained at resolutions of 54 and 6 km. Specifically, at both model resolutions, the

**FIGURE 17** (a) The BS scores, (b) the ROC curves, and (c) the reliability diagrams, of the track strike probability computed based on the ensembles generated by the orthogonal nonlinear forcing singular vectors (O-NFSVs) (blue bars and lines), the stochastic kinetic-energy backscatter (SKEB) (purple bars and lines), and the stochastically perturbed parametrization tendency (SPPT) (green bars and lines) for the six TC cases using the WRF model with 2 km resolution.



ensembles generated by the O-NFSVs usually possess a much larger spread and smaller ensemble mean forecast errors than those generated by the SKEB and SPPT, particularly during the intensification period (including the RI processes). This leads to smaller discrepancies between the ensemble spread and the ensemble mean forecast error and points to a higher reliability of the ensembles generated by the O-NFSVs for TC intensity forecasting. The ensembles based on the O-NFSVs also show lower BSs for intensity and track and better ROC curves for track strike probability. These results illustrate that the probability forecast skills for TC intensity and track are higher than those of SKEB and SPPT. Furthermore, the reliability diagram for the track forecasting is close to the diagonal for the O-NFSVs for both resolutions. All these results indicate that the O-NFSVs are more appropriate than SKEB and SPPT in representing the model uncertainties of WRF at resolutions of 54 and 6 km and in improving the TC forecasting skill.

When a comparison is made between the coarse (i.e. EX-54 km) and finer (i.e. EX-6 km) resolutions, we find that finer resolutions augment the importance of using O-NFSVs in improving TC forecasting. For the intensity

forecasts, the finer resolutions provide a considerable decrease in the errors in the control forecast, which, combined with the O-NFSVs, further reduce these errors. However, the use of both SKEB and SPPT does not reduce them any further. For the track forecasts, SKEB and SPPT both decrease the errors in the control forecasts with finer resolutions, but the O-NFSVs tend to reduce them by much more.

In the convection-permitting model, the role of the O-NFSVs is further tested. They show similar performances as for the coarser resolution, except for a less good reliability of the TC intensity forecasting than the SPPT and SKEB schemes, as measured by the spread-skill relationship. The O-NFSVs scheme, however, shows superior performance than the SKEB and SPPT for ensemble forecasting of the TC track and intensity while comparing other skill scores.

Nevertheless, the reliability of the ensembles generated by the O-NFSVs needs to be improved, especially when the convection-permitting model is used. These structures only represent model uncertainty errors. In fact, TC forecasts are contaminated by both initial errors and model errors. A natural extension for the current

analysis will be to explore the approach proposed in Duan *et al.* (2022). They developed the C-NFSVs ensemble forecasting method that consists of optimally combining initial perturbations and model perturbations in a nonlinear setting. In particular, the approach allows us to take into account the interaction between the sources of initial and model errors. Duan *et al.* (2022) in particular show that the C-NFSVs yield an ensemble spread that is almost identical to the ensemble mean forecasting error and provide highly reliable ensembles in the context of an idealized model. It is therefore expected that when the O-NFSVs are properly combined with initial perturbations, new highly reliable ensembles will be available. This aspect will be explored in the future for TC forecasting.

Multiple perturbation schemes are often adopted to achieve higher forecasting skill (Jankov, 2017; Melhauser *et al.*, 2017; Xu *et al.*, 2022a, 2022b). For instance, Xu *et al.* (2022a, 2022b) superimposed the first NFSV on SPPT perturbations and further improved the probabilistic skill of convective-scale systems. In fact, when comparing the O-NFSVs with SKEB and SPPT, we found that with the change in model resolution, they differentially impact the TC forecasting skill. This interesting phenomenon suggests that further exploration of the usefulness of the hybrid scheme of O-NFSVs, SKEB and/or SPPT is worth pursuing in a future study.

In addition to SPPT and SKEB, a large body of research discusses the usefulness of the Stochastically Perturbed Parameterization scheme (i.e. SPP) to emulate model uncertainty (Ollinaho, 2017). The SPP focuses on perturbing one or a set of parameters with stochastic noise. The O-NFSVs, on the other hand, tend to describe the impact of tendency errors, which is similar to SKEB and SPPT. Therefore, O-NFSVs, as a new ensemble forecasting method, naturally compares with SKEB and SPPT.

Although we have shown the effectiveness of the O-NFSVs in improving TC forecasting skill, the current approach requires a massive amount of computer time. In this study, we use the SPG2 algorithm to compute the O-NFSVs, which calls the adjoint of the WRF model during the iteration process. This requires many computing resources. Furthermore, the NFSVs must be solved one by one. More precisely, 1–2 hr is necessary to compute one NFSV with 144 computing kernels in parallel. It is also noted that when calculating the O-NFSVs, perturbations to wind field or higher vertical resolution would cause the integration to break down. We believe that this problem is due to the limited computing resources rather than the sensitivity of O-NFSVs. We are currently working on the development of a new efficient algorithm, which is similar to the Lanczos algorithm (Simon, 1984), for calculating singular vectors (SVs: Mureau *et al.*, 1993; Buizza

and Palmer, 1995; Molteni *et al.*, 1996) or forcing singular vectors (FSVs: Barkmeijer *et al.*, 2003). We have used an idealized model to test it, and the preliminary results are encouraging. Its advantage is that it allows for computing all the NFSVs at once. In other words, we would be able to use this algorithm to generate all the above 26 O-NFSVs of WRF in 1–2 hours, even less time in the future. This new efficient algorithm for the O-NFSVs may help address the above difficulty and generalize the potential applications of the O-NFSVs, which will be reported in a follow-up article.

## AUTHOR CONTRIBUTIONS

**Yichi Zhang:** Conceptualization; data curation; formal analysis; investigation; software; validation; visualization; writing – original draft; writing – review and editing. **Wansuo Duan:** Conceptualization; formal analysis; funding acquisition; investigation; methodology; supervision; validation; writing – original draft; writing – review and editing. **Stéphane Vannitsem:** Formal analysis; methodology; writing – review and editing. **Han Zhang:** Conceptualization; data curation; software; writing – review and editing.

## ACKNOWLEDGEMENTS

The authors appreciate the editor and the two anonymous reviewers very much for their insightful comments and suggestions. This work was supported by the National Natural Science Foundation of China (Grant No. 41930971; 42288101).

## CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

## ORCID

Yichi Zhang  <https://orcid.org/0000-0002-3587-618X>

Wansuo Duan  <https://orcid.org/0000-0002-0122-2794>

Stéphane Vannitsem  <https://orcid.org/0000-0002-1734-1042>

## REFERENCES

- Barkmeijer, J., Iversen, T. and Palmer, T.N. (2003) Forcing singular vectors and other sensitive model structures. *Quarterly Journal of the Royal Meteorological Society*, 129, 2401–2423. <https://doi.org/10.1256/qj.02.126>.
- Berner, J., Fossell, K.R., Ha, S.-Y., Hacker, J.P. and Snyder, C. (2015) Increasing the skill of probabilistic forecasts: understanding performance improvements from model-error representations. *Monthly Weather Review*, 143, 1295–1320. <https://doi.org/10.1175/mwr-d-14-00091.1>.
- Berner, J., Ha, S.-Y., Hacker, J.P., Fournier, A. and Snyder, C. (2011) Model uncertainty in a mesoscale ensemble prediction system: stochastic versus multiphysics representations. *Monthly*

- Weather Review*, 139, 1972–1995. <https://doi.org/10.1175/2010mwr3595.1>.
- Berner, J., Shutts, G.J., Leutbecher, M. and Palmer, T.N. (2009) A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system. *Journal of the Atmospheric Sciences*, 66, 603–626. <https://doi.org/10.1175/2008jas2677.1>.
- Bhatia, K.T. and Nolan, D.S. (2015) Prediction of intensity model error (PRIME) for Atlantic Basin tropical cyclones. *Weather and Forecasting*, 30, 1845–1865. <https://doi.org/10.1175/WAF-D-15-0064.1>.
- Birgin, E.G., Martínez, J.M. and Raydan, M. (2000) Nonmonotone spectral projected gradient methods on convex sets. *SIAM Journal on Optimization*, 10, 1196–1211. <https://doi.org/10.1137/S1052623497330963>.
- Brier, G.W. (1950) Verification of Forecasts Expressed in Terms of Probability. *Monthly Weather Review*, 78, 1–3. [https://doi.org/10.1175/1520-0493\(1950\)078<0001:Vofeit>2.0.Co;2](https://doi.org/10.1175/1520-0493(1950)078<0001:Vofeit>2.0.Co;2).
- Buckingham, C., Marchok, T., Ginis, I., Rothstein, L. and Rowe, D. (2010) Short- and medium-range prediction of tropical and transitioning cyclone tracks within the NCEP global ensemble forecasting system. *Weather and Forecasting*, 25, 1736–1754. <https://doi.org/10.1175/2010waf222398.1>.
- Buizza, R. (2019) Introduction to the special issue on “25 years of ensemble forecasting”. *Quarterly Journal of the Royal Meteorological Society*, 145, 1–11. <https://doi.org/10.1002/qj.3370>.
- Buizza, R. and Palmer, T.N. (1995) The singular-vector structure of the atmospheric global circulation. *Journal of Atmospheric Sciences*, 52, 1434–1456. [https://doi.org/10.1175/1520-0469\(1995\)052<1434:Tsvsot>2.0.Co;2](https://doi.org/10.1175/1520-0469(1995)052<1434:Tsvsot>2.0.Co;2).
- Carr III, L.E. and Elsberry, R.L. (2000) Dynamical tropical cyclone track forecast errors. Part I: tropical region error sources. *Weather and Forecasting*, 15, 641–661. [https://doi.org/10.1175/1520-0434\(2000\)015<0641:Dtctfe>2.0.Co;2](https://doi.org/10.1175/1520-0434(2000)015<0641:Dtctfe>2.0.Co;2).
- Chan, J.C.L. (1985) Identification of the steering flow for tropical cyclone motion from objectively analyzed wind fields. *Monthly Weather Review*, 113, 106–116. [https://doi.org/10.1175/1520-0493\(1985\)113<0106:Iotsff>2.0.Co;2](https://doi.org/10.1175/1520-0493(1985)113<0106:Iotsff>2.0.Co;2).
- Chan, J.C.L. and Gray, W.M. (1982) Tropical cyclone movement and surrounding flow relationships. *Monthly Weather Review*, 110, 1354–1374. [https://doi.org/10.1175/1520-0493\(1982\)110<1354:Tcmasf>2.0.Co;2](https://doi.org/10.1175/1520-0493(1982)110<1354:Tcmasf>2.0.Co;2).
- Davis, C.A. (2018) Resolving tropical cyclone intensity in models. *Geophysical Research Letters*, 45, 2082–2087. <https://doi.org/10.1002/2017GL076966>.
- Demaeyer, J. and Vannitsem, S. (2018) Stochastic parameterization of subgrid-scale processes: a review of recent physically based approaches. In: Tsonis, A.A. (Ed.) *Advances in Nonlinear Geosciences*. Cham: Springer International Publishing, pp. 55–85.
- DeMaria, M., Sampson, C.R., Knaff, J.A. and Musgrave, K.D. (2014) Is tropical cyclone intensity guidance improving? *Bulletin of the American Meteorological Society*, 95, 387–398. <https://doi.org/10.1175/bams-d-12-00240.1>.
- Duan, W. and Huo, Z. (2016) An approach to generating mutually independent initial perturbations for ensemble forecasts: orthogonal conditional nonlinear optimal perturbations. *Journal of the Atmospheric Sciences*, 73, 997–1014. <https://doi.org/10.1175/jas-d-15-0138.1>.
- Duan, W., Ma, J. and Vannitsem, S. (2022) An ensemble forecasting method for dealing with the combined effects of the initial errors and model errors and a potential deep learning implementation. *Monthly Weather Review*, 150, 2959–2976. <https://doi.org/10.1175/mwr-d-22-0007.1>.
- Duan, W. and Zhou, F. (2013) Non-linear forcing singular vector of a two-dimensional quasi-geostrophic model. *Tellus A: Dynamic Meteorology and Oceanography*, 65, 18452. <https://doi.org/10.3402/tellusa.v65i0.18452>.
- Duda, J.D., Wang, X., Kong, F., Xue, M. and Berner, J. (2016) Impact of a stochastic kinetic energy backscatter scheme on warm season convection-allowing ensemble forecasts. *Monthly Weather Review*, 144, 1887–1908. <https://doi.org/10.1175/mwr-d-15-0092.1>.
- Dudhia, J. (1989) Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. *Journal of Atmospheric Sciences*, 46, 3077–3107. [https://doi.org/10.1175/1520-0469\(1989\)046<3077:Nsocod>2.0.Co;2](https://doi.org/10.1175/1520-0469(1989)046<3077:Nsocod>2.0.Co;2).
- Eckel, F.A. and Mass, C.F. (2005) Aspects of effective mesoscale, short-range ensemble forecasting. *Weather and Forecasting*, 20, 328–350. <https://doi.org/10.1175/waf843.1>.
- Emanuel, K. and Zhang, F. (2016) On the predictability and error sources of tropical cyclone intensity forecasts. *Journal of the Atmospheric Sciences*, 73, 3739–3747. <https://doi.org/10.1175/jas-d-16-0100.1>.
- Ha, S., Berner, J. and Snyder, C. (2015) A comparison of model error representations in mesoscale ensemble data assimilation. *Monthly Weather Review*, 143, 3893–3911. <https://doi.org/10.1175/mwr-d-14-00395.1>.
- Hamill, T.M., Whitaker, J.S., Kleist, D.T., Fiorino, M. and Benjamin, S.G. (2011) Predictions of 2010’s tropical cyclones using the GFS and ensemble-based data assimilation methods. *Monthly Weather Review*, 139, 3243–3247. <https://doi.org/10.1175/mwr-d-11-00079.1>.
- Hong, S.-Y., Noh, Y. and Dudhia, J. (2006) A new vertical diffusion package with an explicit treatment of entrainment processes. *Monthly Weather Review*, 134, 2318–2341. <https://doi.org/10.1175/mwr3199.1>.
- Jankov, I. and Coauthors. (2017) A performance comparison between multiphysics and stochastic approaches within a North American RAP ensemble. *Monthly Weather Review*, 145, 1161–1179. <https://doi.org/10.1175/mwr-d-16-0160.1>.
- Judt, F., Chen, S.S. and Berner, J. (2016) Predictability of tropical cyclone intensity: scale-dependent forecast error growth in high-resolution stochastic kinetic-energy backscatter ensembles. *Quarterly Journal of the Royal Meteorological Society*, 142, 43–57. <https://doi.org/10.1002/qj.2626>.
- Kain, J.S. (2004) The Kain–Fritsch convective parameterization: an update. *Journal of Applied Meteorology*, 43, 170–181. [https://doi.org/10.1175/1520-0450\(2004\)043<0170:Tkcpcou>2.0.Co;2](https://doi.org/10.1175/1520-0450(2004)043<0170:Tkcpcou>2.0.Co;2).
- Kalnay, E. (2003) *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge: Cambridge University Press, p. 341.
- Krishnamurti, T., Pattnaik, S., Stefanova, L., Tallapragada, V., Mackey, B., O’Shay, A. and Pasch, R. (2005) The hurricane intensity issue. *Monthly Weather Review*, 133, 1886–1912. <https://doi.org/10.1175/MWR2954.1>.
- Lang, S.T.K., Leutbecher, M. and Jones, S.C. (2012) Impact of perturbation methods in the ECMWF ensemble prediction system on tropical cyclone forecasts. *Quarterly Journal of the Royal Meteorological Society*, 138, 2030–2046. <https://doi.org/10.1002/qj.1942>.



- Lang, S.T.K., Lock, S.-J., Leutbecher, M., Bechtold, P. and Forbes, R.M. (2021) Revision of the stochastically perturbed parametrisations model uncertainty scheme in the integrated forecasting system. *Quarterly Journal of the Royal Meteorological Society*, 147, 1364–1381. <https://doi.org/10.1002/qj.3978>.
- Leutbecher, M. and Coauthors. (2017) Stochastic representations of model uncertainties at ECMWF: state of the art and future vision. *Quarterly Journal of the Royal Meteorological Society*, 143, 2315–2339. <https://doi.org/10.1002/qj.3094>.
- Lin, Y.-L., Farley, R.D. and Orville, H.D. (1983) Bulk parameterization of the snow field in a cloud model. *Journal of Applied Meteorology and Climatology*, 22, 1065–1092. [https://doi.org/10.1175/1520-0450\(1983\)022<1065:Bpotsf>2.0.Co;2](https://doi.org/10.1175/1520-0450(1983)022<1065:Bpotsf>2.0.Co;2).
- Lorenz, E.N. (1969) The predictability of a flow which possesses many scales of motion. *Tellus*, 21, 289–307. <https://doi.org/10.1111/j.2153-3490.1969.tb00444.x>.
- Lorenz, E.N. (1982) Atmospheric predictability experiments with a large numerical model. *Tellus*, 34, 505–513. <https://doi.org/10.3402/tellusa.v34i6.10836>.
- Magnusson, L. and Coauthors. (2019) ECMWF activities for improved hurricane forecasts. *Bulletin of the American Meteorological Society*, 100, 445–458. <https://doi.org/10.1175/bams-d-18-0044.1>.
- Melhauser, C., Zhang, F., Weng, Y., Jin, Y., Jin, H. and Zhao, Q. (2017) A multiple-model convection-permitting ensemble examination of the probabilistic prediction of tropical cyclones: hurricanes Sandy (2012) and Edouard (2014). *Weather and Forecasting*, 32, 665–688. <https://doi.org/10.1175/waf-d-16-0082.1>.
- Miglietta, M.M., Mastrangelo, D. and Conte, D. (2015) Influence of physics parameterization schemes on the simulation of a tropical-like cyclone in the Mediterranean Sea. *Atmospheric Research*, 153, 360–375. <https://doi.org/10.1016/j.atmosres.2014.09.008>.
- Mlawer, E.J., Taubman, S.J., Brown, P.D., Iacono, M.J. and Clough, S.A. (1997) Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research: Atmospheres*, 102, 16663–16682. <https://doi.org/10.1029/97JD00237>.
- Molteni, F., Buizza, R., Palmer, T.N. and Petroliagis, T. (1996) The ECMWF ensemble prediction system: methodology and validation. *Quarterly Journal of the Royal Meteorological Society*, 122, 73–119. <https://doi.org/10.1002/qj.49712252905>.
- Munsell, E.B. and Zhang, F. (2014) Prediction and uncertainty of hurricane Sandy (2012) explored through a real-time cloud-permitting ensemble analysis and forecast system assimilating airborne Doppler radar observations. *Journal of Advances in Modeling Earth Systems*, 6, 38–58. <https://doi.org/10.1002/2013MS000297>.
- Mureau, R., Molteni, F. and Palmer, T.N. (1993) Ensemble prediction using dynamically conditioned perturbations. *Quarterly Journal of the Royal Meteorological Society*, 119, 299–323. <https://doi.org/10.1002/qj.49711951005>.
- Nicolis, C. (2004) Dynamics of model error: the role of unresolved scales revisited. *Journal of the Atmospheric Sciences*, 61, 1740–1753. [https://doi.org/10.1175/1520-0469\(2004\)061<1740:DOMETR>2.0.CO;2](https://doi.org/10.1175/1520-0469(2004)061<1740:DOMETR>2.0.CO;2).
- Nicolis, C., Perdigao, R. and Vannitsem, S. (2009) Dynamics of prediction errors under the combined effect of initial condition and model errors. *Journal of the Atmospheric Sciences*, 66, 766–778. <https://doi.org/10.1175/2008JAS2781.1>.
- Novak, D.R., Bright, D.R. and Brennan, M.J. (2008) Operational forecaster uncertainty needs and future roles. *Weather and Forecasting*, 23, 1069–1084. <https://doi.org/10.1175/2008waf2222142.1>.
- Ollinaho, P. and Coauthors. (2017) Towards process-level representation of model uncertainties: stochastically perturbed parametrizations in the ECMWF ensemble. *Quarterly Journal of the Royal Meteorological Society*, 143, 408–422. <https://doi.org/10.1002/qj.2931>.
- Ono, K., Kunii, M. and Honda, Y. (2021) The regional model-based mesoscale ensemble prediction system, MEPS, at the Japan meteorological agency. *Quarterly Journal of the Royal Meteorological Society*, 147, 465–484. <https://doi.org/10.1002/qj.3928>.
- Palmer, T.N., Buizza, R., Doblas-Reyes, F., Jung, T., Leutbecher, M., Shutts, G., Steinheimer, M. and Weisheimer, A. (2009) Stochastic parametrization and model uncertainty. *ECMWF Tech. Memo.*, 598, 44 <http://www.ecmwf.int/sites/default/files/elibrary/2009/11577-stochastic-parametrization-and-model-uncertainty.pdf>.
- Puri, K., Barkmeijer, J. and Palmer, T.N. (2001) Ensemble prediction of tropical cyclones using targeted diabatic singular vectors. *Quarterly Journal of the Royal Meteorological Society*, 127, 709–731. <https://doi.org/10.1002/qj.49712757222>.
- Qin, X., Duan, W. and Xu, H. (2020) Sensitivity to tendency perturbations of tropical cyclone short-range intensity forecasts generated by WRF. *Advances in Atmospheric Sciences*, 37, 291–306. <https://doi.org/10.1007/s00376-019-9187-6>.
- Reynolds, C.A., McLay, J.G., Goerss, J.S., Serra, E.A., Hodyss, D. and Sampson, C.R. (2011) Impact of resolution and design on the U.S. Navy global ensemble performance in the tropics. *Monthly Weather Review*, 139, 2145–2155. <https://doi.org/10.1175/2011mwr3546.1>.
- Rogers, R. and Coauthors. (2006) The intensity forecasting experiment: a NOAA multiyear field program for improving tropical cyclone intensity forecasts. *Bulletin of the American Meteorological Society*, 87, 1523–1538. <https://doi.org/10.1175/bams-87-11-1523>.
- Romine, G.S., Schwartz, C., Berner, J., Fossell, K., Snyder, C., Anderson, J. and Weisman, M. (2014) Representing forecast error in a convection-permitting ensemble system. *Monthly Weather Review*, 142, 4519–4541. <https://doi.org/10.1175/MWR-D-14-00100.1>.
- Shutts, G. (2005) A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quarterly Journal of the Royal Meteorological Society*, 131, 3079–3102. <https://doi.org/10.1256/qj.04.106>.
- Simon, H.D. (1984) The Lanczos algorithm with partial reorthogonalization. *Mathematics of Computation-Math. Comput.*, 42, 115. <https://doi.org/10.2307/2007563>.
- Torn, R.D. (2016) Evaluation of atmosphere and ocean initial condition uncertainty and stochastic exchange coefficients on ensemble tropical cyclone intensity forecasts. *Monthly Weather Review*, 144, 3487–3506. <https://doi.org/10.1175/mwr-d-16-0108.1>.
- Torn, R.D., Elless, T.J., Papin, P.P. and Davis, C.A. (2018) Tropical cyclone track sensitivity in deformation steering flow. *Monthly Weather Review*, 146, 3183–3201. <https://doi.org/10.1175/mwr-d-18-0153.1>.
- Toth, Z. and Kalnay, E. (1993) Ensemble forecasting at NMC: the generation of perturbations. *Bulletin of the American Meteorological*

- Society, 74, 2317–2330. [https://doi.org/10.1175/1520-0477\(1993\)074<2317:Efantg>2.0.Co;2](https://doi.org/10.1175/1520-0477(1993)074<2317:Efantg>2.0.Co;2).
- Toth, Z. and Kalnay, E. (1997) Ensemble forecasting at NCEP and the breeding method. *Monthly Weather Review*, 125, 3297–3319. [https://doi.org/10.1175/1520-0493\(1997\)125<3297:Efanat>2.0.Co;2](https://doi.org/10.1175/1520-0493(1997)125<3297:Efanat>2.0.Co;2).
- Toth, Z. and Vannitsem, S. (2002) *Model errors and ensemble forecasting*. Operational Systems Workshop. Reading: ECMWF, pp. 146–154.
- Vannitsem, S. and Toth, Z. (2002) Short-term dynamics of model errors. *Journal of the Atmospheric Sciences*, 59, 2594–2604. [https://doi.org/10.1175/1520-0469\(2002\)059<2594:Stdome>2.0.Co;2](https://doi.org/10.1175/1520-0469(2002)059<2594:Stdome>2.0.Co;2).
- Weber, H.C. (2003) Hurricane track prediction using a statistical ensemble of numerical models'. *Monthly Weather Review*, 131, 749–770. [https://doi.org/10.1175/1520-0493\(2003\)131<0749:Htpuas>2.0.Co;2](https://doi.org/10.1175/1520-0493(2003)131<0749:Htpuas>2.0.Co;2).
- Wu, C.C., Huang, T.-S. and Chou, K.-H. (2004) Potential vorticity diagnosis of the key factors affection the motion of Typhoon Sinlaku (2002). *Monthly Weather Review*, 132, 2084–2093 [https://doi.org/10.1175/1520-0493\(2004\)132<2084:Pvdotk>2.0.Co;2](https://doi.org/10.1175/1520-0493(2004)132<2084:Pvdotk>2.0.Co;2).
- Xiao, Q. and Coauthors. (2008) Application of an adiabatic WRF adjoint to the investigation of the May 2004 McMurdo, Antarctica, severe wind event. *Monthly Weather Review*, 136, 3696–3713. <https://doi.org/10.1175/2008mwr2235.1>.
- Xu, Z., Chen, J., Mu, M., Dai, G. and Ma, Y. (2022a) A nonlinear representation of model uncertainty in a convective-scale ensemble prediction system. *Advances in Atmospheric Sciences*, 39, 1432–1450. <https://doi.org/10.1007/s00376-022-1341-x>.
- Xu, Z., Chen, J., Mu, M., Tao, L., Dai, G., Wang, J. and Ma, Y. (2022b) A stochastic and non-linear representation of model uncertainty in a convective-scale ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, 148, 2507–2531. <https://doi.org/10.1002/qj.4322>.
- Yamaguchi, M., Nakazawa, T. and Aonashi, K. (2012) Tropical cyclone track forecasts using JMA model with ECMWF and JMA initial conditions. *Geophysical Research Letters*, 39(9), L09801. <https://doi.org/10.1029/2012GL051473>.
- Yamaguchi, M., Sakai, R., Kyoda, M., Komori, T. and Kadowaki, T. (2009) Typhoon ensemble prediction system developed at the Japan Meteorological Agency. *Monthly Weather Review*, 137, 2592–2604. <https://doi.org/10.1175/2009mwr2697.1>.
- Yao, J., Duan, W. and Qin, X. (2021) Which features of the SST forcing error most likely disturb the simulated intensity of tropical cyclones? *Advances in Atmospheric Sciences*, 38, 581–602. <https://doi.org/10.1007/s00376-020-0073-z>.
- Zhang, X., Huang, X.-Y. and Pan, N. (2013) Development of the upgraded tangent linear and adjoint of the Weather Research and Forecasting (WRF) model. *Journal of Atmospheric and Oceanic Technology*, 30, 1180–1188. <https://doi.org/10.1175/jtech-d-12-00213.1>.
- Zhang, Z. and Krishnamurti, T.N. (1997) Ensemble forecasting of hurricane tracks. *Bulletin of the American Meteorological Society*, 78, 2785–2796. [https://doi.org/10.1175/1520-0477\(1997\)078<2785:Efoht>2.0.Co;2](https://doi.org/10.1175/1520-0477(1997)078<2785:Efoht>2.0.Co;2).
- Zhang, Z., Tallapragada, V., Kieu, C., Trahan, S. and Wang, W. (2014) HWRF based ensemble prediction system using perturbations

from GEFS and stochastic convective trigger function. *Tropical Cyclone Research and Review*, 3, 145–161. <https://doi.org/10.6057/2014TCRR03.02>.

Zhou, X. and Coauthors. (2022) The development of the NCEP global ensemble forecast system version 12. *Weather and Forecasting*, 37, 1069–1084. <https://doi.org/10.1175/waf-d-21-0112.1>.

**How to cite this article:** Zhang, Y., Duan, W., Vannitsem, S. & Zhang, H. (2023) A new approach to represent model uncertainty in the forecasting of tropical cyclones: The orthogonal nonlinear forcing singular vectors. *Quarterly Journal of the Royal Meteorological Society*, 149(755), 2206–2232. Available from: <https://doi.org/10.1002/qj.4502>

## APPENDIX A. STATISTICAL TOOLS TO EVALUATE THE FORECAST SKILL

### A.1 Forecast error

The forecast error of the TC track is determined by the great-circle distance between the forecast TC centre  $F = (a_f, y_f)$  and observed TC centre  $A = (a_o, y_o)$ , where  $a_f$  and  $a_o$  are the longitudes and  $y_f$  and  $y_o$  are the latitudes. Then, the forecast errors of the TC track can be written as

$$E_{\text{track}} = |F - A| = 111.11 \cdot \cos^{-1} \times [\sin y_o \sin y_f + \cos y_o \cos y_f \cos(a_o - a_f)]. \quad (\text{A1})$$

The forecast error of TC intensity (i.e.  $P_{\min}$  and  $V_{\max}$ ) is determined by the difference between the forecast TC intensity (i.e.  $F$ ) and observed TC intensity (i.e.  $A$ ), which can be expressed as:

$$E_{\text{intensity}} = |F - A|. \quad (\text{A2})$$

### A.2 Ensemble spread

The ensemble spread reflects the uncertainty of the forecast. The ensemble spread is described as follows:

$$s = \sqrt{\frac{1}{N-1} \sum_{i=1}^N |F_i - \bar{F}|^2}, \quad (\text{A3})$$

where  $\bar{F} = \frac{1}{N} \sum_{i=1}^N F_i$  is the ensemble mean and  $N$  is the number of members.

For a reliable ensemble forecast system, the ratio of the error of the ensemble mean forecast and the spread of ensemble members should be close to 1.

### A.3 Brier score

The Brier score (BS: Brier, 1950) is the mean square error of the probability forecasts and can be described as:

$$BS = \frac{1}{N} \sum_{j=1}^N (p_j - o_j)^2, \quad (\text{A4})$$

where  $N$  is the number of realizations of the prediction process,  $p_j$  is the probability of a dichotomous event in the  $j$ th prediction process and  $o_j$  is the probability of an observation in the  $j$ th prediction process. When the BS value is closer to 0, the probability forecast skill is higher.

In this study, the BS value is used to evaluate the probability forecast skills of TC track intensity. For the TC track, the BS values of the track strike probability are calculated. For intensity, the BS values of five TC intensity categories of ensemble members are calculated according to the following classification.

Tropical Depression (TD)  $22 \text{ kt} \leq \text{maximum wind} < 34 \text{ kt}$ .

Tropical Storm (TS)  $34 \text{ kt} \leq \text{maximum wind} < 64 \text{ kt}$ .

Typhoon (TY)  $64 \text{ kt} \leq \text{maximum wind} < 80 \text{ kt}$ .

Severe typhoon (STY)  $80 \text{ kt} \leq \text{maximum wind} < 100 \text{ kt}$ .

Super-typhoon (Super)  $\text{maximum wind} \geq 100 \text{ kt}$ .

When an ensemble member for each TC is generated, the relevant TC intensity may fall in different categories at different times during the forecast period. With these intensities and the corresponding observed intensities, one can calculate the BS scores for different intensity categories in the forecast period.

### A.4 Relative operating characteristic curve

The relative operating characteristic curves (ROCs) assess the ability of the forecast to discriminate between event and non-event. By considering whether an event occurs at

TABLE A1 Two-by-two contingency table of a binary event.

| Forecast | Observation |         | Total   |
|----------|-------------|---------|---------|
|          | Yes         | No      |         |
| Yes      | $a$         | $b$     | $a + b$ |
| No       | $c$         | $d$     | $c + d$ |
| Total    | $a + c$     | $b + d$ |         |

every grid and checking forecasts against observations, a two-category contingency table is constructed in Table A1, where  $a$  and  $b$  represent the number of hits and false alarms, respectively, and  $c$  and  $d$  represent the number of misses and correct rejections, respectively.

Then, the hit rate ( $H$ ) and the false alarm rate ( $F$ ) can be calculated as follows:

$$H = a/(a + c)$$

$$F = b/(a + b). \quad (\text{A5})$$

The ROC curve can be obtained by the pairs  $H$  and  $F$ . Perfect skill produces a curve from bottom left to top left to top right, and no skill is indicated by the diagonal line from (0,0) to (1,1).

### A.5 Reliability diagram

The reliability diagram (RD) is used to assess the degree of matching between the forecast probability and observation frequency. The forecast probability can be set to multiple intervals between 0 and 1. Here, we calculate the observation frequency of samples with the interval [0.0–0.1, 0.2–0.3, ..., 0.9–1] for the forecast probability. If the forecast probability is consistent with the frequency of the event, the reliability line would be distributed from (0.0) to (1.1), which indicates that the probabilistic forecasts are credible.