ADVANCES IN ATMOSPHERIC SCIENCES, 2026

• **Original Paper** •

Recognizing Sources of Forecast Uncertainty in Extreme Low-Temperature Events: The C-NFSVs Method

Yuxuan HOU[1,2], Wansuo DUAN[*1,2], and Zhe HAN[3]

[1]*State Key Laboratory of Earth System Numerical Modeling and Application*, *Institute of Atmospheric Physics*, *Chinese Academy of Sciences*, *Beijing* 100029, *China*

[2]*University of Chinese Academy of Sciences*, *Beijing* 100049, *China*

[3]*Laboratory of Regional Climate and Environment for Temperate Asia*, *Institute of Atmospheric Physics*, *Chinese Academy of Sciences*, *Beijing* 100029, *China*

(Received 15 July 2025; revised 27 November 2025; accepted 17 December 2025)

ABSTRACT

The combined nonlinear forcing singular vectors (C-NFSVs) method combines initial and model perturbations and accounts for the collective effect of initial and model uncertainties in ensemble forecasts through the nonlinear forcing singular vector (NFSV; also referred to as CNOP-F) approach. We apply C-NFSVs to the Weather Research and Forecasting (WRF) model and investigate the forecast uncertainty of the 2 m temperature over southern China during four major sequential periods of the 2008 extreme cold event. The results show that compared with scenarios considering only initial or model perturbations, C-NFSVs can provide more reliable ensemble forecasts. Furthermore, the C-NFSVs reveal that the 2 m temperature forecast uncertainties are predominantly sensitive to uncertainties in the upstream circulation system, whereas sensitivity to initial and model uncertainties varies across different periods of the cold event. The early period of the extreme cold event tends to propagate forecast uncertainty, as represented by the C-NFSVs ensemble spread from the upstream circulation to southern China following the background circulation. However, the forecasts of later periods present spread-characterized uncertainty that persists in the upstream circulation while remaining in contact with the background circulation and continuously propagating its effect downstream to southern China. This mechanism indicates that forecast uncertainties are dominated by initial uncertainties in the forecasts of the earlier period, whereas model uncertainties play a much more significant role in the forecasts of later periods. These findings highlight the potential of the C-NFSVs method in identifying the source of forecast uncertainty and delivering skillful forecasts for extreme cold events.

Key words: extreme cold events, forecast uncertainty, optimal perturbation, ensemble forecasting

Citation: Hou, Y. X., W. S. Duan, and Z. Han, 2026: Recognizing sources of forecast uncertainty in extreme low-temperature events: The C-NFSVs method. *Adv. Atmos. Sci.*, https://doi.org/10.1007/s00376-025-5443-0.

Article Highlights:

- C-NFSVs are first applied to the WRF model to study the ensemble forecasts of extreme cold events.
- C-NFSVs provide reliable ensemble forecasts for extreme cold events by capturing forecast uncertainty effectively.
- Analysis of the ensemble spread from C-NFSVs offers a potential approach to identifying sources of forecast uncertainty in extreme cold events.

1. Introduction

Extreme low-temperature events (also referred to as cold waves), as one of the primary weather phenomena during winter, have widespread impacts, often affecting large regions of China. These events typically involve significant temperature drops, along with snow and freezing precipitation, often resulting in considerable economic losses and posing serious threats to public safety (Wang and Ding, 2006; Wei, 2008; Ryti et al., 2016). In recent years, China has experienced a growing number of extreme cold events. Notable examples include the 2008 early-year low-temperature and freezing rain disaster (Tao and Wei, 2008), the nationwide cold wave in January 2016 (Jiang et al., 2016; Yamaguchi et al., 2019), the East Asian extreme cold event during the winter of 2020 (Zhang et al., 2021; Zheng et al., 2022), and the widespread cold wave in early 2024 (Xie et al., 2024; Yu et al., 2024). However, for these events, the European Centre for Medium-Range Weather Forecasts (ECMWF), a

* Corresponding author: Wansuo DUAN
Email: duanws@lasg.iap.ac.cn

leader in weather prediction, gave inadequate estimates of forecast uncertainty, even when the cold signals were captured in the forecasts [see Fig. S1 in the electronic supplementary material (ESM) and section 4]. This limitation highlights the crucial need to identify the underlying sources of forecast uncertainties, improve their estimation, and enhance the ability to predict extreme cold events.

Uncertainty in numerical weather forecasting typically arises from uncertainties in initial conditions and numerical models (Lorenz, 1969, 1982; Toth and Vannitsem, 2002; Vannitsem and Toth, 2002; Nicolis et al., 2009). This inspires us to identify the underlying sources of forecast uncertainty from the perspective of whether they stem from initial or model-related factors. Notably, a reliable ensemble forecast system not only enables the estimation of forecast uncertainty and provides probabilistic forecasts for specific events but also characterizes the uncertainty arising from initial or model errors (Palmer, 2000; Buizza, 2019). Among the above extreme cold events, the 2008 cold wave that struck southern China was one of the most typical and highly influential cases (Li et al., 2022; Qing et al., 2025). It brought about severe low temperatures, freezing precipitation, and widespread snowfall (Ding et al., 2008; Tao and Wei, 2008; Wang et al., 2009), resulting in devastating impacts and extensive losses to the economy, energy supply, and human livelihoods (Li et al., 2008; Ma et al., 2011). In this study, we take the 2008 extreme cold event as a representative case to investigate its forecast uncertainty from an ensemble forecasting perspective and analyze the sources. We attempt to explain the reasons behind the ECMWF's underestimation of forecast uncertainties.

A reliable ensemble forecast generally provides an ensemble spread comparable to the ensemble mean forecast error and accurately reflects the forecast uncertainty (Leutbecher and Palmer, 2008; Fortin et al., 2014; Hopson, 2014). However, for the 2008 cold event, which can be divided into four cold and freezing disaster periods (11–16 January, 18–22 January, 25–29 January, and 31 January to 2 February), the ensemble forecasts produced by the ECMWF (data source: TIGGE database; Bougeault et al., 2010; ECMWF, 2015; Swinbank et al., 2016; Leutbecher et al., 2024) show evident underdispersion. Specifically, the ensemble spreads for these four periods are smaller than the root-mean-square error (RMSE; see Appendix A) of the ensemble mean forecasts of the 2 m temperature. This underdispersion reflects a notable underestimation of forecast uncertainty (see Fig. S1 in the ESM). In the ECMWF ensemble forecasting system, the initial perturbation method used was singular vectors (SVs; Buizza and Palmer, 1995; Molteni et al., 1996), while the model perturbation method employed was the stochastic perturbation of physical tendencies (Buizza et al., 1999). SVs are insufficient to fully capture the amplification of initial errors in a nonlinear regime, whereas the model perturbation method may struggle to represent the rapid nonlinear growth of model errors (Anderson, 1997; Hamill et al., 2000; Duan et al., 2023a; Zhang et al., 2023a; Zhang et al., 2023b). The underdispersion may therefore stem from the limitations of both the initial and model perturbation methods.

To overcome the linear limitations of SVs, Duan and Huo (2016) proposed the method of orthogonal conditional nonlinear optimal perturbations (O-CNOPs). This approach extends SVs to nonlinear systems and identifies a set of mutually orthogonal initial perturbations, which tend to exhibit the greatest nonlinear evolution within their respective subphase spaces (see also Huo et al., 2019; Huo and Duan, 2019; Zhang et al., 2023a). Compared with traditional methods such as random perturbations, breeding vectors, and SVs, this method has already demonstrated advanced forecast skill and better characterization of forecast uncertainty for tropical cyclone (TC) tracks (Huo et al., 2019; Duan et al., 2023a, 2023b; Zhang et al., 2023a). In fact, the O-CNOPs method has been recognized as an important approach in an ensemble forecast handbook, alongside traditional methods (Du et al., 2019).

With respect to model perturbations, as previously discussed, perturbation methods based on stochastic physics schemes have limitations in capturing the nonlinear growth of model errors, which contributes to the issue of underdispersion in ensemble forecasts (Duan et al., 2023a, 2023b; Zhang et al., 2023b). To address this awkwardness, the method of orthogonal nonlinear forced singular vectors (O-NFSVs) was subsequently developed (Duan et al., 2022; Zhang et al., 2023b). O-NFSVs represent a set of mutually orthogonal model tendency perturbations that produce the fastest growth within their respective subspaces. This method has already been applied to the ensemble forecasting of TCs, demonstrating better forecast skill for both intensity and track than traditional stochastic physics schemes (Duan et al., 2023a, 2023b; Zhang et al., 2023b).

Real forecast systems always contain both initial errors and model errors simultaneously. Therefore, ignoring either one or their interactions limits the ability of ensemble forecasts to accurately represent forecast uncertainty and subsequently diminishes the forecast skill (Nicolis et al., 2009). In light of this, Duan et al. (2022) proposed the combined nonlinear forcing singular vectors (C-NFSVs) method. This method generates a group of rapidly growing perturbations that account for the combined effects of initial and model errors based on the NFSV framework while ensuring their coordinated growth. For the 2008 extreme cold event, as previously mentioned, the ECMWF ensemble forecasts generate initial and model perturbations separately and do not consider the interactions between them. This lack of dynamic coordination of initial and model perturbations may also contribute to the underdispersion issue in the ECMWF forecast for the 2008 cold event. In contrast, if the C-NFSVs approach is adopted, it may have the potential to better capture the coordinated growth of both types of errors, thereby offering a possible advantage in generating a more ideal set of ensemble members. In this study, we use the Weather Research and Forecasting (WRF) model to compute the C-NFSVs ensemble during the 2008 extreme cold event and validate its effec-

tiveness. We focus on assessing whether this method can better capture the forecast uncertainty of 2 m temperature over southern China and enhance the reliability of ensemble forecasts. By applying this approach, we aim to identify the exact source of forecast uncertainty and explore the distinct roles of initial and model perturbations, eventually interpreting the reason why the ECMWF forecasts tended to underestimate uncertainty during this event.

The rest of the paper is organized as follows. Section 2 introduces the WRF model, the data sources, and the C-NFSVs approach. Section 3 outlines the experimental strategy and the relative configurations. Section 4 presents the ability of the ensemble forecasts generated by C-NFSVs to capture the forecast uncertainty during the 2008 event and evaluates both their deterministic and probabilistic forecast skills, along with the underlying mechanisms. Section 5 highlights the role of the ensemble spread generated by C-NFSVs in identifying the sources of forecast uncertainty while also examining the distinct effects of initial and model perturbations. Finally, further discussion and conclusions are provided in sections 6 and 7, respectively.

2. Model, data, and methods

This study, as mentioned in the introduction, explores the forecast uncertainties of cold waves by applying the C-NFSVs method in the WRF model, with the 2008 extreme cold event over southern China serving as a representative example. To assist readers, we provide in this section an overview of the WRF model and associated data, as well as the C-NFSVs method proposed by Duan et al. (2022).

2.1. *Model and data*

In this study, we utilize version 3.9.1 of the WRF model. Considering computational efficiency, a coarse-resolution model was used when calculating the C-NFSV perturbations, with a horizontal resolution of 1° (approximately 100 km), covering a grid of 60 × 50 points. For ensemble forecasting, however, a high-resolution configuration was adopted, with a horizontal resolution of 0.25° (approximately 25 km), encompassing a grid of 240 × 200 points. The model includes 45 vertical layers, with the top level set at 50 hPa. The model domain covers the region from 20°N to 70°N and from 60°E to 120°E, including the key circulation systems that influenced the extreme cold event in January 2008, as well as the areas in southern China that were severely affected by the event. Notably, since calculating C-NFSV perturbations requires corresponding gradients, we also utilized the WRFV3.9.1 adjoint model. The initial and boundary conditions for driving the model integration were provided by the operational Global Forecast System (GFS) forecast data, with a resolution of 0.5° × 0.5°. To evaluate the forecast skill, the fifth generation ECMWF atmospheric reanalysis (ERA5) was used as the reference truth, with a resolution of 0.25° × 0.25°.

2.2. *The C-NFSVs method*

Consider a nonlinear system in a domain Ω with initial error $\mathbf{u}_0$ and model tendency errors $\mathbf{f}(\mathbf{x},t)$; according to the definition in Duan et al. (2022), the corresponding forecast error $\mathbf{u}_T$ at time T can be represented by

$$\mathbf{u}_T = M_T(\mathbf{f})(\mathbf{U}_0+\mathbf{u}_0) - M_T(\mathbf{U}_0)\ . \tag{1}$$

where $\mathbf{U}_0$ is the initial state, M_T denotes the nonlinear propagation operator (i.e., numerical model), and $M_T(\mathbf{f})$ represents the propagation operator in the presence of model tendency error $\mathbf{f}$. On this basis, the maximization problem for calculating the C-NFSVs can be expressed as follows [see Duan et al. (2022)]:

$$J\left(\mathbf{f}_j^*\right) = \max_{\mathbf{f}_j\in\Omega_j} \|M_T\left(\mathbf{f}_j\right)\left(\mathbf{U}_0+r\mathbf{f}_j\right) - M_T(\mathbf{U}_0)\|_{\mathrm{b}}\ , \tag{2}$$

where

$$\Omega_j = \begin{cases} \mathbf{f}_1 \in R^n, \|\mathbf{f}_1\|_{\mathrm{a}} \le \sigma_{\mathbf{f}}, \\ \{\mathbf{f}_j \in R^n \,|\, \|\mathbf{f}_j\|_{\mathrm{a}} \le \sigma_{\mathbf{f}}, \mathbf{f}_j \perp \Omega_k, k=1,\ldots,j-1\}, j>1, \end{cases} \tag{3}$$

$\|\cdot\|_{\mathrm{a}}$ and $\|\cdot\|_{\mathrm{b}}$ represent the norms used to measure the amplitudes of the perturbations and the objective function, respectively. Note that to simplify the problem, we assume that the initial and model perturbations share the same pattern and have the same growth direction, i.e., $\mathbf{u}_0 = r\mathbf{f}$. This simplification is supported by the fact that model errors are also amplified by the chaotic mechanism that governs the growth of initial errors, after which the evolution of model perturbations approaches that of the initial perturbations over time (Vannitsem, 2006; Nicolis et al., 2009). The ratio of the amplitude of the initial perturbation to that of the model perturbation is denoted as r, defined as $r = \sigma_{\mathbf{I}}/\sigma_{\mathbf{f}}$. Thus, the C-NFSVs are defined by the combined modes $\left(r\mathbf{f}_j^*, \mathbf{f}_j^*\right)$, which consist of two components: the initial perturbation component $r\mathbf{f}_j^*$ and the model perturbation component $\mathbf{f}_j^*$. In the ensemble forecasts, the former is added once to the control forecast at the initial time ("one-time" addition), while the latter is continuously superimposed throughout the entire forecast. To solve the maximization problem shown in Eq. (2), we use the Spectral Projected Gradient 2 (SPG2; Birgin et al., 2000) algorithm, with the gradient of the objective function $J\left(\mathbf{f}_j^*\right)$ obtained through the adjoint WRF model.

3. Experimental strategy and configuration

To calculate the C-NFSVs that characterize the uncertainty in 2 m temperature forecasts during the 2008 event over southern China, the objective function here is defined within this region (25°–35°N, 105°–120°E), covering a grid of 57 × 41 points. It represents the difference in 2 m temperature between the perturbed and control forecasts at prediction time T, as expressed in Eq. (2), where M_T denotes the WRF

model. The perturbed variables include the meridional and zonal wind fields, as well as the potential temperature, which play key roles in the advection of temperature, the primary factor contributing to surface cooling during the event (Ding et al., 2008; Tao and Wei, 2008; Wang et al., 2009). To measure the model perturbation, we adopt the energy norm, given by the following equation:

$$\|\mathbf{f}\|_{\mathrm{a}} = \int_D \left[\mathbf{u}'^2 + \mathbf{v}'^2 + \left(\frac{g}{\overline{N}\theta} \right)^2 \boldsymbol{\theta}'^2 \right] \mathrm{d}D \,, \tag{4}$$

where $\mathbf{u}'$, $\mathbf{v}'$, and $\boldsymbol{\theta}'$ represent the perturbations in the zonal wind field, meridional wind field, and potential temperature, respectively. g is the acceleration due to gravity, $\overline{N}$ represents the Brunt–Väisälä frequency, and $\theta = 300$ K is a physical constant.

The reliability of the ensemble forecasts generated by C-NFSVs depends on several factors, including the number of ensemble members, the optimization time [i.e., T in Eq. (2)], and the amplitudes of both the initial and model perturbations (Duan and Huo, 2016; Duan et al., 2022). The selection of these parameters is based on experimentation. The ensemble members consist of the control forecast and perturbation forecast pairs created by adding and subtracting the C-NFSVs. We found that when calculating up to the fifth C-NFSV, the ensemble already exhibited a sufficient spread to capture the forecast uncertainty, leading us to select 11 ensemble members. With respect to the optimization time T, an excessively long optimization time can cause instability in the adjoint model, whereas a short optimization time fails to ensure sufficient spread in the later stages of the forecast. In light of these considerations, we chose an optimization time of three days. For the amplitude of the initial perturbations (Zhang et al., 2023a), we referred to the magnitude of the initial analysis errors, defined as the difference between the GFS initial fields and the corresponding ERA5 reanalysis. On the basis of the L^2 norm in Eq. (4), the magnitude of these errors is approximately smaller than 10 J kg^{-1}. A small amplitude would not allow the perturbations to develop sufficiently, whereas a large amplitude could cause numerical instability in the model. To balance sufficient perturbation growth with model stability, the initial perturbation amplitude was set to 1 J kg^{-1} through sensitivity experiments. For the amplitude of the model perturbations (Zhang et al., 2023b), we estimated the magnitude using the difference between two consecutive integration steps of the perturbed variables and took (10% – 20%) of this difference. On the basis of this approach, we selected three model perturbation amplitudes: $\sigma_{\mathbf{f1}} = 1 \times 10^{-10}$ J kg^{-1} s^{-1}, $\sigma_{\mathbf{f2}} = 5 \times 10^{-11}$ J kg^{-1} s^{-1}, and $\sigma_{\mathbf{f3}} = 3 \times 10^{-11}$ J kg^{-1} s^{-1}. Combined with the predetermined initial perturbation amplitude, the ratio r was then determined through the definition $r = \sigma_{\mathbf{I}}/\sigma_{\mathbf{f}}$ [see Eq. (2)].

From the above discussion, the ensemble parameters are predetermined: the number of ensemble members is 11, the optimization time is 3 days, and the initial perturbation amplitude is 1 J kg^{-1}. It is obvious that only the model perturbation amplitude still remains uncertain when $\sigma_{\mathbf{f1}} = 1 \times 10^{-10}$ J kg^{-1} s^{-1}, $\sigma_{\mathbf{f2}} = 5 \times 10^{-11}$ J kg^{-1} s^{-1}, and $\sigma_{\mathbf{f3}} = 3 \times 10^{-11}$ J kg^{-1} s^{-1} are considered. Therefore, to test which amplitude of model perturbation ensures a much more reliable ensemble forecast, we conduct 12 sensitivity experiments by applying C-NFSVs of model perturbation amplitudes $\sigma_{\mathbf{f1}}$, $\sigma_{\mathbf{f2}}$, and $\sigma_{\mathbf{f3}}$ with other predetermined ensemble parameters to the forecasts of the four periods in the 2008 cold event. Table 1 lists the 12 experiments $E_1 \sim E_{12}$, where for example, E_1, E_4, E_7, and E_{10} in the first column represent the experiments in which the C-NFSVs are calculated using the model perturbation amplitude $\sigma_{\mathbf{f1}}$ for each of the forecast periods.

Notably, for computational efficiency, C-NFSVs are calculated at a coarse resolution of 1°. Then, bilinear interpolation is applied to upscale the perturbations from 1° to 0.25° for use in the ensemble forecasts, following a similar approach as in Zhang et al. (2023a, b). The ratio of the ensemble spread to the RMSE of the ensemble mean forecast averaged over the four periods as a function of the forecast lead time is shown in Fig. 1. A ratio closer to 1 indicates a more reliable ensemble forecast, where the ensemble spread better represents the forecast uncertainty. As shown in Fig. 1, we selected $\sigma_{\mathbf{f3}}$ as the optimal model perturbation amplitude to

Table 1. Twelve groups of experiments for selecting the optimal model perturbation amplitude.

Forecast period	Model perturbation amplitude		
	$\sigma_{\mathbf{f1}}$	$\sigma_{\mathbf{f2}}$	$\sigma_{\mathbf{f3}}$
20080109T00–16T00	E_1	E_2	E_3
20080115T00–22T00	E_4	E_5	E_6
20080122T00–29T00	E_7	E_8	E_9
20080128T00–0204T00	E_{10}	E_{11}	E_{12}

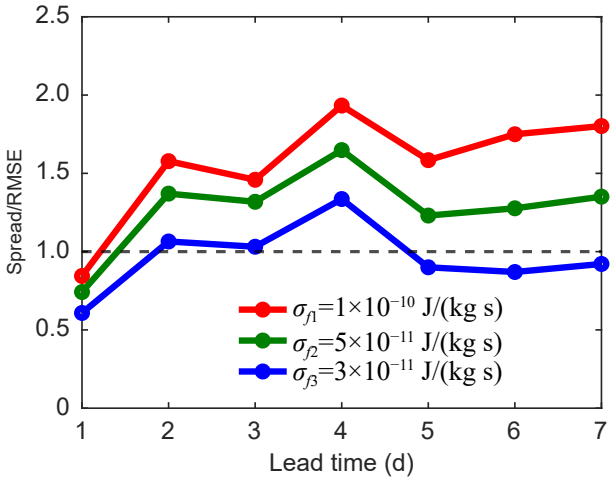

Fig. 1. Temporal variability of the spread–RMSE ratio for the C-NFSVs ensemble forecasts of 2 m temperature in southern China, as averaged for the four periods of the 2008 event. The three colored dotted lines represent different model perturbation amplitudes.

further investigate the ability of the ensemble forecasts generated by C-NFSVs to capture the forecast uncertainty. The ensemble forecast reliability here involves deterministic forecast reliability, measured using the spread–RMSE ratio as described earlier, and probabilistic forecast reliability, assessed using reliability diagrams (RDs). In addition, we examine the deterministic and probabilistic forecast skill of the C-NFSVs ensemble for this event, including improvements in the RMSE for deterministic forecasts and the Brier score (BS; Brier, 1950) and relative operating characteristic (ROC; Mason, 1982) curves for probabilistic forecasts. Detailed definitions for these metrics are provided in Appendix B.

4. Capability of the ensemble spread generated by C-NFSVs in representing forecast uncertainty

In this section, we generate reliable ensemble forecasts using the C-NFSVs method and analyze the uncertainty in the 2008 extreme cold event forecasts.

4.1. *Reliability of ensemble forecasts generated by C-NFSVs*

Under the optimal parameter configuration identified in section 3, Fig. 2 presents the temporal evolution of the spread–RMSE ratio (see Appendix A) for the ensemble forecasts generated by the C-NFSVs. The results are averaged over the four forecast periods associated with the 2008 extreme cold event (see Table 1). We also provide in Fig. 2 the spread–RMSE ratio of the ensemble forecasts generated by separately adding the initial and the model perturbation components of the C-NFSVs. It is evident that the C-NFSV perturbations achieve a spread–RMSE ratio much closer to 1 throughout the entire forecast stage. In contrast, the initial component of the C-NFSVs yields a ratio close to 1 in the early stage but deviates substantially in the later stage, whereas the model component exhibits the opposite behavior. Obviously, the initial and model perturbation components of the C-NFSVs provide reasonable dynamics (Nicolis et al., 2009; Duan et al., 2022) for ensemble forecasts in explaining forecast uncertainties. Consequently, the C-NFSVs maintain strong ensemble reliability throughout the forecast and provide a spread that represents forecast uncertainty much more effectively over time. Moreover, we find that the ensemble forecasts generated by the C-NFSVs show the greatest improvement in the RMSE during the four periods of the 2008 event. The improvement is evident over the latter four days of each forecast, with the ensemble mean closer to the truth than in the other scenarios. Averaged over these latter four days, the improvements in the RMSE relative to the control forecast reach 5.04% for the C-NFSVs, 3.39% for their initial component, 3.37% for their model component, and 2.90% for the O-CNOPs (Fig. 2). Although the C-NFSV ensemble forecasts did not effectively reduce the RMSE in the early forecast stage, especially at the three-day lead time (Fig. 2b), their spread–RMSE ratio remained close to 1, as shown in Fig. 2a. In contrast, the ratios in the scenarios with only one type of perturbation deviated much more from 1. This finding indicates that the C-NFSVs provide a more accurate estimation of forecast uncertainty and better ensemble reliability. It should be noted that good reliability does not necessarily imply a lower RMSE, as reliability reflects the ability of an ensemble to represent the forecast uncertainty rather than its absolute accuracy.

Apart from deterministic forecasts, the reliability and skill of probabilistic forecasts are also crucial metrics for evaluating ensemble forecasts. During the 2008 extreme cold event, as introduced in section 1, freezing precipitation caused significant impacts in southern China, in addition to the damage caused by low temperatures. Some studies suggest that a 2 m temperature below 0°C might serve as a critical indicator of the occurrence of freezing precipitation (Zhao and Sun, 2008; Qi, 2012). As another important aspect of the ability of the ensemble to capture the forecast uncertainty, we evaluate its probabilistic forecast reliability. This evaluation is based on the binary event of whether the 2 m temperature at each grid point in the region of interest decreases below 0°C. In this assessment, we focus on the fore-

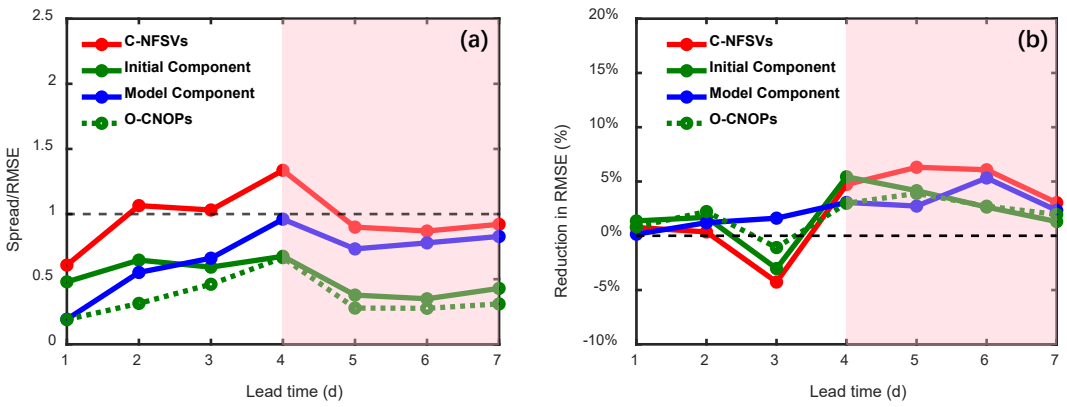

Fig. 2. (a) Spread–RMSE ratio of the ensemble forecasts generated by C-NFSVs, their initial and model perturbation components, and O-CNOPs for the 2 m temperature in southern China. (b) Reduction in the RMSE relative to the control forecast under different scenarios.

casts for the last four days of each forecast period, which correspond to the four freezing disaster periods highlighted in the introduction when freezing precipitation was most concentrated. Figure 3 suggests that C-NFSVs provide the most reliable ensemble forecasts, with the RD line closest to the diagonal. This finding indicates that their forecast probabilities for grid points where the 2 m temperature falls below 0°C most closely align with the observed frequencies compared with scenarios considering only initial or model perturbations. Additionally, the ROC curves (Fig. 3), used to assess the ability of the forecast to discriminate between events and non-events, show that the ensemble forecasts generated by C-NFSVs achieve the highest hit rates and the lowest false alarm rates for predicting the location of the 0°C isotherm, although the improvements over the other scenarios are not significant.

The BS, averaged over four periods of the 2008 event and southern China, is calculated using a sample size of 4 × 4 × 57 × 41, corresponding to four forecast periods, four days per period, and 57 × 41 grid points within the region of interest, as shown in Fig. 4. Its spatial distribution (with a sample size of 4 × 4 per grid point, corresponding to the four periods and four days per period), also shown in Fig. 4, is presented as the difference between the forecasts generated by the initial and model perturbation components of C-NFSVs and those generated by the full C-NFSVs ensemble. The ensemble forecasts generated by C-NFSVs exhibit the smallest BS, indicating relatively high probabilistic forecasting skills for the binary event described above. Furthermore, the spatial distribution shows that the BS differences between the single-perturbation scenarios and the C-NFSVs are mostly positive near the 0°C isotherm of 2 m temperature in the ERA5 reanalysis. This means that the C-NFSVs ensemble forecasts generally achieve lower BS values in this region. This further highlights their relatively high probability of accurately predicting the location of the 0°C isotherm of 2 m temperature.

Notably, as shown in Fig. 2, when averaged over the four periods of the 2008 event, the ensemble forecasts generated by the model perturbation component of C-NFSVs exhibit a spread–RMSE ratio that nearly approaches that of the C-NFSVs. This ratio progressively surpasses that of the ensemble forecasts considering only the initial perturbation component while also leading to an improvement in the deterministic forecast skill, although the improvement is not significant. Moreover, as illustrated in Figs. 3 and 4, the probabilistic reliability and forecast skill of the ensemble forecasts incorporating only model perturbations are also better than those considering only the initial perturbation component of the C-NFSVs and approaches those of the C-NFSVs. These findings suggest that model perturbations seem to play a more dominant role in estimating forecast uncertainty and enhancing ensemble forecast skill. This may shed light on the fact that the uncertainties in the 2008 extreme cold event forecasts generated by the WRF model are attributable mainly to model error effects. Given this, we further calculated the spread–RMSE ratio for the ensemble forecasts generated by O-CNOPs (see section 2) using the same parameter configuration as in the C-NFSVs, which is considered capable of enhancing the ensemble spread and improving the spread–RMSE relationship (Duan and Huo, 2016). The results reveal that the ensemble forecasts generated by O-CNOPs performed similarly to those of the initial component of the C-NFSVs and still exhibited weak reliability in the later stages of the forecast. Therefore, even if the initial perturbations are theoretically optimized to have the greatest growth, they, similar to the initial perturbation components in the C-NFSVs, still considerably underestimate the forecast uncertainty.

4.2. *Why do C-NFSVs provide high ensemble reliability?*

In this section, the underlying mechanisms responsible

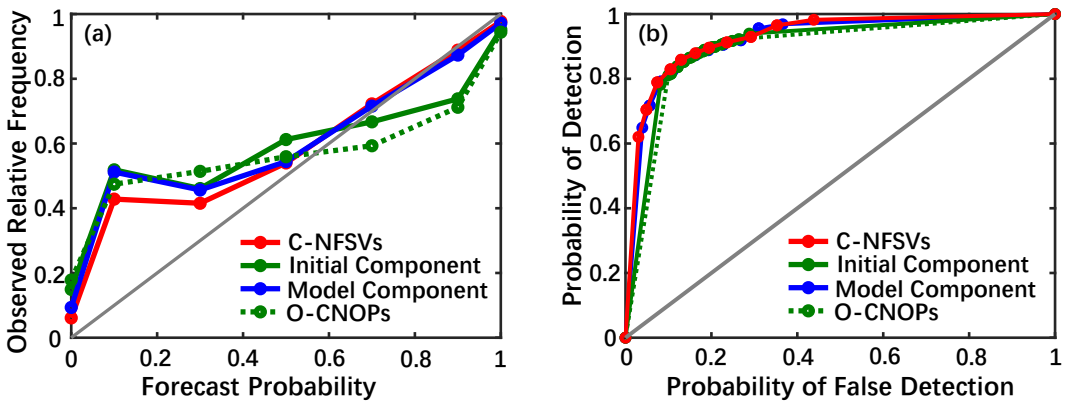

Fig. 3. (a) Ensemble reliability curves under different scenarios, where closer proximity to the diagonal line indicates higher reliability. The average distances of the curves to the diagonal are 0.06, 0.10, 0.07, and 0.11, corresponding to the ensemble forecasts generated by C-NFSVs, the initial perturbation component of C-NFSVs, the model perturbation component of C-NFSVs, and O-CNOPs, respectively. (b) ROC curves under different scenarios, with ROCA values of 0.93, 0.92, 0.90, and 0.89 corresponding to the ensemble forecasts generated by C-NFSVs, the initial perturbation component of C-NFSVs, the model perturbation component of C-NFSVs, and O-CNOPs, respectively.

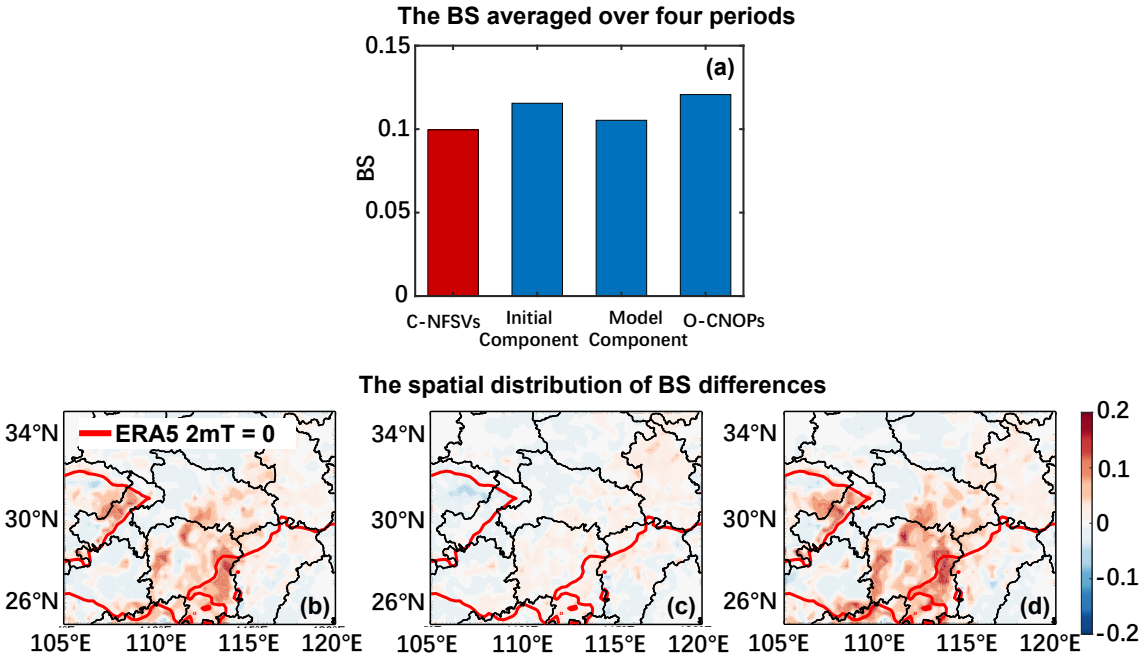

Fig. 4. (a) The Brier score (BS) averaged over southern China during the four periods under different scenarios; relative to the C-NFSVs scenario, the BS values increase by 16.6%, 5.72%, and 21.2% for the other scenarios, respectively. Spatial distribution of BS differences between the (b) initial components of C-NFSVs and C-NFSVs, (c) model components of C-NFSVs and C-NFSVs, and (d) O-CNOPs and C-NFSVs. The red line indicates the location of the 0°C isotherm in the ERA5 reanalysis averaged over the four periods. The BS values are calculated on the basis of the binary event of whether the 2 m temperature at each grid point in southern China decreases below 0°C.

for the high reliability of the C-NFSVs ensemble are investigated. Atmospheric circulation anomalies are the direct reasons for the extreme cold events in January and February 2008. The main anomalies include the prolonged Ural–Siberian blocking, a transverse trough, and an active southern branch trough (Yang et al., 2008; Zhao and Sun, 2008; Wang et al., 2009; Zuo et al., 2016). These anomalies are typically characterized by a 500 hPa geopotential height. The anomalies over mid-to-high latitudes enhance northerly winds, lead to the accumulation of cold air, and influence South China via the decay of the transverse trough and southward movement of the Siberian high, as reflected in sea level pressure (SLP). The active southern branch trough facilitated the eastward transmission of anomalies over the Tibetan Plateau, which also influenced the SLP over southern China. The convergence of cold air further contributes to the substantial temperature decline in the region (Zuo et al., 2016), aside from influencing water transport. To illustrate these atmospheric circulation systems, a schematic diagram is shown in Fig. 5.

To understand why the C-NFSVs ensemble effectively represents forecast uncertainty, we examine the structure of the C-NFSVs and the evolution of the corresponding ensemble spreads for relevant variables. All these are compared with the errors of control forecasts to determine whether the spreads capture the errors. From the structure of the C-NFSVs, represented by the ensemble spread of the 500 hPa wind and temperature components (Fig. 6), the perturbations are often concentrated in the upstream regions of the atmospheric circulation. These regions include areas ahead of the midlatitude ridge and near the southern branch trough. This suggests that the forecast uncertainty for 2 m temperature in southern China is sensitive to the combined effect of initial and model uncertainties located upstream in the circulation. Nevertheless, we find that the upstream temperature and wind uncertainties affect the forecasts of the cold event in South China in different ways across the four periods. In the first period, the ensemble spread generated by the C-NFSVs, which represents the forecast uncertainties, propagates downstream from the upstream circulation to southern China along the background flow. In contrast, for the later three periods, the spread representing uncertainty remains upstream but continues to influence southern China through sustained interaction with the background circulation. Further details are provided below.

The evolution of both the ensemble spread and the control forecast error of the 500 hPa geopotential height for different lead times in the forecasts of the first period of the cold event, from 9 to 16 January 2008, are shown in Fig. 7a. When the C-NFSVs are superimposed onto the control forecast, the geopotential height field responds to the perturbations of the wind and temperature components. The ensemble spread is concentrated in regions ahead of the midlatitude ridge and the southern branch trough at the initial stage of the forecast, similar to the perturbation structures shown in Fig. 6. As the forecast time progresses, the temperature and

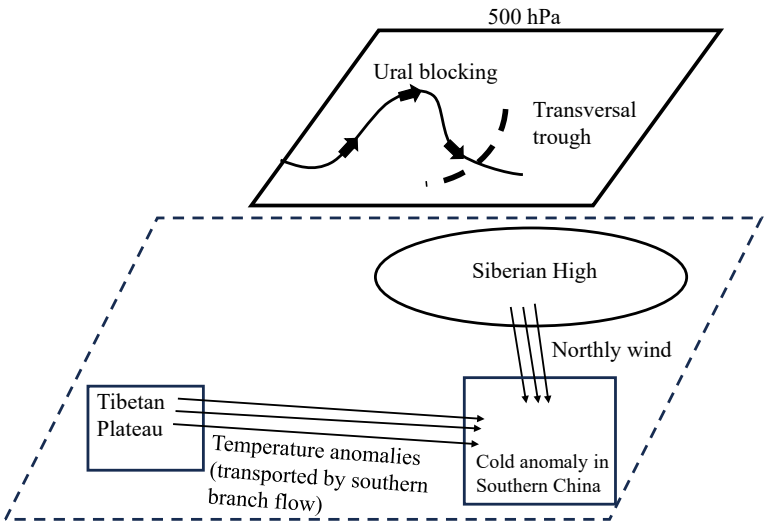

Fig. 5. Schematic diagram illustrating the atmospheric circulations leading to the extreme cold event in South China in early 2008.

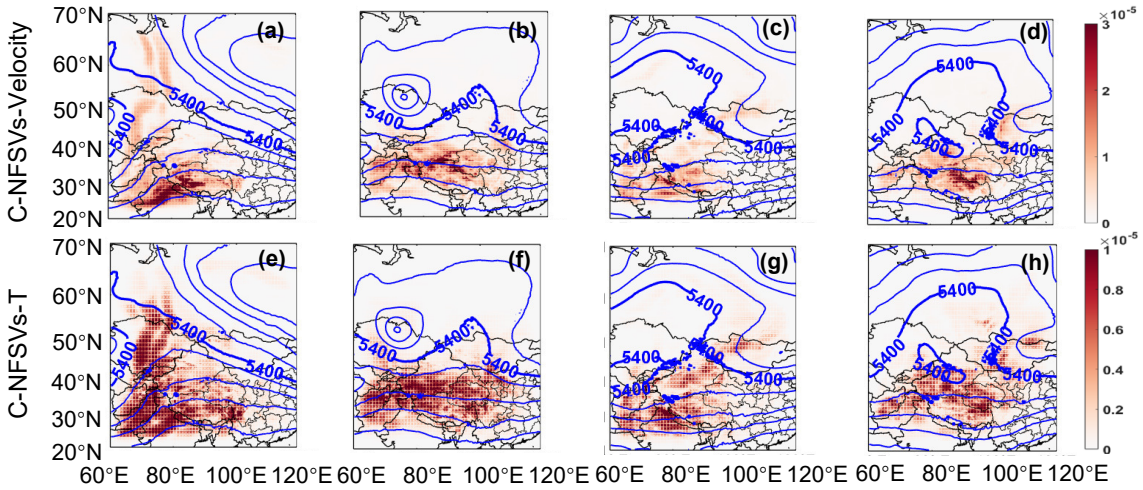

Fig. 6. C-NFSV perturbation structures: (a–d) ensemble spread of the velocity of wind perturbations at 500 hPa in the four forecasts (units: m s^{-1}) and (e–h) ensemble spread of temperature perturbations at 500 hPa in the same forecasts (units: K). The background field represents the geopotential height from the control forecast at the initial time in each forecast (blue contours; interval: 100 gpm).

wind perturbations propagate downstream under the influence of the background circulation. The ensemble spread region of the geopotential height field then moves to the low-pressure trough during the middle stage of the forecast, affecting southern China. The regions with a concentrated ensemble spread closely align with areas of the control forecast error at the corresponding forecast time. In the later stages of the forecast, the uncertainty of the 500 hPa geopotential height, as characterized by its ensemble spread, is transported eastward by the background circulation and moves out of the model domain. This reduces its influence on the 2 m temperature forecast uncertainty over southern China.

As previously mentioned, during this event, southern China was dominated by high-pressure systems centered over Siberia, whose intensity and spatial extent are typically represented by SLP. We also examined the temporal evolution of the ensemble spread and control forecast errors in SLP over the forecast lead time (Fig. 7b). At the early forecast stages, the SLP ensemble spread is concentrated over western China, closely resembling the spread patterns of the upper-level circulation shown in Fig. 7a. As the forecast progresses, this spread shifts southeastward and intensifies over southern China, indicating a baroclinic influence of upper-level systems on near-surface pressure. Notably, the regions of concentrated spread in SLP closely coincide with areas of forecast errors in the control over southern China before decreasing in the later stages as the upper-level systems exit the model domain. These results demonstrate that the C-NFSVs ensemble spread effectively captures the evolution and location of forecast errors in SLP. This, in turn, directly contributes to a reliable representation of 2 m temperature forecast errors and improves the forecasting skill over south-

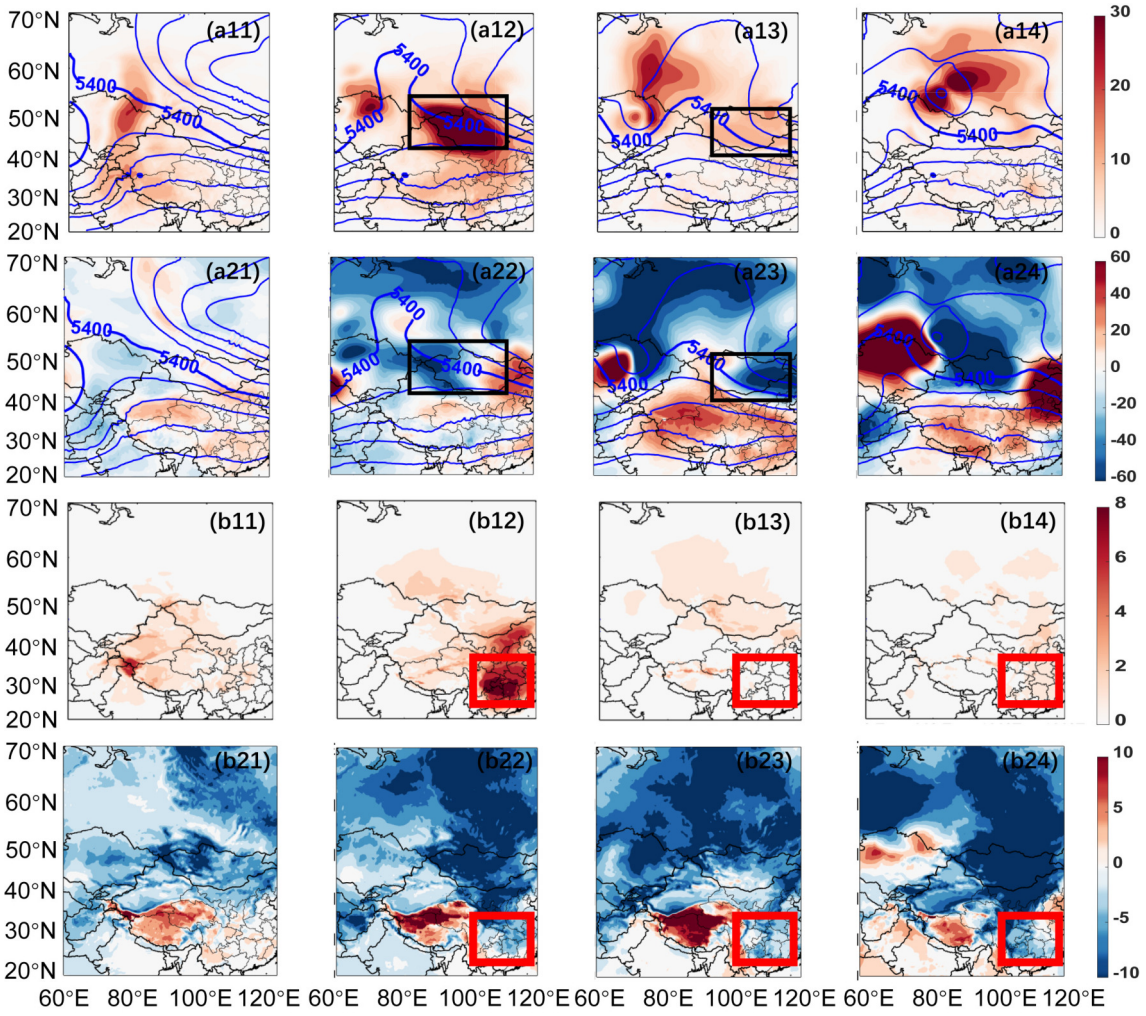

Fig. 7. (a11–a14) Ensemble spread of the 500 hPa geopotential height (units: gpm) from the C-NFSV forecasts at 0109T12, 0111T12, 0113T12, and 0115T12. (a21–a24) Forecast errors of the 500 hPa geopotential height from the control forecast at the corresponding times. In each panel, blue contours (interval: 100 gpm) show the 500 hPa geopotential height from the control forecast, and the black box indicates the low-pressure trough. Panels (b11–b14) and (b21–b24) are the same as in (a11–a14) and (a21–a24), but for sea level pressure (SLP; units: hPa). The red box indicates the southern China region.

ern China.

The above analysis indicates that the ensemble spread generated by the C-NFSVs effectively captures the forecast uncertainty in both the upper-level 500 hPa geopotential height and the near-surface SLP. As previously discussed, the southward movement and intensification of SLP toward southern China are generally guided by the 500 hPa circulation. This suggests that the high ensemble reliability of C-NFSVs in predicting 2 m temperature (Fig. 2) mainly comes from its ability to represent the uncertainties embedded in these processes. Similar results are obtained for the other three periods as well. For simplicity, we summarize the conclusions here without providing a detailed analysis. In general, the structure of the C-NFSV perturbations suggests that the areas contributing to the forecast uncertainty of the 2 m temperature over southern China lie upstream of the key circulation processes. These processes include downstream propagation from the mid-latitude high-pressure ridge to the low-pressure trough during the first period and eastward propagation along the southern branch flow during the subsequent three periods. Furthermore, through an analysis of the ensemble spread distributions in both upper-level circulation and near-surface SLP, we show that the C-NFSV ensemble forecasts can capture these circulation processes and the associated forecast uncertainties well. This ability may explain the high reliability of the C-NFSVs ensemble and its superior ability to improve the forecast skill.

4.3. *Reasons for the underestimation of the forecast uncertainty of the ECMWF with respect to the 2 m temperature over southern China*

As shown in Fig. S1 in the ESM, the ensemble forecast generated by ECMWF underestimates the forecast uncertainty in predicting the 2 m temperature over southern China. This underestimation may result from the limited skill of its ensemble spread in representing the actual forecast

errors and in capturing the key circulation processes that drive the temperature drop. As shown in Fig. S2 in the ESM, during the first period, the 500 hPa geopotential height ensemble spread of the ECMWF does not sufficiently concentrate ahead of the midlatitude ridge and the southern branch trough in the early stages of the forecast. As the forecast progresses, the spread is not effectively propagated and accumulated downstream into the low-pressure trough by the background circulation. This results in poor correspondence with the control forecast errors in this region and consequently limits its ability to adequately reflect the forecast uncertainty near the trough. In fact, since the ECMWF ensemble primarily captures circulation uncertainties at higher latitudes, such as Mongolia and Siberia, the SLP spread is mostly confined to these areas (Fig. S3 in the ESM). It thus offers limited representation of the Siberian high intensity at lower latitudes. As a result, the ability of the ensemble to effectively capture forecast errors over southern China is limited, which may explain its underestimation of forecast uncertainty in predicting 2 m temperature in this region. Evidently, the ECMWF ensemble spread shows a weaker correspondence with the control forecast errors. The spread of the upper-level circulation field remains confined primarily to higher latitudes, showing limited sensitivity to forecast uncertainties over mid- to low-latitude regions, with similar results observed in the subsequent three periods. This spatial mismatch further restricts the ability of the ensemble to capture the near-surface SLP forecast errors over southern China. Consequently, its effectiveness in representing the forecast uncertainty of 2 m temperature in this region is limited.

These explanations underscore that, compared with the ECMWF ensemble forecasts, those generated by the C-NFSVs more effectively characterize the forecast uncertainty. As mentioned in the introduction, this may be due to the SV initial perturbation method and stochastic physical perturbation scheme used in the ECMWF ensemble. These perturbations may not capture the interaction between initial and model errors and may not fully represent the nonlinear evolution of forecast errors. This limitation could hinder their ability to capture key circulation processes that determine forecast uncertainty. In contrast, the C-NFSVs method accounts for nonlinear error growth and generates coordinated initial and model perturbations, allowing a more realistic representation of forecast uncertainty.

5. Identification of forecast uncertainty sources

In section 4.1, we report that when averaged over the four periods, the forecast uncertainty of the 2 m temperature appears to be influenced primarily by the model component of the C-NFSVs, especially in the later stages of the forecasts. Compared with that generated by the initial components, the ensemble spread generated by the model components is better at characterizing the forecast uncertainty. However, through the mechanistic analysis in section 4.2, it is evident that during the first period, the development of spread-characterized uncertainty exhibits downstream propagation guided by the background circulation. This propagation corresponds well with the regions where forecast errors accumulate and enables the ensemble spread to effectively reflect the forecast errors. Such a mechanism indicates that the forecast uncertainty during this period is driven primarily by initial uncertainties concentrated upstream of the background circulation.

To verify this, Fig. 8 shows the temporal evolution of the spread–RMSE ratio for the ensemble forecasts generated by separately incorporating the initial and model perturbation components of the C-NFSVs across the four periods of the 2008 cold event. As shown, for the first period, initial perturbations contribute more significantly to the ensemble reliability than model perturbations. The spread–RMSE ratio of forecasts using only initial perturbations reaches a level comparable to that of the C-NFSVs ensemble. This is consistent with the earlier finding that initial uncertainties play a dominant role during this period.

To further investigate this, the mechanisms behind the differing contributions of initial and model perturbations to forecast uncertainty warrant closer examination. Following the approach used in section 4.2, we analyze the spatial evolution of the ensemble spread for the 500 hPa geopotential height and SLP in the ensemble forecasts generated by incorporating only the initial or model perturbations.

As illustrated in Fig. 9, for the first period, under the scenario with only initial perturbations, which are "one-time" additions, the perturbations are transported downward by the background flow. Those located ahead of the high-latitude ridge shift eastward and concentrate around the low-pressure trough. Moreover, perturbations associated with the southern branch trough are transported to higher latitudes because of the convergence of the northern and southern streams, ultimately converging in the same low-pressure trough region. In contrast, under the scenario with only model perturbations, which are consistently added with small magnitude, the ensemble spread remains quite small during the early forecast stages. As the forecast progresses, the high-latitude ridge in the background flow gradually shifts eastward, blocking the persistently added midlatitude perturbations and preventing their further downstream propagation. Additionally, the perturbations associated with the southern branch trough accumulate in the convergence area of the northern and southern streams. However, they do not propagate effectively toward the downstream trough because the convergence location is situated too far west. To more clearly illustrate the difference in the ensemble spread between the two cases, the mean spread was calculated over the key region (see the black boxes in Figs. 9a12 and a22, representing the trough). The mean spread is 24.96 gpm when only initial perturbations are added, whereas it is 9.74 gpm when only model perturbations are added. These results suggest that the ensemble spread from initial perturbations more effectively captures the circulation processes driving the 2 m temperature forecast uncertainty in southern China. In contrast, the continuously added model perturbations, constrained by the background

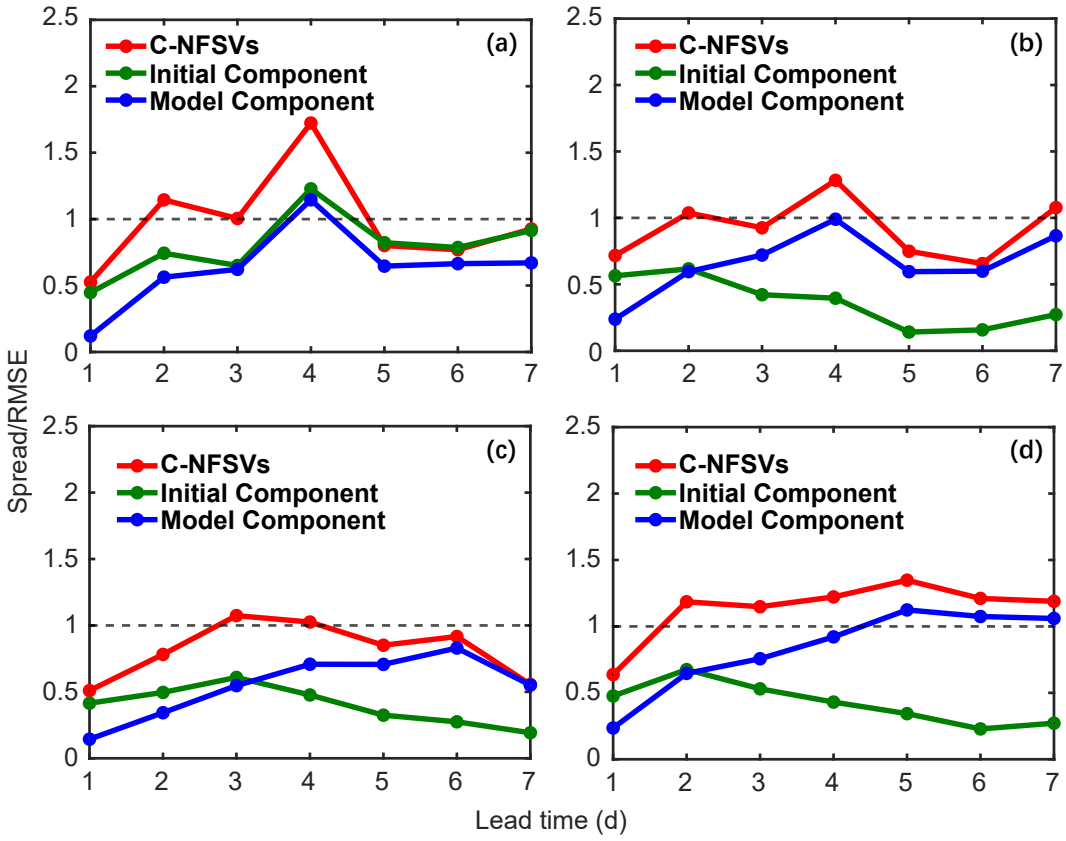

Fig. 8. Temporal spread–RMSE ratio over southern China for the ensemble forecasts generated by the C-NFSVs, along with their initial perturbation and model perturbation components, for the forecast periods (a) 20080109T00–20080116T00, (b) 20080115T00–20080122T00, (c) 20080122T00–20080129T00, and (d) 20080128T00–20080204T00, as shown in Table 1.

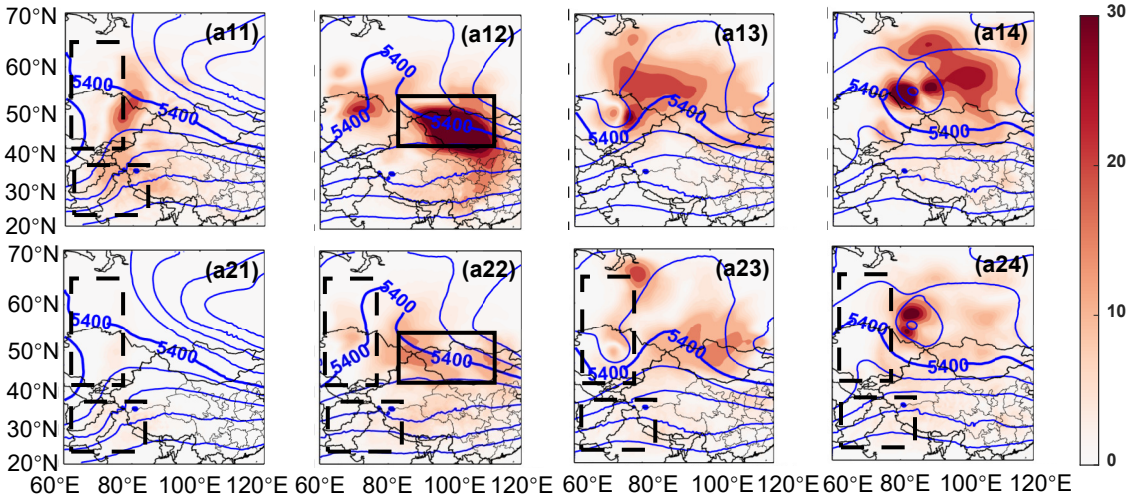

Fig. 9. Spatial distribution of the ensemble spread of the 500 hPa geopotential height (units: gpm) at 0109T12, 0111T12, 0113T12, and 0115T12. Panels (a11–a14) and (a21–a24) show the ensemble spreads of the 500 hPa geopotential height from the initial and model components of the C-NFSVs, respectively, with blue contours denoting the control forecast geopotential height (interval: 100 gpm). The solid and dashed black boxes indicate the low-pressure trough and the region of superimposed perturbations, respectively.

flow, fail to propagate toward the critical downstream regions and thus cannot adequately represent the associated forecast uncertainty, consistent with Fig. 8a.

For the other three periods, the situation clearly reverses, as shown in Fig. 8. Compared with those generated using only initial perturbations, ensemble forecasts that incorporate only model perturbations yield a significantly improved spread–RMSE ratio, more closely aligning with

that of the C-NFSVs ensemble. These findings indicate that during these periods, the ensemble reliability is largely attributable to the impact of model perturbations, suggesting that forecast uncertainty primarily arises from model uncertainty. Such insight leads to an inquiry into the distinct roles of initial and model perturbations in shaping ensemble reliability across these periods and how their respective influences differ.

Taking the third period as an example, we further examine the mechanisms through which model uncertainty serves as the primary contributor to forecast uncertainty. During this period, the evolution of the ensemble spread generated by the C-NFSVs is very similar to that produced by their model perturbation component. Therefore, we directly present the spread evolution of the C-NFSVs model component and compare it with that of the initial component. The perturbations at the 500 hPa level are primarily concentrated near the southern branch trough (Fig. 10). The perturbations predominantly propagate along the southern branch during the early forecast stages. In the case where only initial perturbations are considered, the ensemble spread is gradually transported to higher latitudes during the mid-forecast period under the influence of the background flow. The initial perturbations cause the overall spread to shift northward. Combined with the weak northerly wind in the later stage, this leads to a weaker impact on southern China. As a result, the ensemble spread generated by considering only initial perturbations fails to capture forecast errors over southern China. To further quantify the differences caused by the initial uncertainty component and model uncertainty component, the mean spread was calculated over the key region influencing southern China (the black boxes in Figs. 8a3 and 8b3). The mean spread is 2.82 gpm when only initial perturbations are added, whereas it is 4.46 gpm when only model perturbations are added, suggesting that model perturbations grow more effectively over regions with large forecast errors than initial perturbations. In Fig. 11, we present the distribution of SLP spread under the addition of both types of perturbations. Compared with the scenario with only initial perturbations, the ensemble spread with model perturbations clearly becomes more distinctly concentrated in southern China during the later stage, aligning better with the forecast error in that region. This enables better capture of the 2 m temperature forecast uncertainty and identifies model errors as its primary source during this period, which is also confirmed in Fig. 8c.

Through the above analysis, we find that the sensitivity of each process to initial and model perturbations depends primarily on the evolution of the background flow field. In the first period, the background circulation favors the downstream transport of initial perturbations toward key regions. However, as time progresses, the background flow itself changes considerably, with substantial shifts in the wind direction, making it difficult for subsequently added model perturbations at the same location to further propagate down-

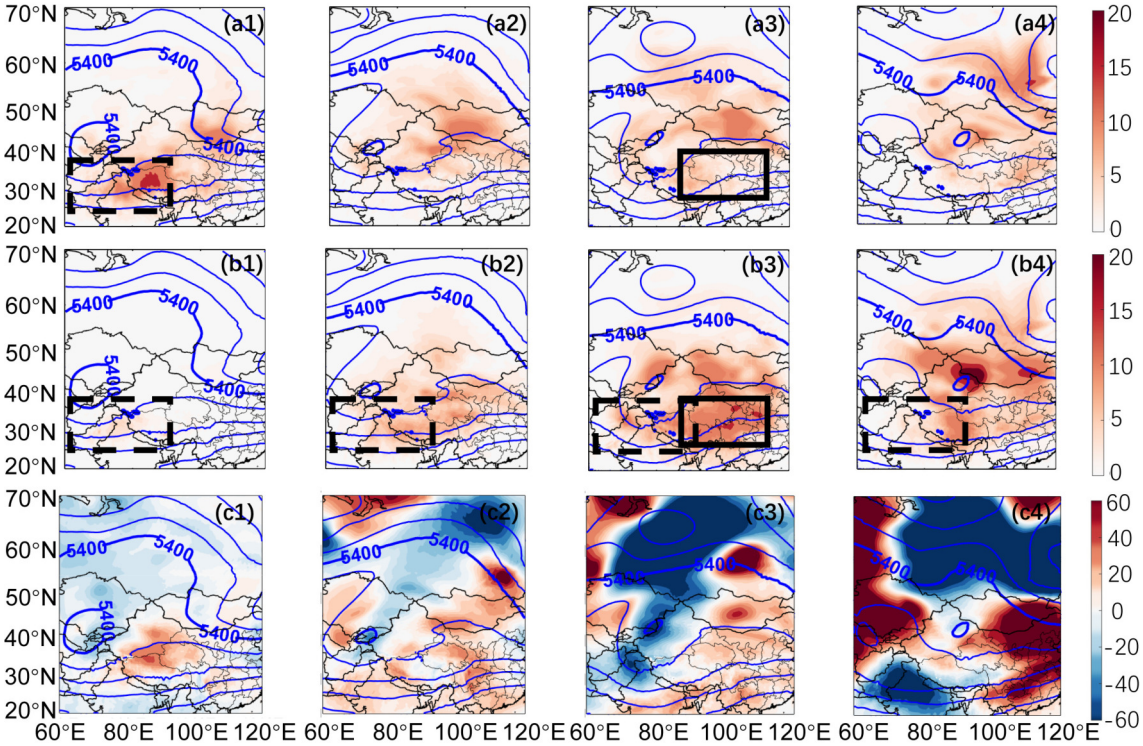

Fig. 10. Spatial distribution of the ensemble spread for the 500 hPa geopotential height (units: gpm) at 0122T12, 0124T12, 0126T12, and 0128T12: (a1–a4) ensemble forecasts with initial perturbations of C-NFSVs and (b1–b4) ensemble forecasts with model perturbations of C-NFSVs. (c1–c4) Control forecast errors at the corresponding times. The background field represents the geopotential height from the control forecast at the respective times (blue contours, interval: 100 gpm). The dashed black boxes indicate the regions of superimposed perturbations.

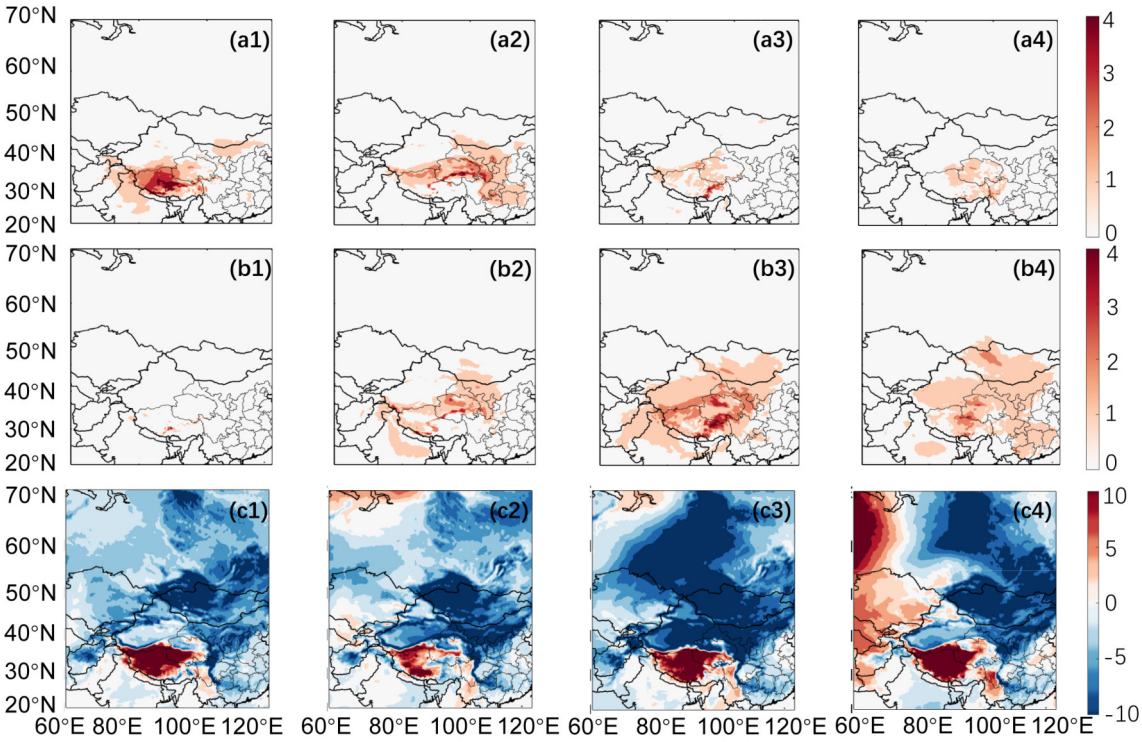

Fig. 11. As in Fig. 10 but for SLP (units: hPa).

stream. In contrast, during the third period, the background flow exhibited smaller directional changes, allowing the continuously added model perturbations to be steadily transported downstream and influence the study region.

In summary, considering only one type of perturbation may lead to an inadequate representation of forecast uncertainty in ensemble forecasts. In contrast, the C-NFSVs method, which accounts for both initial and model perturbations, has the potential to provide a more accurate estimate of forecast uncertainty. Moreover, the analysis of these two periods reveals that examining the spatial evolution of the ensemble spread generated by the C-NFSVs is useful for investigating forecast uncertainty. This offers a practical approach for identifying the exact underlying sources of forecast uncertainty in extreme cold events.

6. Discussion

In the above analysis, we focused on the extreme cold event that occurred in 2008 and used the C-NFSVs ensemble forecasting approach to estimate the forecast uncertainty of the 2 m temperature during the event. In section 4.1, we evaluated the C-NFSVs ensemble forecasts in terms of reliability using the spread–RMSE ratio, deterministic skill through the ensemble-mean RMSE relative to the control forecast, and probabilistic skill. While the C-NFSVs method improves the ensemble reliability, the improvements in the deterministic and probabilistic skill remain relatively small. In ensemble forecasting, accurately representing forecast uncertainty is a key function, and a close match between the ensemble spread and the ensemble-mean forecast error indicates good reliability. However, this does not necessarily translate into a large improvement in the ensemble-mean forecast over the control forecast, which is consistent with the study of Feng et al. (2019), who further verified this through experiments with the idealized Lorenz-96 model. For probabilistic skill, if the ensemble size is sufficient, good reliability reflected by the spread–RMSE ratio may suggest reasonable probabilistic performance. In the current study, the relatively small improvement in probabilistic skill is likely due to the limited number of ensemble members, despite a reasonable spread–RMSE ratio. In future work, we plan to use larger ensemble sizes to further examine the C-NFSVs forecasting skill and determine the number of members needed to enhance the forecast skill while maintaining good reliability.

In fact, similar analyses were also conducted for the extreme cold events that occurred in 2016 and 2024, including one period from 21 to 25 January 2016 and three periods from 20 to 22 January, 13 to 15 February, and 17 to 22 February 2024. The results demonstrate the ability of the C-NFSVs-based ensemble forecasts to characterize the forecast uncertainty and identify its sources. The forecast uncertainty is caused primarily by initial uncertainty in the first period in 2024. In contrast, model uncertainty plays a dominant role in the other two periods in 2024 and in the first period in 2016. These results are further confirmed by the comparison shown in Fig. S4 in the ESM. Analysis of the extreme cold events in 2016 and 2024 also demonstrated that the spatial evolution of the ensemble spread generated by the C-NFSVs can be used to identify the sources of forecast uncertainty, suggesting its potential applicability to future cases.

Notably, beyond the 2008 event (Fig. S1 in the ESM), the ECMWF ensemble forecasts also underestimated the 2 m temperature forecast uncertainty for the 2016 and 2024 extreme cold events. On the basis of this analysis, it appears that the ECMWF generally underestimates the forecast uncertainty for extreme cold events in southern China. A comparison with the results from C-NFSVs ensemble forecasts suggests that the limitations of the SV initial perturbation method and stochastic physical perturbation scheme of the ECMWF, as previously mentioned, are likely the primary causes of this underestimation of forecast uncertainty. Nevertheless, this is only an inference and should be further examined in the future. In addition, in the present study of the 2008 extreme cold event, the forecast length was restricted to 7 days because of the 7.5-day lateral boundary conditions provided by the GFS data used to drive the WRF model. For consistency, we maintained a 7-day forecast window for all events (including the 2016 and 2024 cases), even though longer boundary conditions were available for the latter two. Moving forward, the C-NFSVs method, which combines initial condition and model perturbations, is expected to yield more reliable uncertainty estimates and greater forecast skill improvements, particularly for extended lead times.

7. Conclusion

In this study, we explore the application of the C-NFSVs ensemble forecast method in identifying the forecast uncertainty for extreme cold events that occurred in China. The 2008 extreme cold event over southern China is adopted as a representative case to determine how the ensemble forecast generated by C-NFSVs is applied to characterize the forecast uncertainty and help identify the sources of uncertainty.

With the optimal configuration for the C-NFSVs determined by sensitivity experiments, the ensemble members with high reliability, as indicated by a spread–RMSE ratio closest to 1, are obtained. On this basis, the ensemble forecasts were conducted by incorporating the C-NFSVs as well as their initial and model perturbation components separately. The results show that when averaged across the four forecast periods of the 2008 extreme cold event, the ensemble forecasts with only the initial or model perturbation are insufficient to adequately represent forecast uncertainty. This behavior is consistent with the dynamically reasonable expectations for ensemble forecasting systems in representing the evolution of forecast uncertainty. It contributes to the C-NFSVs sustaining ensemble reliability throughout the forecast and offers a spread that meaningfully reflects the temporal variation in forecast uncertainty.

Our results further suggest that the evolution of the ensemble spread generated by the C-NFSVs can offer insights into the possible sources of forecast uncertainty—whether stemming from initial or model-related uncertainties—and their locations. Specifically, in events where the spread-characterized uncertainty propagates downstream from the upstream region and then remains concentrated in key areas, such as the first period of the 2008 events, this spread behavior is clearly observed. Forecast uncertainty in these cases is driven primarily by initial errors located upstream of the background circulation. These findings are further confirmed by the ensemble reliability derived from forecasts generated using only the initial or model components of the C-NFSVs. The results are consistent with the identified sources of forecast uncertainty. In contrast, for events where the ensemble spread persists in the upstream region and is continuously transported downstream by the background flow, it gradually accumulates in key areas and influences the 2 m temperature. Accordingly, the forecast uncertainty is attributed mainly to model errors. In such cases, the ensemble spread generated using only the initial perturbation component is rapidly advected out of the study area, resulting in a weaker influence on the 2 m temperature. This pattern is also confirmed by the ensemble reliability of forecasts generated with individual perturbation components, which exhibit a notable underestimation of forecast uncertainty when only the initial component is applied. The findings from the 2016 and 2024 extreme cold events provide further support for these results. These cases collectively suggest that analyzing the spatial evolution of the ensemble spread generated by the C-NFSVs may offer an approach for identifying the exact sources of forecast uncertainty in extreme cold events.

As the first application of the C-NFSVs ensemble forecasting method within a realistic numerical model, this study focused on extreme cold events over southern China. Notably, the approach is not only effective for current cases but also has potential for longer-lead forecasts and other high-impact weather events. By providing valuable insights into the estimation and sources of forecast uncertainty, it may help improve forecast performance and support broader efforts in disaster risk reduction.

Acknowledgements. The authors greatly appreciate the insightful comments from the two reviewers. This study was supported by the National Natural Science Foundation of China (Grant No. 42330111) and the Meteorological Joint Fund (Grant No. U2442221).

Data availability. The TIGGE data are available at https://apps.ecmwf.int/datasets/. The GFS forecast data are available at https://www.ncei.noaa.gov/products/weather-climate-models/global-forecast. The ERA5 reanalysis data are available at https://cds.climate.copernicus.eu/datasets.

Electronic supplementary material**:** Supplementary material is available in the online version of this article at https://doi.org/10.1007/s00376-025-5443-0.

APPENDIX A

A1. Root-Mean-Square Error of the Ensemble Mean and Ensemble Spread

For an ensemble forecast, the ensemble mean is calcu-

lated as the arithmetic average of all members. Let $\overline{X}_n$ represent the ensemble mean of any variable at the nth grid point, which is determined as follows:

$$\overline{X}_n = \frac{1}{M}\sum_{m=1}^{M} X_n^m , \tag{A1}$$

where X_n^m denotes the 2 m temperature of the mth ensemble member at the nth grid point and $m = 1,2,\ldots,M$. The ensemble mean root-mean-square error (RMSE) is the difference between the ensemble mean and the observation value O_n (in this study, the ERA5 data). It is defined as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(\overline{X}_n - O_n\right)^2} , \tag{A2}$$

where N represents the total number of grid points in the study region.

The ensemble spread characterizes the dispersion among the ensemble members and is defined as follows:

$$\mathrm{SPREAD} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\frac{1}{M-1}\sum_{m=1}^{M}\left(X_n^m - \overline{X}_n\right)^2} . \tag{A3}$$

when averaging the RMSE and spread over multiple forecast periods, the forecasts and corresponding observations from all periods are first combined. The RMSE and spread are then calculated from these combined data for each period.

APPENDIX B

B1. Brier Score

The Brier score (BS; Brier 1950) describes the mean squared error of probabilistic forecasts and is defined as follows:

$$\mathrm{BS} = \frac{1}{N}\sum_{n=1}^{N}(f_n - o_n)^2 , \tag{A4}$$

where N is the total number of grid points in the study region, and f_n and o_n represent the forecast and observation probabilities, respectively, at the nth grid point. A smaller BS indicates higher skill in the probabilistic forecast.

B2. Relative Operating Characteristic Curve Area

The relative operating characteristic curve area (ROCA; Mason 1982) is another commonly used statistic to assess the skill of probabilistic forecasts. For a binary event, on the basis of the forecast and observation of the occurrence of the event at each grid point, a 2 × 2 contingency table (see Table B1) can be constructed.

In Table B1, a and c represent the numbers of hits and false alarms, respectively, and b and d represent the numbers of misses and correct rejections, respectively.

Table B1. The 2 × 2 contingency table of a binary event.

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

The ROCA is calculated on the basis of the area under the receiver operating characteristic (ROC) curve, where a larger area indicates better probabilistic forecast skill. Then, the hit rate (HR) and false alarm rate (FR) can be expressed as follows:

$$\mathrm{HR} = \frac{a}{a+b} ; \tag{A5}$$

$$\mathrm{FR} = \frac{c}{c+d} . \tag{A6}$$

By selecting different forecast probability thresholds between 0 and 1, multiple HR and FR pairs can be obtained, which are subsequently used to construct the ROC curve. The area under the ROC curve, integrated over the interval [0,1], is the ROCA. A larger ROCA indicates higher probabilistic forecast skill. A ROCA value greater than 0.5 is generally considered to indicate skill in the probabilistic forecast.

B3. Reliability Diagram

The reliability diagram (RD) is used to evaluate the match between forecast probabilities and observed frequencies. For a binary event, forecast probabilities are divided into several small intervals within the range [0,1], and the observed frequencies are calculated within each corresponding interval. If the forecast probability is consistent with the event frequency, the reliability curve will be closer to the diagonal, indicating higher reliability of the probabilistic forecast.

REFERENCES

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

Birgin, E. G., J. M. Martínez, and M. Raydan, 2000: Nonmonotone spectral projected gradient methods on convex sets. *SIAM Journal on Optimization*, **10**, 1196–1211, https://doi.org/10.1137/S1052623497330963.

Bougeault, P., and Coauthors, 2010: The THORPEX interactive grand global ensemble. *Bull. Amer. Meteor. Soc.*, **91**, 1059–1072, https://doi.org/10.1175/2010BAMS2853.1.

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

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

Buizza, R., M. Milleer, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887–2908, https://doi.org/10.1002/qj.49712556006.

Ding, Y. H., Z. Y. Wang, Y. F. Song, and J. Zhang, 2008: The unprecedented freezing disaster in January 2008 in southern China and its possible association with the global warming. *J. Meteor. Res.*, **22**, 538–558.

Du, J., and Coauthors, 2019: Ensemble methods for meteorological predictions. *Handbook of Hydrometeorological Ensemble Forecasting*, Q. Y. Duan et al., Eds., Springer, 99–149, https://doi.org/10.1007/978-3-642-39925-1_13.

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.

Duan, W. S., J. J. Ma, and S. Vannitsem, 2022: An ensemble forecasting method for dealing with the combined effects of the initial and model errors and a potential deep learning implementation. *Mon. Wea. Rev.*, **150**, 2959–2976, https://doi.org/10.1175/MWR-D-22-0007.1.

Duan, W. S., L. C. Yang, M. Mu, B. Wang, X. S. Shen, Z. Y. Meng, and R. Q. Ding, 2023a: Recent advances in China on the predictability of weather and climate. *Adv. Atmos. Sci.*, **40**, 1521–1547, https://doi.org/10.1007/s00376-023-2334-0.

Duan, W. S., L. C. Yang, Z. Z. Xu, and J. Chen, 2023b: Conditional nonlinear optimal perturbation: Applications to ensemble forecasting of high-impact weather systems. *Numerical Weather Prediction: East Asian Perspectives*, S. K. Park, Ed.,Springer, 441–460, https://doi.org/10.1007/978-3-031-40567-9_17.

ECMWF, 2015: Part V: Ensemble prediction system. *IFS Documentation CY41R1*. , ECMWF, 25 pp., https://doi.org/10.21957/eow1lonc.

Feng, J., J. P. Li, J. Zhang, D. Q. Liu, and R. Q. Ding, 2019: The relationship between deterministic and ensemble mean forecast errors revealed by global and local attractor radii. *Adv. Atmos. Sci.*, **36**, 271–278, https://doi.org/10.1007/s00376-018-8123-5.

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

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

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

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

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

Jiang, Q., X. K. Ma, and F. Wang, 2016: Analysis of the January 2016 atmospheric circulation and weather. *Meteor. Mon.*, **42**, 514–520, https://doi.org/10.7519/j.issn.1000-0526.2016.04.016. (in Chinese with English abstract)

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

Leutbecher, M., S. Lang, S.-J. Lock, C. D. Roberts, and A. Tsiringakis, 2024: Improving the physical consistency of ensemble forecasts by using SPP in the IFS. Newsletter No. 181, 8 pp., https://doi.org/10.21957/mlz238dk1p.

Li, C. Y., H. Yang, and W. Gu, 2008: Cause of severe weather with cold air, freezing rain and snow over South China in January 2008. *Climatic and Environmental Research*, **13**, 113–122, https://doi.org/10.3878/j.issn.1006-9585.2008.02.01. (in Chinese with English abstract)

Li, X., Y.-J. Zhang, H. Gao, and T. Ding, 2022: Extreme cold wave in early November 2021 in China and the influences from the meridional pressure gradient over East Asia. *Advances in Climate Change Research*, **13**, 797–802, https://doi.org/10.1016/j.accre.2022.11.011.

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.

Lorenz, E. N., 1982: Atmospheric predictability experiments with a large numerical model. *Tellus*, **34**, 505–513, https://doi.org/10.3402/tellusa.v34i6.10836.

Ma, N., Y.-F. Li, and J.-H. Ju, 2011: Intraseasonal oscillation characteristics of extreme cold, snowy and freezing rainy weather in southern China in early 2008. *Plateau Meteorology*, **30**, 318–327. (in Chinese with English abstract)

Mason, I., 1982: A model for assessment of weather forecasts. *Aust. Meteor. Mag.*, **30**, 291–303, https://doi.org/10.1071/es82036.

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

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

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

Qi, L. B., 2012: Formation mechanism and forecast on freezing rain and ice pellet in winter of China. *Meteor. Mon.*, **38**, 769–778, https://doi.org/10.7519/j.issn.1000-0526.2012.07.001. (in Chinese with English abstract)

Qing, Y. P., J. Y. Wu, and J.-J. Luo, 2025: Characteristics and subseasonal prediction of four types of cold waves in China. *Theor. Appl. Climatol.*, **156**, 192, https://doi.org/10.1007/s00704-025-05420-3.

Ryti, N. R. I., Y. M. Guo, and J. J. K. Jaakkola, 2016: Global association of cold spells and adverse health effects: A systematic review and meta-analysis. *Environmental Health Perspectives*, **124**, 12–22, https://doi.org/10.1289/ehp.1408104.

Swinbank, R., and Coauthors, 2016: The TIGGE project and its achievements. *Bull. Amer. Meteor. Soc.*, **97**, 49–67, https://

doi.org/10.1175/BAMS-D-13-00191.1.

Tao, S. Y., and J. Wei, 2008: Severe snow and freezing-rain in January 2008 in the southern China. *Climatic and Environmental Research*, **13**, 337–350, https://doi.org/10.3878/j.issn.1006-9585.2008.04.01. (in Chinese with English abstract)

Toth, Z., and S. Vannitsem, 2002: Model errors and ensemble forecasting. *Operational Systems Workshop*, ECMWF, Reading, 146–154.

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

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

Wang, D. H., and Coauthors, 2009: A preliminary analysis of features and causes of the snow storm event over the southern areas of China in January 2008. *J. Meteor. Res.*, **23**, 374–384.

Wang, Z. Y., and Y. H. Ding, 2006: Climate change of the cold wave frequency of China in the last 53 years and the possible reasons. *Chinese Journal of Atmospheric Sciences*, **30**, 1068–1076, https://doi.org/10.3878/j.issn.1006-9895.2006.06.02. (in Chinese with English abstract)

Wei, F. Y., 2008: The changing characteristics of China's cold wave disaster under the background of climate warming. *Progress in Natural Science*, **18**, 289–295. (in Chinese)

Xie, C., H. L. Gui, and Y. You, 2024: Analysis of the January 2024 atmospheric circulation and weather. *Meteor. Mon.*, **50**, 514–520, https://doi.org/10.7519/j.issn.1000-0526.2024.032901. (in Chinese with English abstract)

Yamaguchi, J., Y. Kanno, G. X. Chen, and T. Iwasaki, 2019: Cold air mass analysis of the record-breaking cold surge event over East Asia in January 2016. *J. Meteor. Soc. Japan. Ser. II*, **97**, 275–293, https://doi.org/10.2151/jmsj.2019-015.

Yang, G. M., Q. Kong, D. Y. Mao, F. H. Zhang, Z. M. Kang, and Z. P. Zong, 2008: Analysis of the long-lasting cryogenic freezing rain and snow weather in the beginning of 2008. *Acta Meteorologica Sinica*, **66**, 836–849, https://doi.org/10.11676/qxxb2008.076. (in Chinese with English abstract)

Yu, X. D., H. Y. Fei, and X. M. Wang, 2024: Some issues concerning the large area rainstorm, snowstorm, sleet, freezing rain and severe convection event over China from 17 to 23 February 2024. *Meteor. Mon.*, **50**, 1033–1042, https://doi.org/10.7519/j.issn.1000-0526.2024.080902. (in Chinese with English abstract)

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

Zhang, Y. C., W. S. Duan, S. Vannitsem, and H. Zhang, 2023b: A new approach to represent model uncertainty in the forecasting of tropical cyclones: The orthogonal nonlinear forcing singular vectors. *Quart. J. Roy. Meteor. Soc.*, **149**, 2206–2232, https://doi.org/10.1002/qj.4502.

Zhang, Y. J., Z. C. Yin, H. J. Wang, and S. P. He, 2021: 2020/21 record-breaking cold waves in east of China enhanced by the 'Warm Arctic-Cold Siberia' pattern. *Environmental Research Letters*, **16**, 094040, https://doi.org/10.1088/1748-9326/ac1f46.

Zhao, S. W., and J. H. Sun, 2008: Multi-scale systems and conceptual model on freezing rain and snow storm over southern China during January-February 2008. *Climatic and Environmental Research*, **13**, 351–367, https://doi.org/10.3878/j.issn.1006-9585.2008.04.02. (in Chinese with English abstract)

Zheng, F., and Coauthors, 2022: The 2020/21 extremely cold winter in China influenced by the synergistic effect of La Niña and warm Arctic. *Adv. Atmos. Sci.*, **39**, 546–552, https://doi.org/10.1007/s00376-021-1033-y.

Zuo, Q. J., S. T. Gao, and X. G. Sun, 2016: Effects of the upstream temperature anomaly on freezing rain and snowstorms over Southern China in early 2008. *J. Meteor. Res.*, **30**, 694–705, https://doi.org/10.1007/s13351-016-5253-5.