Received: 1 July 2024 | Revised: 21 October 2024 | Accepted: 6 November 2024

DOI: 10.1002/qj.4913

Quarterly Journal of the Royal Meteorological Society RMetS

RESEARCH ARTICLE

# Optimal distributions of growing-type initial perturbations for ensemble forecasts: Theory and application in the Lorenz-96 model

**Chaopeng Ji**[1,2,3] | **Bo Qin**[1,2] | **Mu Mu**[1,2,3,4] | **Junjie Ma**[5] | **Xianghui Fang**[1,2,3] | **Jie Feng**[1,2,6] | **Qiang Wang**[7,8]

[1]Department of Atmospheric and Oceanic Sciences and Institute of Atmospheric Sciences, Fudan University, Shanghai, China

[2]Shanghai Key Laboratory of Ocean–Land–Atmosphere Boundary Dynamics and Climate Change, Fudan University, Shanghai, China

[3]Shanghai Frontiers Science Center of Atmosphere-Ocean Interaction, Fudan University, Shanghai, China

[4]Key Laboratory of Polar Atmosphere-ocean-ice System for Weather and Climate, Ministry of Education, Fudan University, Shanghai, China

[5]School of Mathematics, North University of China, Taiyuan, China

[6]Shanghai Academy of Artificial Intelligence for Science, Shanghai, China

[7]Key Laboratory of Marine Hazards Forecasting, Ministry of Natural Resources, Hohai University, Nanjing, China

[8]College of Oceanography, Hohai University, Nanjing, China

**Correspondence**
Bo Qin, Department of Atmospheric and Oceanic Sciences and Institute of Atmospheric Sciences, Fudan University, Shanghai, China.
Email: boqin@fudan.edu.cn

**Funding information**
National Natural Science Foundation of China, Grant/Award Numbers: 42288101, 42405147; China National Postdoctoral Program for Innovative Talents, Grant/Award Number: BX20230071; TianHe Qingsuo Project - special fund project in the field of climate, meteorology, and ocean, National Supercomputer Center in Tianjin, China

## Abstract

Ensemble forecasts are frequently utilized to assess the uncertainties of prediction systems. There is a consensus that generating initial perturbations with specific structures is more conducive to characterizing the growth dynamics of analysis errors, demonstrating higher forecast skills. However, the widely used methods, such as linear singular vectors (SVs) and orthogonal conditional nonlinear optimal perturbations (O-CNOPs) exhibit strong linear-assumption dependence and the overestimation of growing properties for analysis errors, respectively, severely limiting the ability to capture analysis errors and forecast skills. To tackle these challenges, a theoretical framework is established to solve the optimal distribution of the nonlinear growing-type initial perturbations by variational inference (VI) incorporated with the concept of CNOPs, marked as VI-CNOPs. As the distribution is obtained, diverse initial perturbations for ensemble forecasts can be easily sampled in it. To evaluate the reliability of VI-CNOPs, a series of ensemble forecast experiments are then conducted using the Lorenz-96 model. We compare the deterministic and probabilistic forecast skills of VI-CNOPs, O-CNOPs, and SVs under various optimization durations. The results reveal that, as the optimization durations extend, the forecast skills of VI-CNOPs progressively improve, consistently outperforming O-CNOPs and SVs. This trend remains consistent across various forecast lead times. Further analysis reveals that VI-CNOPs more effectively capture the covariance matrix of analysis errors, aligning with the fundamental concept of perturbation generation methods for ensemble forecasts. Moreover, unlike O-CNOPs and SVs, VI-CNOPs do not require the utilization of adjoint and tangent linear models, largely expanding its application. These results indicate the novelty and efficacy of VI-CNOPs for ensemble forecasts.

**KEYWORDS**

conditional nonlinear optimal perturbation, ensemble forecast, optimization, probability distribution, variational inference

†Chaopeng Ji and Bo Qin contributed equally to this study.

# 1 INTRODUCTION

Due to the chaotic nature of the atmospheric system, two states with minor differences at an initial state gradually diverge over time, ultimately evolving into entirely distinct states (Lorenz, 1963a,b). Therefore, a single deterministic forecast merely represents a possible scenario of the future atmospheric state, inherently susceptible to uncertainty (Buizza, 2008, 2019; Leutbecher & Palmer, 2008). To quantify forecast uncertainties and improve forecast skills, the ensemble forecast has arisen. Epstein (1969) elucidated that the methodology of stochastic dynamic prediction holds promise. By leveraging the probability distribution function (PDF) of initial conditions within the "perfect" model scenario, rather than deterministic values, the future PDF of atmospheric or oceanic states is delineated, thereby considering some problems arising from insufficient or inaccurate observations. While this approach might theoretically characterize forecast errors and uncertainties in more detail, numerical weather prediction systems have millions of degrees of freedom, making it computationally infeasible (Leith, 1974). Therefore, based on the concept of acknowledging initial uncertainties, Leith (1974) proposed the Monte Carlo Forecasting (MCF) method, wherein random initial perturbations (IPs) are superposed on the initial analysis field to characterize the analysis errors (AEs) and generate a collection of ensemble members aimed at estimating the PDF of an event. The MCF method facilitates the transition of ensemble forecasts from theoretical research to practical applications. However, random IPs fail to capture the spatially continuous structures of AEs and do not guarantee coherence between the initial conditions and the numerical models (Leith, 1978; Murphy, 1988). Besides, the resulting ensemble spread is also limited, which always requires a large number of samples, consuming huge computational costs.

To address this problem, numerous research endeavors have proposed diverse methodologies aimed at more effectively delineating the characteristics of AEs through innovative IP generation techniques. Besides, with an in-depth comprehension of the evolution of AEs, it becomes evident that the initial AEs include both rapid-growing-type and high-energy elements, alongside slower or non-growing-type with lower-energy elements (Toth & Kalnay, 1993). In scenarios where a forecast incorporates both types of AEs, slower or non-growing-type AEs only manifest a projection toward error growing in later stages, while growing-type AEs rapidly amplify and assert dominance over the entire error field in the early stage (Mureau et al., 1993; Toth & Kalnay, 1997). This underscores the greater significance of growing-type AEs compared to non-growing-type AEs for forecast accuracy. Consequently, for ensemble forecasts, superposing growing-type IPs on the initial analysis field effectively delineates the growing characteristics of AEs, and amplifies the likelihood of ensemble members deviating from the control forecast to closely approximate and cover the truth state (Toth & Kalnay, 1997).

Derived from this fundamental principle, Toth and Kalnay (1993) developed the breeding vectors (BVs) method, which represent an extension of linear local Lyapunov vectors (LLVs) under nonlinearity (Toth & Kalnay, 1993, 1997). BVs have been applied to the National Centers for Environmental Prediction (NCEP) operational ensemble forecast system since 1992 and have demonstrated high forecast skills. Similarly, Feng et al. (2014) extended nonlinear LLVs (NLLVs) to ensemble forecasts, employing orthogonal NLLVs to improve the efficacy of BVs in ensemble forecasts (Feng et al., 2016, 2018). Both BVs and NLLVs possess dynamic growth structures that contribute to increasing ensemble spread, thereby enhancing the estimation of forecast uncertainty. In parallel, the European Centre for Medium-Range Weather Forecasts (ECMWF) employs linear SVs (Lorenz, 1965) to construct IPs (Buizza, 1997; Buizza et al., 2008; Gelaro et al., 1998; Lorenz, 1965; Molteni et al., 1996; Mureau et al., 1993) and achieve notable success in constructing ensemble forecasts. The core idea behind SVs is to employ finite-time instability theory in nonlinear dynamics, as well as tangent linear and adjoint models, to determine the direction of the phase space with the largest linear development during the forecast period.

However, SVs have limited capability in describing optimal growing perturbations in nonlinear systems. SVs rely on the IPs being sufficiently small and their development being approximated by a tangent linear model, thus failing to capture the nonlinear characteristics of atmospheric and oceanic variables fully (Anderson, 1997; Gilmour & Smith, 1997). These studies underscore the necessity of considering nonlinear optimal perturbations in atmospheric and oceanic ensemble forecasts.

Recognizing the limitations inherent in the linear theory of SVs, Mu and Duan (2003) introduced the conditional nonlinear optimal perturbation (CNOP), which is an extension of the leading SV (LSV) in the nonlinear regime. Physically, CNOP represents the IP that exhibits the maximum nonlinear evolution under given physical constraints. Leveraging this concept, CNOP can be employed to identify the optimal precursor signals for atmospheric and oceanic high-impact weather and climate events, thereby elucidating the mechanisms behind these events, but can also be studied the fastest-growing initial errors in the forecasting of such events and their associated predictability properties (see the review papers of Wang et al., 2020 and Duan et al., 2023). Considering the effect of nonlinearity in ensemble forecasts, Mu and Jiang (2008) replaced the LSV with CNOP, and proposed the CNOP + SVs method, achieving higher performance than pure SVs in ensemble forecasts with a quasi-geostrophic model. This method has also been successfully adopted for the ensemble forecasts of El Niño-Southern Oscillation (Zhou et al., 2021). Those results underscore the advantage of identifying IPs with rapid nonlinear growing when constructing ensemble forecasts (Jiang et al., 2009).

Furthermore, to consider the multiple nonlinear processes within the development of IPs for ensemble forecasts, Duan and Huo (2016) proposed the orthogonal CNOPs (O-CNOPs) method, which is engineered to generate mutually independent, nonlinear optimal growing IPs in orthogonal subspaces, denoted as 1st-CNOP, 2nd-CNOP, 3rd-CNOP, … , $n$th-CNOP. These can be regarded as the nonlinear extension of orthogonal SVs. Idealized experiments using the Lorenz-96 model illustrate the advantages of O-CNOPs over SVs and CNOP + SVs in ensemble forecast (Duan & Huo, 2016; Huo & Duan, 2019). Furthermore, Huo and Duan (2019) and Zhang et al. (2023) applied O-CNOPs to the fifth-generation mesoscale model (MM5) and the weather research and forecasting (WRF) model, respectively, for constructing ensemble forecasts of typhoon tracks. The results indicate that O-CNOPs exhibit a wider horizontal distribution and offer an improved depiction of forecast uncertainties and skills when compared to SVs and CNOP + SVs, further underscoring the practical effectiveness of O-CNOPs.

While O-CNOPs offer a more comprehensive characterization of the nonlinear growing of AEs compared to SVs and increase ensemble spread more effectively, their role as the predominant IPs poses challenges. For example, O-CNOPs mathematically represent the IPs with the largest growths within each orthogonal perturbation subspace, yet expose a risk of overestimating the growing characteristics of AEs in ensemble forecasts, resulting in considerable deviations of individual members from the truth state and amplification of ensemble mean forecast errors (Wang & Duan, 2019). In summary, SVs and O-CNOPs still have certain limitations in characterizing the nonlinear growing of AEs. A consensus is that the ability of the IPs to more accurately characterize the nonlinear growing of AEs is a necessary condition for a reliable ensemble forecast system. In other words, if the IPs are sampled from the PDF of the growing-type IPs (hereafter denoted as conditional probability distribution $p(\mathbf{z}|\mathbf{x})$, where $\mathbf{x}$ represents a certain initial analysis field and $\mathbf{z}$ represents the corresponding growing-type IPs, showing that the growing-type IPs are related to the given initial analysis field), it is more likely to improve ensemble forecast skills further. To this end, considering that CNOP represents a deterministic IP with the largest nonlinear development under certain constraints, we utilize the concept of CNOP to construct a theoretical framework with a clearly-designed nonlinear optimization problem to obtain $p(\mathbf{z}|\mathbf{x})$.

More precisely, we use a known distribution (marked as $q_\theta(\mathbf{z})$, $\theta$ is the parameter of the distribution) with the given analytical expression, such as Gaussian distribution in this paper, to approximate $p(\mathbf{z}|\mathbf{x})$ (i.e., optimizing to minimize the divergence between $q_\theta(\mathbf{z})$ and $p(\mathbf{z}|\mathbf{x})$). The variational inference (VI) approach is applied for this optimization, which is a common and effective way for recognizing complex or even black-box distributions in statistics and machine learning, particularly posterior distributions in Bayesian inference (Jordan et al., 1999). We mark our proposed framework as VI-CNOPs and illustrate its effectiveness on the Lorenz-96 model with varied metrics. VI-CNOPs methods aim to delineate the growth characteristics of initial AEs more efficiently, thereby reducing forecast errors and providing a more appropriate spread for forecast errors estimation (see Section 3 for more details).

The rest of this paper is organized as follows. The next section describes the Lorenz-96 model employed in this study. The following section introduces a method to compute the PDF of growing-type IPs. Subsequent sections outline the experimental design, followed by a presentation of the results of the ensemble forecast experiments. A summary and discussion are provided in the final section.

## 2 | LORENZ-96 MODEL

In this paper, the experiments are conducted within the framework of the Lorenz-96 model, which is a conceptual dynamical model proposed by Lorenz (1996). Although the model is relatively simple, it captures the key features of the atmospheric system, and its dynamical properties are consistent with those of actual numerical forecast models (Lorenz & Emanuel, 1998). It has been widely utilized for evaluating the effectiveness of new theories and methods in data assimilation (Amemiya et al., 2023; Brajard et al., 2020; Stanley et al., 2021; Terasaki & Miyoshi, 2014), adaptive observations (Khare & Anderson, 2006; Lorenz & Emanuel, 1998), and ensemble forecasts (Duan et al., 2022; Duan & Huo, 2016; Feng et al., 2014; Revelli et al., 2010).

The model is governed by the dynamical equations:

$$\frac{dX_j}{dt} = \left(X_{j+1} - X_{j-2}\right)X_{j-1} - X_j + F, \tag{1}$$

where $j = 1, \ldots, m, m \in \mathbb{N}$ is the dimension of the state space. The variables in the system are all dimensionless. This system satisfies the cyclic boundary conditions, that is, $X_{-1} = X_{m-1}, X_0 = X_m, X_{m+1} = X_1$. The state variables $X_j$ can be interpreted as non-dimensional meteorological variables uniformly distributed on the same latitude circle (e.g., temperature, pressure, vorticity, or gravitational potential). The right-hand side of Equation (1) comprises nonlinear terms, linear terms, and constant terms denoted by $F$, which represent the advection, internal dissipation, and external forcing of the atmosphere, respectively.

The nonlinear behavior of this system varies with the total number of variables ($m$), and the external forcing $F$. When $m = 40$ and $F = 8$, the model exhibits chaos. The configured Lorenz-96 model undergoes numerical integration utilizing a fourth order Runge–Kutta scheme with a nondimensional time-step of 0.05 units (roughly 6 h in the real world). Furthermore, upon reaching its attractor, the mean and standard deviation of variable $X_j$ are approximately equal to 2.3 and 3.6, respectively. The same configuration is also employed for subsequent ensemble forecast experiments.

## 3 | MATHEMATICAL THEORETICAL FRAMEWORK FOR OBTAINING THE OPTIMAL PDF OF GROWING-TYPE INITIAL PERTURBATIONS

As stated in the introduction, the growing-type AEs exert the most significant influence on forecast results. Therefore, the IPs superposed on the analysis field should effectively capture the characteristics of the growing components of AEs, thereby sampling the possible development of AEs to the fullest extent and improving forecast skills. This section first establishes a theoretical framework with a clearly designed nonlinear optimization problem for solving the PDF of growing-type IPs, followed by the solution approach for this optimization problem, namely variational inference.

As mentioned earlier, CNOP represents the dynamically conditioned fastest-growing initial perturbation under a certain constraint. However, since CNOP is a deterministic IP given the initial reference state, it cannot meet the requirements of ensemble forecasts which are initialized from a set of equally distributed perturbed states. To deal with this problem, we aim to solve the PDF of growing-type IPs based on the fundamental concept of CNOP, and then randomly sample from the PDF to generate the IPs for ensemble forecast.

According to the definition of CNOP (see Appendix B for more details), a natural idea is to establish the following optimization problem:

$$J(p(\mathbf{z}|\mathbf{x})) = \max_{\|\mathbf{z}\|_i \le \delta} \left\{ \mathbb{E}_p \left( \|M_\tau(\mathbf{x}+\mathbf{z}) - M_\tau(\mathbf{x})\|_f \right) \right\}, \quad (2)$$

where $\mathbb{E}_p$ denotes the mathematical expectation with respect to $p$; In mathematical terms, Equation (2) represents that the mathematical expectation of the difference, between all possible initial perturbations $\mathbf{z}$ sampled from $p(\mathbf{z}|\mathbf{x})$, superimposed on a given initial condition $\mathbf{x}$, and the reference state (without perturbations), achieves its maximum value. In other words, as long as an IP exhibits significant development, it is an element of $p(\mathbf{z}|\mathbf{x})$. However, solving this problem in practice is challenging due to its fundamental aim of characterizing the temporal evolution of the distribution of growing-type IPs and the excessive computational complexity involved.

Therefore, our objective is to identify an optimal surrogate PDF $q_\theta^*(\mathbf{z})$, which is a simple distribution accompanied by an analytical expression and parameter $\theta$ from a known-form probability distribution family $Q$, that is, $q_\theta^*(\mathbf{z})$ is the distribution closest to $p(\mathbf{z}|\mathbf{x})$ in the entire family $Q$, which means that the distance between $q_\theta(\mathbf{z})$ and $p(\mathbf{z}|\mathbf{x})$ is minimized, under a given metric. In this paper, we select $Q$ as Gaussian distributions, which is a commonly used distribution family in VI methods (Damianou et al., 2016; Tomczak et al., 2018; Tran et al., 2016). That is, we seek the optimal Gaussian distribution $q^*(\mathbf{z}; \boldsymbol{\mu}, \boldsymbol{\sigma})$ to approximate the unknown distribution $p(\mathbf{z}|\mathbf{x})$, where $\boldsymbol{\mu}$,$\boldsymbol{\sigma}$ represent the mean and standard deviation (i.e., the variational parameters $\theta$), respectively. Besides, Kullback–Leibler (KL) divergence ($D_{\mathrm{KL}}$; Kullback & Leibler, 1951), is used for measuring the difference between two PDFs. Then, we can get growing-type IPs for ensemble forecast by sample from $q_\theta^*(\mathbf{z})$. The mathematization is as follows.

$$\begin{aligned} J\left(q_\theta^*(\mathbf{z})\right) &= \min_{q_\theta(\mathbf{z})\in Q} D_{KL}(q_\theta(\mathbf{z}) \| p(\mathbf{z}|\mathbf{x})) \\ &= \min_{q_\theta(\mathbf{z})\in Q} - \int_{\mathbf{z}} q_\theta(\mathbf{z}) \log \left[ \frac{p(\mathbf{z}|\mathbf{x})}{q_\theta(\mathbf{z})} \right] d\mathbf{z}. \end{aligned} \quad (3)$$

By solving the optimization problem determined by Equation (3), we can obtain the optimal approximation $q_\theta^*(\mathbf{z})$ for $p(\mathbf{z}|\mathbf{x})$.

Solving the optimal $q_\theta^*(\mathbf{z})$ to approximate the unknow $p(\mathbf{z}|\mathbf{x})$ in Equation (3) relies on VI, which is a method employed in Bayesian statistics and machine learning to approximate complex or even black-box probability distributions (Jordan et al., 1999). VI approximates the true posterior distribution with a simpler distribution that is easier to compute (as shown in Figure 1a). It is highly valued for its scalability and adaptability in handling large datasets and complex models (Blei et al., 2017; Cheng et al., 2023; Ganguly & Earp, 2021; Sjölund, 2023). To apply this method to solve the optimization problem in Equation (3), which involves an unknown distribution $p(\mathbf{z}|\mathbf{x})$, it is necessary to reformulate this problem into an equivalent form.

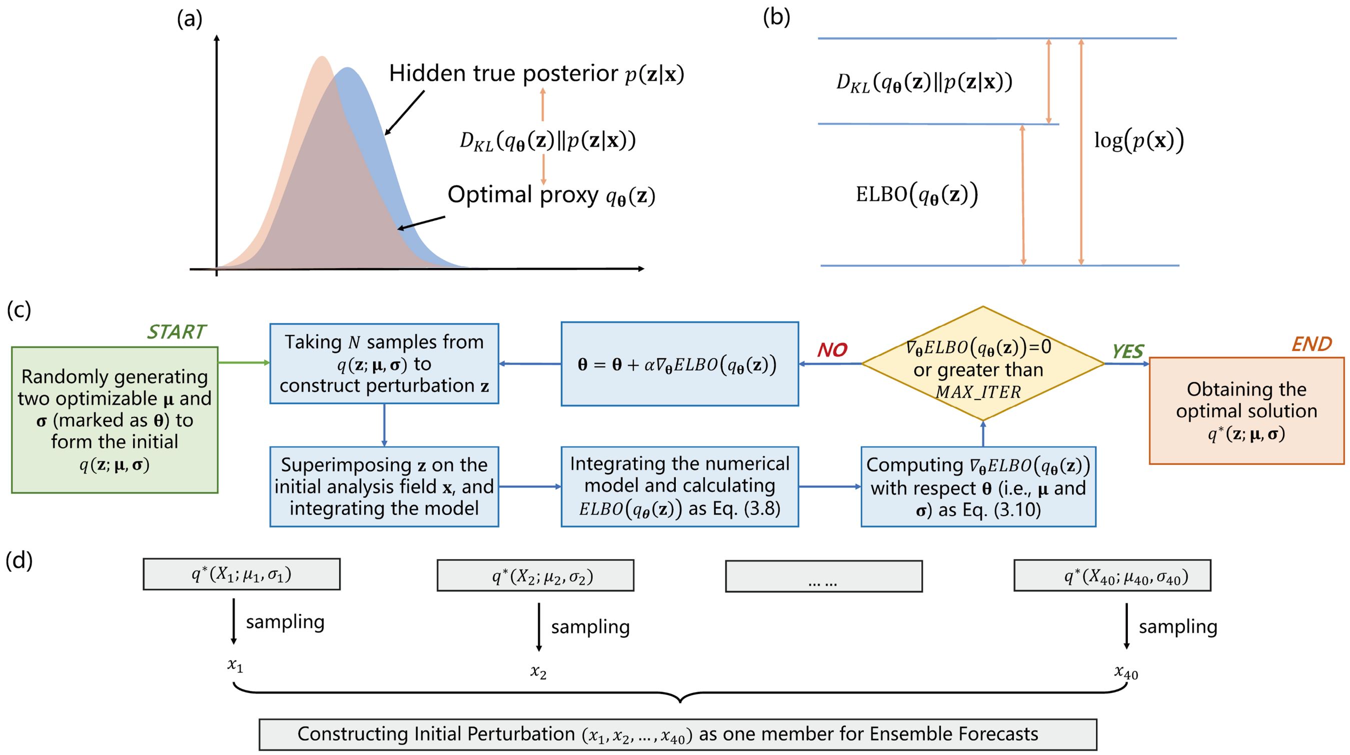


**FIGURE 1** Conceptual diagram illustrating the (a) variational inference, an optimal probability distribution $q_\theta(\mathbf{z})$ is sought to approximate an unknown probability distribution $p(\mathbf{z}|\mathbf{x})$ by minimizing the Kullback–Leibler (KL) divergence between them; (b) the equivalence of minimizing KL divergence and maximizing evidence lower bound (ELBO); (c) algorithm flow chart for solving $q_\theta(\mathbf{z})$. $N$ is the sample number for the estimation of objective function and derivation, *MAX*_ITER and $\alpha$ represent the configured maximum iteration step and learning rate, respectively; and (d) sampling from the optimal distribution to form the initial perturbations (IPs) for ensemble forecasts. [Colour figure can be viewed at wileyonlinelibrary.com]

Firstly, expand the $D_{\mathrm{KL}}$ term as follows:

$$\begin{aligned} D_{\mathrm{KL}}(q_\theta(\mathbf{z})\|p(\mathbf{z}|\mathbf{x})) &= -\int_{\mathbf{z}} q_\theta(\mathbf{z})\log\left[\frac{p(\mathbf{z}|\mathbf{x})}{q_\theta(\mathbf{z})}\right]d\mathbf{z} \\ &= \int_{\mathbf{z}} q_\theta(\mathbf{z})\log(q_{\boldsymbol\theta}(\mathbf{z}))d\mathbf{z} \\ &\quad - \int_{\mathbf{z}} q_\theta(\mathbf{z})\log(p(\mathbf{z}|\mathbf{x}))d\mathbf{z}, \end{aligned} \tag{4}$$

where the first term on the right-hand side involves the determination of the mathematical expectation with respect to $q_\theta(\mathbf{z})$, that is, $\int_{\mathbf{z}} q_\theta(\mathbf{z})f(\mathbf{z},\cdot)\,d\mathbf{z} = \mathbb{E}_q[f(\mathbf{z},\cdot)]$. Then, we have,

$$\begin{aligned} D_{\mathrm{KL}}(q_\theta(\mathbf{z})\|p(\mathbf{z}|\mathbf{x})) &= \int_{\mathbf{z}} q_\theta(\mathbf{z})\log(q_{\boldsymbol\theta}(\mathbf{z}))\mathbf{dz} \\ &\quad - \int_{\mathbf{z}} q_\theta(\mathbf{z})\log(p(\mathbf{z}|\mathbf{x}))\mathbf{dz} \\ &= \mathbb{E}_q\left[\log(q_\theta(\mathbf{z}))\right] - \mathbb{E}_q[\log(p(\mathbf{z}|\mathbf{x}))] \\ &= \mathbb{E}_q\left[\log(q_\theta(\mathbf{z}))\right] - \mathbb{E}_q[\log(p(\mathbf{x},\mathbf{z}))] \\ &\quad + \mathbb{E}_q[\log(p(\mathbf{x}))]. \end{aligned} \tag{5}$$

As $\log(p(\mathbf{x}))$ is independent of $q_\theta(\mathbf{z})$, the expectation operator can be directly removed. So,

$$\begin{aligned} D_{\mathrm{KL}}(q_\theta(\mathbf{z})\|p(\mathbf{z}|\mathbf{x})) &= \underbrace{\mathbb{E}_q\left[\log(q_\theta(\mathbf{z}))\right] - \mathbb{E}_q[\log(p(\mathbf{x},\mathbf{z}))]}_{-\mathrm{ELBO}} \\ &\quad + \log(p(\mathbf{x})), \end{aligned} \tag{6}$$

where the first two terms are commonly regarded as the negative evidence lower bound (ELBO), which serves as a lower bound on the log marginal likelihood of the data. $Log(p(\mathbf{x}))$ is a constant as it pertains to statistical information inherent in the given conditions, termed the "evidence." In fact, in numerical models, $\log(p(\mathbf{x}))$ is solely related to the given initial analysis field $\mathbf{x}$, that is, the reference state. Then, we have,

$$\log(p(\mathbf{x})) = \mathrm{ELBO}(q_\theta(\mathbf{z})) + D_{\mathrm{KL}}(q_\theta(\mathbf{z})\|p(\mathbf{z}|\mathbf{x})). \tag{7}$$

Therefore, minimizing the KL divergence is equivalent to maximizing the ELBO with respect to $q_\theta(\mathbf{z})$ (as shown in Figure 1b), that is,

$$J\left(q_\theta^*(\mathbf{z})\right) = \min_{q_\theta(\mathbf{z})\in\mathcal{Q}} D_{\mathrm{KL}}\left( q_\theta(\mathbf{z}) \| \overbrace{p(\mathbf{z}|\mathbf{x})}^{\text{unknown}} \right) = \max_{q_\theta(\mathbf{z})\in\mathcal{Q}} \mathrm{ELBO}(q_\theta(\mathbf{z})). \tag{8}$$

Furthermore, according to the conditional probability formula, the $\mathrm{ELBO}(q_\theta(\mathbf{z}))$ is given by:

$$\begin{aligned}\mathrm{ELBO}(q_\theta(\mathbf{z})) &= \mathbb{E}_q[\log(p(\mathbf{x},\mathbf{z}))] - \mathbb{E}_q\left[\log(q_\theta(\mathbf{z}))\right] \\ &= \mathbb{E}_q[\log(p(\mathbf{x}|\mathbf{z})p(\mathbf{z}))] - \mathbb{E}_q\left[\log(q_\theta(\mathbf{z}))\right] \\ &= \underbrace{\mathbb{E}_q[\log(p(\mathbf{x}|\mathbf{z}))]}_{\text{Reconstruction term}} \\ &\quad \underbrace{+\mathbb{E}_q[\log(p(\mathbf{z}))] - \mathbb{E}_q\left[\log(q_\theta(\mathbf{z}))\right]}_{\text{Regularization terms}},\end{aligned} \tag{9}$$

where the first term on the right-hand side is the reconstruction term. It represents the expected log-likelihood of reconstructing the data $\mathbf{x}$ under the approximated posterior distribution of the latent variables $\mathbf{z}$. A larger value of this term indicates a better fit of the variational distribution $q_\theta(\mathbf{z})$ to the unknown distribution $p(\mathbf{z}|\mathbf{x})$. From a physical perspective, a larger $\mathbb{E}_q[\log(p(\mathbf{x}|\mathbf{z}))]$ implies that the IPs represented by $\mathbf{z}$ more closely matches the given analysis field $\mathbf{x}$, thereby better satisfying the flow-dependent nature of the IPs. During the practical optimization, we use the sampling strategy to approximate this term. That is, we first draw a number of samples from $q_\theta(\mathbf{z})$ and then superpose them into the reference state of numerical model to calculate the logarithm of their average forecast errors.

The second and third terms on the right-hand side are known as regularization. They constrain the difference between the approximate posterior distribution $q_\theta(\mathbf{z})$ and the prior distribution $p(\mathbf{z})$, thereby preventing overfitting. During the practical optimization, we usually assume the prior distribution $p(\mathbf{z})$ as a standard Gaussian distribution $\mathcal{N}(0,1)$, which is a common manner. $\log(p(\mathbf{z}))$ and $\log(q_\theta(\mathbf{z}))$ can be calculated by Equation (10), where $d$ is the dimension of $\mathbf{z}$, $\mu_j$, and $\sigma_j$ $(j = 1, \dots, d)$ are respectively the mean and standard deviation.

$$\log(\phi(\mathbf{z})) = -\frac{1}{2}\left( d\log(2\pi) + \sum_{j=1}^{d}\log\left(\sigma_j^2\right) + \sum_{j=1}^{d}\frac{\left(z_j-\mu_j\right)^2}{\sigma_j^2} \right). \tag{10}$$

Besides, we employ the gradient information to optimize problem Equation (8) by VI, which can obtain the optimal parameters of the $q_\theta(\mathbf{z})$. The gradient of $\mathrm{ELBO}(q_\theta(\mathbf{z}))$ with respect to $\theta$ can be written as:

$$\begin{aligned}&\nabla_\theta \mathrm{ELBO}(q_\theta(\mathbf{z})) \\ &= \nabla_\theta\left(\mathbb{E}_q\left[\log(p(\mathbf{x}|\mathbf{z})) + \log(p(\mathbf{z})) - \log(q_{\boldsymbol{\theta}}(\mathbf{z}))\right]\right) \\ &= \mathbb{E}_q\left(\nabla_\theta \log q_\theta(\mathbf{z})\left(\log(p(\mathbf{x}|\mathbf{z})) + \log(p(\mathbf{z})) - \log(q_\theta(\mathbf{z}))\right)\right).\end{aligned} \tag{11}$$

The log-derivative trick is used to convert the first line to the second line.

Equation (11) indicates there is no necessity for the gradient of the numerical model concerning the perturbations, eliminating the requirement for tangent linear and adjoint models. Based on the above derivation, the problem in Equation (8) can be solved iteratively, and the optimal parameter $\theta$ of $q_\theta(\mathbf{z})$ can be obtained. The flowchart of the optimization process is shown in Figure 1c, where the derivation determines the optimization direction and *MAX_*ITER represents the configured maximum iteration step. It is worth noting that the formulation of the derivation in Equation (11) relies on the calculation of an expectation. In the practical implementation, we use the idea of Monte Carlo to estimate the expectation, where $N$ samples are generated from $q_\theta(\mathbf{z})$ in the current iteration and their averaged derivation can be considered as the gradient of $\mathrm{ELBO}(q_\theta(\mathbf{z}))$. This is also a common manner in VI. We set $N = 20$ in our case study for approximating the stable expectation (the results are roughly consistent with the larger $N$). Even though our proposed theoretical framework combines the fundamental concepts of CNOP and VI, the solving process is quite convenient compared with the previous method that relied on tangent linear and adjoint models.

In the Lorenz-96 model, to prevent $q^*(\mathbf{z};\boldsymbol{\mu},\boldsymbol{\sigma})$ from collapsing to a point and ensure diversity in the IPs sampling, $\boldsymbol{\sigma}$ is set with a lower bound. During the optimization process, $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are continuously adjusted to maximize the objective function in Equation (8), that is, the optimization variables are $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$. For each state variable $X_j$ $(j = 1, \dots, 40)$ in the Lorenz-96 model, we can obtain a $q^*\left(X_j;\mu_j,\sigma_j\right)$. Subsequently, sampling randomly from $q^*\left(X_j;\mu_j,\sigma_j\right)$ enables the composition of IPs for ensemble forecasts as shown in Figure 1d, denoted as VI-CNOPs.

The optimization of the 40 variables $X_j$ is conducted independently. However, one may argue that there is a correlation between variables at different grid points. Importantly, the IP must satisfy Equation (2), which aims to maximize the mathematical expectation of the perturbation development. It is widely accepted that only IPs with a specific spatial structure are prone to exhibit significant nonlinear development. Thus, the optimization process inherently incorporates constraints on the correlation between grid variables.

## 4 | EXPERIMENTAL STRATEGY

Similar to Duan and Huo (2016), we adopt an experimental design. Specifically, after a spin-up run of 4000 time-steps (i.e., 1000 model days), the integration of the Lorenz-96 model extends for additional 146,000-time-steps (i.e., 100 model years), yielding time series data for state variables $X_j$ (j = 1, … , 40). From this time series, we extract $X_j$ at intervals of 1460 time-steps (i.e., one model year) to serve as initial conditions for subsequent integration over 44 time-steps (i.e., 264 model hours, also the longest lead times for the following forecast experiments). These sequential integrations yield 100 "truths," each representing future states being forecasted.

To forecast the truths, we generate the initial analysis fields through the four-dimensional variational data assimilation (4DVAR). The corresponding "observations" are derived by adding random-noise (observational) errors, characterized by a standard normal distribution $\mathcal{N}(0, I)$ into the truths at six-hour intervals (i.e., one time-step). Notably, the standard deviation of these observational errors amounts to 27.7% of the standard deviation of $X_j$, rendering the magnitude of the observational errors deemed acceptable.

During the forecasting process of the truth run, 4DVAR is employed to generate the initial analysis by assimilating observations from both the current and subsequent time-steps. Leveraging this initial analysis field, we integrate the Lorenz-96 model to create a forecast for the truth (i.e., control forecast). The disparity between the initial analysis field and the initial value of the truth constitutes the initial analysis error. Subsequently, the associated forecast errors are gauged by evaluating discrepancies between the forecasts and the truth. In this article, the norm used to measure the AE's amplitude and its development is the L2 norm (see Appendix A).

For each of the selected 100 truths (i.e., cases), the control forecast serves as a reference state to compute SVs, O-CNOPs (see Appendixes A and B for more details), and VI-CNOPs. For simplicity, the mean AEs amplitude size (Equation 12) was chosen as the constraint radius $\delta$ for the analysis of the 100 cases.

$$\text{AEs amplitude} = \|\text{analysis} - \text{truth}\|_{\text{L2}}. \tag{12}$$

We also set the $\delta$ to match the size of each AE corresponding to each case, which does not affect the results below (Figures S1–S6). Then, computations were conducted across various optimization durations (i.e., 48, 96, 144, 192, and 264 hr).

Wang and Duan (2019) demonstrated that optimal ensemble forecast skills in the Lorenz-96 model are typically achieved with an ensemble size ranging from 20 to 40 members. Therefore, to ensure a fair basis for comparison and maintain consistency with Duan and Huo (2016), allowing for comparative analysis among VI-CNOPs, O-CNOPs, and SVs, we randomly sampled 15 VI-CNOPs from $q^*(\mathbf{z}; \boldsymbol{\mu}, \boldsymbol{\sigma})$ (15 O-CNOPs were also calculated) and superposed them to the initial analysis fields (i.e., reference state) in pairs of positive and negative perturbations, which yielded 30 perturbation forecasts. Including the control forecast, 31 forecasts can be obtained for the truth. Furthermore, we retained the first ten normalized SVs (accounting for approximately 91% of the cumulative explained variance) and used them as the basis to randomly linearly combine 15 IPs (Equation (13); Mureau et al., 1993). The perturbation forecast follows the methodology akin to VI-CNOPs and O-CNOPs.

$$\mathbf{P}_{\text{sv}} = \sum_{i=1}^{10} \alpha_i \mathbf{SV}_i, \tag{13}$$

where $\mathbf{P}_{\text{sv}}$ represents the IPs spanned by SVs; $\alpha_i$ are random numbers obeying the standard normal distribution.

It is worth noting that, unlike SVs, which are often used as a linear combination, we superpose VI-CNOPs and O-CNOPs directly onto the initial analysis field as IPs. This distinction arises from the inherent nonlinear dynamics of both VI-CNOPs and O-CNOPs, making their linear combinations mathematically and physically meaningless.

Here, we refer to Duan et al. (2022) to define the following two classifications of events: high-frequency (HF) events with $X_j > 2$; low-frequency (LF) events with $X_j > 5$; and the two events occur with frequencies of 0.50 and 0.17, respectively.

## 5 | RESULTS

In this section, we evaluate the forecast quality of VI-CNOPs, O-CNOPs, and SVs in terms of both deterministic and probabilistic forecasts. The quality of ensemble forecasts is frequently assessed using several metrics including the anomaly correlation coefficient (ACC), root-mean-square error (RMSE), spread, Brier score (BS; Brier, 1950), continuous ranked probability score (CRPS), relative operating characteristic curve area (ROCA), reliability diagram (RD), Talagrand diagram (Talagrand et al., 1997), and perturbation versus error correlation analysis (PECA; Wei & Toth, 2003). For a comprehensive understanding of their calculation methodologies, please refer to Appendix C.

### 5.1 | Ensemble forecast performance

For all 100 cases, we calculated the ensemble mean ACC, RMSE, BS, and ROCA of the three ensemble forecast

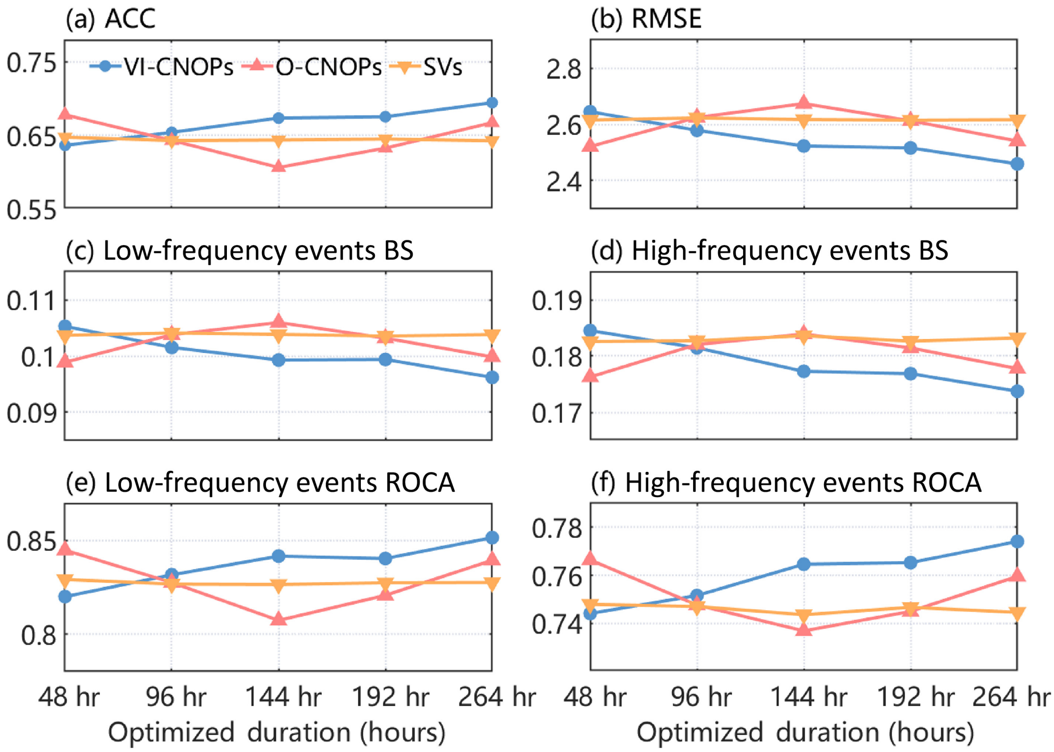


**FIGURE 2** The ensemble mean forecast skills of the three ensemble forecast schemes (variational inference-conditional nonlinear optimal perturbations (VI-CNOPs) in blue, orthogonal conditional nonlinear optimal perturbations [O-CNOPs] in orange, and singular vectors [SVs] in yellow) for various optimization durations (48, 96, 144, 192, and 264 hr; represented on the horizontal axis), averaged over 100 truth runs and all lead times (from 6 to 264 hr), as measured by (a) anomaly correlation coefficient (ACC), (b) root-mean-square error (RMSE), (c,d) Brier score (BS), and (e,f) relative operating characteristic curve area (ROCA). In particular, (c,e) denote the BS and ROCA for low-frequency (LF) binary events, whereas (d,f) denote high-frequency (HF) events. ACC and ROCA are positively oriented, indicating that higher values correspond to better forecasting skills, whereas BS and RMSE are negatively oriented, meaning that smaller values correspond to better forecasting skills. [Colour figure can be viewed at wileyonlinelibrary.com]

schemes, VI-CNOPs, O-CNOPs, and SVs, averaged over all forecast lead times (from 6 to 264 hr) for different optimization durations (48, 96, 144, 192, and 264 hr). As shown in Figure 2, across all assessment metrics, the VI-CNOPs scheme consistently exhibits the highest forecast skills among the three schemes. Moreover, a distinct trend emerges wherein as the optimization duration extends, the forecast skills of VI-CNOPs undergoes a gradual improvement. Nonetheless, in scenarios where the optimization duration is relatively short, the benefits of the VI-CNOPs scheme exhibit less prominence, resulting in relatively poor forecast performance compared to other schemes. The potential reasons for this phenomenon will be explained in detail in the Section "The impact of short optimization duration on VI-CNOPs".

Next, we take the optimization duration of 264 hr as an example to further analyze the forecast performance of the three ensemble forecast schemes, alongside evaluating the ability of the IPs in delineating the AEs.

We examined the probabilistic forecast performance of the three ensemble forecast schemes for LF and HF events based on the RD in the consistency between the probability forecasted by the model and the actual observed frequency. The accuracy and bias of the three schemes at different probability thresholds can be visually displayed (Figure 3). The results show that the probabilistic forecast skills of VI-CNOPs are relatively the most reliable among the three schemes (i.e., RD is closest to the 45° diagonal) for both LF and HF events, which further clarifies the reliability of the VI-CNOPs scheme.

Furthermore, we evaluated three ensemble forecast schemes using CRPS and PECA (Figure 4). The former measures the difference between the cumulative distribution function (CDF) of the truth states and the CDF derived from forecasted values; while the latter describes the capacity of ensemble perturbations to elucidate the control forecast errors.

The findings indicate that VI-CNOPs exhibit notably poorer PECA score during shorter forecast lead times (Figure 4b). As previously mentioned, VI-CNOPs are initialized by randomly sampling from the optimized PDF $q^*(\mathbf{z};\boldsymbol{\mu},\boldsymbol{\sigma})$. From a statistical perspective, this random sampling typically leads to a clustering of samples around

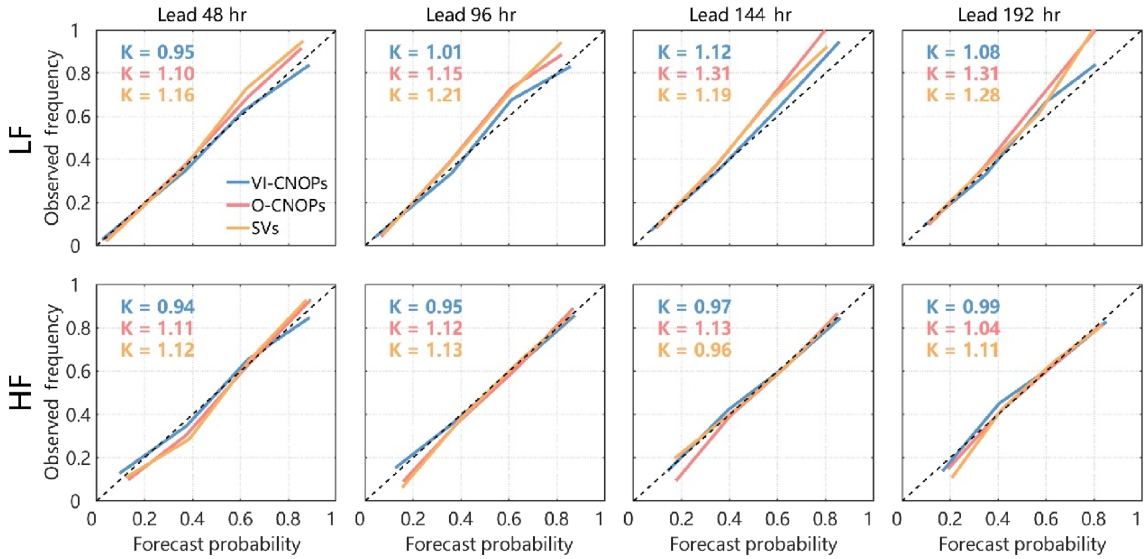


**FIGURE 3** When the optimization duration spans 264 hr, the reliability diagram (RD) of the variational inference-conditional nonlinear optimal perturbations (VI-CNOPs) (blue), orthogonal conditional nonlinear optimal perturbations (O-CNOPs) (orange), and singular vectors (SVs) (yellow) ensemble forecast schemes at different lead times (i.e., lead 48, 96, 144, and 192 hr). The first line is the low-frequency (LF) event and the second line is the high-frequency (HF) event. The horizontal axis represents the forecast probability, and the vertical axis represents the observation frequency. For example, if 10 out of 31 members forecast that an event will occur, then the forecast probability is 10/31. Observation frequency is calculated similarly. The number $K$ on the panel represents the slope of linear fitting using the points that constitute RD. The closer RD is to the 45° diagonal (the closer $K$ is to 1), the higher the forecast skills. [Colour figure can be viewed at wileyonlinelibrary.com]

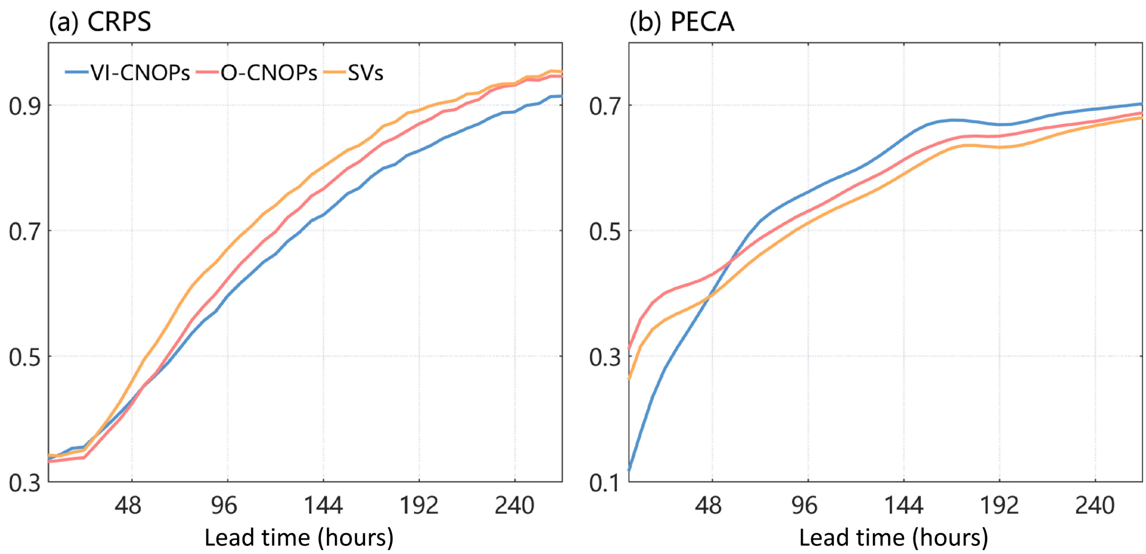


**FIGURE 4** The (a) continuous ranked probability score (CRPS) and (b) perturbation versus error correlation analysis (PECA) of the variational inference-conditional nonlinear optimal perturbations (VI-CNOPs) (blue), conditional nonlinear optimal perturbations (O-CNOPs) (orange), and singular vectors (SVs) (yellow) ensemble forecast schemes are plotted as functions of lead times. continuous ranked probability score (CRPS) PECA is negatively (positively) oriented, that is, the smaller (larger) the value, the higher the forecast skills. [Colour figure can be viewed at wileyonlinelibrary.com]

the mean ($\mu$), resulting in relatively concentrated IPs for VI-CNOPs at shorter lead times, thereby yielding a smaller ensemble spread (Figure 5). As lead times extend, nonlinear effects gradually manifest, amplifying the diversity among ensemble members of VI-CNOPs. This phenomenon, in turn, enhances the scheme's capability to capture control forecast errors, reflected by an increased PECA. Simultaneously, it contributes to improved ensemble reliability, as indicated by a spread-to-RMSE ratio closer to 1 (Figure 5).

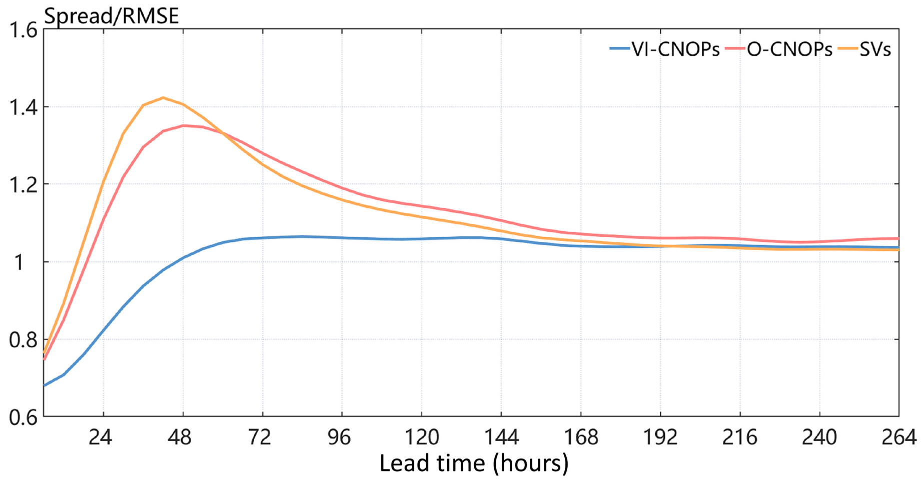


**FIGURE 5** The ratio of spread to root-mean-square error (RMSE) of the variational inference-conditional nonlinear optimal perturbations (VI-CNOPs) (blue), orthogonal conditional nonlinear optimal perturbations (O-CNOPs) (orange), and singular vectors (SVs) (yellow) ensemble forecast schemes are plotted as functions of lead times. A ratio closer to 1 signifies greater reliability in the ensemble forecast system. [Colour figure can be viewed at wileyonlinelibrary.com]

The above results have shown that VI-CNOPs can provide higher ensemble forecast skills. In addition, reliability is another important aspect in evaluating an ensemble forecast system (see the review paper of Feng et al., 2024). A reliable ensemble forecast system should ideally have a ratio of the ensemble spread to the ensemble mean RMSE close to 1 during the entire forecast period, signifying that the spread serves as a reliable indicator of the ensemble mean forecast errors (Fortin et al., 2014; Hopson, 2014). As displayed in Figure 5, due to the orthogonality of O-CNOPs and SVs, its ensemble spread is rapidly amplified in the early stage, overestimating the forecast uncertainty (i.e., the spread is too large). Subsequently, as the model is adjusted, the increase in spread slows down and eventually converges at the RMSE level. Conversely, VI-CNOPs exhibit a different trajectory. As shown in Figure 5, the growing influence of nonlinear effects leads to a gradual increase and eventual stabilization of the spread-to-RMSE ratio near 1 for VI-CNOPs, thereby providing a more consistent and reliable estimation of the forecast uncertainty.

Naturally, one may argue that VI-CNOPs grow slightly slower in the early stages and exhibit a smaller spread. Admittedly, this is a limitation of the VI-CNOPs method, as we currently approximate the solution using only a simple Gaussian distribution. In reality, the analysis error may follow a more complex distribution. In the future, we plan to explore the use of multiple distributions, such as a combination of different statistical models, to solve optimization problem Equation (8) and increase the diversity of the IPs.

Besides, the Talagrand histograms serve as an additional approach to evaluate the reliability of ensemble forecast systems (Talagrand et al., 1997). Ideally, the Talagrand histogram should show a flat shape, which means that there is no statistically significant difference between the truth and the forecast values, suggesting that the ensemble forecast system is reliable (Candille & Talagrand, 2005; Talagrand et al., 1997). Figure 6 illustrates the Talagrand diagram and its flatness depicting the VI-CNOPs, O-CNOPs, and SVs across varying lead times. The observed flatness of the VI-CNOPs histogram suggests that ensemble forecasts using VI-CNOPs offer greater reliability compared to those based on O-CNOPs and SVs, especially for longer forecast lead times. This underscores the robustness of VI-CNOPs in capturing uncertainty while maintaining reliable forecast performance.

## 5.2 | Characterization of analysis errors by the initial perturbations

An effective ensemble forecast scheme can aptly characterize AEs (Bowler, 2006; Molteni et al., 1996). Given that the initial AEs (derived from the difference between the initial analysis field and the truth) is known in the ideal experiment, we can examine the capability of the three IPs to capture AEs. Besides, Wang and Bishop (2003) emphasized that the primary concept behind the generation of IPs for ensemble forecasts is to ensure that the covariance matrix (CM) formed by the IPs approximates the CM of AEs at the initial time. Therefore, we delve deeper into assessing the capability of the IPs generated by VI-CNOPs, O-CNOPs, and SVs to characterize the CM of AEs.

Firstly, to estimate the CM of the initial AEs, 500 distinct random perturbations are superposed on the truth value to generate "observations" for a given case. Subsequently, 500 different analysis fields are derived using 4DVar. The CM of the initial AEs for a given case is then estimated based on these 500 distinct analysis fields.

As shown in Figure 7, although the CM formed by the IPs of VI-CNOPs (Figure 7b) still exhibits discrepancies compared to the CM of AEs (Figure 7a), which

Quarterly Journal of the Royal Meteorological Society RMetS



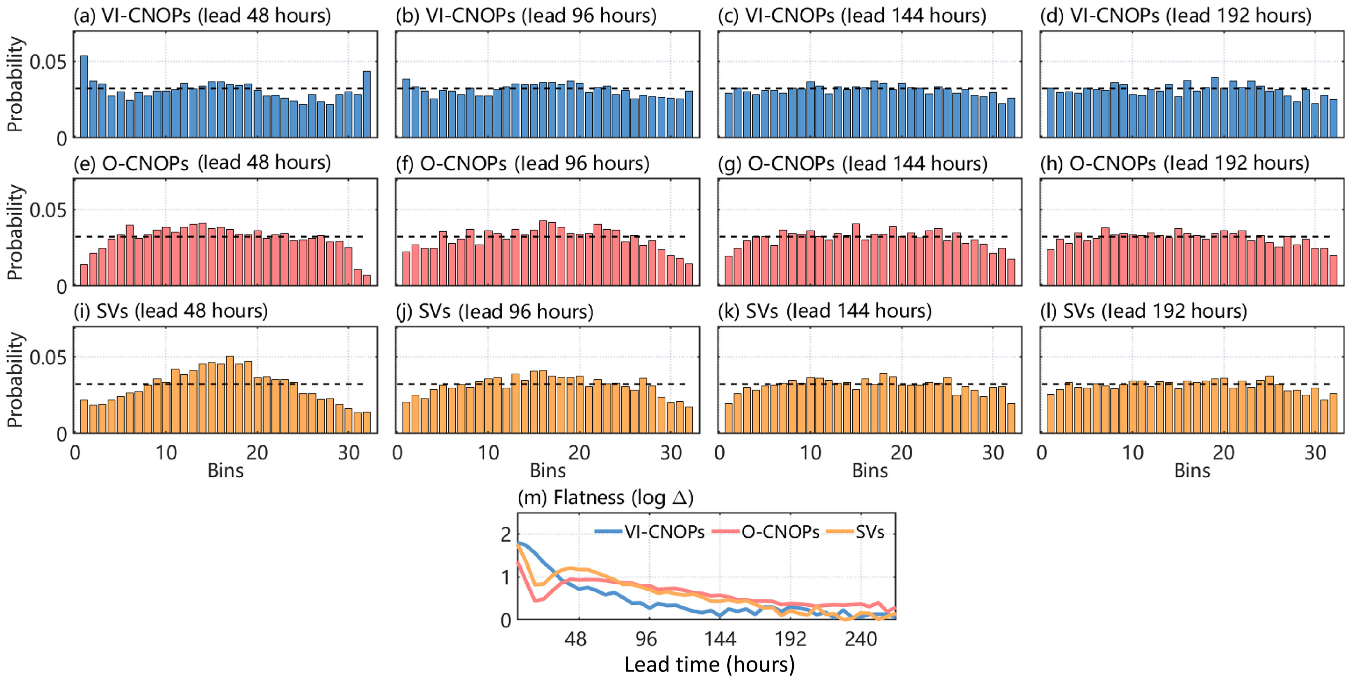


**FIGURE 6** Talagrand histograms for the ensemble forecasts made by (a–d, first row) variational inference-conditional nonlinear optimal perturbations (VI-CNOPs), (e–h, second row) orthogonal conditional nonlinear optimal perturbations (O-CNOPs), and (i–l, third row) singular vectors (SVs) at different lead times, and (m) their flatness (logarithmically transformed for clarity) as a function of lead times, with VI-CNOPs, O-CNOPs and SVs represented in blue, orange, and yellow, respectively. [Colour figure can be viewed at wileyonlinelibrary.com]

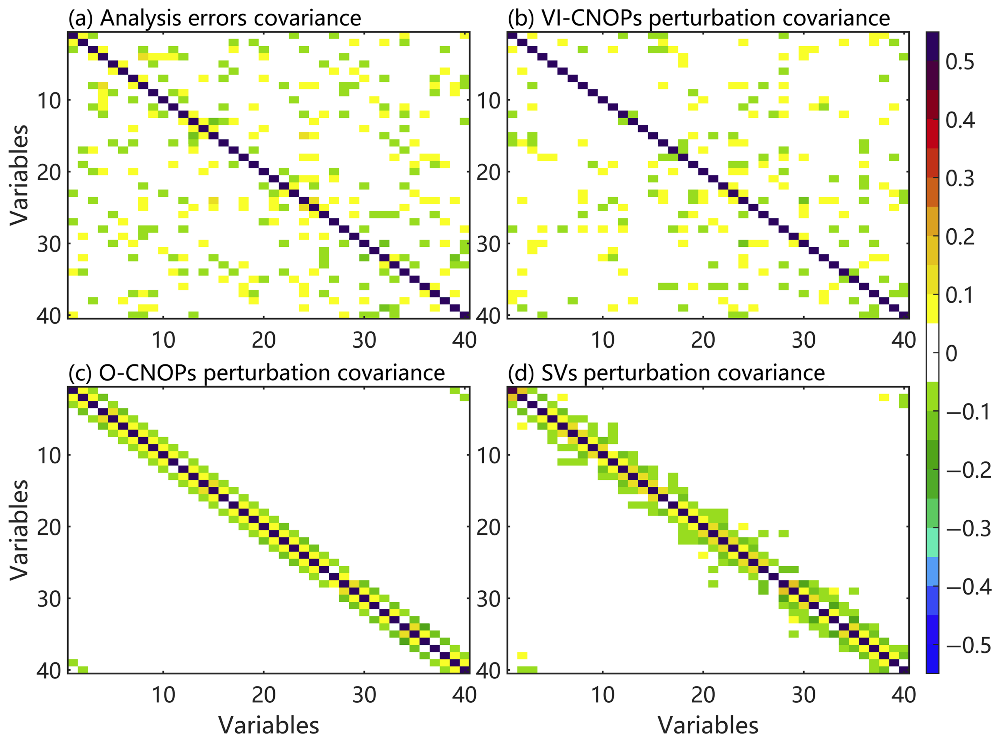


**FIGURE 7** (a) The covariance matrix (CM) of analysis errors (AEs), and the initial perturbations (IPs) formed by the IPs of (b) variational inference-conditional nonlinear optimal perturbations (VI-CNOPs), (c) orthogonal conditional nonlinear optimal perturbations (O-CNOPs), and (d) singular vectors (SVs) at the initial time, averaged over 100 cases. [Colour figure can be viewed at wileyonlinelibrary.com]

could be attributed to VI-CNOPs solely characterizing the growing-type AEs, while other types of error such as stochastic errors, slow-growing, or even non-growing errors are present within AEs (Toth & Kalnay, 1993). Nonetheless, in comparison to O-CNOPs and SVs, VI-CNOPs demonstrate closer proximity to the CM of AEs. Conversely, the CMs formed by the IPs of O-CNOPs and SVs (Figure 7c,d) result in spurious correlations among

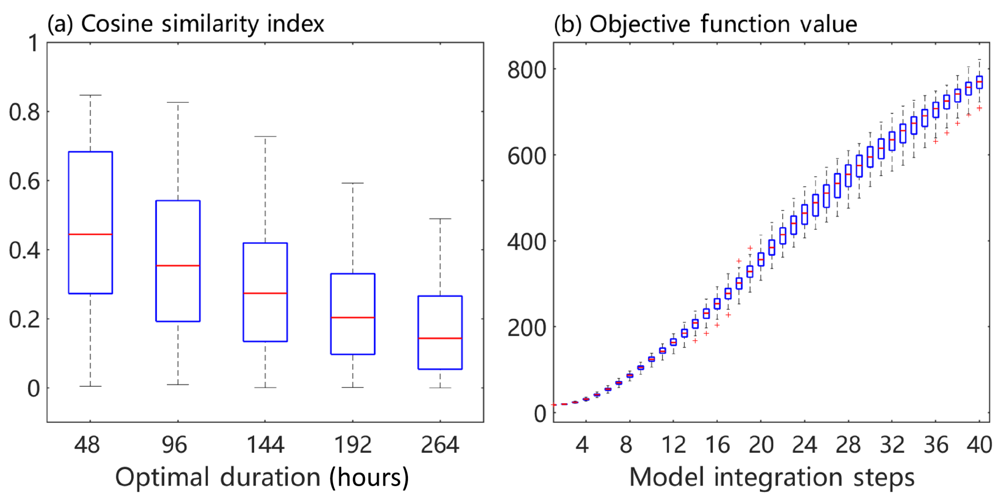


**FIGURE 8** For those 100 true cases, (a) the cosine similarity index (CSI) between $\boldsymbol{X}_{\text{VI-CNOP}}$ and the 1st-nonlinear optimal perturbations (1st-CNOPs) under different optimization durations, (b) the objective function value ($\|M_\tau(\mathbf{x}+\Delta) - M_\tau(\mathbf{x})\|$) obtained by adding 100 sets of random perturbations $\Delta$ to the initial analysis field as a function of the model integration steps. [Colour figure can be viewed at wileyonlinelibrary.com]

adjacent grid points, leading to slight deviations in depicting the spatial structure of AEs. Therefore, VI-CNOPs scheme exhibit higher ensemble forecast skills.

## 5.3 | The impact of short optimization duration on VI-CNOPs

As mentioned in the Section "Ensemble forecast performance," the forecast skills of the VI-CNOPs scheme are the worst among the three schemes when the optimization duration is short (Figure 2). This section will analyze the possible reasons for this result.

First, regarding $q^*(\mathbf{z};\boldsymbol{\mu},\boldsymbol{\sigma})$, we designate the 40 variables of the Lorenz-96 model as the mean $\boldsymbol{\mu}$, thus forming a 40-dimensional vector $\mathbf{X}_{\text{VI-CNOP}}$. Then, we calculate the cosine similarity index (CSI; a widely used metric for evaluating the similarity between two vectors; Duan & Huo, 2016 — see Appendix D for details) between $\mathbf{X}_{\text{VI-CNOP}}$ and the first CNOP (1st-CNOP; see Appendix B) in O-CNOPs.

As depicted in Figure 8a, the highest similarity between $\mathbf{X}_{\text{VI-CNOP}}$ and the 1st-CNOP is evident when optimized for 48 h. As the optimization duration increases, the similarity gradually diminishes, which suggests that with shorter optimization durations, $q^*(\mathbf{z};\boldsymbol{\mu},\boldsymbol{\sigma})$ tends to cluster near the 1st-CNOP. Consequently, the IPs sampled from $q^*(\mathbf{z};\boldsymbol{\mu},\boldsymbol{\sigma})$, that is, VI-CNOPs, predominantly cluster around the 1st-CNOP, limiting VI-CNOPs' characterization of information on AEs and consequently reducing ensemble forecast skills. Conversely, with a longer optimization duration, VI-CNOPs can capture more extensive nonlinear growing information of AEs, thereby improving ensemble forecast performance.

Furthermore, Figure 8b illustrates the objective function values ($\|M_\tau(\mathbf{x}+\Delta) - M_\tau(\mathbf{x})\|$) derived from the development of 100 random perturbations superposed on the analysis field. With shorter integration times, owing to weaker nonlinear effects, there is negligible difference in the objective function values, indicating that the development of these perturbations is largely similar. Considering an extreme scenario where all $\mathbf{z}$ in Equation (2) are identical, the distribution $q^*(\mathbf{z};\boldsymbol{\mu},\boldsymbol{\sigma})$ would collapse into a single 1st-CNOP. This elucidates why with shorter optimization durations, $q^*(\mathbf{z};\boldsymbol{\mu},\boldsymbol{\sigma})$ tends to converge around the 1st-CNOP.

Additionally, from the perspective of optimization algorithms, the "robustness" of $q^*(\mathbf{z};\boldsymbol{\mu},\boldsymbol{\sigma})$ optimization results weaken with shorter optimization durations, a facet beyond the scope of this study.

## 6 | SUMMARY AND DISCUSSION

In this paper, drawing on the CNOP concept, which embodies rapidly growing IP, combined with VI technology, we first construct a theoretical framework for solving the optimal PDF of the growing-type IPs. This theory was then applied to the generation of IPs in ensemble forecasts, and its effectiveness was verified on the Lorenz-96 model. Specifically, recognizing the considerable impact of growing-type AEs in the analysis field on forecast results, we utilize VI methods to derive an optimal approximate distribution $q^*(\mathbf{z};\boldsymbol{\mu},\boldsymbol{\sigma})$ for the unknown growing-type IPs PDF $p(\mathbf{z}|\mathbf{x})$. Crucially, this approach dispenses with the need for employing adjoint and tangent linear models, thus representing a notable advantage over O-CNOPs and SVs methods.

The ensemble forecast experiments demonstrate that VI-CNOPs exhibit higher forecast skills compared to O-CNOPs and SVs methods including metrics like ACC, RMSE, CRPS, and this skill increases with the extension of optimization durations. However, it is important to acknowledge that VI-CNOPs perform the poorest among the three methods during short optimization durations. Further analysis indicates that during the short optimization durations, the weaker nonlinear effects of the model cause $q^*(\mathbf{z}; \boldsymbol{\mu}, \boldsymbol{\sigma})$ to concentrate more closely around the 1st-CNOP, resulting in a diminished capture of growing-type AEs' information by the IPs sampled from $q^*(\mathbf{z}; \boldsymbol{\mu}, \boldsymbol{\sigma})$, consequently leading to a decrease in forecast skills. Additionally, similar issues arise when forecast lead times are short, attributable to VI-CNOPs exhibiting a lower ensemble spread in the early stages.

As highlighted by Wang and Bishop (2003), the fundamental concept behind ensemble forecast IPs generation methods is to ensure that the CM of the IPs at the initial time closely approximates the CM of AEs. To further illustrate the effectiveness of the VI-CNOPs method, considering computer simulation experiments where the initial AEs are known, we compared the similarity between the CMs of IPs generated by VI-CNOPs, O-CNOPs, and SVs schemes and the CM of AEs. The results indicate that the CM of the VI-CNOPs scheme more closely resembles the CM of AEs, which is also the reason for its higher forecast skills. There are still certain differences between the CM comprised of VI-CNOPs and the CM of AEs. This is due to the fact that $q^*(\mathbf{z}; \boldsymbol{\mu}, \boldsymbol{\sigma})$ only depicts growing-type AEs. Overall, VI-CNOPs represents an effective new method for ensemble forecasts.

It should be acknowledged that this work still requires further research. Firstly, Lorenz-96 serves merely as a basic conceptual model. In the future, we will verify the effectiveness of VI-CNOPs in more complex models. Secondly, we only consider initial uncertainties, while in reality, perfect models do not exist, and model perturbations in ensemble forecasts are also crucial (Buizza et al., 1999; Duan et al., 2019). For example, Duan and Zhou (2013) proposed the nonlinear forcing singular vector (NFSV) to characterize the impact of model uncertainty on forecast results. Further, Duan et al. (2022) developed the joint perturbations that can characterize the coordinated growth of initial and model errors, namely, C-NFSV, and clarified the model errors' importance in ensemble forecasts. Thirdly, in solving for $q^*(\mathbf{z}; \boldsymbol{\mu}, \boldsymbol{\sigma})$, this study optimized the 40 variables of the Lorenz-96 model individually and utilized a Gaussian distribution to fit $p(\mathbf{z}|\mathbf{x})$. However, real-world IPs may exhibit a more complex distribution. Hence, exploring combinations of different distributions may provide a more effective pattern to characterize $p(\mathbf{z}|\mathbf{x})$. Previous studies have shown that the CNOPs are very valuable for applications to ensemble forecasting of high-impact weather and climate events (Duan et al., 2023). VI-CNOPs inherit the growing characteristics of CNOPs while simplifying the original computational complexity. Therefore, we believe that VI-CNOPs have great potential as in ensemble forecasts as well.

Looking ahead, VI does not depend on the adjoint model. Consequently, we plan to develop this method into a software package that will facilitate the construction of IPs for ensemble forecasts across various numerical models. Moreover, it is important to note that VI may have broader applications beyond generating IPs for ensemble forecasts. For instance, VI can calculate the optimal approximation of the background PDF of real data, followed by random sampling from it to generate additional data that retains similar or identical statistical properties as the real data. This can effectively address the current shortage of training data for artificial intelligence models.

## FUNDING INFORMATION

This work is supported in part by the National Natural Science Foundation of China (42288101, 42405147), and in part by the China National Postdoctoral Program for Innovative Talents (BX20230071). This study is carried out at National Supercomputer Center in Tianjin, China, supported by TianHe Qingsuo Project-special fund project in the field of climate, meteorology, and ocean.

## CONFLICT OF INTEREST STATEMENT

The authors declare that they have no competing interests.

## DATA AVAILABILITY STATEMENT

The datasets generated and/or analyzed during the study are stored on computers at the High-end Computing Center Cluster, Fudan University (https://fudancluster.fudannhpcc.com.cn) and will be available to researchers upon request.

## ORCID

*Bo Qin* https://orcid.org/0000-0001-7093-6531
*Jie Feng* https://orcid.org/0000-0002-2480-2003

## REFERENCES

Amemiya, A., Shlok, M. & Miyoshi, T. (2023) Application of recurrent neural networks to model bias correction: idealized experiments with the Lorenz-96 model. *Journal of Advances in Modeling Earth Systems*, 15(2), e2022MS003164. Available from: https://doi.org/10.1029/2022MS003164

Anderson, J.L. (1997) The impact of dynamical constraints on the selection of initial conditions for ensemble predictions: low-order perfect model results. *Monthly Weather Review*, 125(11), 2969–2983. Available from: https://doi.org/10.1175/1520-0493(1997)125<2969:TIODCO>2.0.CO;2

Blei, D.M., Kucukelbir, A. & McAuliffe, J.D. (2017) Variational inference: a review for statisticians. *Journal of the American Statistical Association*, 112(518), 859–877. Available from: https://doi.org/10.1080/01621459.2017.1285773

Bowler, N.E. (2006) Comparison of error breeding, singular vectors, random perturbations and ensemble Kalman filter perturbation strategies on a simple model. *Tellus Series A: Dynamic Meteorology and Oceanography*, 58(5), 538–548. Available from: https://doi.org/10.1111/J.1600-0870.2006.00197.X

Brajard, J., Carrassi, A., Bocquet, M. & Bertino, L. (2020) Combining data assimilation and machine learning to emulate a dynamical model from sparse and noisy observations: a case study with the Lorenz 96 model. *Journal of Computational Science*, 44, 101171. Available from: https://doi.org/10.1016/j.jocs.2020.101171

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

Buizza, R. (1997) Potential forecast skill of ensemble prediction and spread and skill distributions of the ECMWF ensemble prediction system. *Monthly Weather Review*, 125(1), 99–119. Available from: https://doi.org/10.1175/1520-0493(1997)125<0099:PFSOEP>2.0.CO;2

Buizza, R. (2008) The value of probabilistic prediction. *Atmospheric Science Letters*, 9(2), 36–42. Available from: https://doi.org/10.1002/asl.170

Buizza, R. (2019) Introduction to the special issue on '25 years of ensemble forecasting'. *Quarterly Journal of the Royal Meteorological Society*, 145(S1), 1–11. Available from: https://doi.org/10.1002/qj.3370

Buizza, R., Leutbecher, M. & Isaksen, L. (2008) Potential use of an ensemble of analyses in the ECMWF ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, 134(637), 2051–2066. Available from: https://doi.org/10.1002/qj.346

Buizza, R., Milleer, M. & Palmer, T.N. (1999) Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, 125(560), 2887–2908. Available from: https://doi.org/10.1002/QJ.49712556006

Candille, G. & Talagrand, O. (2005) Evaluation of probabilistic prediction systems for a scalar variable. *Quarterly Journal of the Royal Meteorological Society*, 131(609), 2131–2150. Available from: https://doi.org/10.1256/QJ.04.71

Cheng, J., Liu, T. & Tan, S. (2023) Score priors guided deep Variational inference for unsupervised real-world single image Denoising. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 12937–12948.

Damianou, A.C., Titsias, M.K., Gr, M., Lawrence, N.D. & Storkey, A. (2016) Variational inference for latent variables and uncertain inputs in Gaussian processes. *Journal of Machine Learning Research*, 17, 1–62.

Duan, W. & 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(3), 997–1014. Available from: https://doi.org/10.1175/JAS-D-15-0138.1

Duan, W., Ma, J. & Vannitsem, S. (2022) An ensemble forecasting method for dealing with the combined effects of the initial and model errors and a potential deep learning implementation. *Monthly Weather Review*, 150(11), 2959–2976. Available from: https://doi.org/10.1175/MWR-D-22-0007.1

Duan, W., Wang, Y., Huo, Z. & Zhou, F. (2019) Ensemble forecast methods for numerical weather forecast and climate prediction: thinking and Prospect. *Climatic and Environmental Research*, 24(3), 396–406. Available from: https://doi.org/10.3878/j.issn.1006-9585.2018.18133

Duan, W., Yang, L., Xu, Z. & Chen, J. (2023) Conditional nonlinear optimal perturbation: applications to ensemble forecasting of high-impact weather systems. In: *Numerical Weather Prediction: East Asian Perspectives*. Cham: Springer International Publishing, pp. 441–460.

Duan, W. & Zhou, F. (2013) Non-linear forcing singular vector of a two-dimensional quasi-geostrophic model. *Tellus A: Dynamic Meteorology and Oceanography*, 65(1), 18452. Available from: https://doi.org/10.3402/tellusa.v65i0.18452

Epstein, E.S. (1969) Stochastic dynamic prediction. *Tellus*, 21(6), 739–759. Available from: https://doi.org/10.1111/j.2153-3490.1969.tb00483.x

Feng, J., Ding, R., Li, J. & Liu, D. (2016) Comparison of nonlinear local Lyapunov vectors with bred vectors, random perturbations and ensemble transform Kalman filter strategies in a barotropic model. *Advances in Atmospheric Sciences*, 33, 1036–1046. Available from: https://doi.org/10.1007/s00376-016-6003-4

Feng, J., Ding, R., Liu, D. & Li, J. (2014) The application of nonlinear local Lyapunov vectors to ensemble predictions in Lorenz systems. *Journal of the Atmospheric Sciences*, 71(9), 3554–3567. Available from: https://doi.org/10.1175/JAS-D-13-0270.1

Feng, J., Li, J., Ding, R. & Toth, Z. (2018) Comparison of nonlinear local Lyapunov vectors and bred vectors in estimating the spatial distribution of error growth. *Journal of the Atmospheric Sciences*, 75(4), 1073–1087. Available from: https://doi.org/10.1175/JAS-D-17-0266.1

Feng, J., Toth, Z., Zhang, J. & Peña, M. (2024) Ensemble forecasting: a foray of dynamics into the realm of statistics. *Quarterly Journal of the Royal Meteorological Society*, 150, 2537–2560.

Fortin, V., Abaza, M., Anctil, F. & Turcotte, R. (2014) Why should ensemble spread match the RMSE of the ensemble mean? *Journal of Hydrometeorology*, 15(4), 1708–1713. Available from: https://doi.org/10.1175/JHM-D-14-0008.1

Ganguly, A. & Earp, S.W.F. (2021) An introduction to Variational inference. arXiv preprint arXiv: 2108.13083.

Gelaro, R., Buizza, R., Palmer, T.N. & Klinker, E. (1998) Sensitivity analysis of forecast errors and the construction of optimal perturbations using singular vectors. *Journal of the Atmospheric Sciences*, 55(6), 1012–1037. Available from: https://doi.org/10.1175/1520-0469(1998)055<1012:SAOFEA>2.0.CO;2

Gilmour, I. & Smith, L.A. (1997) Enlightenment in shadows. In: *Applied Nonlinear Dynamics and Stochastic Systems near the Millennium*. American Institute of Physics, pp. 335–340.

Hopson, T.M. (2014) Assessing the ensemble spread–error relationship. *Monthly Weather Review*, 142(3), 1125–1142. Available from: https://doi.org/10.1175/MWR-D-12-00111.1

Huo, Z. & Duan, W. (2019) The application of the orthogonal conditional nonlinear optimal perturbations method to typhoon track ensemble forecasts. *Science China Earth Sciences*, 62(2), 376–388. Science in China Press. Available from: https://doi.org/10.1007/s11430-018-9248-9

Jiang, Z.N., Mu, M. & Wang, D.H. (2009) Ensemble prediction experiments using conditional nonlinear optimal perturbation. *Science in China Series D: Earth Sciences*, 52(4), 511–518. Available from: https://doi.org/10.1007/S11430-009-0042-Y

Jordan, M., Ghahramani, Z., Jaakkola, T., learning, L.S.-M. & undefined. (1999) An introduction to variational methods for graphical models. *Machine Learning*, 37, 183–233. Available from: https://doi.org/10.1023/A:1007665907178

Khare, S.P. & Anderson, J.L. (2006) An examination of ensemble filter based adaptive observation methodologies. *Tellus Series A: Dynamic Meteorology and Oceanography*, 58(2), 179–195. Available from: https://doi.org/10.1111/j.1600-0870.2006.00163.x

Kullback, S. & Leibler, R.A. (1951) On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1), 79–86. Available from: https://doi.org/10.1214/aoms/1177729694

Leith, C.E. (1974) Theoretical skill of Monte Carlo forecasts. *Monthly Weather Review*, 102(6), 409–418.

Leith, C.E. (1978) Objective methods for weather prediction. *Annual Review of Fluid Mechanics*, 10(1), 107–128. Available from: https://doi.org/10.1146/annurev.fl.10.010178.000543

Leutbecher, M. & Palmer, T.N. (2008) Ensemble forecasting. *Journal of Computational Physics*, 227(7), 3515–3539. Available from: https://doi.org/10.1016/j.jcp.2007.02.014

Lorenz, E.N. (1963a) Section of planetary sciences: The predictability of hydrodynamic flow*, †. *Transactions of the New York Academy of Sciences*, 25(4 Series II), 409–432. Available from: https://doi.org/10.1111/j.2164-0947.1963.tb01464.x

Lorenz, E.N. (1963b) Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20(2), 130–141. Available from: https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2

Lorenz, E.N. (1965) A study of the predictability of a 28-variable atmospheric model. *Tellus A: Dynamic Meteorology and Oceanography*, 17(3), 321–333. Available from: https://doi.org/10.3402/tellusa.v17i3.9076

Lorenz, E.N. (1996) Predictability: a problem partly Lorenz96. *Proceedings Workshop on Predictability*, ECMWF, Reading, United Kingdom. pp. 1–18. https://www.ecmwf.int/node/10829

Lorenz, E.N. & Emanuel, K.A. (1998) Optimal sites for supplementary weather observations: simulation with a small model. *Journal of the Atmospheric Sciences*, 55(3), 399–414. Available from: https://doi.org/10.1175/1520-0469(1998)055<0399:OSFSWO>2.0.CO;2

Mason, S.J. & Graham, N.E. (1999) Conditional probabilities, relative operating characteristics, and relative operating levels. *Weather and Forecasting*, 14(5), 713–725. Available from: https://doi.org/10.1175/1520-0434(1999)014<0713:CPROCA>2.0.CO;2

Molteni, F., Buizza, R., Palmer, T.N. & Petroliagis, T. (1996) The ECMWF ensemble prediction system: methodology and validation. *Quarterly Journal of the Royal Meteorological Society*, 122(529), 73–119. Available from: https://doi.org/10.1002/qj.49712252905

Mu, M. & Duan, W. (2003) A new approach to studying ENSO predictability: conditional nonlinear optimal perturbation. *Chinese Science Bulletin*, 48(10), 1045–1047. Available from: https://doi.org/10.1360/02wd0389

Mu, M. & Jiang, Z.N. (2008) A new approach to the generation of initial perturbations for ensemble prediction: conditional nonlinear optimal perturbation. *Chinese Science Bulletin*, 53(13), 2062–2068. Available from: https://doi.org/10.1007/S11434-008-0272-Y

Mureau, R., Molteni, F. & Palmer, T.N. (1993) Ensemble prediction using dynamically conditioned perturbations. *Quarterly Journal of the Royal Meteorological Society*, 119(510), 299–323. Available from: https://doi.org/10.1002/QJ.49711951005

Murphy, J.M. (1988) The impact of ensemble forecasts on predictability. *Quarterly Journal of the Royal Meteorological Society*, 114(480), 463–493. Available from: https://doi.org/10.1002/QJ.49711448010

Revelli, J.A., Rodríguez, M.A. & Wio, H.S. (2010) The use of rank histograms and MVL diagrams to characterize ensemble evolution in weather forecasting. *Advances in Atmospheric Sciences*, 27(6), 1425–1437. Available from: https://doi.org/10.1007/s00376-009-9153-6

Sjölund, J. (2023) A tutorial on parametric Variational inference. https://doi.org/10.48550/arXiv.2301.01236.

Stanley, Z., Grooms, I. & Kleiber, W. (2021) Multivariate localization functions for strongly coupled data assimilation in the bivariate Lorenz 96 system. *Nonlinear Processes in Geophysics*, 28(4), 565–583. Available from: https://doi.org/10.5194/npg-28-565-2021

Talagrand, O., Vautard, R. & Straus, B. (1997) *Evaluation of probabilistic prediction systems*. ECMWF Workshop on Predictability, Reading, United Kingdom, ECMWF. pp. 1–25.

Terasaki, K. & Miyoshi, T. (2014) Data assimilation with error-correlated and non-orthogonal observations: experiments with the Lorenz-96 model. *Scientific Online Letters on the Atmosphere*, 10(1), 210–213. Available from: https://doi.org/10.2151/sola.2014-044

Tomczak, M., Swaroop, S. & Turner, R. (2018) Efficient low rank gaussian variational inference for neural networks. *Advances in Neural Information Processing Systems*, 33, 4610–4622.

Toth, Z. & Kalnay, E. (1993) Ensemble forecasting at NMC: the generation of perturbations. *Bulletin of the American Meteorological Society*, 74(12), 2317–2330.

Toth, Z. & Kalnay, E. (1997) Ensemble forecasting at NCEP and the breeding method. *Monthly Weather Review*, 125(12), 3297–3319.

Tran, D., Ranganath, R. & Blei, D.M. (2016) The variational Gaussian process. https://doi.org/10.48550/arXiv.1511.06499.

Wang, Q., Mu, M. & Sun, G. (2020) A useful approach to sensitivity and predictability studies in geophysical fluid dynamics: conditional non-linear optimal perturbation. *National Science Review*, 7(1), 214–223. Available from: https://doi.org/10.1093/nsr/nwz039

Wang, X. & Bishop, C.H. (2003) A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *Journal of the Atmospheric Sciences*, 60(9), 1140–1158. Available from: https://doi.org/10.1175/1520-0469(2003)060<1140:ACOBAE>2.0.CO;2

Wang, Y. & Duan, W. (2019) Influences of initial perturbation amplitudes and ensemble sizes on the ensemble forecasts made by CNOPs method. *Chinese Journal of Atmospheric Sciences*, 43(4), 915–929. Available from: https://doi.org/10.3878/j.issn.1006-9895.1810.18219

Wei, M. & Toth, Z. (2003) A new measurement of ensemble performance: perturbation versus error correlation analysis (PECA). *Monthly Weather Review*, 131, 1549–1565. Available from: https://doi.org/10.1175//1520-0493(2003)131<1549:ANMOEP>2.0.CO;2

Zhang, H., Duan, W. & Zhang, Y. (2023) Using the orthogonal conditional nonlinear optimal perturbations approach to address the uncertainties of tropical cyclone track forecasts generated by the WRF model. *Weather and Forecasting*, 38(10), 1907–1933. Available from: https://doi.org/10.1175/WAF-D-22-0175.1

Zhou, Q., Chen, L., Duan, W., Wang, X., Zu, Z., Li, X. et al. (2021) Using conditional nonlinear optimal perturbation to generate initial perturbations in ENSO ensemble forecasts. *Weather and Forecasting*, 36(6), 2101–2111. Available from: https://doi.org/10.1175/waf-d-21-0063.1

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Ji, C., Qin, B., Mu, M., Ma, J., Fang, X., Feng, J. *et al.* (2025) Optimal distributions of growing-type initial perturbations for ensemble forecasts: Theory and application in the Lorenz-96 model. *Quarterly Journal of the Royal Meteorological Society*, 151:e4913. Available from: https://doi.org/10.1002/qj.4913

## APPENDIX A

Assuming that the state vector is $\mathbf{U}$, its evolution can be described by the following nonlinear partial differential equations:

$$\begin{cases} \frac{\partial \mathbf{U}}{\partial t} = A(\mathbf{U}(\mathbf{x}, t)) \\ \mathbf{U}|_{t=0} = \mathbf{U_0} \end{cases}, \tag{A1}$$

where $\mathbf{U_0}$ and $\mathbf{U}(\mathbf{x}, t)$ represent the initial field and its evolution in time, respectively, and $A$ is a nonlinear differential operator. The solution $\mathbf{U}(\mathbf{x}, \tau)$ of the equations at time $\tau$ can be given:

$$\mathbf{U}(\mathbf{x}, \tau) = M_\tau(\mathbf{U_0}), \tag{A2}$$

where $M_\tau$ is the nonlinear propagator of Equation (A1) from 0 to $\tau$. For a sufficiently small perturbation $\mathbf{u}$, the time evolution trajectory $\mathbf{u}(t)$ within a finite time interval can be described by a linearized equation:

$$\begin{cases} \frac{\partial \mathbf{u}}{\partial t} = A_l(\mathbf{U})\mathbf{u} \\ \mathbf{u}|_{t=0} = \mathbf{u_0} \end{cases}. \tag{A3}$$

$A_l(\mathbf{U}) = \frac{\delta A}{\delta \mathbf{U}}$ is the tangent linear operator of $A$. Let $L(t, t_0)$ denote the tangent forward propagator of the linear equation. The solution $\mathbf{u}(t)$ of Equation (C3) from $t_0$ to $t$ can be written:

$$\mathbf{u}(t) = L(t, t_0)\mathbf{u_0}. \tag{A4}$$

The norm $\| \, \|_d$ is introduced to measure the development of the perturbation at the final time $t$:

$$\begin{aligned} \|\mathbf{u}(t)\|_d &= \langle \mathbf{u}(t), \mathbf{u}(t) \rangle = \langle L\mathbf{u}(t_0), L\mathbf{u}(t_0) \rangle \\ &= \langle \mathbf{u}(t_0), L^* L\mathbf{u}(t_0) \rangle, \end{aligned} \tag{A5}$$

where $L^*$ is the adjoint propagator of $L$. Therefore, SVs is the solution of the following eigenvalue problem:

$$L^* L\mathbf{u}_i(t_0) = \sigma_i^2 \mathbf{u}_i(t_0), \tag{A6}$$

where $\mathbf{u}_i$ represents the $i$th eigenvector of $L^*L$, that is, $SV_i$, $\sigma_i^2$ are the corresponding eigenvalue.

In this paper, we employ the L2 norm (the same as below) to quantify the magnitude of both AEs and corresponding forecast errors:

$$\|\mathbf{x}\|_{L2} = \sqrt{\sum_{i=1}^{n} x_i^2}. \tag{A7}$$

## APPENDIX B

If $\mathbf{u_0}$ denotes the initial errors, the resultant prediction errors $u_\tau$ at prediction time $\tau$ can be approximated by

$$\mathbf{u}_\tau = M_\tau(\mathbf{U_0} + \mathbf{u_0}) - M_\tau(\mathbf{U_0}). \tag{B1}$$

Based on Equations (A2) and (B1), Mu and Duan (2003) defined the CNOP. An initial perturbation $\mathbf{u}_{0\delta}$ is called CNOP if and only if

$$J(\mathbf{u}_{0\delta}) = \max_{\|\mathbf{u_0}\|_i \le \delta} \|M_\tau(\mathbf{U_0} + \mathbf{u_0}) - M_\tau(\mathbf{U_0})\|_f, \tag{B2}$$

where $\| \cdot \|_i$ denotes a norm used to quantify the magnitude of the initial errors, while $\|\mathbf{u}_0\|_i \le \delta$ represents the constraint condition governing the magnitude of the initial errors. $\| \cdot \|_f$ signifies the norm employed to evaluate the magnitude of the forecast errors at the forecast time $\tau$.

Subsequently, Duan and Huo (2016) introduced the Orthogonal CNOPs (O-CNOPs). Their mathematical definition is as follows:

The $j$th-CNOP can be calculated from Equation (B2), and satisfies:

$$\boldsymbol{\Omega}_j = \begin{cases} \left\{ \mathbf{u}_{0j} \in \mathbb{R}^n \middle| \|\mathbf{u}_{0j}\| \le \delta \right\}, j = 1 \\ \left\{ \mathbf{u}_{0j} \in \mathbb{R}^n \middle| \|\mathbf{u}_{0j}\| \le \delta, \mathbf{u}_{0j} \perp \boldsymbol{\Omega}_k, k = 1, \ldots, j-1 \right\}, \\ j > 1 \end{cases} \tag{B3}$$

where $\mathbf{u}_{0j}$ represents the initial perturbation in the subspace $\boldsymbol{\Omega}_j$. The $j$th-CNOP epitomizes the initial perturbation evolving with maximum nonlinearity within the subspace $\boldsymbol{\Omega}_j$, while remaining orthogonal to the 1st-CNOP, 2nd-CNOP, 3rd-CNOP, …, up to the $(j-1)$th-CNOP. These CNOPs exhibit mutual orthogonality.

Quarterly Journal of the Royal Meteorological Society RMetS



## APPENDIX C

The anomaly correlation coefficient (ACC) and root-mean-square error (RMSE) were used to evaluate the forecast skills:

$$\mathrm{ACC} = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2}\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i-\bar{y})^2}}, \quad \text{(C1)}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-y_i)^2}. \quad \text{(C2)}$$

The ensemble spread delineates the magnitude of fluctuations among ensemble members relative to the ensemble mean, which serves as an estimation of forecast uncertainty:

$$\mathrm{Spread} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2}, \quad \text{(C3)}$$

where $x_i$ and $y_i$ represent the truth and forecast values, respectively; $\bar{x}$ and $\bar{y}$ denote the mean of the truth and forecast, respectively. $N$ represents the total number of samples.

The Brier Score (BS; Brier, 1950) quantifies the accuracy and reliability of forecasts based on the disparity between predicted probability values and truth. Its computational formula is expressed as follows:

$$\mathrm{BS} = \frac{1}{N}\sum_{i=1}^{N}(f_i-o_i)^2, \quad \text{(C4)}$$

where $f_i$ and $o_i$ denote the forecast probability and observed probability of the $i$th sample, respectively. A lower BS indicates higher predictive accuracy.

The Continuous Ranked Probability Score (CRPS) quantifies the discrepancy between the cumulative distribution function (CDF) of true values and the predicted CDF. Lower CRPS values signify higher predictive accuracy:

$$\mathrm{CRPS} = \int_{-\infty}^{+\infty}\left(F(x)-1_{\{y\le x\}}\right)^2 dx, \quad \text{(C5)}$$

where $F(x)$ is the predicted CDF; $1_{\{y\le x\}}$ denotes the indicator function, representing 1 when the observed value $y$ is less than or equal to the predicted value $x$, and 0 otherwise.

The Relative Operating Characteristic (ROC) curve is commonly utilized to evaluate the performance of probability forecast models. Considering the occurrence of a weather event, the model's predictions are verified against the truth, leading to one of the following results (Mason & Graham, 1999): true positive (TP), false negative (FN), false positive (FP), or true negative (TN). Based on these results, a binary contingency Table C1 can be obtained:

**TABLE C1** Contingency table.

| | Observation | |
|---|---|---|
| **Forecast** | **Occurrence** | **Non-occurrence** |
| Occurrence | TP | FP |
| Non-occurrence | FN | TN |

The hit rate and false alarm rate are defined as:

$$H = \frac{TP}{TP+FN}, \quad \text{(C6)}$$

$$F = \frac{FP}{FP+TN}. \quad \text{(C7)}$$

Subsequently, plotting a curve with H as the vertical axis and F as the horizontal axis yields the ROC curve. The area under the ROC curve (ROCA) encapsulates the comprehensive assessment of the model:

$$\mathrm{ROCA} = \int_0^1 H(x)dx = \sum_{i=1}^{M}\frac{1}{2}(H_{i+1}+H_i)(F_{i+1}-F_i), \quad \text{(C8)}$$

where $M$ represents the number of categories relative to the probability threshold. A higher ROCA value implies greater precision in ensemble forecasting.

The Reliability Diagram is a visualization tool used to assess the accuracy of probability forecast. The computation involves three steps:

1. The predicted probability values are grouped or binned, each encompassing a certain range of predicted probability values.
2. For each group of predicted probability values, the frequency of observed events occurring within that group is calculated.
3. The Reliability Diagram is plotted, with the horizontal axis representing the predicted probability values and the vertical axis representing the frequency of actual observed events occurring.

If the Reliability Diagram exhibits good symmetry near the 45° diagonal line, it indicates that the model's probability forecasts are reliable and accurate. Otherwise, it suggests the presence of systematic bias or inaccuracy in the model.

Another crucial method for assessing the reliability of ensemble forecast systems is the Talagrand diagram

(Talagrand et al., 1997). Suppose there are $J$ grid points in the verification domain, the ensemble forecast comprises $M$ ensemble members, and $N$ experiments are conducted. The forecast values of ensemble members at each grid point are arranged in ascending order as $x_1 < x_2 < \dots < x_M$, dividing these $M$ forecast values into $M+1$ intervals. We have a total of $S = J \times N$ valid samples. $S_m$ denote the frequency of observations falling into the $m$th interval. The probability that the observation falls in the $m$th interval is:

$$P_m = \frac{S_m}{S} \tag{C9}$$

Then, the degree of deviation of $S_m$ from its expected value can be denoted as:

$$\Delta = \frac{1}{\mathrm{M}} \sum_{m=1}^{M+1} \left( S_m - \frac{\mathrm{S}}{\mathrm{M}+1} \right)^2 \tag{C10}$$

A flat shape (smaller $\Delta$) in the Talagrand histogram signifies negligible statistical disparity between the truth and the forecast values, thereby indicating the reliability of the ensemble forecast.

Perturbation versus Error Correlation Analysis (PECA) is a commonly used for assessing the quality of ensemble forecasts. Ensemble perturbation is defined as the difference between perturbed forecasts and their corresponding control forecasts, which can be expressed by the following:

$$\mathbf{P}_i(t) = \mathbf{F}_i(t) - \mathbf{F}_{\mathrm{Ctrl}}(t), \tag{C11}$$

where $\mathbf{P}_i(t)$, $\mathbf{F}_i(t)$, $\mathbf{F}_{\mathrm{Ctrl}}(t)$ represent the $i$th ensemble perturbation, the $i$th ensemble forecast member, and the control forecast, respectively.

The forecast errors $\mathbf{E}(t)$ are defined as the difference between the control forecast $\mathbf{F}_{\mathrm{Ctrl}}(t)$ and the verifying analysis $\mathbf{F}(t)$, that is, the truth runs:

$$\mathbf{E}(t) = \mathbf{F}_{\mathrm{Ctrl}}(t) - \mathbf{F}(t). \tag{C12}$$

For an ensemble forecast system with $M$ perturbation members, there exist $M$ optimal combinations of ensemble perturbations that minimize the difference with the control forecast errors:

$$\min \left\| \mathbf{E} - \sum_{i=1}^{M} \alpha_i \mathbf{P}_i \right\|_2^2. \tag{C13}$$

To solve this optimization problem, we obtain the optimal estimate values of $\alpha_i$ denoted as $\hat{\alpha}_i$, along with the combination vector $\mathbf{P}_{\mathrm{Opt}}$ under the optimal estimation:

$$\mathbf{P}_{\mathrm{Opt}} = \sum_{i=1}^{M} \hat{\alpha}_i \mathbf{P}_i. \tag{C14}$$

Then, the PECA can be calculated:

$$\mathrm{PECA} = \frac{\mathrm{cov}\left(\mathbf{E}, \mathbf{P}_{\mathrm{Opt}}\right)}{(\mathrm{cov}(\mathbf{E}, \mathbf{E}))^{1/2} \times \left(\mathrm{cov}\left(\mathbf{P}_{\mathrm{Opt}}, \mathbf{P}_{\mathrm{Opt}}\right)\right)^{1/2}}, \tag{C15}$$

where $\mathrm{cov}(\cdot)$ is the covariance.

A higher PECA value signifies a heightened correlation between perturbations and errors, indicating that the ensemble perturbations more faithfully capture the analysis errors characteristics, which augments the reliability and efficacy of the ensemble forecast.

## APPENDIX D

The cosine similarity index is defined as follows:

$$\mathrm{CSI} = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \cdot \|\mathbf{y}\|}, \tag{D1}$$

where $\langle \cdot, \cdot \rangle$ is the Euclidean inner product, and $\| \cdot \|$ is the L2 norm.