Performance and Regime of Ensemble Perturbations in Sampling Initial and Forecast Errors at Multiple Spatial Scales in the NCEP Global Ensemble Forecast System

Li MA (马丽)[1], Jie FENG (冯杰)[1,2,3*], Jingchen PU (浦景晨)[1], and Liangying LIU (刘靓颖)[1]

1 *Department of Atmospheric and Oceanic Sciences and Institute of Atmospheric Sciences, Fudan University, Shanghai* 200438
2 *Shanghai Key Laboratory of Ocean–land–atmosphere Boundary Dynamics and Climate Change, Shanghai* 200438
3 *Shanghai Academy of Artificial Intelligence for Science, Shanghai* 200003

(Received 9 June 2025; revised 10 August 2025; accepted 6 September 2025)

ABSTRACT

One of the keys to ensemble forecasting lies in the effective generation of initial ensemble perturbations, which aims to represent the probability distribution of analysis errors. Most previous studies have focused on assessing the predictive skills of ensemble forecasts, offering only indirect assessments of the quality of initial ensemble conditions. This study evaluates the performance of ensemble perturbations in sampling errors at both the initial and forecast stages within the operational ensemble forecast system at NCEP. In particular, the ensemble sampling behavior across different spatial scales is assessed quantitatively, with consideration of both the magnitude and direction. The primary results indicate that, at the initial time, smaller scales exhibit a higher spatial correlation between ensemble spread and error magnitude compared to larger scales. This is related to the instability of small-scale perturbations and errors associated with strong localized forcings. However, ensemble perturbations at larger scales demonstrate superior performance in capturing the spatially coherent structures of analysis errors, exhibiting a higher perturbation–error correlation than those at smaller scales. This can be attributed to the substantially lower effective degrees of freedom (DOF) associated with larger-scale perturbations and errors. These differences in sampling performance across scales persist throughout the entire forecast lead times. The contrasting trends in sampling behavior of the magnitude and direction across various scales suggest that the keys for effective ensemble generation vary with spatial scales.

Key words: ensemble forecasting, perturbation sampling, multiple spatial scales, directional sampling

Citation: Ma, L., J. Feng, J. C. Pu, et al., 2026: Performance and regime of ensemble perturbations in sampling initial and forecast errors at multiple spatial scales in the NCEP global ensemble forecast system. *J. Meteor. Res.*, **40**(1), 56–74, https://doi.org/10.1007/s13351-026-5143-4.

1. Introduction

A single forecast, typically used to provide deterministic forecast information, cannot estimate the uncertainties inherent in the forecast. This limitation has driven the wide adoption and development of ensemble forecasting in theory, methodology, and application. Ensemble forecasting generates multiple perturbed forecasts instead of a single one by perturbing the initial conditions and model-related dynamics and physics (Epstein, 1969; Leith, 1974; Molteni and Palmer, 1993; Toth and Kalnay, 1993; Zhou et al., 2022). These distinct forecast members facilitate the quantification of uncertainties in both initial and forecast states, as well as the identification of possible scenarios of atmospheric evolution (Barker, 1991; Houtekamer and Derome, 1995; Molteni et al., 1996; Buizza, 1997; Toth and Kalnay, 1997; Hopson, 2014). Therefore, ensemble forecasting has become the dominant mode in forecast centers globally.

The pioneering research by Epstein (1969) conducted the prediction of the probability distribution of atmospheric variables by solving the Liouville equation, a continuity equation with the probability density function (PDF) as the prognostic variable. His study emphasized the critical importance of accurately determining the PDF of initial conditions for predicting the PDF of future

Supported by the National Natural Science Foundation of China (42375058) and Joint Research Project for Meteorological Capacity Improvement (24NLTSZ004).

*Corresponding author: fengjiefj@fudan.edu.cn

states. Building on this foundation, Leith (1974) proposed the Monte Carlo method to approximate initial state uncertainty using a limited number of randomly generated states. This study introduced the ensemble initialization framework used in modern ensemble forecast systems, which involves generating a set of state samples to represent the PDF of the true initial state.

Over the past three decades, significant progress has been made in the methodology and application of ensemble forecasting, driven by improvements in numerical model performance and computational capabilities, and advances in observational networks and data assimilation. In operational practice, initial ensemble perturbations are routinely generated and centered on the initial control analysis or the posterior ensemble mean state provided by an operational data assimilation system. While various ensemble initialization schemes have been developed across different operational centers, these initial perturbations aim to represent the possibility of analysis errors (i.e., the difference between the analysis and reality which can only be implicitly estimated). One type of method is dynamic-based, including the bred vector (Toth and Kalnay, 1993, 1997) previously used at NCEP, and the singular vector (Molteni and Palmer, 1993; Molteni et al., 1996; Buizza, 1997) currently employed at ECMWF. These methods focus on capturing perturbations that reside in the dynamically unstable or growing subspace of the system, which majorly contributes to the evolution of forecast errors (Mu et al., 2003; Palmer et al., 2006; Duan and Huo, 2016; Feng et al., 2018; Huo and Duan, 2018). The other type incorporates statistical algorithms to refine dynamically-evolved perturbations. These methods combine observational information with prior ensemble forecasts attempting to estimate the analysis error covariance matrix and generate posterior ensemble samples that reflect the updated statistical characteristics (Evensen, 1994; Houtekamer and Derome, 1995; Houtekamer and Mitchell, 1998; Wang and Bishop, 2003; Wei et al., 2008). A representative example of this type is the Ensemble Kalman filter (EnKF; Houtekamer and Mitchell, 1998; Whitaker and Hamill, 2002; Evensen, 2003; Zhou et al., 2022), which is currently operationally implemented at NCEP and various other operational centers.

Given the critical role of ensemble initialization in the performance of ensemble forecasting, it is essential to conduct an in-depth and comprehensive evaluation of the ability of initial ensemble perturbations in sampling initial analysis errors. However, such studies are scarce, largely due to the limited knowledge of specific analysis errors and a deficient theoretical understanding of statistical sampling. Most existing studies evaluate the efficacy of different ensemble initialization schemes indirectly by assessing ensemble forecast skill. For instance, it is conventional to evaluate the root-mean-square error (RMSE) and pattern anomaly correlation of ensemble mean forecasts (e.g., Palmer and Tibaldi, 1988; Stanski et al., 1989; Buizza, 1997; Atger, 1999; Zhang et al., 2015; Schwartz et al., 2022). Additionally, diagnostics of probabilistic forecasting scores of ensemble forecasting, including the rank histogram, continuous ranked probability score, and relative operating characteristics, are frequently calculated and compared (Brier, 1950; Murphy, 1973; Mason, 1982; Hamill, 2001; Bradley and Schwartz, 2011; Schwartz et al., 2020). However, these metrics are influenced not only by the ensemble perturbation schemes, such as dynamical methods or ensemble-based data assimilation, but also by the model performance. Consequently, these metrics provide only indirect and partial indicators of the sampling performance of initial perturbations in the ensemble forecast system.

Numerous studies have investigated the relationship between ensemble perturbations and errors in the ensemble mean. The majority of these studies have focused on the spatiotemporal consistency between ensemble spread and ensemble mean error magnitude, known as the spread–error correlation. For example, Whitaker and Loughe (1998) and Molteni and Buizza (1999) examined how the spatial relationship between ensemble spread and ensemble mean absolute error fields varies with lead time (Whitaker and Loughe, 1998; Buizza et al., 2005; Wei et al., 2006). However, the spread–error relationship only reflects the statistical consistency between the amplitudes (or standard deviation) of gridded perturbations and errors. They do not address the correlation between perturbations and errors in their spatially coherent structures, or in other words, the directional relationship of perturbation and error vectors.

Unlike most studies on the spread–error relationship, Wei and Toth (2003) introduced a novel metric known as Perturbation versus Error Correlation Analysis (PECA). This metric has proven to be a useful quantitative metric to assess the effectiveness of ensemble perturbation vectors in capturing forecast error vectors in a high-dimensional phase space (Wei et al., 2006, 2008; Feng et al., 2016; Hou et al., 2018). It is important to note that a higher level of PECA at initial time does not necessarily indicate an improved ensemble prediction skill. However, the ensemble with higher initial PECA indicates that the physical constraints on perturbations are more consistent with the characteristics of analysis errors, increasing the potential for ensemble forecasts to capture the possibility of future states. Nevertheless, an import-

ant characteristic of the atmosphere, i.e., the multi-scale nature, has not been considered in their metric, and the correlation between perturbations and errors at different spatial scales has remained unexplored.

The atmosphere is characterized by motions across multiple scales, ranging from large-scale phenomena, such as planetary and synoptic systems, to smaller scales, including mesoscale and convective processes (Lorenz, 1969; Tribbia and Baumhefner, 2004). Numerous studies have demonstrated the multi-scale nature of forecast error propagation both theoretically and practically (e.g., Privé and Errico, 2015). In recent years, advancements in convection-allowing ensemble forecasts have emerged as a frontier in the area of ensemble forecasting (e.g., Romine et al., 2013a, b; Schwartz et al., 2014, 2021; Keresturi et al., 2019; Loken et al., 2019). A major challenge in this field is managing the complicated interactions between different scales, exemplified by phenomena such as convective storm cells intensifying within synoptic-scale squall lines. Therefore, it is essential to account for the multi-scale evolution of ensemble perturbations and forecast errors, as well as their relationships across various spatial scales, when developing high-resolution ensemble forecast systems. Recent efforts, such as those by Schwartz et al. (2021), have incorporated multi-scale initialization techniques into convection-allowing ensemble forecasts, combining high-resolution EnKF analysis with coarser global fields. However, these methods have often been empirical and lacked a solid theoretical basis. Several unresolved theoretical challenges remain, particularly regarding the different characteristics and impacts of multi-scale sampling of initial perturbations. Addressing these foundational issues is critical to enhancing the efficacy of ensemble initialization strategies and improving forecast accuracy.

Given the limitations of existing research on ensemble sampling, this study aims to investigate the multi-scale characteristics of ensemble perturbations in sampling errors for both initial analyses and forecasts. Additionally, the theoretical mechanisms underlying the performance of ensemble sampling will be explored and discussed. Operational forecast data from the global ensemble forecast system (GEFS) at NCEP will be utilized for evaluation in this study. The GEFS employs the EnKF scheme for ensemble initialization, recognized as one of the most advanced initial ensemble schemes and widely applied worldwide.

This paper is organized in the following manner. Section 2 and Section 3 describe the data and methodology, respectively. In Section 4, we first evaluate the sample-mean RMSE and spread of the ensemble forecasts. Section 5 assesses and analyzes the initial sampling performance of full fields (i.e., the non-decomposed field) in terms of magnitude and direction. Similar sampling performance at multiple spatial scales, as well as their temporal evolution, is further analyzed in Section 6 and Section 7, respectively. Conclusions and discussion are presented in Section 8.

2. Data

This study utilized ensemble forecast data from the operational Global Ensemble Forecast System version 12 (GEFSv12) at NCEP. Specifically, the GEFSv12 product consists of 30 perturbed ensemble members and one control forecast, all at a horizontal resolution of 0.5° (approximately 70 km grid spacing; available at https://registry.opendata.aws/collab/noaa/). The initial ensemble members are centered on the initial analysis produced by the global data assimilation system (GDAS) at NCEP. The perturbed forecasts are generated by integrating the EnKF posterior states centered on the control analysis state (Zhou et al., 2022). The GEFSv12 data at NCEP was selected due to its initialization with the EnKF scheme (Whitaker and Hamill, 2002; Zhou et al., 2022), which is recognized as one of the most advanced ensemble generation schemes worldwide. The GEFSv12 adopts the dynamic core of the Finite Volume Cubed-Sphere (FV3) model (Zhou et al., 2022). The 70 km grid spacing allows for the effective resolving of spatial scales ranging from over 10,000 km to about 280 km (4 × 70 km) (Ji and Toepfer, 2016; Zhou et al., 2024). Although the convective scale is not resolved in this forecast data, the range of the horizontal scales is sufficient to address the differences and regimes of ensemble sampling performance at various scales.

The operational GEFS forecast data are produced every 6 hours (i.e., 4 times per day) and each forecast extends up to 16 days ahead. For simplicity, this study evaluated only the ensemble and control forecasts initialized at 0000 UTC daily from 1 December 2020 to 28 February 2021, resulting in a total of 90 samples. The forecast output interval is 12 h. Since initial ensemble perturbations are superimposed on the model control variables, e.g., zonal wind (U) and temperature (T), to initialize ensemble forecasts, U and T at the mid-level (500 hPa) were selected for analysis.

3. Methodology

3.1 *Defining analysis error and ensemble perturbations*

The true state of the atmosphere (hereafter referred to as "truth", denoted by $\boldsymbol{T}$) can never be precisely known due to the sparseness and errors in monitoring or ob-

serving systems. Consequently, both historical and current atmospheric states are typically estimated by using data assimilation (DA) algorithms, such as the optimal interpolation and EnKF. These DA schemes combine model first guess fields (or background) with observational data to generate an optimal estimate of the atmospheric truth, referred to as the analysis state $\boldsymbol{A}$. Due to the unavoidable errors in observational systems and model first guess, and deficiencies in DA schemes, the analysis $\boldsymbol{A}$ generally differs from the true state $\boldsymbol{T}$. This difference is defined as the analysis error $\boldsymbol{e}$:

$$\boldsymbol{e} = \boldsymbol{A} - \boldsymbol{T}, \tag{1}$$

where $\boldsymbol{e} = (e_1, e_2, e_3, ..., e_M)^{\mathrm{T}}$, M represents the dimension of error vector $\boldsymbol{e}$. Here, M includes all analysis variables and vertical levels, and it also corresponds to the number of grid points of spatial variables.

The initial M-dimensional states are denoted as $\boldsymbol{P}_1$, $\boldsymbol{P}_2$, …, $\boldsymbol{P}_N$, (N is the number of ensemble members), and are centered around the optimally estimated state $\boldsymbol{A}$, i.e., $\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{P}_i = \boldsymbol{A}$. For an ensemble-based DA scheme, such as the EnKF, the analysis state $\boldsymbol{A}$ may be replaced by the mean state of posterior ensemble. The initial ensemble perturbations $\boldsymbol{p}_i$ are defined as the deviations of the individual ensemble members from their mean state and are given by:

$$\boldsymbol{p}_i = \boldsymbol{P}_i - \boldsymbol{A}, \tag{2}$$

where $\boldsymbol{p}_i = (p_{i,1}, p_{i,2}, p_{i,3}, ..., p_{i,M})^{\mathrm{T}}$.

3.2 *Defining analysis error covariance and ensemble perturbation covariance*

Let us assume that the analysis error is unbiased, i.e., $E(\boldsymbol{e}) = 0$, where $E(\cdot)$ represents the mathematical expectation. The analysis error covariance matrix denoted by $\boldsymbol{B}^{\mathrm{a}}$ is expressed as:

$$\boldsymbol{B}^{\mathrm{a}} = \begin{pmatrix} \overline{e_1e_1} & \overline{e_1e_2} & \cdots & \overline{e_1e_M} \\ \overline{e_2e_1} & \overline{e_2e_2} & \cdots & \overline{e_2e_M} \\ \vdots & \vdots & \ddots & \vdots \\ \overline{e_Me_1} & \overline{e_Me_2} & \cdots & \overline{e_Me_M} \end{pmatrix}, \tag{3}$$

where $\overline{e_je_j}$ represents the analysis error variance at grid point j and $\overline{e_je_k}$ represents the analysis error covariance between grids j and k.

The ensemble perturbation covariance matrix denoted by $\boldsymbol{B}^{\mathrm{p}}$ is similarly expressed as:

$$\boldsymbol{B}^{\mathrm{p}} = \begin{pmatrix} \overline{p_1p_1} & \overline{p_1p_2} & \cdots & \overline{p_1p_M} \\ \overline{p_2p_1} & \overline{p_2p_2} & \cdots & \overline{p_2p_M} \\ \vdots & \vdots & \ddots & \vdots \\ \overline{p_Mp_1} & \overline{p_Mp_2} & \cdots & \overline{p_Mp_M} \end{pmatrix}, \tag{4}$$

where $\overline{p_jp_j}$ represents the ensemble perturbation variance perturbations at individual grid point j, and $\overline{p_jp_k}$ represents the ensemble perturbation covariance between grid points j and k measuring the spatial correlation of perturbations between grid points, reflecting their coherent variability. Both quantities are computed by using ensemble members at a given time. Specifically, $\overline{p_jp_j} = \frac{1}{N-1}\sum_{i=1}^{N} p_{i,j}^2$ and $\overline{p_jp_k} = \frac{1}{N-1}\sum_{i=1}^{N} p_{i,j}p_{i,k}$.

3.3 *Ensemble sampling*

The primary objective of ensemble forecasting is to provide probabilistic future scenarios and quantify uncertainties in forecasts. This relies critically on the ability of the initial ensemble members to reflect the PDF of the true state (Epstein, 1969; Leith, 1974; Toth and Kalnay, 1997; Palmer et al., 2006). This challenge is fundamentally a problem of ensemble sampling, which is traditionally framed within the perturbation space. In the context of our study, ensemble sampling is defined as the ability of ensemble perturbations to represent the forecast errors of the ensemble mean. For effective ensemble initialization, it is essential that the initial ensemble perturbation covariance matrix closely approximates the analysis error covariance matrix. Given that the atmosphere is a high-dimensional, spatially continuous fluid, ensemble sampling must account not only for the magnitude of perturbations and errors but also for their spatially coherent structures, i.e., the spatial covariance (Palmer et al., 2006; Feng et al., 2020). This will be elaborated on in the following subsections.

3.3.1 *One-dimensional sampling*

In a one-dimensional scenario, $\boldsymbol{T}$, $\boldsymbol{A}$, and $\boldsymbol{P}_i$ are all scalar states. Assuming that analysis error $\boldsymbol{e}$ follows a Gaussian distribution, given control analysis state $\boldsymbol{A}$ as the maximum likelihood estimation (i.e., the mathematical expectation) and the standard deviation of analysis error $\boldsymbol{e}$, the PDF of the true state $\boldsymbol{T}$ can be accurately depicted (see the blue curve in Fig. 1a). If a set of perturbations is generated according to the same PDF as the analysis error and added to the control analysis state $\boldsymbol{A}$, i.e., $\overline{e_je_j} = \overline{p_jp_j}$ at a random grid point j, these perturbed ensemble members will serve as representative samples of true state $\boldsymbol{T}$. In the one-dimensional context, ensemble sampling considers only the magnitude of the perturbations and errors, without accounting for direction.

3.3.2 *Multi-dimensional sampling*

In multi-dimensional space, analysis error $\boldsymbol{e}$ is a vector with both magnitude and direction (the red arrow in Fig. 1b). Perturbation vectors $\boldsymbol{p}_i$ (the blue arrows) are generated to sample the analysis error within the phase

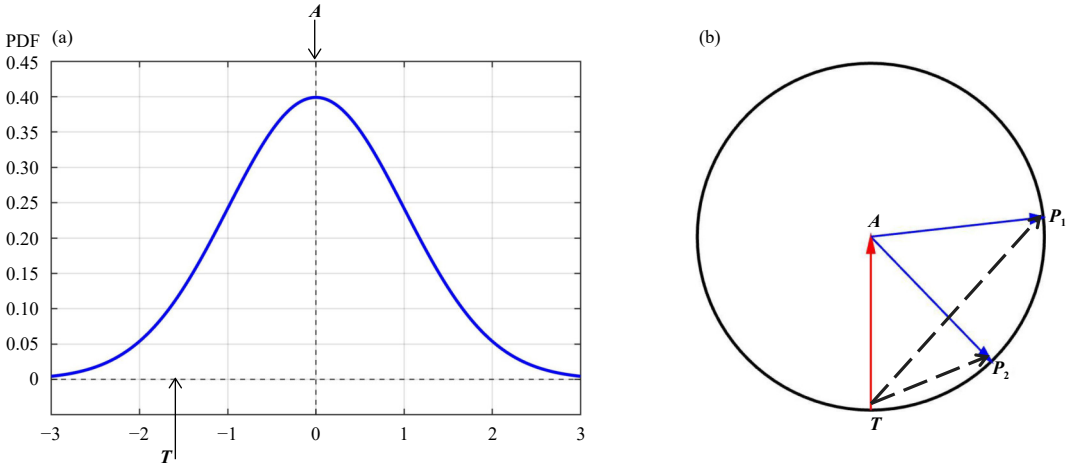

Fig. 1. Schematic diagrams of (a) one-dimensional sampling and (b) multi-dimensional sampling. See text for details on ***T***, ***A***, and ***P***.

space (black circle), encompassing two aspects.

(1) *Magnitude of analysis error*. This aspect is analogous to one-dimensional sampling, where the analysis error variance and initial ensemble perturbation variance at individual grid points should be consistent. This means that the diagonal components of $\boldsymbol{B}^{\mathrm{a}}$ and $\boldsymbol{B}^{\mathrm{p}}$ are equal, i.e., $\overline{e_j e_j} = \overline{p_j p_j}$ (j = 1, 2, …, M). In the case of perfect sampling, the perturbed members should lie on the same circle (the black circle) as the true state.

(2) *Direction of analysis error*. Unlike one-dimensional sampling, the multi-dimensional sampling considers whether the ensemble perturbations can represent the possible direction of the analysis error vector. The core of the directional sampling is the ability of the ensemble perturbation fields to capture the spatially coherent structures of the analysis error fields. For example, if analysis errors exhibit a negative–positive dipole pattern between two adjacent areas, the ensemble perturbations should present a similar or opposite-phase (i.e., positive–negative) pattern in the same regions. Theoretically, effective directional sampling implies that the covariances of analysis errors and ensemble perturbations are similar, i.e., $\overline{e_j e_k} = \overline{p_j p_k}$ (j = 1, 2, …, M; k = 1, 2, …, M; $j \neq k$). Estimating the analysis error covariance matrix ($\boldsymbol{B}^{\mathrm{a}}$) is challenging in practice due to the unknown truth. In ensemble-based DA, such as the EnKF, the perturbation covariance of posterior ensemble (i.e., $\boldsymbol{B}^{\mathrm{p}}$) serves as an effective approximation for $\boldsymbol{B}^{\mathrm{a}}$. One of the primary objectives of this study is to assess the performance of $\boldsymbol{B}^{\mathrm{p}}$ in approximating $\boldsymbol{B}^{\mathrm{a}}$.

3.4 *Scale decomposition of perturbation and error*

The primary objective of this study is to evaluate and analyze the performance of the multi-scale ensemble sampling at both initial and forecast times within a state-of-the-art operational ensemble prediction system. To achieve this, we conducted a Fourier decomposition of forecast errors and ensemble perturbations for selected variables (i.e., 500-hPa U and T) to derive their multi-scale components. This study focuses exclusively on zonal perturbations and errors; hence, a zonal decomposition of the variables within the 30° to 70° latitude band in the Northern Hemisphere (NH) was performed and assessed. This focus is due to the mid- to high-latitude circulation in the NH being predominantly characterized by zonal structures and dynamics (Zhang et al., 2019). The perturbation and error fields were zonally decomposed into 180 waves, covering scales from the meso-β scale (~100 km) to the planetary scale (~10,000 km) (Zhu et al., 2000). For the variables T and U, we performed a Fourier expansion in the zonal direction at each latitude. The decomposition followed the Discrete Fourier Transform. Taking T as an example, for a fixed latitude profile, the discrete Fourier transform can be written as:

$$\hat{T}(k,\varphi) = \sum_{n=0}^{N-1} T(\lambda_n,\varphi)\cdot \mathrm{e}^{-2\pi ikn/N},\ k = 0,1,2,....,N-1. \quad (5)$$

Here, $\hat{T}$ represents the Fourier coefficient of the kth wave number of T; k represents the wave number, which is dimensionless; λ_n indicates the longitude position of the nth grid point and φ represents latitude. Moreover, N represents the number of grid points in the zonal direction; n indicates the grid point index along the longitude direction; $\mathrm{e}^{-2\pi ikn/N}$ is a complex exponential function, representing the Fourier basis functions. Based on the above formula, forward transformation was performed to obtain the coefficients for each wave number, followed by inverse transformation to sum these coefficients and reconstruct the original field.

To clearly examine the properties of sampling performance at specific spatial scales, we categorized the

wavelengths and defined the wavebands for large, intermediate, and small scales. The large-scale wavebands span the wave numbers from 1 to 14 (approximately 2000 to 25,000 km), the intermediate-scale wavebands span wave numbers from 15 to 100 (400 to 2000 km), and the small-scale wavebands span wave numbers from 101 to 180 (200 to 400 km), as defined by Ying and Zhang (2017).

3.5 *Metrics to evaluate sampling performance*

To quantitatively assess the statistical relationship between full and multi-scale perturbation and error fields, two main metrics were utilized. The first metric is the widely used spread–error correlation, measured by the pattern correlation coefficient (PCC) between the ensemble spread field and the ensemble mean error magnitude field. The ensemble spread field denoted as $\boldsymbol{s}^{\mathrm{p}}$ is defined as:

$$\boldsymbol{s}^{\mathrm{p}} = (\sqrt{\overline{p_1 p_1}}, \sqrt{\overline{p_2 p_2}}, ..., \sqrt{\overline{p_M p_M}})^{\mathrm{T}}, \tag{6}$$

where the superscript p represents perturbation. The ensemble mean error magnitude field $\boldsymbol{s}^{\mathrm{e}}$ is defined as:

$$\boldsymbol{s}^{\mathrm{e}} = (|e_1|, |e_2|, ..., |e_M|)^{\mathrm{T}}, \tag{7}$$

where the superscript e represents error, and $|\cdot|$ denotes the absolute value. Note that each element of $\boldsymbol{s}^{\mathrm{e}}$, i.e., the error magnitude $|e_j|$ represents a realization that adheres to the statistic of analysis error variance $\overline{e_j e_j}$. In the absence of direct estimation of analysis error variance, the spread–error correlation is used as an approximate to measure the consistency between the ensemble perturbation variance and ensemble mean error variance, i.e., the magnitude sampling, as demonstrated in previous studies (e.g., Whitaker and Loughe, 1998; Magnusson et al., 2008; Cui et al., 2021). The spread–error correlation can be computed by:

$$r^{\mathrm{m}} = \frac{(\boldsymbol{s}^{\mathrm{p}} - \overline{\boldsymbol{s}^{\mathrm{p}}})(\boldsymbol{s}^{\mathrm{e}} - \overline{\boldsymbol{s}^{\mathrm{e}}})}{\|\boldsymbol{s}^{\mathrm{p}} - \overline{\boldsymbol{s}^{\mathrm{p}}}\| \, \|\boldsymbol{s}^{\mathrm{e}} - \overline{\boldsymbol{s}^{\mathrm{e}}}\|}, \tag{8}$$

where the superscript m represents magnitude; $\|\cdot\|$ denotes the L_2 norm; and $\overline{\boldsymbol{s}^{\mathrm{p}}} = \frac{1}{M}\sum_{j=1}^{M}\sqrt{\overline{p_j p_j}}$, while $\overline{\boldsymbol{s}^{\mathrm{e}}} = \frac{1}{M}\sum_{j=1}^{M}|e_j|$.

A higher value of the metric r^{m} indicates better performance of the ensemble spread in representing analysis and forecast uncertainties (Whitaker and Loughe, 1998; Magnusson et al., 2008; Cui et al., 2021). Despite its extensive application, the spread–error correlation has limitations, as it provides an overall measure of grid-point spread–error consistency without accounting for the spatial coherence between grid points.

To address the limitations of the spread–error correlation, we adopted an additional metric to assess the directional relationship between perturbations and errors in high-dimensional space. Specifically, we used the PCC between ensemble perturbation fields and ensemble mean error fields to evaluate the ability of perturbation fields in capturing the spatial coherence of errors, i.e., the directional sampling. The perturbation–error correlation for the ith ensemble perturbation denoted by r_i^{d} is computed by:

$$r_i^{\mathrm{d}} = \frac{(\boldsymbol{p}_i - \overline{\boldsymbol{p}_i})(\boldsymbol{e} - \overline{\boldsymbol{e}})}{\|\boldsymbol{p}_i - \overline{\boldsymbol{p}_i}\| \, \|\boldsymbol{e} - \overline{\boldsymbol{e}}\|}, \tag{9}$$

where the superscript d represents direction; and $\overline{\boldsymbol{p}_i} = \frac{1}{M}\sum_{j=1}^{M} p_{i,j}$, while $\overline{\boldsymbol{e}} = \frac{1}{M}\sum_{j=1}^{M} e_j$.

The perturbation–error correlation is similar to the PECA in Wei and Toth (2003) and has been rarely addressed in preceding studies. Both metrics assess the ability of ensemble perturbations to explain forecast error variance[1] [i.e., $(r_i^{\mathrm{d}})^2$]. A higher explained variance indicates a large potential of the ensemble perturbations for constructing the forecast error covariance matrices. The main distinction between the two metrics is that the perturbation–error correlation evaluates the performance of individual perturbations, while PECA assesses the cumulative explained variance across all ensemble perturbations. We employ the perturbation–error correlation specifically to examine the range of correlations between forecast errors and each individual perturbation, including the maximum, minimum, and mean.

Assuming that ensemble perturbations have the same magnitude as the analysis error, the perturbation–error correlation (i.e., the angle between red and blue arrows in Fig. 1b) modulates the accuracy of initial perturbed members (i.e., the error amplitude of $\boldsymbol{P}_i$ from $\boldsymbol{T}$, as indicated by black dashed arrows in Fig. 1b). For example, a close-to-zero perturbation–error correlation (i.e., $\boldsymbol{P}_1\boldsymbol{A}$ and $\boldsymbol{AT}$) corresponds to a larger error in ensemble members than those with a high negative correlation (i.e., $\boldsymbol{P}_2\boldsymbol{A}$ and $\boldsymbol{AT}$).

Since the "truth" is unknown, the control analysis (i.e., the initial condition of the control forecast) is used as a proxy for the "truth" in the computation of forecast er-

[1]A higher perturbation–error correlation indicates that the perturbation vector has a greater projection onto the direction of the error vector, thereby accounting for a larger portion of the variance contained in the error vector. The explained variance equals the square of the perturbation–error correlation.

rors throughout all evaluations in this study. The control analysis is consistent with the mean of the initial ensemble members, as all members are centered around the control analysis in the NCEP ensemble forecast system. The assessment of the initial ensemble sampling is a key focus of this study. However, when the analysis is used as the "truth", the analysis errors are zero. Therefore, the performance of the ensemble sampling at 12 h is evaluated as an approximation of the initial ensemble sampling. The rationale behind this approach is discussed in detail in Section 4.

4. Sample-mean RMSE and spread

The generation of initial perturbed members significantly affects the quality of individual perturbed forecasts and their mean forecasts. Consequently, before examining the sampling performance of ensemble perturbations, we first assessed the sample-mean RMSE for 500-hPa U and T of the control forecast, ensemble mean forecast, individual perturbed forecasts, as well as the ensemble spread (Fig. 2).

Figure 2 illustrates that the RMSE of the ensemble mean forecast is substantially lower than that of the control forecast for both variables (cf. red and blue solid lines) as documented in numerous previous studies (e.g., Leith, 1974; Atger, 1999; Swinbank et al., 2016). Additionally, the mean RMSE of individual perturbed forecasts, averaged across all ensemble members, is generally higher than that of the control forecast (cf. black and blue solid lines). This suggests that, on average, the perturbed ensemble forecasts exhibit lower forecast accuracy compared to the control forecast (Palmer et al., 2006; Feng et al., 2024). This lower accuracy is primarily due to the initial ensemble sampling, which is further discussed below. However, certain ensemble members demonstrate the capability to surpass the control forecast in terms of accuracy after approximately four days (see the lower boundary of the gray shading) due to the nonlinear error growth.

Notably, the ensemble spread is generally close to, albeit slightly below, the ensemble mean error (cf. red dashed and solid lines), indicating a reasonable spread–error relationship for the GEFS (Toth et al., 1998; Zhou et al., 2022). The RMSE for both the control and ensemble mean (blue and red solid lines) at short lead times may be underestimated due to errors in the reference analyses. However, as shown in our previous study (Feng et al., 2023), this underestimation becomes less significant over long time. Note that the ensemble spread intersects with the ensemble mean error between 12 and 24 h. Consequently, the relationship between ensemble perturbations and ensemble mean error at 12 h is critically analyzed as a proxy for evaluating the initial ensemble sampling in subsequent results. Since the control analysis is consistently employed as the reference for all evaluations across all lead times in our study, this approach may be considered justifiable. The evaluation of the ensemble sampling performance at 24 h shows consistent conclusions.

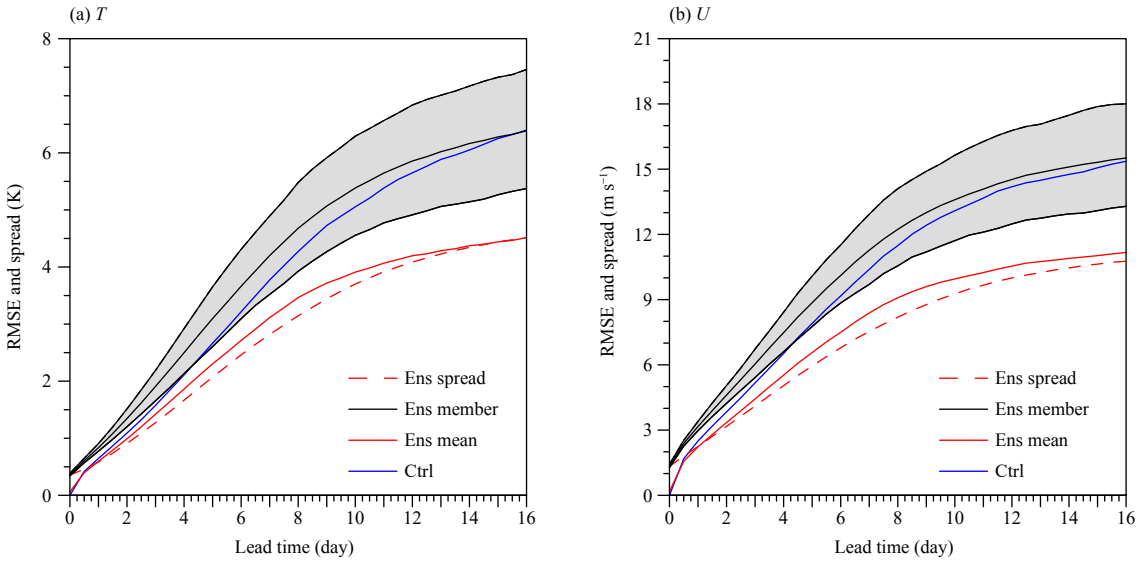

Fig. 2. Sample mean root mean square error (RMSE) of control (blue solid line), ensemble mean (red solid line), and the mean RMSE of individual perturbation members (black solid lines) for forecasts over all 90 cases: (a) temperature (T) and (b) zonal wind (U) fields at 500 hPa over the NH. The ensemble spread is represented by red dashed lines. The upper and lower bounds of the light blue shading represent the maximum and minimum RMSE values of the ensemble members for each lead time and case, averaged across all 90 forecast cases.

5. Initial sampling of full fields

5.1 *Magnitude*

As discussed in Section 3.3.2, sampling in high-dimensional space involves assessing the magnitude and direction of perturbation and error vectors (or fields). Figure 2 demonstrates that at 12 h, the spatial-mean ensemble spread and error magnitude are close. It is important to further compare the spatial distribution of gridpoint ensemble spread and analysis error. Figure 3 illustrates the initial analysis error magnitude and ensemble spread over the NH for 500-hPa T and U in a randomly selected case. It reveals that the initial ensemble spread field can roughly approximate the distribution of analysis error uncertainties.

The spread–error correlations [Eq. (8)] for 500-hPa U and T in this case are 0.41 and 0.44, respectively, which are close to the average spread–error correlations of 0.40 and 0.43 over all samples. This suggests the effectiveness of the initial EnKF ensemble in quantifying flow-dependent uncertainties in analysis errors. This principle also fundamentally motivates the application of ensemble-based estimation of background error covariance in DA (Houtekamer and Mitchell, 1998; Whitaker and Hamill, 2002; Evensen, 2003; Zhou et al., 2022), which could improve the quality of initial analysis. It is noteworthy that the initial ensemble spread field exhibits smoother structures compared to the initial analysis error. This is because the calculation of ensemble spread involves averaging deviations of all ensemble members from their mean state, whereas the ensemble mean error (i.e., the analysis error in this context) represents a single realization of the diagonal component of the analysis error covariance. The spread–error correlation is expected to be somewhat lower when the actual initial conditions at time zero, rather than at 12 h, are considered, owing to the errors inherent in the control analysis.

5.2 *Direction*

Similar results as in Fig. 3 have been reported in numerous preceding studies as a fundamental measure of the performance of initial perturbation sampling. However, it is important to emphasize that the spread–error correlation primarily offers a gridpoint assessment of the consistency between the initial ensemble spread and the magnitude of analysis error. In other words, this assessment involves comparing only the diagonal components of the ensemble perturbation covariance and the analysis error covariance matrices, while neglecting the spatial covariance of perturbations and errors as discussed in Section 3.5. This section explores the directional relationship between the initial ensemble perturbations and the analysis errors, providing an additional perspective on the assessment of initial ensemble sampling.

Figure 4, which resembles the schematic of sampling in high-dimensional space shown in Fig. 1b, is derived based on the operational forecast data of GEFSv12 for 500-hPa T and U over the NH at 12 h. The true state is

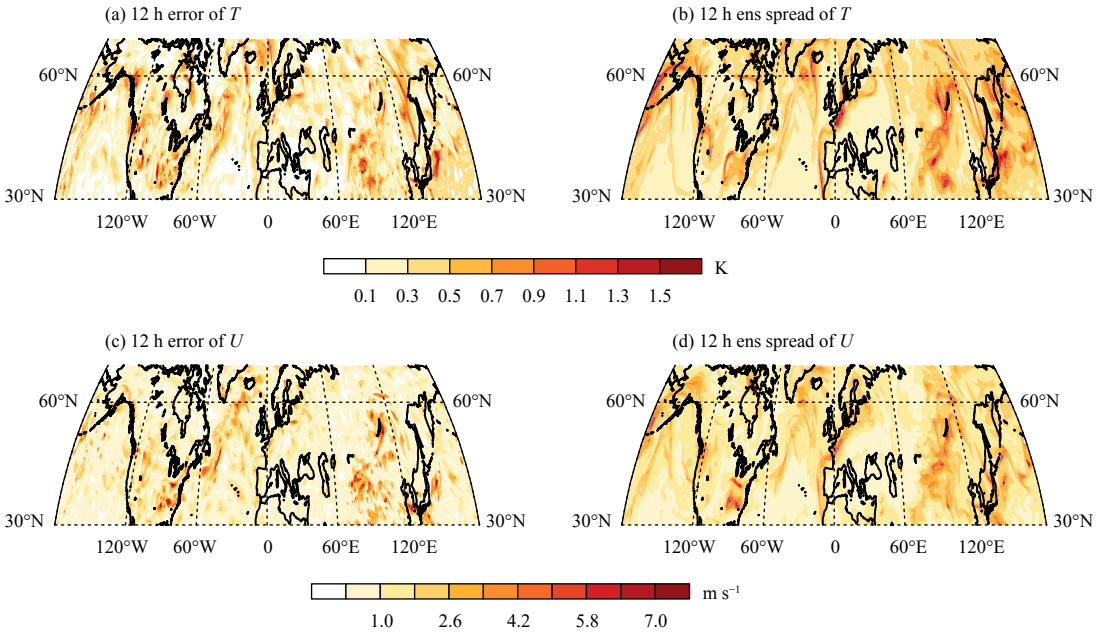

Fig. 3. (a) Absolute error field of the ensemble mean and (b) ensemble spread field for 500-hPa T at 12-h lead time. (c, d) are the same as (a, b), respectively, but for 500-hPa U. The initial time of this case is 0000 UTC 25 February 2021.

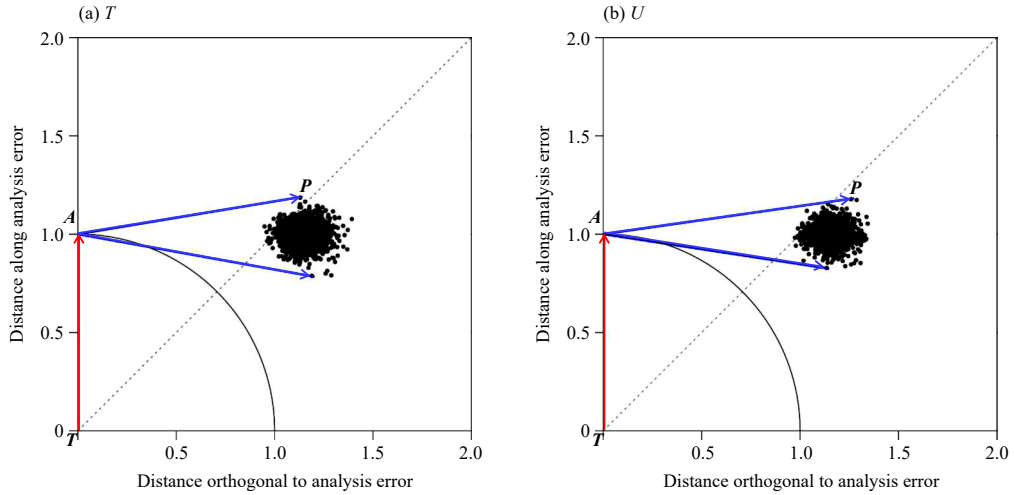

Fig. 4. Scatter plots of standardized RMSE of all ensemble members (***P***, black dot) from the corresponding true reference (***T***) in all 90 cases for 500-hPa (a) *T* and (b) *U* at 12-h lead time. The *x*- and *y*-axes represent the projection components of perturbation member errors on the orthogonal direction of the analysis error and on the analysis error, respectively. The ensemble perturbations and analysis errors are depicted by blue and red arrows, respectively; and ***A*** represents the analysis. The black arc with a radius of 1 represents the amplitude of analysis error.

denoted by ***T*** (the control analysis in this context) at the coordinate (0, 0), and the analysis state is denoted by ***A*** (the ensemble mean in this context) at the coordinate (0, 1), indicating that the analysis error is oriented along the *y*-axis (red arrows). The black scatter points represent all ensemble members over all 90 cases (i.e., 30 × 90 = 2700) at 12 h. The positions of these points are determined by the components of the perturbed member errors (i.e., perturbed states $\boldsymbol{P}_i$ minus the true state ***T***) projected onto the analysis error (*y*-axis) and the orthogonal direction (*x*-axis). The distance between individual black dots (***P***) from the state ***T*** measures the RMSE of individual ensemble members from the "truth" reference, and the distance between ***P*** and ***A*** measures the magnitude of ensemble perturbations. Note that these error components of the perturbed members in Fig. 4 are standardized by the analysis error in each case.

The results in Fig. 4 indicate that for both 500-hPa *T* and *U*, the initial ensemble perturbation vectors (blue arrows) in high-dimensional space (i.e., the NH) exhibit very low PCCs with the analysis error (red arrow) ranging from −0.18 to 0.17, using temperature *T* as an example. This result is consistent with Feng et al. (2024) based on the "perfect" ensemble forecast data (using a random member as the ground "truth") and can be attributed to the "curse of dimensionality". The superposition of initial ensemble perturbations on the analysis state in high-dimensional space results in all initial perturbed states deviating further away from the true state. This is evidenced by the fact that all ensemble member states (black scatter points) fall outside the black arc, each with errors exceeding the analysis error (i.e., greater than 1). These outcomes underline the challenges of initial sampling in high-dimensional space and explain the larger errors observed in perturbed members relative to the control and ensemble mean at 12 h, as depicted in Fig. 2.

6. Sampling at multiple spatial scales

6.1 *Decomposition of perturbation and error*

In addition to the sampling performance of full fields, it is important to examine the distinct sampling performance of ensemble perturbations at multiple spatial scales. Figure 5 provides an example of the full fields of initial perturbations of 500-hPa *T* (i.e., 12-h perturbations as a proxy) for the same case as Fig. 3, along with their respective components decomposed into large, intermediate, and small scales defined in Section 3.4. The results show that the full initial perturbation at 500 hPa of NH is primarily contributed by large- and intermediate-scale components, which exhibit higher perturbation strength than the small scales (see results in Section 6.2). The spatial correlation coefficients between the full and large-scale fields of initial perturbations are 0.66 for 500-hPa *T* and 0.64 for 500-hPa *U* (figures not shown), which are slightly lower than those between the full and intermediate-scale fields. The large-scale perturbation displays wave-like patterns (Fig. 5b) associated with synoptic-scale baroclinic instabilities, whereas the small-scale perturbation exhibits more localized features that may be linked to regional mesoscale instabilities (Tan et al., 2004; Sun, 2017).

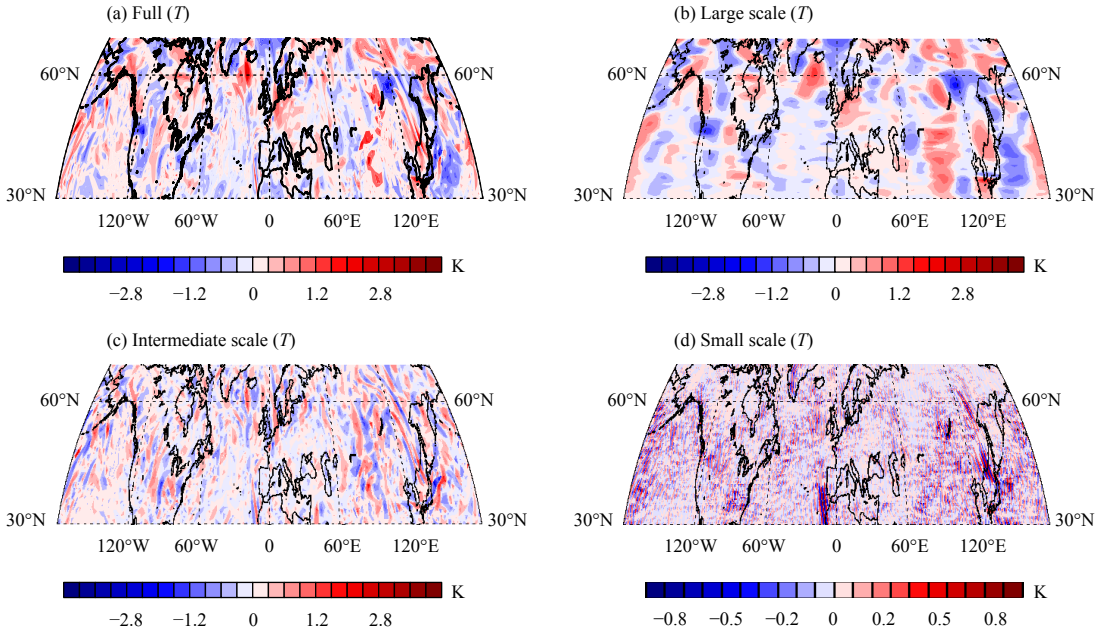

Fig. 5. (a) Full perturbation field of a randomly selected ensemble member for 500-hPa *T* at 12-h lead time in the same case as Fig. 3 and its (b) large-, (c) intermediate-, and (d) small-scale components. Note that the color bar for the small scale differs from those for the other scales.

6.2 *Magnitude of perturbation and error at multiple scales*

Due to the multi-scale nature of the atmosphere, well-generated initial ensemble perturbations should reflect the multi-scale characteristics of analysis error. Beyond comparing the full perturbation and error fields (Fig. 3), we further examine the initial ensemble spread and the analysis error magnitude for large, intermediate, and small scales for 500-hPa *T* and *U* (Fig. 6). Those for 500-hPa *U* are qualitatively similar and thus are not shown. While the initial error magnitude and ensemble spread fields at large scales show some similarities (Fig. 6a vs. 6b), their PCCs (0.28 for 500-hPa *T* and *U*) are considerably lower than those for full fields and other scales. This could be attributed to the notable phase errors of large-scale perturbations and errors in global space (Feng et al., 2020; Jankov et al., 2022), such as northeastern China in Figs. 6a, b (highlighted by the blue box). In contrast, the PCCs for intermediate-scale perturbations and errors are 0.40 for 500-hPa *T* and 0.44 for 500-hPa *U*, which is similar to those of full fields, 0.40 and 0.43 respectively. Noticeably, the spread–error correlation for the small scale is the highest among all defined scales, despite their larger randomness (Fig. 5). This high correlation is likely due to its strongly localized features of perturbation and error. For example, aside from a few regions exhibiting remarkable strength (highlighted by the blue box in Figs. 6e, f), the intensity of perturbation and error in other areas across the globe is significantly weaker. The sample-mean results are qualitatively simiar to those in this case (figures omitted).

Figure 7 compares the spectrum of perturbation and error variances from 1 to 180 wavenumbers for 500-hPa *T* and *U* at various lead times. At the 12-h lead time, the perturbation variances for both variables are generally larger than the corresponding error variances across nearly all scales (cf. purple solid and dashed lines). The larger perturbation variance may be intentionally designed and tuned to ensure the consistency between the perturbation and error variance over longer lead times, as shown by the results at 24, 48, 192, and 360 h. By 12 h, the ensemble perturbation variance at small scales has already reached the saturation level and shows minimal change thereafter (Sun and Zhang, 2020). Apparent upscale error growth for both the ensemble perturbations and errors can be seen at increasingly larger scales, consistent with findings from previous studies (e.g., Bei and Zhang, 2014; Sun, 2017; Sun and Zhang, 2020). This is clearer when examining the aggregate perturbation and error variances at large, intermediate, and small scales (hollow triangles and circles).

6.3 *Direction of perturbation and error at multiple scales*

In addition to magnitude, the relationship between the direction of perturbation and error is another critical metric for evaluating ensemble sampling. Figure 8 displays the maximum (red), minimum (blue), and mean (black) PCCs between the 12-h ensemble perturbation and error

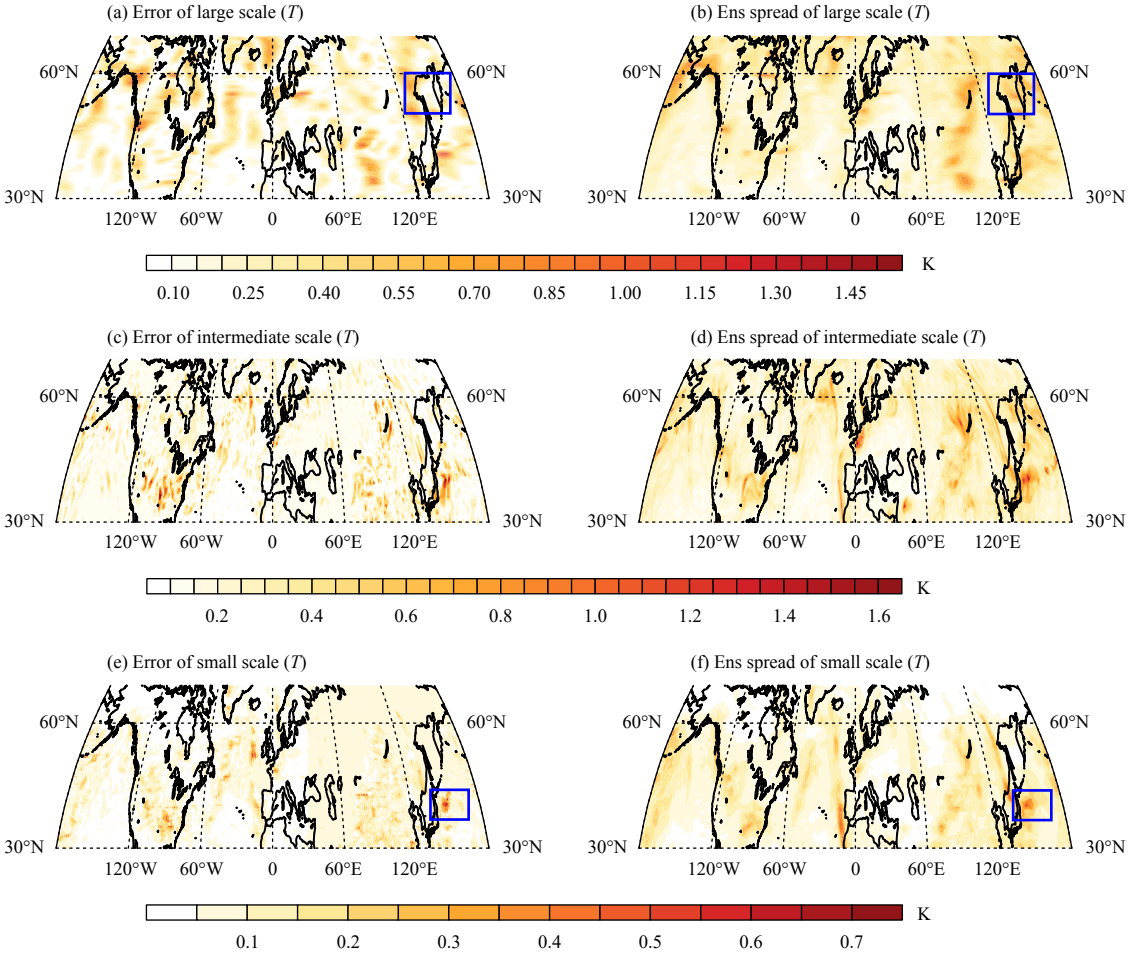

Fig. 6. As in Figs. 3a, b, but for the (a, b) large-, (c, d) intermediate-, and (e, f) small-scale components of error magnitude and ensemble spread of 500-hPa T at 12-h lead time in the same case. Regions with obvious strength are highlighted by blue boxes. The PCCs of error and spread for large, intermediate, and small scales are 0.28, 0.42, and 0.49, respectively. Note that the color bars for the three defined scales are different.

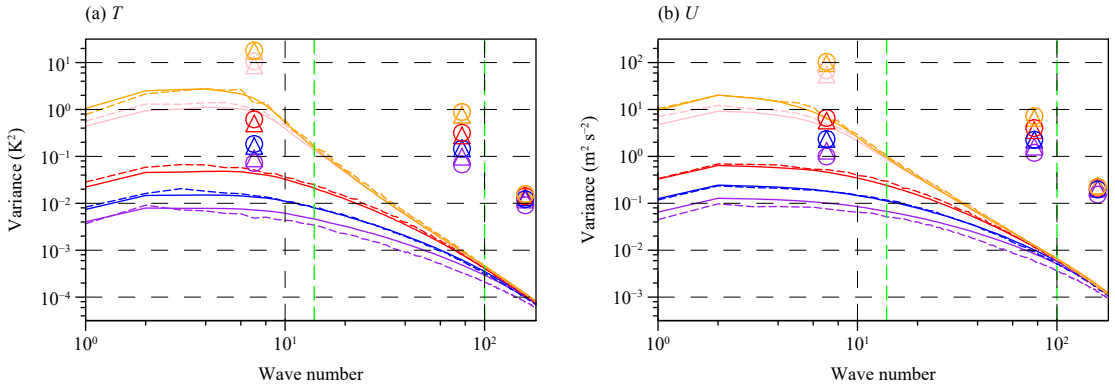

Fig. 7. Spectrum of perturbation (solid lines) and ensemble mean error (dashed lines) variances of 500-hPa (a) T and (b) U at 12- (purple), 24- (blue), 48- (red), 192- (pink), and 360-h (orange) lead times. The triangles and circles correspond to the accumulative perturbation and ensemble mean error variances of all wave numbers in the corresponding wavelength range, respectively, for large, intermediate, and small scales from left to right (divided by green dashed lines) in each panel.

over the NH (solid curves) and eastern China (corresponding colors but dashed lines, with details discussed in Section 6.4) across individual wave numbers, averaged over all samples. The PCC range over the NH between the maximum and minimum gradually narrows from larger to smaller scales, varying from about ±0.4 for

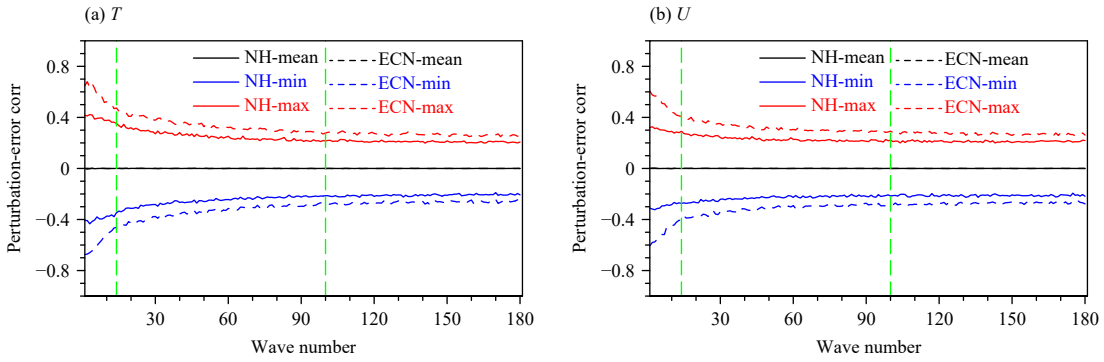

Fig. 8. Maximum (red), minimum (blue), and mean (black) perturbation–error correlations in each case for 500-hPa (a) *T* and (b) *U* at 12-h lead time, averaged over all 90 cases. The solid and dotted lines represent the NH and Eastern China (ECN), respectively. The green dashed reference lines divide the large, intermediate, and small scales.

wavenumber 1 to about ±0.25 for wavenumber 180. This implies that the initial perturbations (i.e., 12-h perturbations as a proxy) at larger scales can potentially capture a greater variance in analysis error at these corresponding spatial scales compared to smaller scales. In other words, the initial perturbation fields are more effective in simulating or representing analysis error fields for larger-scale, spatially coherent structures. This is essentially attributed to the reduced effective degrees of freedom (DOF) in larger-scale atmospheric dynamics. Moreover, the positive and negative PCCs are nearly symmetric across all wave numbers, resulting in a mean PCC close to zero (see the black solid curve). This highlights the efficacy of the EnKF initial perturbations at NCEP in the randomness and equal possibility of the sampling, despite the EnKF scheme not using the positive and negative paired initial perturbations (Wei and Toth, 2003; Wang et al., 2014). Such performance makes the identification of possible future scenarios in an equally likely and unbiased manner.

It is noteworthy that the error magnitudes in individual ensemble members, relative to the true reference (i.e., the control analysis), present an overall narrower range moving from larger to smaller scales (Fig. 9), which is closely related to the directional sampling performance of initial perturbations in Fig. 8. For larger-scale variable fields, it is more likely to generate perturbed initial members with comparable or even higher accuracy relative to the control analysis (see also Fig. 10) due to the enhanced directional sampling (i.e., higher perturbation–error correlation). In contrast, the initial sampling of small-scale errors is more like a random sampling with much less knowledge and information about small-scale analysis error covariance in a higher dimensional space. This underscores the considerable challenges involved in generating initial perturbations for meso- and convective-scale ensemble forecasting, where accurate representation of small-scale dynamics is crucial for short-term forecasts.

Similar to Fig. 4, the scatter plots present in Fig. 10 exhibit the relationship between initial (12 h as a proxy) ensemble perturbations and errors over the NH in their magnitude and direction between all members and samples for wave numbers 1 (Fig. 10a), 30 (Fig. 10b), and 180 (Fig. 10c) of 500-hPa *T*, while 500-hPa *U* presents qualitatively similar results (figures not shown).

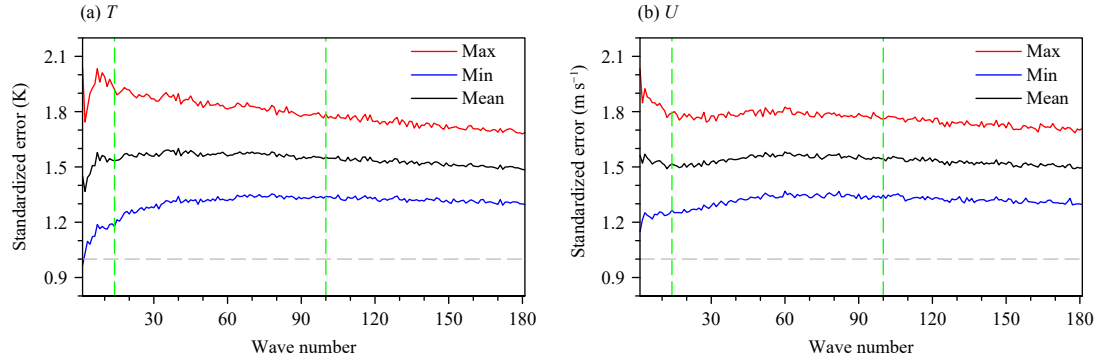

Fig. 9. As in Fig. 8, but for the standardized RMSE of perturbation members from the true reference over the NH. The gray dashed line is the reference line for RMSE of one.

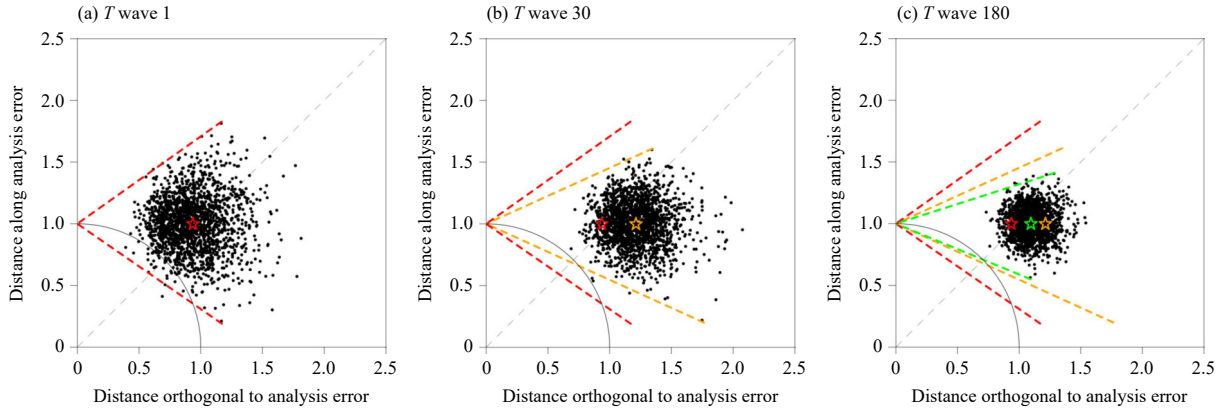

Fig. 10. As in Fig. 4, but for wave numbers (a) 1, (b) 30, and (c) 180 of 500-hPa *T*. Their maximum and minimum perturbation–error correlations are highlighted by red (wave number 1), orange (wave number 30), and green (wave number 180) dashed lines. The stars represent the mean values of the *x*-axis and *y*-axis components of all perturbation members for wave numbers 1 (red), 30 (orange), and 180 (green).

The distance between individual black dots and the point (0, 1) represents the magnitude of the initial perturbation. The results show that while the mean magnitudes of initial perturbations are overall close to one (see the stars), i.e., the magnitude of analysis error, there is considerable variability in perturbation magnitudes across members. For the selected wave numbers, the perturbation [i.e., distance between black scatters and the point (0, 1)] components along the analysis error (*y*-axis, predominantly less than 0.5) are markedly smaller than those orthogonal to analysis error (*x*-axis, mostly from 0.5 to 1.5). This indicates that the wave components in initial perturbed members are overall less accurate than the corresponding wave components in the control analysis, similar to the full initial fields in Fig. 4. However, the initial perturbations at larger scales (e.g., wave number 1) show an overall greater projection (38.87%, 27.09%, and 22.27% for wavenumbers 1, 30, and 180, respectively) on the analysis error vector (i.e., a larger expansion angle by the red dashed lines), resulting in a somewhat higher number of perturbed members exhibiting lower error compared to the control analysis (1.89%, 0%, and 0% for wavenumbers 1, 30, and 180, respectively), i.e., more black dots are located within the black arc. This result is consistent with those in Figs. 8, 9, indicating that the initial ensemble members perform better in representing the truth at larger scales than at smaller scales.

6.4 *Sensitivity of initial sampling performance to domain size*

This subsection explores the initial sampling performance within a localized area, specifically Eastern China (ECN, ranging from 21° to 38.5°N and 113° to 123°E), which encompasses approximately 11% of the NH region. The range of PCCs between initial perturbations and analysis errors broadens for all wave scales when focusing on this smaller region. This effect is particularly notable for larger wave scales compared to smaller ones (see dashed lines in Fig. 8). This expansion in range suggests that the performance of directional sampling, or the explained error variance by perturbations is highly dependent on the dimension of perturbation and error vectors. Furthermore, the comparative analysis of the NH and ECN in Fig. 8 implies that the same initial perturbation may perform better in directional sampling in certain regions while performing worse in others. Such behavior appears to be random. The discrepancy in sampling effectiveness between larger and smaller scales within a confined region underscores the potential need for blending different schemes for various scales in the generation of perturbations within a regional ensemble prediction system. Taking convection-scale ensemble forecasting as an example, generating initial large scale to mesoscale perturbations may focus on better representing the physically consistent and spatially coherent error structures of weather systems. This can provide reasonable environmental forcing for convection development and improve the accuracy of longer-range ensemble forecasts. For smaller scales such as convection storms, however, it might be more critical to sample its uncertainties in intensity and position. This can be achieved through strategies such as assimilating radar observations and incorporating fine-scale local topographical data.

7. Ensemble sampling of forecasts

7.1 *Temporal variation of spread–error correlation*

One of the most critical roles of ensemble forecasts is to quantify forecast uncertainties through the ensemble spread. Figure 11 shows the spread–error correlation

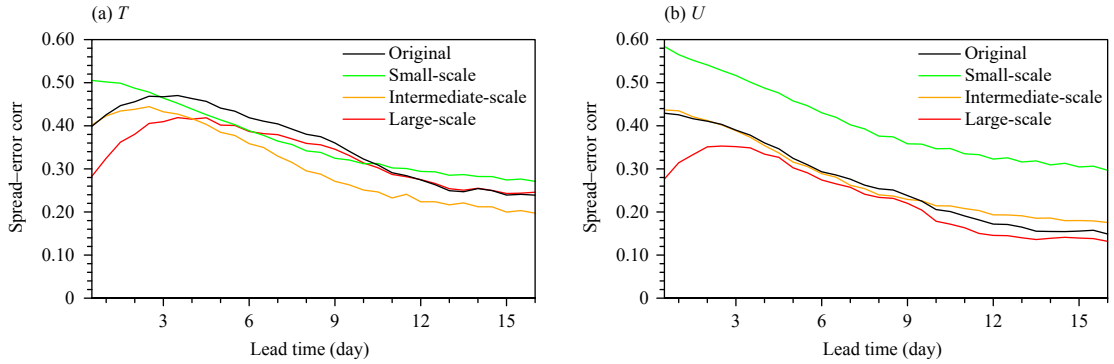

Fig. 11. PCCs between ensemble spread and ensemble mean absolute error over the NH as a function of lead time for full fields (black solid lines), and the large- (red), intermediate- (orange), and small-scale (green) components for 500-hPa (a) *T* and (b) *U*.

across large (red), intermediate (orange), and small (green) scales of 500-hPa *T* and *U* as a function of lead time. The spread–error correlation is computed as the sample mean of the spatial correlation over the NH. As demonstrated in Barker (1991), the spread–error correlation is generally below 0.5 even in the mid-level troposphere of the NH extratropics, where there is no model systematic errors and model dynamics and physics are accurately represented. However, the ensemble forecasts across the three wave bands exhibit distinct characteristics of the spread–error relationship. At a 12-h lead time, the large scale demonstrates the lowest spread–error correlation, followed by the intermediate and small scales for both variables, which is similar to that in Fig. 6. The initial spread–error correlation for the full fields and the intermediate scale is similar, ranging from approximately 0.40 to 0.45. This correlation is much higher than that using the bred vector as initial perturbations [about 0.13 as shown in Whitaker and Loughe (1998)], suggesting that there have been significant improvements in capturing the statistics of analysis errors in ensemble systems since 1998. It could be possible that the EnKF perturbations may be more effective than the bred vectors in capturing the statistics of analysis errors.

For 500-hPa *T*, the spread–error correlation of the full ensemble fields (black line) increases in the short range, peaks at about 0.50 around day 3.5, and then gradually declines to about 0.25 by day 16. This behavior is qualitatively similar to Fig. 4 in Whitaker and Loughe (1998) and is predominantly contributed by the large- and intermediate-scale ensemble. The large- and intermediate-scale ensembles exhibit qualitatively similar temporal variations in spread–error relationship. However, the large-scale ensemble shows a peak at approximately day 4.5, while the intermediate-scale ensemble reaches its peak at around day 2.5. This suggests that the short-range increase in the spread–error correlation is likely due to the quasi-linear error growth regime of initial perturbations and errors with similar dynamic instabilities. Beyond this stage, the influence of nonlinear dynamics becomes apparent, leading to a gradual reduction in the similarity between spread and error. Notably, the variation in the spread–error correlation of large-scale ensemble, due to its dominant variance, aligns closely with that of full ensemble fields beyond five days (cf. red and black lines). In contrast, owing to the much shorter predictability at small scales (typically a few hours), their spread–error correlation decreases from the beginning but remains the highest through all the lead times. This is possibly because these small-scale perturbations and errors become random noises, the frequency and amplitude of which are modulated by larger-scale local forcing like topography (Berner et al., 2009; Wang et al., 2014; Bai et al., 2021), and air–sea breeze (Coutinho et al., 2004; Chen et al., 2016; Yang et al., 2024).

The results for 500-hPa *U* are generally similar to those for 500-hPa *T* (Fig. 11). However, there is no short-range increase in spread–error correlation for intermediate-scale and full ensemble fields in the case of 500-hPa *U*. This is possibly due to the much shorter range of linear dynamics for zonal wind compared to temperature (Sardeshmukh and Hoskins, 1988; Ting, 1996). Additionally, the spread–error correlation for small scales of 500-hPa *U* is substantially higher than other scales through all lead times, which is more pronounced compared to the difference for 500-hPa *T*.

7.2 *Temporal variation of perturbation–error relationship*

Figure 12 shows the temporal variation of the maximum, minimum, and mean PCCs between ensemble perturbation and ensemble mean error fields over NH for large, intermediate, and small scales. The PCC range for

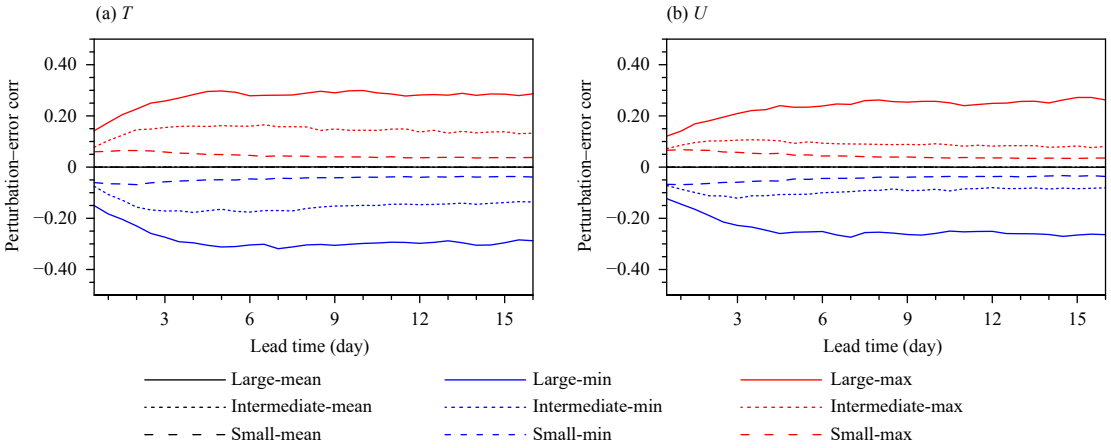

Fig. 12. Maximum (red), minimum (blue), and mean (black) perturbation–error correlations for 500-hPa (a) *T* and (b) *U* over the NH as a function of lead time, averaged over all cases. The solid, short dashed, and long dashed lines represent the large, intermediate, and small scales, respectively. Note that all the black lines, representing the mean values for different scales, are superimposed on top of each other.

both large and intermediate scales increases during the initial few days (about 4.5 days for large scales and 2.5 days for intermediate scales). This could be attributed to the initial transient linear growth regimes of perturbations and errors, which drive their projection onto the leading Lyapunov vector (Kalnay, 2002). This is evidenced by the fact that the timing at which the ensembles reach a saturated level closely corresponds to the peak time of their spread–error correlation (Fig. 11). Subsequently, the PCC range for large scales maintains an approximately saturated level for both variables, despite a loss of predictability. This could be explained by the gradual decrease of the effective DOF of perturbations and errors, as their energy accumulates at larger scales. In contrast, the PCC range for small scales decreases steadily from the beginning to the end at a rather slow rate.

8. Conclusions and discussion

This study examines the multi-scale characteristics of ensemble perturbations in sampling errors in both initial analyses and forecasts. The ensemble forecast data analyzed are from the operational GEFS at NCEP, covering the period from 1 December 2020 to 28 February 2021. Two representative variables, zonal wind (*U*) and temperature (*T*), were evaluated at 500 hPa in the midlatitudes over NH. For a quantitative assessment, we employed two metrics, including the PCC between ensemble spread and ensemble mean error magnitude (i.e., spread–error correlation) and the PCC between ensemble perturbations and ensemble mean errors (i.e., perturbation–error correlation). These metrics assess the performance of ensemble sampling in terms of both magnitude and direction, respectively.

The results indicate that the sampling performance of initial ensemble perturbations varies significantly across different spatial scales. For sampling in terms of magnitude, the spread–error correlation is higher at smaller scales compared with larger scales. This can be attributed to the highly localized nature of small-scale perturbation and error, whereas large-scale perturbations and errors tend to cover broader regions and exhibit notable phase deviations in certain regions. In contrast, the directional sampling shows an opposite trend. The initial perturbations at larger scales can better capture or represent analysis error for spatially coherent, large-scale structures compared to smaller scales. This difference in multi-scale sampling becomes even more pronounced for smaller domains. The decrease in perturbation–error correlation at smaller scales is likely due to an increase in the effective DOF. The contrast between sampling in magnitude and direction suggests that, for large-scale analysis errors, it may be more important to accurately sample the spatial covariance and coherence of these analysis errors, benefiting longer-range ensemble forecasts. On the other hand, for small-scale analysis errors, the focus may need to be on representing the magnitude or strength of these errors, which tend to be more random, as seen in phenomena like convective storms.

Our results indicate that, in addition to initial sampling, ensemble sampling during the forecast also varies across different spatial scales. In terms of spread–error correlation, both large- and intermediate-scale components exhibit similar behavior to the full ensemble fields, characterized by a gradual increase during the early stages, followed by a decline at longer lead times. Notably, the spread–error correlation peaks later for large scales (approximately day 4.5) compared to intermediate

scales (around day 2.5), which is attributable to differences in the duration of linear error growth dynamics. In contrast, due to the significantly shorter predictability at small scales (typically a few hours), their spread–error correlation decreases from the beginning of the forecast but remains the highest across all spatial scales throughout the forecast period. This persistent high correlation may be attributed to the influence of predictable larger-scale environmental circulation, as well as localized forcing mechanisms that contribute to small-scale uncertainties. These include mesoscale wind shear and jet, topographic influences, and ocean–land thermodynamic effects. The temporal evolution of spread–error correlation of the ensemble forecast is initially dominated by intermediate-scale components during the first few days and subsequently by large-scale components as the lead time increases.

Unlike spread–error correlation, the correlation between perturbations and errors consistently exhibits higher values for larger spatial scales throughout the forecast lead times. For both large and intermediate scales, the perturbation–error correlations increase gradually during the initial stages, eventually stabilizing as the DOF in the perturbation and error fields decrease over longer lead times. In contrast, the perturbation–error correlation for small scales declines steadily from the beginning and remains the lowest among all scales throughout the entire forecast period.

From a multi-scale sampling perspective, this study offers a systematic and quantitative assessment of the relationship between ensemble perturbations and errors in terms of both magnitude and direction at initial and forecast times. The findings suggest that the keys for effective ensemble generation differ across spatial scales. For larger scales, it is both possible and necessary to sample the long-distance spatial covariance of initial condition errors by improving the accuracy of ensemble perturbation covariance. This enhances the ability of initial ensemble members to represent the true atmospheric state, thereby producing an ensemble forecast that better captures the actual conditions. However, for small scales, sampling the spatial coherence of errors is highly challenging due to the large DOF. In this case, the critical factor is estimating the statistical characteristics of errors, such as the magnitude of error in the intensity and location of convective storms. This requires the development of high-resolution models and improved convective-scale simulation capabilities. Future research could extend this work by applying convection-resolving models to high-resolution ensemble forecasting of real multi-scale weather events, such as squall lines. Furthermore, it is worth noting that the keys for effective ensemble generation across different spatial scales were identified using the NCEP GEFS data. In the future, validation using data from other ensemble forecasting systems will be necessary to strengthen and generalize these conclusions.

Acknowledgments. The authors appreciate the computing resources provided by Fudan University to ensure the completion of this work.

Data availability. This study utilized ensemble forecast data from the operational Global Ensemble Forecast System (GEFS) at NCEP. The data are available at https://registry.opendata.aws/collab/noaa/. Codes used for data analysis in this study are available upon request to the corresponding author.

REFERENCES

Atger, F., 1999: The skill of ensemble prediction systems. *Mon. Wea. Rev.*, **127**, 1941–1953, https://doi.org/10.1175/1520-0493(1999)127<1941:TSOEPS>2.0.CO;2.

Bai, L. Q., G. X. Chen, Y. P. Huang, et al., 2021: Convection initiation at a coastal rainfall hotspot in South China: Synoptic patterns and orographic effects. *J. Geophys. Res. Atmos.*, **126**, e2021JD034642, https://doi.org/10.1029/2021JD034642.

Barker, T. W., 1991: The relationship between spread and forecast error in extended-range forecasts. *J. Climate*, **4**, 733–742, https://doi.org/10.1175/1520-0442(1991)004<0733:TRBSAF>2.0.CO;2.

Bei, N. F., and F. Q. Zhang, 2014: Mesoscale predictability of moist baroclinic waves: Variable and scale-dependent error growth. *Adv. Atmos. Sci.*, **31**, 995–1008, https://doi.org/10.1007/s00376-014-3191-7.

Berner, J., G. J. Shutts, M. Leutbecher, et al., 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.

Bradley, A. A., and S. S. Schwartz, 2011: Summary verification measures and their interpretation for ensemble forecasts. *Mon. Wea. Rev.*, **139**, 3075–3089, https://doi.org/10.1175/2010MWR3305.1.

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., 1997: Potential forecast skill of ensemble prediction and spread and skill distributions of the ECMWF ensemble prediction system. *Mon. Wea. Rev.*, **125**, 99–119, https://doi.org/10.1175/1520-0493(1997)125<0099:PFSOEP>2.0.CO;2.

Buizza, R., P. L. Houtekamer, G. Pellerin, et al., 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.

Chen, X. C., F. Q. Zhang, and K. Zhao, 2016: Diurnal variations of the land–sea breeze and its related precipitation over South China. *J. Atmos. Sci.*, **73**, 4793–4815, https://doi.org/10.1175/JAS-D-16-0106.1.

Coutinho, M. M., B. J. Hoskins, and R. Buizza, 2004: The influ-

ence of physical processes on extratropical singular vectors. *J. Atmos. Sci.*, **61**, 195–209, https://doi.org/10.1175/1520-0469(2004)061<0195:TIOPPO>2.0.CO;2.

Cui, J., S. Y. Yang, and T. Li, 2021: Intraseasonal variability of summertime surface air temperature over mid–high-latitude Eurasia and its prediction skill in S2S models. *J. Meteor. Res.*, **35**, 815–830, https://doi.org/10.1007/s13351-021-1131-x.

Duan, W. S., 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.

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

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res. Oceans*, **99**, 10,143–10,162, https://doi.org/10.1029/94JC00572.

Evensen, G., 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.*, **53**, 343–367, https://doi.org/10.1007/s10236-003-0036-9.

Feng, J., R. Q. Ding, J. P. Li, et al., 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.

Feng, J., J. P. Li, R. Q. Ding, et al., 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.

Feng, J., J. Zhang, Z. Toth, et al., 2020: A new measure of ensemble central tendency. *Wea. Forecasting*, **35**, 879–889, https://doi.org/10.1175/WAF-D-19-0213.1.

Feng, J., J. Wang, G. Dai, et al., 2023: Spatiotemporal estimation of analysis errors in the operational global data assimilation system at the China Meteorological Administration using a modified SAFE method. *Quart. J. Roy. Meteor. Soc.*, **149**, 2301–2319, https://doi.org/10.1002/qj.4507.

Feng, J., Z. Toth, J. Zhang, et al., 2024: Ensemble forecasting: A foray of dynamics into the realm of statistics. *Quart. J. Roy. Meteor. Soc.*, **150**, 2537–2560, https://doi.org/10.1002/qj.4745.

Hamill, T. M., 2001: Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev.*, **129**, 550–560, https://doi.org/10.1175/1520-0493(2001)129<0550:IORHFV>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, Z. L., J. P. Li, R. Q. Ding, et al., 2018: The application of nonlinear local Lyapunov vectors to the Zebiak–Cane model and their performance in ensemble prediction. *Climate Dyn.*, **51**, 283–304, https://doi.org/10.1007/s00382-017-3920-6.

Houtekamer, P. L., and J. Derome, 1995: Methods for ensemble prediction. *Mon. Wea. Rev.*, **123**, 2181–2196, https://doi.org/10.1175/1520-0493(1995)123<2181:MFEP>2.0.CO;2.

Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811, https://doi.org/10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2.

Huo, Z. H., and W. S. Duan, 2018: 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.

Jankov, I., Z. Toth, and J. Feng, 2022: Initial-value vs. model-induced forecast error: A new perspective. *Meteorology*, **1**, 377–393, https://doi.org/10.3390/meteorology1040024.

Ji, M., and F. Toepfer, 2016: Dynamical Core Evaluation Test Report for NOAA's Next Generation Global Prediction System (NGGPS). NOAA, 93 pp. https://doi.org/10.25923/ztzy-qn82.

Kalnay, E., 2002: *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, Cambridge, 314 pp.

Keresturi, E., Y. Wang, F. Meier, et al., 2019: Improving initial condition perturbations in a convection-permitting ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **145**, 993–1012, https://doi.org/10.1002/qj.3473.

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.

Loken, E. D., A. J. Clark, M. Xue, et al., 2019: Spread and skill in mixed- and single-physics convection-allowing ensembles. *Wea. Forecasting*, **34**, 305–330, https://doi.org/10.1175/WAF-D-18-0078.1.

Lorenz, E. N., 1969: The predictability of a flow which possesses many scales of motion. *Tellus*, **21**, 289–307, https://doi.org/10.3402/tellusa.v21i3.10086.

Magnusson, L., M. Leutbecher, and E. Källén, 2008: Comparison between singular vectors and breeding vectors as initial perturbations for the ECMWF ensemble prediction system. *Mon. Wea. Rev.*, **136**, 4092–4104, https://doi.org/10.1175/2008MWR2498.1.

Mason, I., 1982: A model for assessment of weather forecasts. *Aust. Meteor. Mag.*, **30**, 291–303, https://search.informit.org/doi/10.3316/anr-ia.s000511.

Molteni, F., and T. N. Palmer, 1993: Predictability and finite-time instability of the northern winter circulation. *Quart. J. Roy. Meteor. Soc.*, **119**, 269–298, https://doi.org/10.1002/qj.49711951004.

Molteni, F., and R. Buizza, 1999: Validation of the ECMWF ensemble prediction system using empirical orthogonal functions. *Mon. Wea. Rev.*, **127**, 2346–2358, https://doi.org/10.1175/1520-0493(1999)127<2346:VOTEEP>2.0.CO;2.

Molteni, F., R. Buizza, T. N. Palmer, et al., 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., W. S. Duan, and B. Wang, 2003: Conditional nonlinear optimal perturbation and its applications. *Nonlin. Processes Geophys.*, **10**, 493–501, https://doi.org/10.5194/npg-10-493-2003.

Murphy, A. H., 1973: A new vector partition of the probability score. *J. Appl. Meteor. Climatol.*, **12**, 595–600, https://doi.org/10.1175/1520-0450(1973)012<0595:ANVPOT>2.0.CO;2.

Palmer, T. N., and S. Tibaldi, 1988: On the prediction of forecast skill. Workshop on Predictability in the Medium and Extended Range, ECMWF, Reading, United Kingdom, 16–18.

Palmer, T., R. Buizza, R. Hagedorn, et al., 2006: Ensemble prediction: A pedagogical perspective. *ECMWF Newsletter*, **106**, 10–17, https://doi.org/10.21957/ab129056ew.

Privé, N. C., and R. M. Errico, 2015: Spectral analysis of forecast error investigated with an observing system simulation experiment. *Tellus*, **67**, 25977, https://doi.org/10.3402/tellusa.v67.25977.

Romine, G. S., C. S. Schwartz, J. Berner, et al., 2013a: Representing forecast error in a convection-permitting ensemble system. *Mon. Wea. Rev.*, **142**, 4519–4541, https://doi.org/10.1175/MWR-D-14-00100.1.

Romine, G. S., C. S. Schwartz, C. Snyder, et al., 2013b: Model bias in a continuously cycled assimilation system and its influence on convection-permitting forecasts. *Mon. Wea. Rev.*, **141**, 1263–1284, https://doi.org/10.1175/MWR-D-12-00112.1.

Sardeshmukh, P. D., and B. J. Hoskins, 1988: The generation of global rotational flow by steady idealized tropical divergence. *J. Atmos. Sci.*, **45**, 1228–1251, https://doi.org/10.1175/1520-0469(1988)045<1228:TGOGRF>2.0.CO;2.

Schwartz, C. S., G. S. Romine, K. R. Smith, et al., 2014: Characterizing and optimizing precipitation forecasts from a convection-permitting ensemble initialized by a mesoscale ensemble Kalman filter. *Wea. Forecasting*, **29**, 1295–1318, https://doi.org/10.1175/WAF-D-13-00145.1.

Schwartz, C. S., G. S. Romine, and D. C. Dowell, 2021: Toward unifying short-term and next-day convection-allowing ensemble forecast systems with a continuously cycling 3-km ensemble Kalman filter over the entire conterminous United States. *Wea. Forecasting*, **36**, 379–405, https://doi.org/10.1175/WAF-D-20-0110.1.

Schwartz, C. S., J. Poterjoy, J. R. Carley, et al., 2022: Comparing partial and continuously cycling ensemble Kalman filter data assimilation systems for convection-allowing ensemble forecast initialization. *Wea. Forecasting*, **37**, 85–112, https://doi.org/10.1175/WAF-D-21-0069.1.

Schwartz, C. S., M. Wong, G. S. Romine, et al., 2020: Initial conditions for convection-allowing ensembles over the conterminous United States. *Mon. Wea. Rev.*, **148**, 2645–2669, https://doi.org/10.1175/MWR-D-19-0401.1.

Stanski, H. R., L. J. Wilson, and W. R. Burrows, 1989: Survey of Common Verification Methods in Meteorology. WMO World Weather Watch Technical Report No.8, WMO/TD No. 358. WMO, Geneva, 114 pp.

Sun, Y. Q., 2017: Scale interaction and mid-latitude atmospheric predictability: Impacts of moist convection and gravity waves. Ph.D. dissertation, The Pennsylvania State University, Stetkolich, 147 pp.

Sun, Y. Q., and F. Q. Zhang, 2020: A new theoretical framework for understanding multiscale atmospheric predictability. *J. Atmos. Sci.*, **77**, 2297–2309, https://doi.org/10.1175/JAS-D-19-0271.1.

Swinbank, R., M. Kyouda, P. Buchanan, et al., 2016: The TIGGE project and its achievements. *Bull. Amer. Meteor. Soc.*, **97**, 49–67, https://doi.org/10.1175/BAMS-D-13-00191.1.

Tan, Z. M., F. Q. Zhang, R. Rotunno, et al., 2004: Mesoscale predictability of moist baroclinic waves: Experiments with parameterized convection. *J. Atmos. Sci.*, **61**, 1794–1804, https://doi.org/10.1175/1520-0469(2004)061<1794:MPOMBW>2.0.CO;2.

Ting, M., 1996: Steady linear response to tropical heating in barotropic and baroclinic models. *J. Atmos. Sci.*, **53**, 1698–1709, https://doi.org/10.1175/1520-0469(1996)053<1698:SLRTTH>2.0.CO;2.

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.

Toth, Z., and E. Kalnay, 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.

Toth, Z., Y. Zhu, T. Marchok, et al., 1998: Verification of the NCEP global ensemble forecasts. Preprints of the 12th Conference on Numerical Weather Prediction, Phoenix, AZ, Amer. Meteor. Soc., 286–289.

Tribbia, J. J., and D. P. Baumhefner, 2004: Scale interactions and atmospheric predictability: An updated perspective. *Mon. Wea. Rev.*, **132**, 703–713, https://doi.org/10.1175/1520-0493(2004)132<0703:SIAAPA>2.0.CO;2.

Wang, H., Y. L. Luo, and B. J. D. Jou, 2014: Initiation, maintenance, and properties of convection in an extreme rainfall event during SCMREX: Observational analysis. *J. Geophys. Res. Atmos.*, **119**, 206–213,232, https://doi.org/10.1002/2014JD022339.

Wang, X. G., and C. H. Bishop, 2003: A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *J. Atmos. Sci.*, **60**, 1140–1158, https://doi.org/10.1175/1520-0469(2003)060<1140:ACOBAE>2.0.CO;2.

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.

Wei, M. Z., Z. Toth, R. Wobus, et al., 2006: Ensemble transform Kalman filter–based ensemble perturbations in an operational global prediction system at NCEP. *Tellus A: Dyn. Meteor. Oceanogr.*, **58**, 28–44, https://doi.org/10.1111/j.1600-0870.2006.00159.x.

Wei, M. Z., Z. Toth, R. Wobus, et al., 2008: Initial perturbations based on the ensemble transform (ET) technique in the NCEP global operational forecast system. *Tellus A: Dyn. Meteor. Oceanogr.*, **60**, 62–79, https://doi.org/10.1111/j.1600-0870.2007.00273.x.

Whitaker, J. S., and A. F. Loughe, 1998: The relationship between ensemble spread and ensemble mean skill. *Mon. Wea. Rev.*, **126**, 3292–3302, https://doi.org/10.1175/1520-0493(1998)126<3292:TRBESA>2.0.CO;2.

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.

Yang, H. P., Y. Du, Z. J. Chen, et al., 2024: Could developing frontal rainfall influence warm-sector rainfall? *Geophys. Res. Lett.*, **51**, e2024GL110430, https://doi.org/10.1029/2024GL110430.

Ying, Y., and F. Q. Zhang, 2017: Practical and intrinsic predictability of multiscale weather and convectively coupled equatorial waves during the active phase of an MJO. *J. Atmos. Sci.*, **74**, 3771–3785, https://doi.org/10.1175/JAS-D-17-0157.1.

Zhang, F. Q., Y. Q. Sun, L. Magnusson, et al., 2019: What is the predictability limit of midlatitude weather? *J. Atmos. Sci.*, **76**, 1077–1091, https://doi.org/10.1175/JAS-D-18-0269.1.

Zhang, H. B., J. Chen, X. F. Zhi, et al., 2015: Study on multi-scale blending initial condition perturbations for a regional ensemble prediction system. *Adv. Atmos. Sci.*, **32**, 1143–1155, https://doi.org/10.1007/s00376-015-4232-6.

Zhou, L. J., L. Harris, J. H. Chen, et al., 2024: Bridging the gap between global weather prediction and global storm-resolving simulation: Introducing the GFDL 6.5-km SHiELD. *J. Adv. Model. Earth Syst.*, **16**, e2024MS004430, https://doi.org/10.1029/2024MS004430.

Zhou, X. Q., Y. J. Zhu, D. C. Hou, et al., 2022: The development of the NCEP global ensemble forecast system version 12. *Wea. Forecasting*, **37**, 1069–1084, https://doi.org/10.1175/waf-d-21-0112.1.

Zhu, Q. G., J. R. Lin, S. W. Shou, et al., 2000: *Principles and Methods of Weather Science*. China Meteorological Press, Beijing, 649 pp. (in Chinese)