An Ensemble Forecasting Method for Dealing with the Combined Effects of the Initial and Model Errors and a Potential Deep Learning Implementation

WANSUO DUAN,[a,b] JUNJIE MA,[a,b] AND STÉPHANE VANNITSEM[c]

[a] *LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China*
[b] *University of Chinese Academy of Sciences, Beijing, China*
[c] *Royal Meteorological Institute of Belgium, Brussels, Belgium*

ABSTRACT: In this paper, a new nonlinear forcing singular vector (NFSV) approach is proposed to provide mutually independent optimally combined modes of initial perturbations and model perturbations (C-NFSVs) in ensemble forecasts. The C-NFSVs are a group of optimally growing structures that take into account the impact of the interaction between the initial errors and the model errors effectively, generalizing the original NFSV for simulations of the impact of the model errors. The C-NFSVs method is tested in the context of the Lorenz-96 model to demonstrate its potential to improve ensemble forecast skills. This method is compared with the orthogonal conditional nonlinear optimal perturbations (O-CNOPs) method for estimating only the initial uncertainties and the orthogonal NFSVs (O-NFSVs) for estimating only the model uncertainties. The results demonstrate that when both the initial perturbations and model perturbations are introduced in the forecasting system, the C-NFSVs are much more capable of achieving higher ensemble forecasting skills. The use of a deep learning approach as a remedy for the expensive computational costs of the C-NFSVs is evaluated. The results show that learning the impact of the C-NFSVs on the ensemble provides a useful and efficient alternative for the operational implementation of C-NFSVs in forecasting suites dealing with the combined effects of the initial errors and the model errors.

SIGNIFICANCE STATEMENT: A new ensemble forecasting method for dealing with combined effects of initial errors and model errors, i.e., the C-NFSVs, is proposed, which is an extension of the NFSV approach for simulating the model error effects in ensemble forecasts. The C-NFSVs provide mutually independent optimally combined modes of initial perturbations and model perturbations. This new method is tested for generating ensemble forecasts in the context of the Lorenz-96 model, and there are indications that the optimally growing structures may provide reliable ensemble forecasts. Furthermore, it is found that a hybrid dynamical–deep learning approach could be a potential avenue for real-time ensemble forecasting systems when perturbations combine the impact of the initial and the model errors.

KEYWORDS: Uncertainty; Model errors

---

1. Introduction

Due to the chaotic nature of the atmosphere, the initial condition and numerical modeling uncertainties rapidly amplify (Nicolis et al. 2009). Ensemble forecasting is an important way to estimate the prediction uncertainty and also to provide probabilistic information on the occurrence of certain events (e.g., Buizza 2019). Traditional ensemble forecasting methods mainly address the impact of uncertainties on the initial conditions. The earliest method, given this context, is the Monte Carlo forecasting method (MCF), which imposes random initial perturbations on the initial analysis field and produces a set of ensemble members to estimate the probability distribution of an event (Epstein 1969; Leith 1974). The MCF was an important step forward in moving the ensemble forecasting approach from basic research to operational practice. Subsequently, some studies suggested that growing-type initial perturbations can offer a good description of the initial analysis errors (Mureau et al. 1993; Toth and Kalnay 1993, 1997), which overcome the problem of the initial perturbation

decrease experienced by the MCF. Toth and Kalnay (1993) developed, in particular, the breeding method to find growing-type initial perturbations, the bred vectors (BVs), and applied it in the ensemble forecasting system at the National Centers for Environmental Prediction (NCEP) in 1992. Similarly, nonlinear local Lyapunov vectors (Feng et al. 2014; Feng et al. 2016; Feng et al. 2018) and backward Lyapunov vectors (e.g., Vannitsem and Duan 2020; Demaeyer et al. 2022) were successfully applied in the development of ensemble forecasts. The European Centre for Medium-Range Weather Forecasts (ECMWF) proposed an alternative method based on singular vectors (SVs; Mureau et al. 1993; Buizza and Palmer 1995; Molteni et al. 1996) and produced ensemble forecasts with great success. The SVs capture the optimal unstable growth property of the initial analysis errors in the linearized regime. However, they cannot cope with the impact of nonlinear physical processes on the amplification of the initial perturbations (Anderson 1997; Hamill et al. 2000).

Considering the limitations of the linear theory of SVs, Mu et al. (2003) proposed the conditional nonlinear optimal perturbation (CNOP), which is an extension of the leading SV in the nonlinear regime. CNOP fully considers the influence of

nonlinear physical processes and represents the optimally growing initial perturbation in the nonlinear regime. Mu and Jiang (2008) replaced the leading SV with the CNOP to produce the initial perturbations of the ensemble forecasts, which demonstrated higher forecast skills than the SVs (see also Huo and Duan 2019; Zhou et al. 2021). To take into account multiple nonlinear processes in the development of the initial perturbations for ensemble forecasts, Duan and Huo (2016) formulated the orthogonal CNOPs (O-CNOPs) method to produce mutually independent nonlinear optimal initial perturbations for ensemble forecasting. The O-CNOPs have been shown to display a higher ensemble forecast skill than the SVs and BVs and a more reasonable ensemble spread for estimating the uncertainty in a hierarchy of models (Duan and Huo 2016; Huo et al. 2019; Wang and Duan 2019; Wang 2021).

The ensemble forecasting methods mentioned above focus on considering the initial uncertainties, which are valid under a perfect model assumption. However, there is no perfect model in reality. Several studies have indicated that considering model errors is important for improving the skill of numerical weather forecasting and climate prediction (Buizza et al. 1999; Palmer 2000; Orrell et al. 2001; Orrell 2005; Palmer et al. 2009; Duan et al. 2013; Vannitsem 2014). In view of this, new ensemble forecasting methods have been designed to address forecast uncertainties caused by model errors. For example, the ECMWF proposed a stochastically perturbed parameterization tendency scheme (SPPT; Buizza et al. 1999) and stochastic kinetic energy backscatter scheme (SKEB; Shutts 2005), leading to important improvements of the ensemble forecast skill (Berner et al. 2009; Palmer et al. 2009; Du et al. 2018; see also the special issue, Buizza 2019). Hou et al. (2006) developed a stochastic total tendency perturbation scheme (STTP) to emulate model uncertainties in the NCEP global ensemble forecasting system in February 2010 (also see Hou et al. 2008; Hou et al. 2010). In the STTP schemes, the stochastic forcing term, similar to the random initial perturbations in MCF, did not fully capture the rapid unstable growth behavior of model errors. As argued above, the ensemble forecasting system requires growing-type perturbations, which can make the ensemble members deviate from the control forecast and reliably encompass the true value. Therefore, in an ensemble forecasting system incorporating the impact of model errors, the key question is how to generate realistic rapidly growing model perturbations.

To obtain rapidly growing model perturbations, Barkmeijer et al. (2003) proposed using a forcing singular vector (FSV) closely related to the SVs, which represents a rapidly growing constant tendency perturbation in a linear framework. This constant tendency perturbation describes the combined effects of the model systematic errors and parts of state-dependent model errors that are not explicitly described in the model equations (Feng and Duan 2013). Due to the limitation of the linear approximation of FSV, Duan and Zhou (2013) proposed approaching this problem using the nonlinear forcing singular vector (NFSV). The NFSV is a tendency perturbation that makes the forecast deviate from the reference state more significantly; and if it is used in an ensemble forecasting framework, it could better encompass the truth and could provide

more reliable ensembles. However, an ensemble forecasting system requires a set of perturbations. This can be achieved by formulating a new approach based on a set of orthogonal NFSVs, following the idea of O-CNOPs developed in Duan and Huo (2016). These vectors will be referred to as the O-NFSVs in the following sections, see section 2. The O-NFSVs provide mutually independent model tendency perturbations that enable the descriptions of the forecast uncertainties caused by the model errors. However, in realistic forecasting systems, the effect of both the initial errors and the model errors, especially the effect of their interaction, are inevitable (Nicolis et al. 2009). The key question to address in a fully integrated forecasting system is how to combine the initial errors and the model errors correctly to obtain a reliable ensemble. Although ensemble forecasting systems exist that consider both the initial errors and the model error effects (e.g., Buizza et al. 1999; Hou et al. 2010), they are built by superimposing the independent initial perturbations (such as SVs, BVs, or others) and the model tendency perturbations (e.g., SPPT or STTP). To date, no attention has been given to the dynamically coordinated growth of the initial and the model perturbations, which may limit the skill of the ensemble forecasts. To overcome this limitation, we further extend the O-NFSVs by developing C-NFSVs that combine the impacts of the initial errors and the model errors and formulate a novel ensemble forecasting approach, which is tested in a simple setting based on the Lorenz-96 model (Lorenz 1996).

The rest of this paper is organized as follows. In section 2, the C-NFSVs are introduced, together with the particular cases referred to as the O-NFSVs and O-CNOPs. In section 3, the Lorenz-96 model adopted in this study is described. In section 4, the experimental design is detailed, and in section 5, the results of the ensemble forecasting experiments are presented. Then, the possible usefulness of deep learning in emulating this type of ensemble forecast is discussed in section 6. Finally, a summary and discussion are provided in section 7.

2. The C-NFSVs method and its two particular cases of O-NFSVs and O-CNOPs

Let us start with a partial differential equation of the following form:

$$\begin{cases} \dfrac{\partial \mathbf{U}}{\partial t} = F[\mathbf{U}(\mathbf{x}, t)], \\ \mathbf{U}|_{t=0} = \mathbf{U}_0, \end{cases} \quad \text{in } \Omega \times [0, T], \qquad (2.1)$$

where $\mathbf{U}_0$ and $\mathbf{U}(\mathbf{x}, t)$ represent the initial field and its evolution in time, respectively, and F is a nonlinear differential operator. Assuming that the dynamical system [Eq. (2.1)] and its initial field are exactly known, then the state $\mathbf{U}(\mathbf{x}, t)$ at a future time T can be given by

$$\mathbf{U}(\mathbf{x}, T) = M_T(\mathbf{U}_0), \qquad (2.2)$$

where M_T is the nonlinear propagator of Eq. (2.1).

Realistic forecasts are generally contaminated by both the initial errors and the model errors. If one uses initial perturbations

$\mathbf{u}_0$ to represent the initial errors and tendency perturbations $\mathbf{f}(\mathbf{x}, t)$ to describe the combined effect of different kinds of model errors (see Barkmeijer et al. 2003; Duan and Zhou 2013; Duan and Zhao 2014; Tao and Duan 2019; Tao et al. 2020), the equations of the forecast model can be written as

$$\begin{cases} \dfrac{\partial(\mathbf{U} + \mathbf{u})}{\partial t} = F(\mathbf{U} + \mathbf{u}) + \mathbf{f}(\mathbf{x}, t), \\ \mathbf{U} + \mathbf{u}|_{t=0} = \mathbf{U}_0 + \mathbf{u}_0, \end{cases} \quad \text{in } \Omega \times [0, T], \quad (2.3)$$

where Ω represents a domain in R^n. If one uses $M_T(\mathbf{f})(\cdot)$ to denote the propagator of Eq. (2.3) from the initial time $t = 0$ to the prediction time $t = T$, the corresponding prediction error (denoted by $\mathbf{u}_T$) can be expressed as

$$\mathbf{u}_T = M_T(\mathbf{f})(\mathbf{U}_0 + \mathbf{u}_0) - M_T(\mathbf{U}_0), \quad (2.4)$$

where $M_T(\mathbf{U}_0)$ is the reference state (to be predicted), not contaminated by any errors.

Based on Eq. (2.4), Mu et al. (2003) proposed the CNOP by assuming a perfect model (i.e., $\mathbf{f} = 0$), which represents the initial error that causes the largest prediction error at the prediction time. On the other hand, Duan and Zhou (2013) developed the NFSV approach under a perfect initial condition assumption (i.e., $\mathbf{u}_0 = \mathbf{0}$), which describes the model tendency error that caused the largest prediction error at the prediction time. Based on these two approaches and considering time-constant model errors (i.e., tendency perturbation $\mathbf{f}$ is assumed to be constant), we can express the maximization problem leading to the C-NFSVs as follows:

$$J(\mathbf{f}_j^*) = \max_{\mathbf{f}_j \in \Omega_j} \|M_T(\mathbf{f}_j)(\mathbf{U}_0 + r\mathbf{f}_j) - M_T(\mathbf{U}_0)\|_b, \quad (2.5)$$

where

$$\Omega_j = \begin{cases} \mathbf{f}_1 \in R^n, & \|\mathbf{f}_1\|_a \leq \sigma_{\mathbf{f}}, \\ \{\mathbf{f}_j \in R^n \mid \|\mathbf{f}_j\|_a \leq \sigma_{\mathbf{f}}, \ \mathbf{f}_j \perp \Omega_k, \ k = 1, \ldots, j-1\}, & j > 1, \end{cases} \quad (2.6)$$

and $\|r\mathbf{f}_j\|_a \leq \sigma_{\mathbf{I}}$; the symbol $\{\cdot\}$ refers to an ensemble of vectors, and $\perp$ indicates the orthogonality; $\|\cdot\|_a$ and $\|\cdot\|_b$ are the norms that are used to measure the amplitudes of the initial perturbations $r\mathbf{f}_j$ and tendency perturbations $\mathbf{f}_j$ and the departure from the reference state at time T, respectively; $\sigma_{\mathbf{f}}$ and $\sigma_{\mathbf{I}}$ are positive constant numbers that constrain the amplitudes of the tendency perturbations and the initial perturbations, respectively. Note that $r\mathbf{f}_j$ is the initial perturbation, where $r = \sigma_{\mathbf{I}}/\sigma_{\mathbf{f}}$ is the magnitude. The combined modes $(r\mathbf{f}_j^*, \mathbf{f}_j^*)$ define the C-NFSVs.

Here, we assume that the initial errors and the model errors are closely related and amplify along the same direction in phase space. This simplification makes life easier but is also justified by the dynamics of the model errors that are also amplified by the generic chaotic mechanism at the origin of the amplification of the initial errors and which after some time behave as an initial error perturbation (Nicolis 2003; Vannitsem 2006; Nicolis et al. 2009).

The C-NFSVs consist of mutually orthogonal initial perturbations and tendency perturbations, which coherently maximize the cost function at time T in Ω_j. In other words, the C-NFSVs coherently induce perturbation evolution [i.e., $\mathbf{u}_T$ in Eq. (2.4)], which are the largest. Moreover, their corresponding objective function values rank as $J(\mathbf{f}_1^*) > J(\mathbf{f}_2^*) >, \cdots, > J(\mathbf{f}_n^*)$.

O-NFSVs: If the initial fields could be exactly known [i.e., $\mathbf{u}_0 = 0$ in Eq. (2.4) or $r = 0$ in Eq. (2.5)], the prediction errors would only be caused by model errors $\mathbf{f}$. Then, Eq. (2.5) becomes

$$J(\mathbf{f}_j^*) = \max_{\mathbf{f}_j \in \Omega_j} \|M_T(\mathbf{f}_j)(\mathbf{U}_0) - M_T(\mathbf{U}_0)\|_b, \quad (2.7)$$

which, combined with the constraint condition in Eq. (2.6), provides a group of mutually orthogonal model tendency perturbations. These tendency perturbations correspond to the orthogonal NFSVs (O-NFSVs), which generalize the NFSV idea defined in Duan and Zhou (2013) [i.e., the tendency perturbation with $j = 1$ in Eq. (2.7)] to mutually orthogonal subspaces of the model tendencies.

O-CNOPs: If the model is considered perfect (i.e., $\mathbf{f} = 0$) and only initial errors $\mathbf{u}_0$ are considered, Eq. (2.5) can be rewritten as

$$J(\mathbf{u}_{0_j}^*) = \max_{\mathbf{u}_{0j} \in \Omega_j} \|M_T(\mathbf{U}_0 + \mathbf{u}_{0j}) - M_T(\mathbf{U}_0)\|_b. \quad (2.8)$$

Equation (2.8), together with $\|\mathbf{u}_{0j}\|_a \leq \sigma_{\mathbf{I}}$, yields the O-CNOPs defined by Duan and Huo (2016). For $j = 1$, the resultant initial perturbation is the CNOP proposed by Mu et al. (2003). The O-CNOPs have better performance than the singular vectors and bred vectors in ensemble forecasts for skillful typhoon tracks (Huo and Duan 2019; Huo et al. 2019).

To generate the C-NFSVs and their particular cases, one should first solve the optimization problems given in Eqs. (2.5), (2.7), and (2.8). This is done using a nonmonotonic spectral projected gradient solver [SPG2; the details can be seen in Birgin et al. (2000)], where the gradients of the cost function with respect to the initial perturbation and the tendency perturbation are needed. For the calculation of C-NFSVs, we refer to Duan and Zhou (2013) for the computation of the gradient of the cost function with respect to both the initial perturbations and tendency perturbations. This is also reproduced in appendix A. With this gradient information, we can compute the C-NFSVs by descending along the direction of the gradient using the SPG2 solver.

Please note that much accurate adjoint-gradient information is used in the present study to compute the C-NFSVs because of the simplicity of the Lorenz-96 model (see section 3 for the Lorenz-96 model). There exist algorithms that do not need an adjoint to calculate the gradient but can still be used to compute the C-NFSVs for complex models; the particle swarm optimization algorithm and the genetic algorithm are examples. These algorithms have been successfully applied in the calculation of the CNOP and NFSV mentioned above [see the review of Wang et al. (2020)].

3. Lorenz-96 model

In this paper, the experiments are performed in the context of the Lorenz-96 model (Lorenz 1996; Lorenz and Emanuel 1998). The model is governed by the dynamical equations:

$$\frac{dX_j}{dt} = (X_{j+1} - X_{j-2})X_{j-1} - X_j + F, \qquad (3.1)$$

where $j = 1, 2, \ldots, m$ ($m \in \mathbb{N}$) is the dimension of the state space; the variable X_j satisfies the cyclic boundary conditions, i.e., $X_{-1} = X_{m-1}$, $X_0 = X_m$, and $X_{m+1} = X_1$, and they can be thought of as representing nondimensional meteorological quantities (e.g., temperature, pressure, vorticity, or gravitational potential, etc.) that are equally spaced along a latitudinal circle. The linear terms and constant term F describe the internal dissipation of the atmosphere and the external forcing, respectively.

Throughout the present study, the reference model dimension and forcing term are chosen as $m = 40$ and $F = 8$, displaying sensitivity to initial conditions (Lorenz 1996; Van Kekem 2018). The Lorenz-96 model with this configuration is integrated using a fourth-order Runge–Kutta scheme with a nondimensional time step of 0.05 time units. Considering a dimensional time unit of 5 days, the dissipative decay time of the system is approximately one time unit, and the error-doubling time is approximately 0.4 time units. These properties of the Lorenz-96 model are consistent with realistic numerical weather forecast models (Lorenz and Emanuel 1998). Furthermore, when the system has reached its attractor, the mean and standard deviation of the variable X_j are approximately equal to 2.3 and 3.6, respectively. This system has been widely used in theoretical studies of error dynamics (Vannitsem and Toth 2002; Orrell 2005), data assimilation (Whitaker and Hamill 2002; Hunt et al. 2004; Bai et al. 2013) and adaptive observation (Lorenz and Emanuel 1998; Khare and Anderson 2006). This model has also been often regarded as a platform to explore the usefulness of new ensemble forecasting approaches (Descamps and Talagrand 2007; Revelli et al. 2010; Basnarkov and Kocarev 2012; Feng et al. 2016; Grudzien et al. 2020). The Lorenz-96 model with the above configuration is therefore also used here for examining the possible impact of using the C-NFSVs for ensemble forecasting.

4. Experimental strategy

In the present study, the Lorenz-96 model is assumed to be imperfect, with additive model errors. A group of constant tendency perturbations η_j ($j = 1, 2, \ldots, 40$) randomly selected from a dataset of normal distribution $N(0, \mathbf{I})$ are supposed to describe the model errors. Then, the perfect system can be written as follows:

$$\frac{dX_j}{dt} = (X_{j+1} - X_{j-2})X_{j-1} - X_j + F + \eta_j. \qquad (4.1)$$

After a spinup run of the reference given in Eq. (4.1), the system is integrated for 292 000 time steps (dimensional 200 years)

TABLE 1. Twelve combinations of optimization time period T and the constraint radius. $\delta_{\mathbf{I}}$ denotes the magnitude of the initial analysis error measured by the L2-norm, $\delta_{\mathbf{f}}$ is the magnitude of the tendency error measured by the L2-norm, and E_i ($i = 1, 2, 3, \ldots, 12$) represents the combination of the four optimization time periods and three constraint radii.

Optimization time period T (days)	Constraint radius		
	$0.6\delta_{\mathbf{I}}, 0.6\delta_{\mathbf{f}}$	$0.8\delta_{\mathbf{I}}, 0.8\delta_{\mathbf{f}}$	$1.0\delta_{\mathbf{I}}, 1.0\delta_{\mathbf{f}}$
3	E_1	E_2	E_3
4	E_4	E_5	E_6
5	E_7	E_8	E_9
6	E_{10}	E_{11}	E_{12}

to obtain a 200-yr time series of $\mathbf{X}$, where $\mathbf{X} = (X_1, X_2, \ldots, X_{40})$. From the time series, we take the state values of $\mathbf{X}$ every 1460 time steps (i.e., one year) as the initial values and their respective subsequent 12-day (i.e., 48 time steps) evolution as truth runs. Thus, a total of 200 truth runs are selected in the present study.

The "observations" used are artificial and obtained by adding random noises sampled from a standard normal distribution $N(0, \mathbf{I})$ with a standard deviation of the observational errors equal to 27% of the standard deviation of X_j within the time period of the data assimilation cycle. The "observations" are assimilated by applying the four-dimensional variational data assimilation (4D-Var) to the imperfect Lorenz-96 model yielding the optimal initial field at the initial time of the forecast. Starting from these initial fields, the Lorenz 96 model is integrated for 12 days, providing the control forecasts associated with the 200 truth runs. Note that these control forecasts here are contaminated by both the initial analysis errors and model tendency errors.

Let us now consider the C-NFSVs for the description of the uncertainties present in both the initial fields and model tendencies and conduct the ensemble forecasting experiments. The computation of the C-NFSVs depends on their constraint radii [i.e., the positive constant numbers $\sigma_{\mathbf{I}}$ and $\sigma_{\mathbf{f}}$ in Eq. (2.6)] and the optimization time periods $[0, T]$ (see section 2), which are also two important factors affecting the ensemble forecast skills. As such, we take different combinations of them to compute the C-NFSVs (see Table 1). As in Orrell (2002), the maximal amplitude of the tendency perturbations is computed as

$$\delta_{\mathbf{f}} = \frac{1}{T}\sum_{j=0}^{T}\left\|\frac{\mathbf{x}(t_j + \Delta t) - \mathbf{s}(t_j + \Delta t)}{\Delta t}\right\|_2, \qquad (4.2)$$

where $\mathbf{x}(t_j + \Delta t)$ denotes the truth state at time $t_j + \Delta t$ using Eq. (4.1), and $\mathbf{s}(t_j + \Delta t)$ is its forecast at time $t_j + \Delta t$ using Eq. (3.1) with the initial value being the true state $\mathbf{x}(t_j)$, and $\|\cdot\|_2$ is an L2 norm. The maximal amplitude of the initial perturbations is fixed as

$$\delta_{\mathbf{I}} = \|\mathbf{a}(t_0) - \mathbf{x}(t_0)\|_2, \qquad (4.3)$$

where $\mathbf{a}(t_0)$ is the initial analysis field of the control forecast and $\mathbf{x}(t_0)$ is the initial field of the truth run. Three kinds of

constraint radii with $\sigma_{\mathbf{I}}$ and $\sigma_{\mathbf{f}}$, i.e., $0.6\delta_{\mathbf{I}}$ and $0.6\delta_{\mathbf{f}}$, $0.8\delta_{\mathbf{I}}$ and $0.8\delta_{\mathbf{f}}$, and $\delta_{\mathbf{I}}$ and $\delta_{\mathbf{f}}$, are experimentally selected for the initial and tendency perturbations, which, together with the four selected optimization times of 3, 4, 5 and 6 days, lead to a total of 12 combinations (i.e., E_i, $i = 1, 2, 3, \ldots, 12$ in Table 1). Note that the constraint radii above are chosen for the purpose of analyzing their impact, while in a realistic forecast, they should be estimated using Eqs. (4.2) and (4.3) according to the historical observations and their hindcasts by assuming the former are truth runs and the latter are control forecasts. For each of the truth runs, the control forecast is regarded as the basic state around which the 20 orthogonal C-NFSVs are computed for each E_i, where the number of C-NFSVs is experimentally selected as 20, as it almost provides the highest ensemble forecast skill. These 20 C-NFSVs are superimposed to the initial field and model tendency of the control forecast as positive and negative perturbation pairs and integrate in to the model equation (3.1), leading to 40 perturbed forecasts which, together with the control forecast, generate 41 ensemble forecast members for each forecast date.

The evaluation of the quality of the ensemble forecast system is often performed by analyzing the skill of the ensemble mean in a deterministic way and by evaluating the skill of the probabilistic forecast themselves. The information extracted illustrates different aspects of the performance of the ensemble forecast system. In the present study, the root-mean-square error (RMSE) and the anomaly correlation coefficient (ACC) between the mean of the ensemble members and the true state are used to assess the quality of the deterministic forecast. The Brier score (BS; Brier 1950) and the relative operating characteristic curve area (ROCA; Mason 1982) are adopted to measure the probabilistic forecast skill of binary events. Here, we define the following two categories of events: higher frequency events (event 1) with $X_j > 2$ and lower frequency events (event 2) with $X_j > 5$, and the two events occur with frequencies of 0.53 and 0.25, respectively. The details of these four scores are described in appendixes B, C, D, and E. Note that the RMSE and BS are negatively oriented (i.e., the smaller the value is, the higher the ensemble forecast skill is), while the ACC and ROCA are positively oriented (i.e., the larger the value is, the higher the ensemble forecast skill is).

5. Results

In this section, the experimental strategy described in the last section is conducted, and the role of C-NFSVs in improving forecast skills is evaluated. Its performance is then compared with the ensemble forecasting based on O-CNOPs and O-NFSVs, emphasizing the importance of simultaneously considering the initial and model errors in the maximization problem leading to the C-NFSVs.

a. The ensemble forecast skill

The C-NFSVs experiments are listed in Table 1. The O-CNOPs and O-NFSVs have been designed in such a way that they have the same amplitudes as the C-NFSVs. For all forecasting dates and lead times ranging from 6 h to 12 days,

we compute the RMSE and ACC of the ensemble mean (deterministic) forecasts for each E_i, together with the BS and ROCA of the two binary events for the probabilistic forecasts. Figure 1 plots the RMSE, ACC, BS, and ROCA averaged over all the lead times for all the forecast dates.

In Fig. 1, it is noticed that for any kind of perturbation, their associated ensemble forecasts tend to achieve the highest skill when the perturbation amplitudes constrained by E_i are close to the amplitudes of the initial analysis errors defined in Eq. (4.3) and/or the tendency errors defined in Eq. (4.2). This suggests that if one adopts the O-CNOPs to estimate the initial error impact on the ensemble forecasts or the O-NFSVs to measure the model error impact, their amplitudes should be close to those of the initial analysis errors or tendency errors, respectively. If one adopts the C-NFSVs to consider the impact of both the initial and model errors, their amplitudes are slightly smaller than the pure initial analysis errors or tendency errors and provide higher skills as measured by the RMSE and ACC for the deterministic forecasting or measured by the BS and ROCA in the case of probabilistic forecasting.

Figure 1 also indicates that the skill based on the O-NFSVs gradually increases with the optimization time intervals and constraint radius. However, for the ensemble forecasts based on the O-CNOPs, the skill undergoes a fast increase first and then a slow increase when the optimization time intervals are increased for small initial perturbations. For large initial perturbations, there is no substantial change in skill. In fact, the forecast model adopted here includes a model error $-\eta_j$, superimposed at each time step of the model integration, so it accumulates, leading to a more significant impact at long lead times. Therefore, the O-NFSVs with longer optimization time intervals may be better for grasping the impact of model errors for increasing lead times. The corresponding ensemble forecasts with larger optimization time intervals may then exhibit higher skills, as shown in Fig. 1. For the O-CNOPs, since they only optimize the impact of the initial uncertainties, they may not be very sensitive to the impact of the model errors at long lead times, implying a reduced sensitivity of the skill to the optimization time interval.

If we now compare the ensemble forecasting system based on O-CNOPs, O-NFSVs and C-NFSVs for the same optimization time interval, we find that the ensemble forecasts generated by the C-NFSVs often achieve the highest skill. This means that there is a much larger space for the ensemble forecasting generated by C-NFSVs to improve the forecast skill against the control forecast than those based on the O-NFSVs and O-CNOPs, which, is because the C-NFSVs combine the impact of both errors.

To further clarify the dependence of the skill on different lead times, Fig. 2 displays the differences in performance measures at each lead time for the highest skill experiments (see Fig. 1). It is shown that, for all the lead times, the differences between the O-NFSVs (O-CNOPs) and C-NFSVs are positive for RMSE and BS, and negative for ACC and ROCA at most lead times, further indicating that the ensemble forecasting based on the C-NFSVs possess a higher skill than those made by O-CNOPs and O-NFSVs. When comparing O-CNOPs
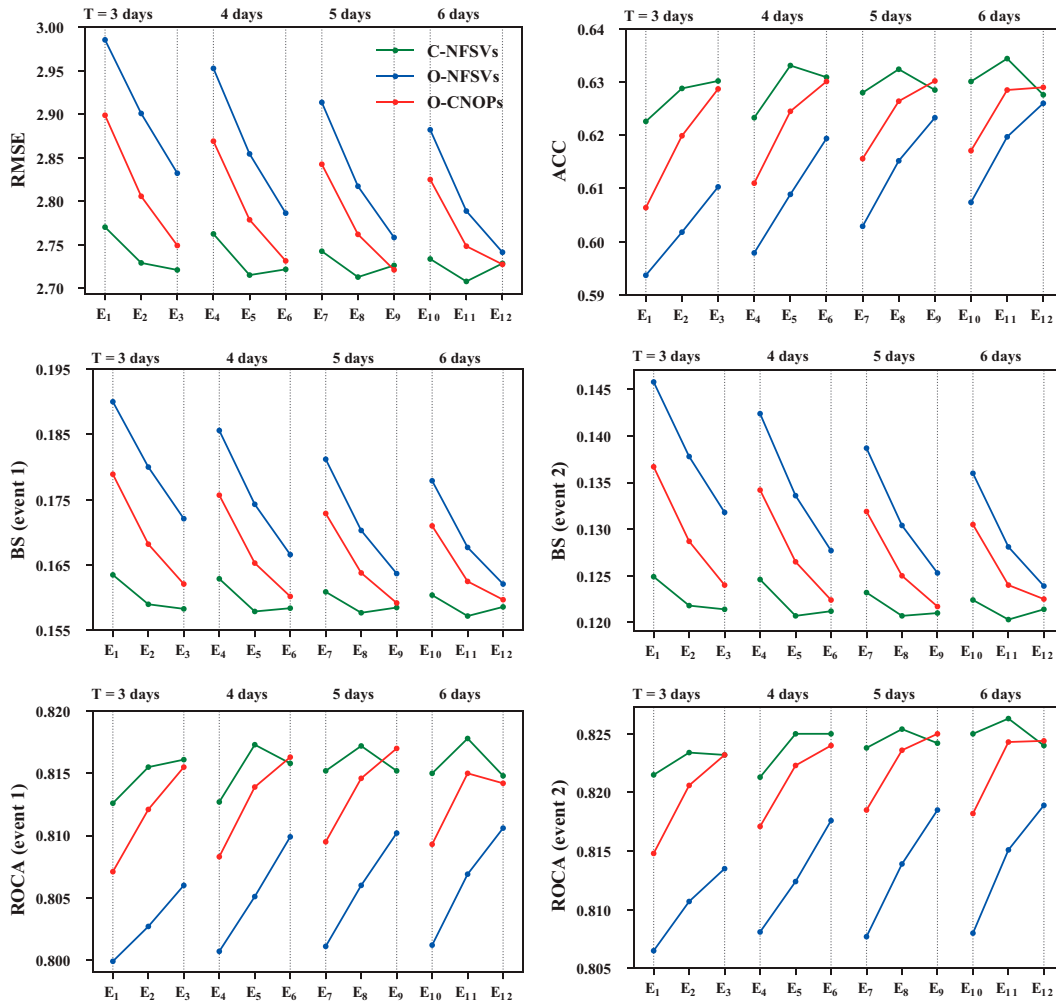
FIG. 1. The skills of the ensemble forecasts generated by C-NFSVs (green), O-NFSVs (blue), and O-CNOPs (red), as measured by the RMSE, ACC, BS, and ROCA. These values are obtained by averaging over 200 truth runs and all lead times over 12 days. The intervals with dashed lines divide the E_i that correspond to the same optimization time period T, and in each interval, the constraint radius increases with i, as shown in Table 1. The vertical axis represents the RMSE, ACC, BS, and ROCA.

and O-NFSVs, Fig. 2 further indicates that in the early stage of the forecasts, the ensembles based on the O-NFSVs have a lower skill than those based on the C-NFSVs and O-CNOPs, while for increasing lead times, the ensemble forecasts based on the O-NFSVs gradually improve compared with those based on the O-CNOPs and the C-NFSVs. This result indicates that the impact of the initial errors dominates the forecast error in the early stage of the forecasts, while when the lead time is large, the impact of the model errors gradually increases, and the O-NFSVs start to exert a stronger influence on the forecast uncertainties with a progressive convergence to the skill of the C-NFSVs.

b. Reliability

The above results have shown that the C-NFSVs allow for obtaining higher skills than the O-NFSVs and O-CNOPs when both the initial and model errors are present. In this

section, we will explore the reliability of the probabilistic aspects of the ensemble forecasting system. In principle, a reliable ensemble should display a ratio between the RMSE of the ensemble mean and the spread of the ensemble members (i.e., the square root of the average of the variances of the ensemble forecast members around their mean; see appendix B) approaching 1 during the entire forecast time period (Bowler 2006; Leutbecher and Palmer 2008; Hopson 2014; Fortin et al. 2014), indicating that the spread is a good proxy for the ensemble mean forecast error. Figure 3 shows the ratio of the ensemble spread to RMSE for C-NFSVs, O-CNOPs, and O-NFSVs when they achieve the highest skill in their respective settings (see Table 1 and Fig. 1), for (i) the O-CNOPs are computed by taking the amplitude δ_I and the optimization time of 5 days; (ii) the O-NFSVs are obtained with the amplitude δ_f and the optimization time of 6 days; and (iii) the C-NFSVs are obtained by using the initial perturbation
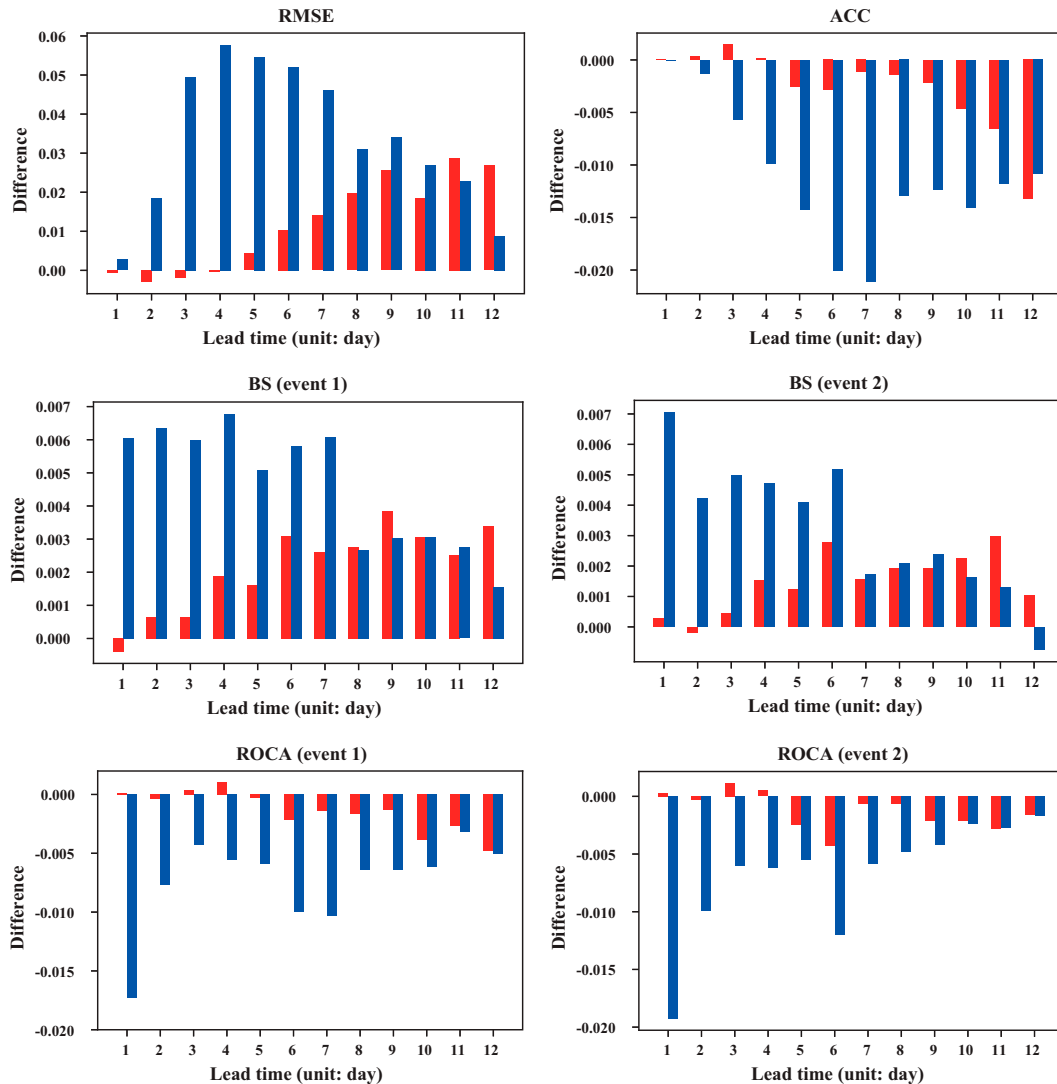
FIG. 2. The difference in skills between the ensemble forecasts generated by O-CNOPs and those generated by C-NFSVs (red) and between the ensemble forecasts made by O-NFSVs and those made by C-NFSVs (blue). The horizontal axis denotes the lead time, and the vertical axis represents the difference in the RMSE, ACC, BS, and ROCA values.

amplitude $0.8\delta_I$ and tendency perturbation amplitude $0.8\delta_f$, together with an optimization time of 6 days. Note that the ratio in Fig. 3a is averaged over all initial states and relevant variables, while in Fig. 3b, the focus is on the individual variables, and the average is taken over the forecasting period. For the latter, it is assumed that the X_j in the Lorenz-96 model describes the variables at the grid points along a latitudinal circle (see section 3), and thus, the figure shows the spatial distribution of the ratio of the ensemble spread to the RMSE. Both ratios in Figs. 3a,b show coherent results. That is, the ensemble spread made by the C-NFSVs is much closer to the RMSE of the ensemble mean forecasts and indicates that the C-NFSVs provide a better estimate of the uncertainty either in time or in space.

The Talagrand diagram (or rank histogram) is also a measure of the reliability for the ensemble forecasting system

(Talagrand et al. 1997; Candille and Talagrand 2005). Specifically, a reliable ensemble has a flat histogram, indicating that the observation is indistinguishable from any member of the ensemble forecast. Figure 4 displays the Talagrand diagram of the C-NFSVs, O-CNOPs, and O-NFSVs at varying lead times. It is shown that the histograms for the C-NFSVs are much flatter than those for the O-NFSVs and O-CNOPs. This suggests that the ensemble forecasting made by the C-NFSVs is more reliable than those based on the O-CNOPs and O-NFSVs in the Lorenz-96 model.

The perturbation versus error correlation analysis (PECA; see appendix F), another popular score, is also used to evaluate the quality of the ensembles (Wei and Toth 2003; Buizza et al. 2005; Wei et al. 2008). The higher the PECA values of the individual (or optimally combined) ensemble members are, the more successful the ensemble is in achieving its goal
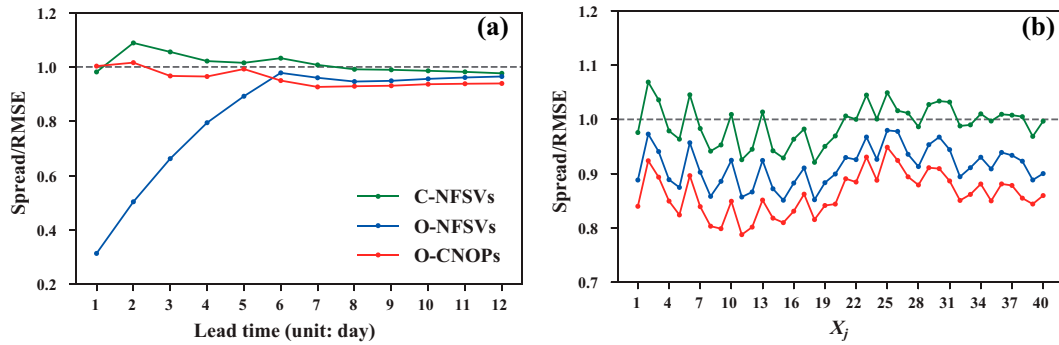
FIG. 3. (a) Temporal and (b) spatial variability of the ratio of the ensemble spread to the RMSE for the ensemble forecasts made by C-NFSVs (green), O-NFSVs (blue), and O-CNOPs (red). The O-CNOPs are calculated by using the amplitude δ_I and the optimization time of 5 days; the O-NFSVs are obtained with the amplitude δ_f and the optimization time of 6 days, while the C-NFSVs are obtained by using the initial perturbation amplitude $0.8\delta_I$ and a tendency perturbation amplitude of $0.8\delta_f$ for an optimization time of 6 days.

of capturing forecast errors. Figure 5 plots the PECA and indicates that, for either the optimally combined ensembles or the individual ones, the C-NFSVs and O-CNOPs tend to explain a larger amount of the forecast errors present in the control forecasts than the O-NFSVs in the early stage of the forecasts. For longer lead times, the ability of the O-NFSVs to explain the forecast error variances overtakes that of the O-CNOPs. In any case, the C-NFSVs almost always possess the largest PECA value and explain the largest amount of the forecast error variances of the control

forecasts. The interpretation provided in section 5a is still valid: the impact of the initial errors dominates the forecast error in the early stage of the forecasts that the O-CNOPs capture well, while for longer lead times, the O-NFSVs start to exert a stronger influence on the forecast uncertainties, with a progressive convergence to the optimal skill of the C-NFSVs. All these results indicate that the ensemble members generated by the C-NFSVs are much better for capturing the forecast errors and lead to an enhanced ensemble quality.
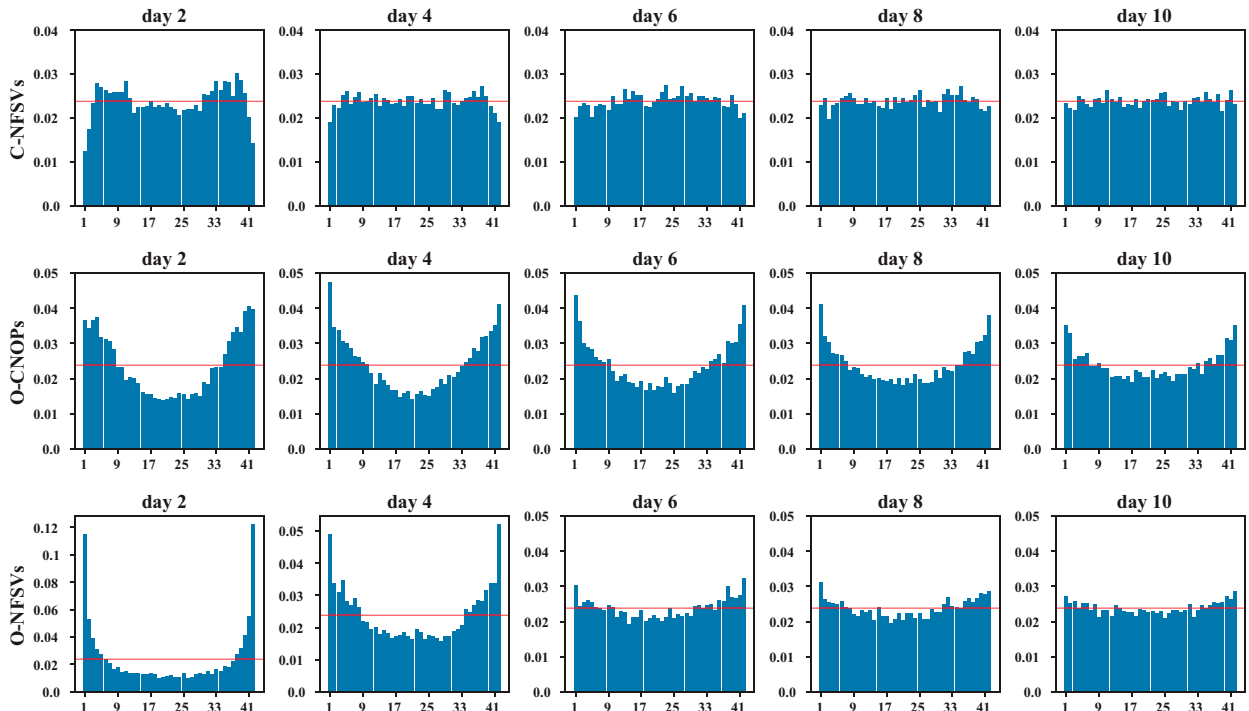

FIG. 4. Talagrand diagrams for the ensemble forecasts made by (top) C-NFSVs, (middle) O-CNOPs, and (bottom) O-NFSVs with configurations of the perturbation amplitudes and optimization times as in Fig. 3a, at lead times of (from left to right) 2, 4, 6, 8, and 10 days, respectively. The red horizontal lines denote the expected value of the probability.
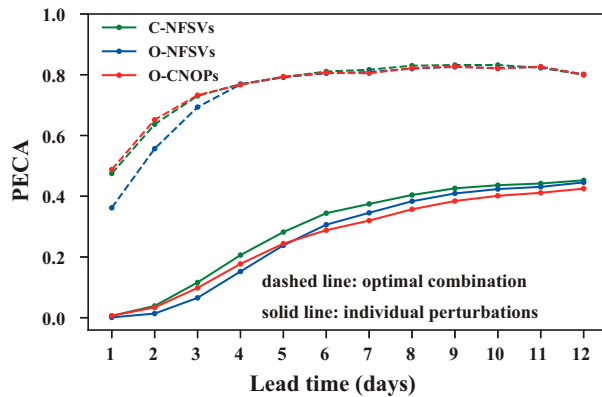
FIG. 5. The PECA values for C-NFSVs (green), O-NFSVs (blue), and O-CNOPs (red). The solid lines are the PECA values averaged over the 40 individual perturbation members, and the dashed lines represent the PECA values for the optimal combinations of the 40 perturbation members.

c. Impact of the dynamics of the C-NFSVs on the forecast skill

The dynamically growing similarity of the initial and tendency perturbations is important in achieving a high skill in the current ensemble forecasting system. To clarify this, we conduct the following two sets of experiments. The first set consists of perturbing the system by combining the O-CNOPs and O-NFSVs in ensemble forecasting mode in configurations in which they provide the highest skill. Specifically, the O-CNOPs are obtained with the initial perturbation amplitudes δ_I and the optimization time interval of 5 days, while the O-NFSVs are the ones for which the tendency perturbation amplitude is constrained by δ_f and the optimization time interval is fixed to 6 days (see Fig. 1). These skills are compared with those of the C-NFSVs in their best configuration (i.e., the initial perturbation amplitude with a radius of $0.8\delta_I$ and tendency perturbation amplitude with a radius of $0.8\delta_f$, together with the optimization time interval of 6 days; see Fig. 1). The second set of experiments combines O-CNOPs and O-NFSVs with the same optimization time interval and amplitude of perturbations as in the C-NFSVs of the first set. Figure 6 shows the comparisons of these two sets of experiments. It is shown that the ensemble forecasts made by the C-NFSVs consistently, on average, provide a smaller RMSE and BS and a larger ACC and ROCA than those made by the combined modes, which indicates that the C-NFSVs possess higher ensemble forecast skills than the above kinds of combined modes. This also suggests that the C-NFSVs are not a simple superposition of O-CNOPs and O-NFSVs but possess dynamical features that lead to a higher forecast skill. In other words, a simple combination of initial perturbations and tendency perturbations may cause inconsistent dynamical behaviors between them and would degrade the ensemble forecasting skill.

6. Tests of the use of deep learning to improve the applicability of C-NFSVs for real-time forecasts

The computation cost of ensemble forecasting systems is a challenge in an operational chain. The ensemble forecast method

introduced in the present study is made of members perturbed with C-NFSVs constructed through an SPG2 algorithm with multiple iterations along the fastest descent direction of the gradient (see section 2), and a large amount of computation time is needed. In this section, we propose an alternative way to estimate the evolution of the ensemble members generated by the C-NFSVs rather than to compute C-NFSVs themselves that would allow for a real-time implementation.

Deep learning (DL) techniques have been applied in weather and climate forecasting (Salman et al. 2015, Scher 2018), statistical postprocessing (Rasp and Lerch 2018; Scheuerer et al. 2020; Veldkamp et al. 2021; Vannitsem et al. 2021, for a recent review), and uncertainty prediction (Scher and Messori 2018), and have shown strong nonlinear fitting and prediction abilities. The DL techniques allow for developing inference functions between an input and an output by learning a training dataset. For the ensemble forecasts generated by the C-NFSVs of section 5, the ensemble members for different truth runs are constructed by perturbing control forecasts with their respective C-NFSVs and are close to the control forecasts. Therefore, we can use the control forecasts (also "hindcasts") of a training set of truth runs and their ensemble forecasting members generated by the C-NFSVs and use the DL approach to learn the dependence relationship of the ensemble forecasting members to the corresponding control forecasts. These inference functions can then be used to produce ensemble forecasting members in the testing period. Clearly, if the DL approach is successful in achieving an ensemble forecasting skill close to that of section 5, it can be used to bypass the computational burden of the C-NFSVs because the training is done offline.

In the present study, the idea is to input the control forecast to a DL model and output the corresponding ensemble members around the control forecast. As the control forecast is made of 40 spatial variables X_j ($j = 1, 2, 3, \ldots, 40$) integrated over a time period of 10 days, it consists of a spatiotemporal forecasting problem. In DL algorithms, the convolutional neural network (CNN) is a tool well adapted to learn the features of multidimensional datasets and is especially popular for dealing with spatial information (LeCun et al. 2015; Krizhevsky et al. 2017), while the long short-term memory neural network (LSTM) is a network able to extract temporal information. The convolutional LSTM [ConvLSTM; Shi et al. (2015)], which is a combination of LSTM and CNN, can then be used for appropriately learning spatial–temporal data. The difference between the ConvLSTM and the LSTM lies in that the fully connected structure of the network is replaced by a convolution operation that can identify spatial features. More precisely, the output of ConvLSTM includes both temporal and spatial dimensions but the memory unit connections of ConvLSTM operate mainly on the temporal dimension, while the convolutions capture the spatial information at a specific time. Then, if the output of the ConvLSTM is used as the input of the CNN, the spatial features of the adjacent time can be further extracted, with the advantage of reducing the number of parameters due to the sparse connectivity and parameter sharing of the CNN. The fully connected layers are then "merging" the information provided by the CNN and produce the output.
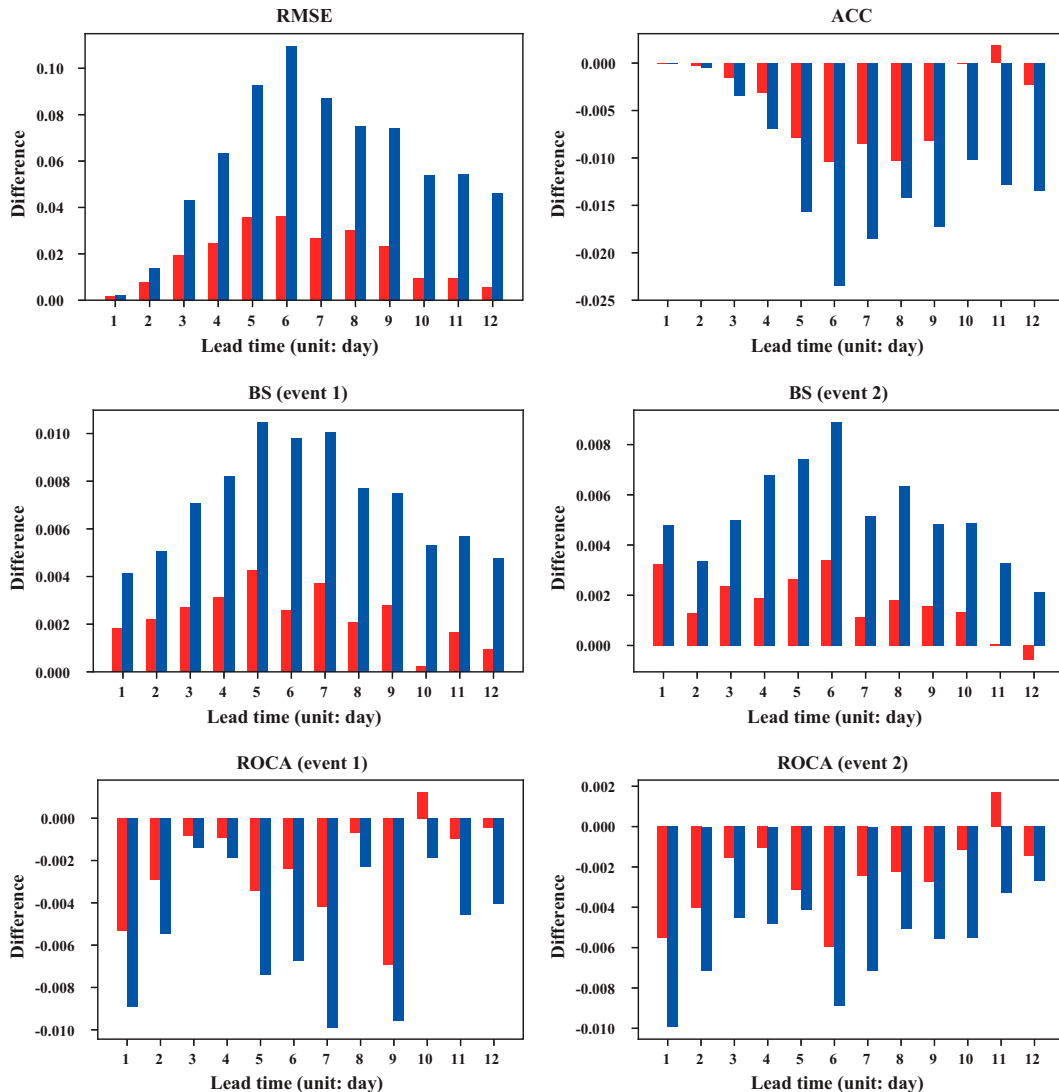
FIG. 6. As in Fig. 2, but showing the skill performance differences between the ensemble forecasts made by the combination of O-CNOPs and O-NFSVs when they obtain the highest skill scores and those made by C-NFSVs (red) and between the combination of O-CNOPs and O-NFSVs with the same optimization time period and perturbation amplitudes as in C-NFSVs and those made by C-NFSVs (blue).

We take the first consecutive 10-yr data (i.e., a total of 14 600 time steps with the time step being equal to 0.05) from the 200-yr integration of the model Eq. (4.1) as truth runs. We divide this time series into two parts: one part contains the first 12 000 time step data as the training period, and the other part includes the remaining 2600 time step data as the test period. The training data for all the lead times are normalized to have an amplitude in the interval $[-1, 1]$ by calculating $z_N = \{[(z - z_{min})/(z_{max} - z_{min})] - 0.5\} \times 2$, where z is the original data, and z_{min}, z_{max}, and z_N are the minimum, maximum and normalized values of the training data z for all the lead times, respectively. These normalized data are the inputs of the LSTM cells. Thanks to the ConvLSTM, the temporal features of the control forecasts are extracted, while the local spatial information is preserved. The output of the ConvLSTM is

then used as the input of the CNNs, allowing for the identification of the deeper spatial features of the control forecast. Then, the fully connected layers are used for merging the features that have been extracted and generating the output of the normalized values of the 40 variables of an ensemble member at a given lead time. Such a strategy is summarized in Fig. 7 describing Model-1. Note that for each member at each lead time, there is a separate Model-1, so that a total of 400 DL models with the architecture of Model-1 are used to generate the 40 ensemble members at lead times of 1, 2, 3, …, 10 days. Note that the system under investigation has only 40 dimensions. In this context, the fully connected layers can be easily used to merge the extracted features and generate the final output. In an operational environment, the spatial dimensions are considerably higher, and other types of layers, such as the deconvolution
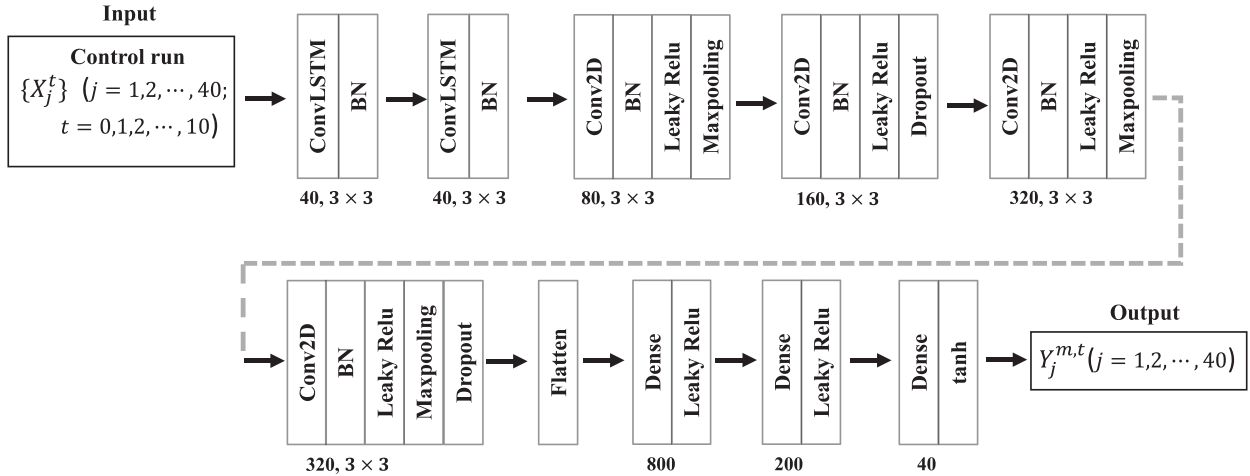
FIG. 7. Sketch of Model-1. Two ConvLSTM layers with filter sizes of 3 × 3 and 40 filters; four 3 × 3 convolutional layers with 80, 160, 320, and 320 filters; three fully connected layers with 800, 200, and 40 nodes; batch normalization (BN) layers; 2 × 2 max pooling layers; and dropout layers with a dropout rate of 0.2 are selected. The input $\{X_j^t\}$ ($j = 1, 2, …, 40; t = 0, 1, 2, …, 10$) represents the control forecast of 40 variables with lead times from 0 to 10 days, and the output $Y_j^{m,t}(j = 1, 2, …, 40)$ represents the values of 40 variables of the mth ensemble member at a given lead time t. For each member at each lead time, a separate Model-1 is used. A total of 400 Model-1s are used to generate the ensemble members.

layers, should be used after the above CNN layers in Model-1 instead of the fully connected layers. This should make the DL models more adaptable to the high-dimensional problems associated with operational forecasts.

In Model-1, we experimentally select two ConvLSTM layers of 40 filters and four convolutional layers of 80, 160, 320 and 320 filters with each filter having the size 3 × 3 (where a padding is made for having the same size of input and output) and three fully connected layers of 800, 200 and 40 nodes. In addition, batch normalization layers and max pooling layers with a filter size of 2 × 2 and a stride of 2 are empirically selected to enhance the efficiency of the training process. Dropout layers with a dropout rate of 0.2 are also used for regularization. Furthermore, the Leaky ReLU active function is also introduced for learning nonlinear features (Maas et al. 2013), and the Adam optimizer with a learning rate of 10^{-5} (Kingma and Ba 2014) is used in the loss function for minimizing the distance between the DL output and the corresponding truth. In Model-1, the size of the mini batch for each epoch in the training process is chosen as 100, and the number of epochs is fixed to 200. All these hyperparameters have been defined by trial and error.

Note that the final ensemble members are generated by running several cycles of Model-1. More specifically, during the training, a control run is injected in Model-1, and the values of 40 variables of one member at a given lead time are produced as output, which are then compared with an ensemble member by calculating a loss function that measures their distance and/or similarity; after this, the Adam Optimizer is applied to find improved parameters based on the batch gradient descending direction. After several cycles, the decrease in the loss function will reach saturation, and the optimal ensemble member is obtained. In the present study, the loss function is defined as

$$\text{Loss} = \alpha \times \text{MSE} - \beta \times \text{CC}, \qquad (6.1)$$

where

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (Y_{\text{truth}} - Y_{\text{pred}})^2 = \frac{1}{n} \sum_{i=1}^{n} [Y_{\text{truth}} - f_\theta(X)]^2, \qquad (6.2)$$

and

$$\text{CC} = \frac{1}{n} \sum_{i=1}^{n} \rho(Y_{\text{truth}}, Y_{\text{pred}}) = \frac{1}{n} \sum_{i=1}^{n} \frac{\text{cov}(Y_{\text{truth}}, Y_{\text{pred}})}{\sqrt{\text{var}(Y_{\text{truth}}) \times \text{var}(Y_{\text{pred}})}}. \qquad (6.3)$$

The loss function consists of a combination of MSE and CC, where the MSE measures the mean square error between the inferred 40 variables of an ensemble member at a given lead time [represented by Y_{pred} in Eq. (6.2)] during the training period and the real values [denoted by Y_{truth} in Eq. (6.2)]; CC is the correlation coefficient between them; and n is the size of the mini batch. The MSE reflects the magnitude of the error and is sensitive to outliers but does not provide the location of the error. Therefore, to make the ensemble members generated by Model-1 more efficient, the parameters α and β are introduced for adjusting the weights of MSE and CC to reach a good balance between the correction of the amplitude and location. The outputs of the DL model are therefore normalized to have an amplitude in the interval $[-1, 1]$ using a hyperbolic tangent function (tanh; see Fig. 7), and the values of α and β are fixed to 0.2 and 0.8 to achieve an optimal quality for the ensemble members.

Figure 8 shows the corresponding RMSE, ACC, BS and ROCA compared to those of the ensemble forecasts of section 5
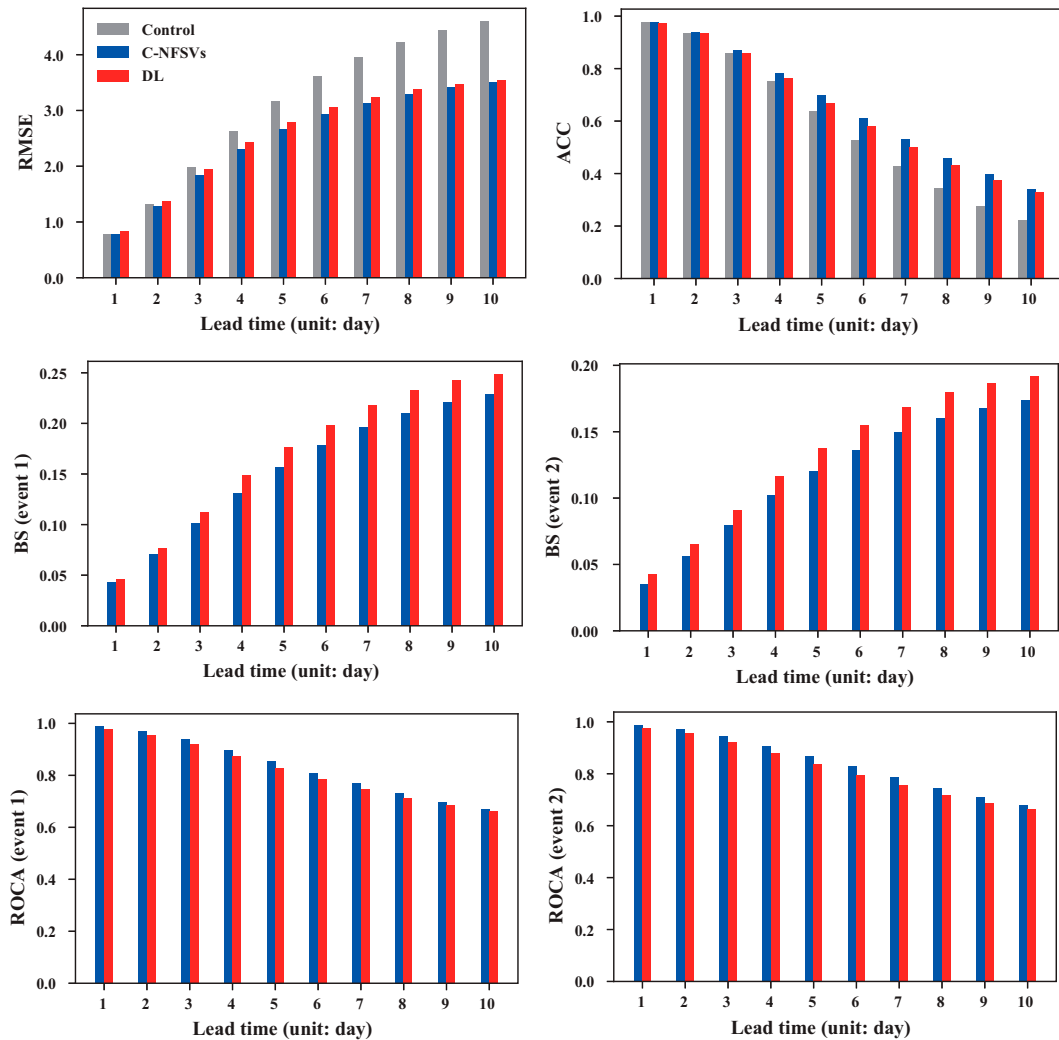
FIG. 8. The skills of the control forecast (gray) and the ensemble forecasts made by the C-NFSVs calculated by the SPG2 (blue) and deep learning (DL) model (i.e., Model-1; red), as measured by RMSE, ACC, BS, and ROCA. These values are obtained by an averaging over 2600 truth runs in the test period.

(for simplicity, hereafter referred to as "original forecasts"). It is found that the ensemble forecasts generated by Model-1 are close to the original forecasts when analyzing the deterministic forecasting skills (RMSE and ACC). For probabilistic forecasting, Model-1 presents a skill approximately 3.0% lower than the original forecast but still acceptable.

It was also found that the ensemble spread inferred by Model-1 is much smaller than the corresponding RMSE of the ensemble mean. This indicates that the ensemble members here, compared to those provided by the C-NFSVs, do lack variability and therefore do not provide an appropriate estimate of the uncertainty around the ensemble mean. Alternative ways should be found. To solve this issue, we further use the architecture of Model-1 but change the model output and associated loss function to directly estimate the RMSE of the individual ensembles (i.e., the RMSE with the realization being $N = 1$ in appendix B), rather than using the corresponding

ensemble spread (i.e., the spread with the realization being $N = 1$ in appendix B). For convenience, we refer to this modified DL model as "Model-2." In this situation, the inputs are still the control forecasts, but the outputs are now the estimations of the RMSE of the individual ensembles at a given lead time (i.e., replace the 40 nodes in the last dense layer with 1 node). For lead times from 1 to 10 days, a total of 10 DL models (i.e., Model-2) are trained. Furthermore, in the loss function MSE, as in Eq. (6.2), Y_{truth} represents the RMSE of the individual ensembles made by Model-1 in the training period, and Y_{pred} is its estimation based on Model-2 and then applied in the testing period. Figure 9 shows the estimation of the RMSE averaged over all the true runs and all the lead times in the testing period. The RMSE forecasted by Model-2 is close to the RMSE of Model-1. As a result, Model-1 combined with Model-2 provides an ensemble forecasting system mirroring that of the
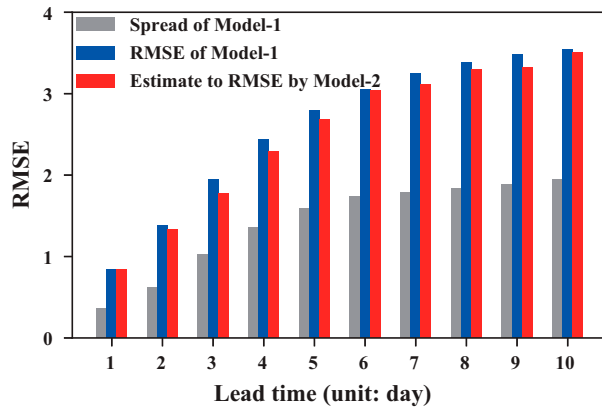
FIG. 9. RMSE of the ensemble forecasts based on Model-1 (blue bars) and the corresponding ensemble spread (gray bars) and the estimation of the RMSE of the ensemble forecasts (red bars), as made by the deep learning model (i.e., the Model-2).

original forecasts, with results close to those predicted by the original forecasting system.

The DL approach to ensemble forecasting based on the C-NFSVs, as proposed in the present work, saves computing time as it no longer needs an optimization process at each forecasting date. However, many uncertainties are still present concerning the construction of the DL model, especially for the choice of the values of the hyperparameters, and other inputs and outputs may also need considering. This needs further investigation.

7. Summary and discussion

In this study, we extend the NFSV approach to address the model error impact on ensemble forecasting and propose a C-NFSV ensemble forecasting method that considers the impact of both the initial and model errors. The C-NFSVs provide a group of optimally growing combined perturbations for both the initial conditions and the model tendencies. Two particular cases were also considered: O-CNOPs and O-NFSVs. The former is an ensemble forecasting method formerly proposed by Duan and Huo (2016), dealing with the initial error impact on the forecasts, while the latter estimates the model error impact. The usefulness of the C-NFSVs is demonstrated in the context of the Lorenz (1996) system.

The results show that the ensemble forecasting based on O-CNOPs has a higher skill than the one based on the O-NFSVs in the early stage of the forecasts, while in the later stage of the forecast, the impact of the model errors becomes more prominent and the ensemble forecasting based on the O-NFSVs excels. In any case, the forecasts based on the C-NFSVs, thanks to their optimization on both the initial and model errors, possess higher skill than the ones based on the O-CNOPs and O-NFSVs. These results justify the interest in using C-NFSVs in building ensemble forecasts.

Considering that developing an ensemble based on C-NFSVs is challenging for ensemble forecasting in an operational environment, we also discuss the possible use of deep learning

algorithms in providing ensemble forecast information similar to that generated by the C-NFSVs. We develop two different DL models, one that learns the link between the control forecasts and the ensemble members and a second that learns the link between the control and the RMSE of the ensemble. These DL models allow bypassing of the optimization problem to obtain the C-NFSVs at each forecast date. In the context of the Lorenz-96 model, we show that the ensemble forecasts made by the DL models have forecasting skills close to those made by computing the C-NFSVs. This suggests that the introduction of DL algorithms could not only greatly reduce the computing time costs of the ensemble forecasting system made by the C-NFSVs but could also achieve a comparable forecast skill. It also illustrates the potential of the DL algorithms in the context of the operational ensemble forecasts but the DL models must then be adapted to the high-dimensional problems of the operational forecasts. Furthermore, a large variety of DL architectures exist and need to be optimized by trial or error. It should also be understood that the DL model requires a very large training dataset during the training period, and then a large amount of computational time is needed for calculating the C-NFSVs before the operational implementation. Therefore, a highly efficient and effective optimization algorithm is still necessary for calculating the C-NFSVs during the hindcast period.

The current results also emphasize the importance of the dynamically coordinated growth of the initial perturbations and model perturbations in improving the ensemble forecast skill. In the same spirit, comparisons between C-NFSVs and the combined modes of other types of initial perturbations (e.g., BVs or SVs) and other types of model perturbations (such as STTP or SPPT) are also worth performing in the future. Another interesting avenue in the development of C-NFSVs is to consider the effect of time-varying stochastic errors; a combined mode of C-NFSVs and random forcing tendency perturbations may cover a broader range of model errors and have potential for further improving the ensemble forecast skill. In addition, given the simplicity of the Lorenz-96 model, more realistic models must be considered to examine the usefulness of the C-NFSVs for ensemble forecasting systems. This will be the subject of a follow-up work. In any case, it is expected that the C-NFSVs, possibly combined with DL algorithms, will play an important role in realistic ensemble forecasts in the future.

Data availability statement. The datasets generated and/or analyzed during the study are stored on computers at the State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics (LASG; https://www. lasg.ac.cn) and will be available to researchers upon request.

APPENDIX A

A Derivation of the Gradients

The cost function associated with the C-NFSVs can be written as follows:

$$J_1(\mathbf{u}_0, \mathbf{f}) = -\frac{1}{2}[J(\mathbf{u}_0, \mathbf{f})]^2 = -\frac{1}{2}\|M_T(\mathbf{f})(\mathbf{U}_0 + \mathbf{u}_0) - M_T(\mathbf{U}_0)\|_b^2$$

$$= -\frac{1}{2}\langle \mathbf{u}(\mathbf{u}_0, \mathbf{f}; T), \mathbf{u}(\mathbf{u}_0, \mathbf{f}; T)\rangle, \tag{A1}$$

where

$$J(\mathbf{u}_0, \mathbf{f}) = \|M_T(\mathbf{f})(\mathbf{U}_0 + \mathbf{u}_0) - M_T(\mathbf{U}_0)\|_b, \tag{A2}$$

$\mathbf{u}(\mathbf{u}_0, \mathbf{f}; T) = M_T(\mathbf{f})(\mathbf{U}_0 + \mathbf{u}_0) - M_T(\mathbf{U}_0)$ [i.e., $\mathbf{u}_T$ in Eq. (2.4) in section 2], and $\langle \cdot \rangle$ is the inner product. By minimizing Eq. (A1) using an existing optimization solver, the C-NFSVs can be obtained, provided the gradients of the cost function with respect to initial perturbations and model tendency perturbations are estimated. Following Duan and Zhou (2013), the gradients can be computed.

The first-order variations of $J_1(\mathbf{u}_0, \mathbf{f})$ are

$$-\delta J_1 = \langle \mathbf{u}(T), \delta\mathbf{u}(T)\rangle = \left\langle \frac{-\partial J_1}{\partial \mathbf{u}_0}, \delta\mathbf{u}_0\right\rangle + \left\langle \frac{-\partial J_1}{\partial \mathbf{f}}, \delta\mathbf{f}\right\rangle, \tag{A3}$$

where $\delta\mathbf{u}(t)$ and $\delta\mathbf{f}$ are governed by the following tangent linear model:

$$\begin{cases} \dfrac{\partial \delta\mathbf{u}}{\partial t} = \dfrac{\partial F[\mathbf{U}(t) + \mathbf{u}(t)]}{\partial \mathbf{u}} \delta\mathbf{u} + \delta\tilde{\mathbf{f}}, \\[2mm] \dfrac{\partial \delta\tilde{\mathbf{f}}}{\partial t} = 0, \\[2mm] \delta\mathbf{u}|_{t=0} = \delta\mathbf{u}_0, \\[2mm] \delta\tilde{\mathbf{f}}|_{t=0} = \delta\mathbf{f}, \end{cases} \tag{A4}$$

By introducing two Lagrangian multipliers λ_1 and λ_2, we obtain

$$-\delta J_1 = \langle \mathbf{u}(T), \delta\mathbf{u}(T)\rangle$$

$$- \int_0^T \left\langle \lambda_1(t), \frac{\partial \delta\mathbf{u}}{\partial t} - \frac{\partial F[\mathbf{U}(t) + \mathbf{u}(t)]}{\partial \mathbf{u}} \delta\mathbf{u} - \delta\tilde{\mathbf{f}}\right\rangle dt$$

$$- \int_0^T \left\langle \lambda_2(t), \frac{\partial \delta\tilde{\mathbf{f}}}{\partial t}\right\rangle dt.$$

With an integration by parts, we can obtain

$$\int_0^T \left\langle \lambda_1(t), \frac{\partial \delta\mathbf{u}}{\partial t}\right\rangle dt = \int_0^T \frac{\partial}{\partial t}\langle \lambda_1(t), \delta\mathbf{u}\rangle dt - \int_0^T \left\langle \frac{\partial \lambda_1(t)}{\partial t}, \delta\mathbf{u}\right\rangle dt$$

$$= \langle \lambda_1(T), \delta\mathbf{u}(T)\rangle - \langle \lambda_1(0), \delta\mathbf{u}(0)\rangle$$

$$- \int_0^T \left\langle \frac{\partial \lambda_1(t)}{\partial t}, \delta\mathbf{u}\right\rangle dt$$

and

$$\int_0^T \left\langle \lambda_2(t), \frac{\partial \delta\tilde{\mathbf{f}}}{\partial t}\right\rangle dt = \int_0^T \frac{\partial}{\partial t}\langle \lambda_2(t), \delta\tilde{\mathbf{f}}\rangle dt - \int_0^T \left\langle \frac{\partial \lambda_2(t)}{\partial t}, \delta\tilde{\mathbf{f}}\right\rangle dt$$

$$= \langle \lambda_2(T), \delta\tilde{\mathbf{f}}(T)\rangle - \langle \lambda_2(0), \delta\tilde{\mathbf{f}}(0)\rangle$$

$$- \int_0^T \left\langle \frac{\partial \lambda_2(t)}{\partial t}, \delta\tilde{\mathbf{f}}\right\rangle dt$$

$$= \langle \lambda_2(T), \delta\mathbf{f}\rangle - \langle \lambda_2(0), \delta\mathbf{f}\rangle$$

$$- \int_0^T \left\langle \frac{\partial \lambda_2(t)}{\partial t}, \delta\tilde{\mathbf{f}}\right\rangle dt.$$

Then, we derive δJ_1 as follows:

$$-\delta J_1 = \int_0^T \left\langle \frac{\partial \lambda_1(t)}{\partial t}, \delta\mathbf{u}\right\rangle dt + \langle \mathbf{u}(T) - \lambda_1(T), \delta\mathbf{u}(T)\rangle + \langle \lambda_1(0), \delta\mathbf{u}(0)\rangle$$

$$+ \int_0^T \left\langle \lambda_1(t), \left[\frac{\partial F(\mathbf{U} + \mathbf{u})}{\partial \mathbf{u}}\right]\delta\mathbf{u}\right\rangle dt + \int_0^T \langle \lambda_1(t), \delta\tilde{\mathbf{f}}\rangle dt$$

$$+ \int_0^T \left\langle \frac{\partial \lambda_2(t)}{\partial t}, \delta\tilde{\mathbf{f}}\right\rangle dt + \langle \lambda_2(0), \delta\mathbf{f}\rangle - \langle \lambda_2(T), \delta\mathbf{f}\rangle$$

$$= \int_0^T \left\langle \frac{\partial \lambda_1(t)}{\partial t}, \delta\mathbf{u}\right\rangle dt + \langle \mathbf{u}(T) - \lambda_1(T), \delta\mathbf{u}(T)\rangle + \langle \lambda_1(0), \delta\mathbf{u}(0)\rangle$$

$$+ \int_0^T \left\langle \left[\frac{\partial F(\mathbf{U} + \mathbf{u})}{\partial \mathbf{u}}\right]^* \lambda_1(t), \delta\mathbf{u}\right\rangle dt + \int_0^T \langle \lambda_1(t), \delta\tilde{\mathbf{f}}\rangle dt$$

$$+ \int_0^T \left\langle \frac{\partial \lambda_2(t)}{\partial t}, \delta\tilde{\mathbf{f}}\right\rangle dt + \langle \lambda_2(0), \delta\mathbf{f}\rangle + \langle 0 - \lambda_2(T), \delta\mathbf{f}\rangle, \tag{A5}$$

where $[\cdot]^*$ denotes the adjoint. By comparing (A5) with (A3), we can obtain the gradients of the cost function with respect to initial perturbations [see Eqs. (A6) and (A8)] and model tendency perturbations [see Eqs. (A7) and (A8)]:

$$\frac{\partial J_1}{\partial \mathbf{u}_0} = -\lambda_1(0), \tag{A6}$$

and

$$\frac{\partial J_1}{\partial \mathbf{f}} = -\lambda_2(0), \tag{A7}$$

where $\lambda_1(t)$ and $\lambda_2(t)$ satisfy

$$\begin{cases} \dfrac{\partial \lambda_1(t)}{\partial t} + \left[\dfrac{\partial F(\mathbf{U} + \mathbf{u})}{\partial \mathbf{u}}\right]^* \lambda_1(t) = 0, \\[2mm] \dfrac{\partial \lambda_2(t)}{\partial t} + \lambda_1(t) = 0, \\[2mm] \lambda_1|_{t=T} = \mathbf{u}(T), \\[2mm] \lambda_2|_{t=T} = 0, \end{cases} \tag{A8}$$

and Eq. (A8) is the adjoint equation of the tangent linear model equation [Eq. (A4)].

Note that the gradients $\partial J_1/\partial \mathbf{u}_0 = -\lambda_1(0)$ are used to calculate O-CNOPs and $\partial J_1/\partial \mathbf{f} = -\lambda_2(0)$ to calculate the O-NFSVs. For the C-NFSVs, since the initial perturbation is taken as $\mathbf{u}_0 = r\mathbf{f}$, the gradient required is $\partial J_1/\partial \mathbf{f} = (\partial J_1/\partial \mathbf{u}_0)(\partial \mathbf{u}_0/\partial \mathbf{f}) + \partial J_1/\partial \mathbf{f} = -r\lambda_1(0) - \lambda_2(0)$.

APPENDIX B

Root-Mean-Square Error of the Ensemble Mean and Ensemble Spread

The ensemble mean is calculated as follows. Let $\overline{X}_j$ be the ensemble mean of the jth component of the M members:

$$\overline{X}_j = \frac{1}{M} \sum_{m=1}^{M} X_j^{(m)},$$

where $X_j^{(m)}$ represents the jth component of the mth ensemble member; $j = 1, 2, ..., J$ and $m = 1, 2, ..., M$. The root-mean-square error (RMSE) of the ensemble mean measures the difference between the ensemble mean $\overline{X}_{j,n}$ and the observations $O_{j,n}$ (i.e., the truth runs in the present study), which is defined as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{n=1}^{N} \frac{1}{J} \sum_{j=1}^{J} (\overline{X}_{j,n} - O_{j,n})^2}, \quad \text{(B1)}$$

where N represents the number of realizations.

The ensemble spread is defined as follows:

$$\text{SPRD} = \sqrt{\frac{1}{N} \sum_{n=1}^{N} \frac{1}{J} \sum_{j=1}^{J} \frac{1}{M-1} \sum_{m=1}^{M} (X_{j,n}^{(m)} - \overline{X}_{j,n})^2}. \quad \text{(B2)}$$

APPENDIX C

Anomaly Correlation Coefficient

The anomaly correlation coefficient (ACC) is used to measure the similarity between the forecasted and observed anomalies. The ACC is formulated as follows:

$$\text{ACC} = \frac{\sum_{j}^{J} (X_j' - \overline{X'})(O_j' - \overline{O'})}{\sqrt{\sum_{j=1}^{J} (X_j' - \overline{X'})^2 (O_j' - \overline{O'})^2}}, \quad \text{(C1)}$$

where $X_j' = X_j - C_j^X$ and $O_j' = O_j - C_j^O$ in which X_j is the forecast value, O_j is the observation, C_j^X is the model climatological state, C_j^O is the observed climatological state, X_j' is the forecasted anomaly, and O_j' is the observed anomaly. The larger the ACC is, the higher the forecast skill.

APPENDIX D

Brier Score

The Brier score (BS; Brier 1950) is the mean square error of the probability forecasts with the following definition:

$$\text{BS} = \frac{1}{N} \sum_{i=1}^{N} (f_i - o_i)^2, \quad \text{(D1)}$$

TABLE E1. 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	$a + c$	$b + d$	

where N is the number of realizations of the prediction process, and f_i and o_i are the probability of forecast and observation for the ith prediction process, respectively. A smaller BS value indicates a better probability forecast skill.

APPENDIX E

Relative Operating Characteristic Curve Area

The relative operating characteristic curve area (ROCA; Mason 1982) is a measure of the resolution of a prediction system. By considering whether an event occurs at every grid and checking the forecasts against the observations, we can construct a two-category contingency table (see Table E1), 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 and the false alarm rate can be calculated as in Eqs. (E1) and (E2):

$$\text{Hit rate}: H = P(\text{event predicted} \mid \text{event occurs})$$
$$= a/(a + c), \quad \text{(E1)}$$

$$\text{False alarm rate}: F = P(\text{event predicted} \mid \text{event does not occur})$$
$$= b/(b + d). \quad \text{(E2)}$$

The ROC curve can be obtained by the pairs H and F, and the area under the ROC curve is called ROCA, which decreases from 1 to 0 as more false alarm rates occur. The ROCA is calculated as in Eq. (E3):.

$$\text{ROCA} = \int_0^1 H(x)dx = \sum_{M}^{M} \frac{1}{2}(H_{i+1} + H_i)(F_{i+1} - F_i), \quad \text{(E3)}$$

where M is the number of categories relative to probability thresholds. A larger ROCA value indicates a better probability forecast. When the ROCA is greater than 0.5, the forecast can be regarded as skillful.

APPENDIX F

Perturbation versus Error Correlation Analysis

The perturbation versus error correlation analysis (PECA) measures how well ensemble perturbations can explain forecast error variances (Wei and Toth 2003). Ensemble perturbations are defined as the differences between perturbed forecasts and the corresponding control forecast:

$$\mathbf{P}_i(t) = \mathbf{F}_i(t) - \mathbf{F}_{\text{ctrl}}(t), \tag{F1}$$

where $\mathbf{P}_i(t)$, $\mathbf{F}_i(t)$, and $\mathbf{F}_{\text{ctrl}}(t)$ are the ith ensemble perturbation, ith perturbed forecast, and the control forecast, respectively. The forecast errors $\mathbf{E}(t)$ are defined as the difference between the control forecast $\mathbf{F}_{\text{ctrl}}(t)$ and the verifying analysis $\mathbf{F}(t)$ (i.e., the truth runs in the present study):

$$\mathbf{E}(t) = \mathbf{F}_{\text{ctrl}}(t) - \mathbf{F}(t). \tag{F2}$$

The a posteriori optimal combination of M perturbations is obtained by solving the least squares problem:

$$\text{Min} \left\| \mathbf{E} - \sum_{i=1}^{M} \alpha_i \mathbf{P}_i \right\|_2^2. \tag{F3}$$

By solving the optimal problem (F3), one obtains an optimal $\hat{\alpha}_i$, and then the optimally combined vector $\mathbf{P}_{\text{opt}}$ can be written as

$$\mathbf{P}_{\text{opt}} = \sum_{i=1}^{M} \hat{\alpha}_i \mathbf{P}_i. \tag{F4}$$

If a pattern anomaly correlation between two vectors $\mathbf{X}$ and $\mathbf{Y}$ is defined as

$$A(\mathbf{X}, \mathbf{Y}) = \frac{\text{cov}(\mathbf{X}, \mathbf{Y})}{\text{cov}(\mathbf{X}, \mathbf{X})^{1/2} \text{cov}(\mathbf{Y}, \mathbf{Y})^{1/2}}, \tag{F5}$$

where $\text{cov}(\cdot)$ is the covariance; then, the PECA can be calculated by taking $\mathbf{X} = \mathbf{P}_i$ (or $\mathbf{X} = \mathbf{P}_{\text{opt}}$) and $\mathbf{Y} = \mathbf{E}$. The square of the correlation A^2 is the explained error variance.

REFERENCES

Anderson, J. L., 1997: The impact of dynamical constraints on the selection of initial conditions for ensemble predictions: Low-order perfect model results. *Mon. Wea. Rev.*, **125**, 2969–2983, https://doi.org/10.1175/1520-0493(1997)125<2969:TIODCO>2.0.CO;2.

Bai, Y. L., X. Li, and C. L. Huang, 2013: Handling error propagation in sequential data assimilation using an evolutionary strategy. *Adv. Atmos. Sci.*, **30**, 1096–1105, https://doi.org/10.1007/s00376-012-2115-7.

Barkmeijer, J., T. Iversen, and T. N. Palmer, 2003: Forcing singular vectors and other sensitive model structures. *Quart. J. Roy. Meteor. Soc.*, **129**, 2401–2423, https://doi.org/10.1256/qj.02.126.

Basnarkov, L., and L. Kocarev, 2012: Forecast improvement in Lorenz 96 system. *Nonlinear Processes Geophys.*, **19**, 569–575, https://doi.org/10.5194/npg-19-569-2012.

Berner, J., G. J. Shutts, M. Leutbecher, and T. N. Palmer, 2009: A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system. *J. Atmos. Sci.*, **66**, 603–626, https://doi.org/10.1175/2008JAS2677.1.

Birgin, E. G., J. M. Martinez, and M. Raydan, 2000: Non-monotone spectral projected gradient methods on convex sets. *SIAM J. Optim.*, **10**, 1196–1211, https://doi.org/10.1137/S1052623497330963.

Bowler, N. E., 2006: Comparison of error breeding, singular vectors, random perturbations and ensemble Kalman filter perturbation strategies on a simple model. *Tellus*, **58A**, 538–548, https://doi.org/10.1111/j.1600-0870.2006.00197.x.

Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.*, **78**, 1–3, https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2.

Buizza, R., 2019: Introduction to the special issue on "25 years of ensemble forecasting". *Quart. J. Roy. Meteor. Soc.*, **145** (Suppl. 1), 1–11, https://doi.org/10.1002/qj.3370.

——, and T. N. Palmer, 1995: The singular-vector structure of the atmospheric global circulation. *J. Atmos. Sci.*, **52**, 1434–1456, https://doi.org/10.1175/1520-0469(1995)052<1434:TSVSOT>2.0.CO;2.

——, M. Miller, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887–2908, https://doi.org/10.1002/qj.49712556006.

——, P. L. Houtekamer, Z. Toth, G. Pellerin, M. Z. Wei, and Y. J. Zhu, 2005: A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Mon. Wea. Rev.*, **133**, 1076–1097, https://doi.org/10.1175/MWR2905.1.

Candille, G., and O. Talagrand, 2005: Evaluation of probabilistic prediction systems for a scalar variable. *Quart. J. Roy. Meteor. Soc.*, **131**, 2131–2150, https://doi.org/10.1256/qj.04.71.

Demaeyer, J., S. G. Penny, and S. Vannitsem, 2022: Identifying efficient ensemble perturbations for initializing subseasonal-to-seasonal prediction. *J. Adv. Model. Earth Syst.*, **14**, e2021MS002828, https://doi.org/10.1029/2021MS002828.

Descamps, L., and O. Talagrand, 2007: On some aspects of the definition of initial conditions for ensemble prediction. *Mon. Wea. Rev.*, **135**, 3260–3272, https://doi.org/10.1175/MWR3452.1.

Du, J., and Coauthors, 2018: Ensemble methods for meteorological predictions. *Handbook of Hydrometeorological Ensemble Forecasting*, Q. Duan et al., Eds., Springer, 1–52.

Duan, W. S., and F. F. Zhou, 2013: Non-linear forcing singular vector of a two-dimensional quasi-geostrophic model. *Tellus*, **65A**, 18452, https://doi.org/10.3402/tellusa.v65i0.18452.

——, and P. Zhao, 2014: Revealing the most disturbing tendency error of Zebiak–Cane model associated with El Niño predictions by nonlinear forcing singular vector approach. *Climate Dyn.*, **44**, 2351–2367, https://doi.org/10.1007/s00382-014-2369-0.

——, and Z. H. Huo, 2016: An approach to generating mutually independent initial perturbations for ensemble forecasts: Orthogonal conditional nonlinear optimal perturbations. *J. Atmos. Sci.*, **73**, 997–1014, https://doi.org/10.1175/JAS-D-15-0138.1.

——, B. Tian, and H. Xu, 2013: Simulations of two types of El Niño events by an optimal forcing vector approach. *Climate Dyn.*, **43**, 1677–1692, https://doi.org/10.1007/s00382-013-1993-4.

Epstein, E. S., 1969: Stochastic dynamic prediction. *Tellus*, **21**, 739–759, https://doi.org/10.3402/tellusa.v21i6.10143.

Feng, F., and W. S. Duan, 2013: The role of constant optimal forcing in correcting forecast models. *Sci. China Earth Sci.*, **56**, 434–443, https://doi.org/10.1007/s11430-012-4568-z.

Feng, J., R. Q. Ding, D. Q. Liu, and J. P. Li, 2014: The application of nonlinear local Lyapunov vectors to ensemble predictions in Lorenz systems. *J. Atmos. Sci.*, **71**, 3554–3567, https://doi.org/10.1175/JAS-D-13-0270.1.

——, ——, J. P. Li, and D. Q. Liu, 2016: Comparison of nonlinear local Lyapunov vectors with bred vectors, random perturbations

and ensemble transform Kalman filter strategies in a barotropic model. *Adv. Atmos. Sci.*, **33**, 1036–1046, https://doi.org/10.1007/s00376-016-6003-4.

——, J. Li, R. Ding, and Z. Toth, 2018: Comparison of nonlinear local Lyapunov vectors and bred vectors in estimating the spatial distribution of error growth. *J. Atmos. Sci.*, **75**, 1073–1087, https://doi.org/10.1175/JAS-D-17-0266.1.

Fortin, V., M. Abaza, F. Anctil, and R. Turcolle, 2014: Why should ensemble spread match the RMSE of the ensemble mean? *J. Hydrometeor.*, **15**, 1708–1713, https://doi.org/10.1175/JHM-D-14-0008.1.

Grudzien, C., M. Bocquet, and A. Carrassi, 2020: On the numerical integration of the Lorenz-96 model, with scalar additive noise, for benchmark twin experiments. *Geosci. Model Dev.*, **13**, 1903–1924, https://doi.org/10.5194/gmd-13-1903-2020.

Hamill, T. M., C. Snyder, and R. E. Morss, 2000: A comparison of probabilistic forecasts from bred, singular-vector, and perturbed observation ensembles. *Mon. Wea. Rev.*, **128**, 1835–1851, https://doi.org/10.1175/1520-0493(2000)128<1835:ACOPFF>2.0.CO;2.

Hopson, T. M., 2014: Assessing the ensemble spread–error relationship. *Mon. Wea. Rev.*, **142**, 1125–1142, https://doi.org/10.1175/MWR-D-12-00111.1.

Hou, D., Z. Toth, and Y. Zhu, 2006: A stochastic parameterization scheme within NCEP global ensemble forecast system. *18th Conf. on Probability and Statistics in the Atmospheric Sciences*, Atlanta, GA, Amer. Meteor. Soc., 4.5, https://ams.confex.com/ams/Annual2006/techprogram/paper_101401.htm.

——, ——, and ——, 2008: The impact of a stochastic perturbation scheme on hurricane prognosis in NCEP global ensemble forecast system. *23rd Conf. on Weather Analysis and Forecasting/19th Conf. on Numerical Weather Prediction*, New Orleans, LA, Amer. Meteor. Soc., 4A.6, https://ams.confex.com/ams/23WAF19NWP/techprogram/paper_154340.htm.

——, ——, ——, W. Yang, and R. Wobus, 2010: A stochastic total tendency perturbation scheme representing model-related uncertainties in the NCEP global ensemble forecast system. NOAA/NCEP/EMC, Camp Springs, MD, 50 pp., http://www.emc.ncep.noaa.gov/gmb/yzhu/gif/pub/Manuscript_STTP_Tellus_A_HOU-1.pdf.

Hunt, B. R., and Coauthors, 2004: Four-dimensional ensemble Kalman filtering. *Tellus*, **56A**, 273–277, https://doi.org/10.3402/tellusa.v56i4.14424.

Huo, Z. H., and W. S. Duan, 2019: The application of the orthogonal conditional nonlinear optimal perturbations method to typhoon track ensemble forecasts. *Sci. China Earth Sci.*, **62**, 376–388, https://doi.org/10.1007/s11430-018-9248-9.

——, ——, and F. F. Zhou, 2019: Ensemble forecasts of tropical cyclone track with orthogonal conditional nonlinear optimal perturbations. *Adv. Atmos. Sci.*, **36**, 231–247, https://doi.org/10.1007/s00376-018-8001-1.

Khare, S. P., and J. L. Anderson, 2006: An examination of ensemble filter based adaptive observation methodologies. *Tellus*, **58A**, 179–195, https://doi.org/10.1111/j.1600-0870.2006.00163.x.

Kingma, D. P., and J. Ba, 2014: Adam: A method for stochastic optimization. *Third Int. Conf. for Learning Representations*, San Diego, CA, ICLR, 1–15, https://arxiv.org/abs/1412.6980.

Krizhevsky, A., I. Sutskever, and G. E. Hinton, 2017: ImageNet classification with deep convolutional neural networks. *Commun. ACM*, **60**, 84–90, https://doi.org/10.1145/3065386.

LeCun, Y., Y. Bengio, and G. Hinton, 2015: Deep learning. *Nature*, **521**, 436–444, https://doi.org/10.1038/nature14539.

Leith, C. E., 1974: Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev.*, **102**, 409–418, https://doi.org/10.1175/1520-0493(1974)102<0409:TSOMCF>2.0.CO;2.

Leutbecher, M., and T. N. Palmer, 2008: Ensemble forecasting. *J. Comput. Phys.*, **227**, 3515–3539, https://doi.org/10.1016/j.jcp.2007.02.014.

Lorenz, E. N., 1996: Predicability: A problem partly Lorenz96. *Proc. Workshop on Predictability*, ECMWF, Reading, United Kingdom, 1–18, https://www.ecmwf.int/node/10829.

——, and K. A. Emanuel, 1998: Optimal sites for supplementary weather observations: Simulation with a small model. *J. Atmos. Sci.*, **55**, 399–414, https://doi.org/10.1175/1520-0469(1998)055<0399:OSFSWO>2.0.CO;2.

Maas, A. L., A. Y. Hannun, and A. Y. Ng, 2013: Rectifier non-linearities improve neural network acoustic models. *Proc. ICML*, Atlanta, GA, PMLR, 1–3, http://ai.stanford.edu/~amaas/papers/relu_hybrid_icml2013_final.pdf.

Mason, I. B., 1982: A model for assessment of weather forecasts. *Aust. Meteor. Mag.*, **30**, 291–303.

Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, 1996: The ECMWF ensemble prediction system: Methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73–119, https://doi.org/10.1002/qj.49712252905.

Mu, M., and Z. N. Jiang, 2008: A new approach to the generation of initial perturbations for ensemble prediction: Conditional nonlinear optimal perturbation. *Chin. Sci. Bull.*, **53**, 2062–2068, https://doi.org/10.1007/s11434-008-0272-y.

——, W. S. Duan, and B. Wang, 2003: Conditional nonlinear optimal perturbation and its applications. *Nonlinear Processes Geophys.*, **10**, 493–501, https://doi.org/10.5194/npg-10-493-2003.

Mureau, R., F. Molteni, and T. N. Palmer, 1993: Ensemble prediction using dynamically conditioned perturbations. *Quart. J. Roy. Meteor. Soc.*, **119**, 299–323, https://doi.org/10.1002/qj.49711951005.

Nicolis, C., 2003: Dynamics of model error: Some generic features. *J. Atmos. Sci.*, **60**, 2208–2218, https://doi.org/10.1175/1520-0469(2003)060<2208:DOMESG>2.0.CO;2.

——, R. A. P. Perdigao, and S. Vannitsem, 2009: Dynamics of prediction errors under the combined effect of initial condition and model errors. *J. Atmos. Sci.*, **66**, 766–778, https://doi.org/10.1175/2008JAS2781.1.

Orrell, D., 2002: Role of the metric in forecast error growth: How chaotic is the weather? *Tellus*, **54**, 350–362, https://doi.org/10.3402/tellusa.v54i4.12159.

——, 2005: Ensemble forecasting in a system with model error. *J. Atmos. Sci.*, **62**, 1652–1659, https://doi.org/10.1175/JAS3406.1.

——, L. Smith, J. Barkmeijer, and T. N. Palmer, 2001: Model error in weather forecasting. *Nonlinear Processes Geophys.*, **8**, 357–371, https://doi.org/10.5194/npg-8-357-2001.

Palmer, T. N., 2000: Predicting uncertainty in forecasts of weather and climate. *Rep. Prog. Phys.*, **63**, 71–116, https://doi.org/10.1088/0034-4885/63/2/201.

——, R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G. J. Shutts, M. Steinheimer, and A. Weisheimer, 2009: Stochastic parametrization and model uncertainty. ECMWF Research Department Tech. Memo. 598, ECMWF, Reading, United Kingdom, 42 pp., https://doi.org/10.21957/ps8gbwbdv.

Rasp, S., and S. Lerch, 2018: Neural networks for postprocessing ensemble weather forecasts. *Mon. Wea. Rev.*, **146**, 3885–3900, https://doi.org/10.1175/MWR-D-18-0187.1.

Revelli, J. A., M. A. Rodriguez, and H. S. Wio, 2010: The use of rank histograms and MVL diagrams to characterize

ensemble evolution in weather forecasting. *Adv. Atmos. Sci.*, **27**, 1425–1437, https://doi.org/10.1007/s00376-009-9153-6.

Salman, A. G., B. Kanigoro, and Y. Heryadi, 2015: Weather forecasting using deep learning techniques. *2015 Int. Conf. on Advanced Computer Science and Information Systems* (ICACSIS), Depok, Indonesia, IEEE, https://doi.org/10.1109/ICACSIS.2015.7415154.

Scher, S., 2018: Toward data-driven weather and climate forecasting: Approximating a simple general circulation model with deep learning. *Geophys. Res. Lett.*, **45**, 12 616–12 622, https://doi.org/10.1029/2018GL080704.

——, and G. Messori, 2018: Predicting weather forecast uncertainty with machine learning. *Quart. J. Roy. Meteor. Soc.*, **144**, 2830–2841, https://doi.org/10.1002/qj.3410.

Scheuerer, M., M. B. Switanek, R. P. Worsnop, and T. M. Hamill, 2020: Using artificial neural networks for generating probabilistic subseasonal precipitation forecasts over California. *Mon. Wea. Rev.*, **148**, 3489–3506, https://doi.org/10.1175/MWR-D-20-0096.1.

Shi, X. J., Z. R. Chen, H. Wang, D. Y. Yeung, W. K. Wong, and W. C. Woo, 2015: Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Proc. 28th Int. Conf. on Neural Information Processing Systems*, Cambridge, MA, MIT Press, https://dl.acm.org/doi/10.5555/2969239.2969329.

Shutts, G., 2005: A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quart. J. Roy. Meteor. Soc.*, **131**, 3079–3102, https://doi.org/10.1256/qj.04.106.

Talagrand, O., R. Vautard, and B. Strauss, 1997: Evaluation of probabilistic prediction systems. *ECMWF Workshop on Predictability*, Reading, United Kingdom, ECMWF, 1–25.

Tao, L. J., and W. S. Duan, 2019: Using a nonlinear forcing singular vector approach to reduce model error effects in ENSO forecasting. *Wea. Forecasting*, **34**, 1321–1342, https://doi.org/10.1175/WAF-D-19-0050.1.

——, ——, and S. Vannitsem, 2020: Improving forecasts of El Nino diversity: A nonlinear forcing singular vector approach. *Climate Dyn.*, **55**, 739–754, https://doi.org/10.1007/s00382-020-05292-5.

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317–2330, https://doi.org/10.1175/1520-0477(1993)074<2317:EFANTG>2.0.CO;2.

——, and ——, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297–3319, https://doi.org/10.1175/1520-0493(1997)125<3297:EFANAT>2.0.CO;2.

Van Kekem, D. L., 2018: *Dynamics of the Lorenz-96 Model: Bifurcations, Symmetries and Waves.* University of Groningen, 193 pp.

Vannitsem, S., 2006: The role of scales in the dynamics of parameterization uncertainties. *J. Atmos. Sci.*, **63**, 1659–1671, https://doi.org/10.1175/JAS3708.1.

——, 2014: Stochastic modelling and predictability: Analysis of a low-order coupled ocean-atmosphere model. *Philos. Trans. Roy. Soc.*, A372, 20130282, https://doi.org/10.1098/rsta.2013.0282.

——, and Z. Toth, 2002: Short-term dynamics of model errors. *J. Atmos. Sci.*, **59**, 2594–2604, https://doi.org/10.1175/1520-0469(2002)059<2594:STDOME>2.0.CO;2.

——, and W. Duan, 2020: On the use of near-neutral backward Lyapunov vectors to get reliable ensemble forecasts in coupled ocean–atmosphere systems. *Climate Dyn.*, **55**, 1125–1139, https://doi.org/10.1007/s00382-020-05313-3.

——, and Coauthors, 2021: Statistical postprocessing for weather forecasts: Review, challenges, and avenues in a big data world. *Bull. Amer. Meteor. Soc.*, **102**, E681–E699, https://doi.org/10.1175/BAMS-D-19-0308.1.

Veldkamp, S., K. Whan, S. Dirksen, and M. Schmeits, 2021: Statistical postprocessing of wind speed forecasts using convolutional neural networks. *Mon. Wea. Rev.*, **149**, 1141–1152, https://doi.org/10.1175/MWR-D-20-0219.1.

Wang, J., 2021: *Study on the Moist Singular Vectors and Nonlinear Initial Perturbation in GRAPES-GEPS.* University of Chinese Academy of Sciences, 129 pp.

Wang, Q., M. Mu, and G. Sun, 2020: A useful approach to sensitivity and predictability studies in geophysical fluid dynamics: Conditional non-linear optimal perturbation. *Natl. Sci. Rev.*, **7**, 214–223, https://doi.org/10.1093/nsr/nwz039.

Wang, Y., and W. Duan, 2019: Influences of initial perturbation amplitudes and ensemble sizes on the ensemble forecasts made by CNOPs method. *Chin. J. Atmos. Sci.*, **43**, 919–933, http://duanws.lasg.ac.cn/ueditor/php/upload/file/20191009/1570608319337097.pdf.

Wei, M. Z., and Z. Toth, 2003: A new measure of ensemble performance: Perturbation versus error correlation analysis (PECA). *Mon. Wea. Rev.*, **131**, 1549–1565, https://doi.org/10.1175//1520-0493(2003)131<1549:ANMOEP>2.0.CO;2.

——, ——, R. Wobus, and Y. J. Zhu, 2008: Initial perturbations based on the ensemble transform (ET) technique in the NCEP global operational forecast system. *Tellus*, **60A**, 62–79, https://doi.org/10.1111/j.1600-0870.2007.00273.x.

Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924, https://doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2.

Zhou, Q., L. Chen, W. Duan, X. Wang, Z. Zu, X. Li, S. Zhang, and Y. Zhang, 2021: Using conditional nonlinear optimal perturbation to generate initial perturbations in ENSO ensemble forecasts. *Wea. Forecasting*, **36**, 2101–2111, https://doi.org/10.1175/WAF-D-21-0063.1.