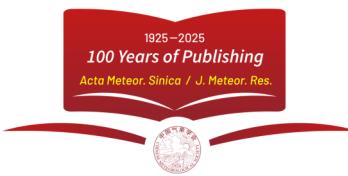


• REVIEW •



A Review on Development, Challenges, and Future Perspectives of Ensemble Forecast

Jing CHEN^{1,2*}, Yuejian ZHU^{1,2}, Wansuo DUAN³, Xiefei ZHI⁴, Jinzhong MIN⁴, Xiaoli LI^{1,2}, Guo DENG^{1,2}, Huiling YUAN⁵, Jie FENG⁶, Jun DU⁷, Qiaoping LI^{1,2}, Jiandong GONG^{1,2}, Xueshun SHEN^{1,2}, and Mu MU⁶

1 *CMA Earth System Modeling and Prediction Centre (CEMC), China Meteorological Administration (CMA), Beijing 100081, China*

2 *State Key Laboratory of Severe Weather Meteorological Science and Technology (LaSW), CEMC, Beijing 100081, China*

3 *Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China*

4 *Nanjing University of Information Science & Technology, Nanjing 210044, China*

5 *Nanjing University, Nanjing 210008, China*

6 *Fudan University, Shanghai 200433, China*

7 *National Centers for Environmental Predictions, National Oceanic and Atmospheric Administration, Maryland 20740, USA*

(Received 5 January 2025; in final form 5 March 2025)

ABSTRACT

This paper reviews the development of ensemble weather forecast and the primary techniques employed in the main ensemble prediction systems (EPSs) designed by China and other countries. Here, the emphasis is placed on the advancements in the China Meteorological Administration (CMA) global and regional ensemble prediction systems (i.e., CMA-GEPS and CMA-REPS), with particular attention to operational technologies such as initial and model perturbation methods and the applications of ensemble forecast. Through comparative verification with EPSs from other leading international numerical weather prediction (NWP) centers, CMA's EPSs demonstrate forecast skills comparable to its global counterparts. As EPSs progress to convective scales and coupled systems between sea, land, air, and ice, the paper addresses some key challenges in ensemble forecast technologies across the aspects of operation, science, integration of artificial intelligence (AI), merging of weather and climate models, and challenging user requirements. Finally, a summary of conclusions and future perspectives on ensemble forecast are provided.

Key words: ensemble forecast, method, China Meteorological Administration (CMA) ensemble forecast, review

Citation: Chen, J., Y. J. Zhu, W. S. Duan, et al., 2025: A review on development, challenges, and future perspectives of ensemble forecast. *J. Meteor. Res.*, **39**(3), 534–558, <https://doi.org/10.1007/s13351-025-4909-4>.

1. Introduction

Numerical weather prediction (NWP) is one of the greatest scientific advancements of the 20th century, often described as a “quiet revolution” (Bauer et al., 2015; Shen et al., 2020). Among its key achievements, four-dimensional variational data assimilation (4DVar) and ensemble forecasting are regarded as two major milestones. The atmospheric system is inherently nonlinear, characterized by chaos that is sensitive to minor initial perturbations. Even with ideal numerical models, uncertainties in the initial states—arising from errors in observations and assimilation systems—lead to significant forecast errors as the forecast lead time increases (Lorenz, 1963; Chou, 1986; Mu et al., 2003, 2011). Additionally, numerical

models employ discretization schemes, sub-grid parameterizations, and other approximations to represent the dynamical and physical processes of the atmosphere. Consequently, single deterministic NWP has inherent forecast uncertainty (Li et al., 2000; Arakawa, 2004; Chen et al., 2004). Quantifying this uncertainty has thus become a critical focus for NWP. The ensemble prediction system (EPS), which may transform single deterministic numerical forecasts into probability density function (PDF) distributions and estimate the uncertainty of deterministic numerical forecasts, was first developed through the Monte Carlo method (Epstein, 1969; Leith, 1974).

Theoretically, the evolution of the PDF over time can be described by the Liouville equation. However, solving this equation is highly challenging and practically in-

Supported by the National Natural Science Foundation of China (U2242213 and 42341209).

*Corresponding author: chenj@cams.gov.cn

© The Chinese Meteorological Society 2025

feasible, even for a nonlinear system with only a few degrees of freedom. To address this issue, the ensemble forecasting method was developed (Molteni et al., 1996). This method utilizes specific mathematical techniques to generate a set of initial conditions, each potentially representing the true state of the atmosphere. These initial conditions are then integrated by a numerical model to produce an ensemble of forecasts, from which the PDF distribution of future weather variables can be derived. When the initial condition ensemble accurately captures the distribution of initial analysis errors and the model has adequate precision, the ensemble forecast can reliably approximate the atmospheric state's PDF.

The core challenge in ensemble forecasting lies in appropriately representing error sources and accurately estimating uncertainties in NWP. The well-documented sources of NWP errors originate from both initial conditions and the model's inherent limitations. Correspondingly, ensemble forecasting methods are derived from the first (initial condition) and second (model) predictability theories. Various perturbation methods are designed to represent initial condition errors and model uncertainties, leading to the development of initial condition and model perturbation methods (Chen et al., 2002; Du et al., 2018). In terms of initial condition perturbation, methods such as the singular vector (SV; Buizza, 1997), the breeding of growing mode (BGM; Toth and Kalnay, 1993), the perturbed observations (Houtekamer and Derome, 1995), the ensemble transform Kalman filter (ETKF; Wang and Bishop, 2003; Wei et al., 2006), the ensemble transform with rescaling (ETR; Wei et al., 2008), the ensemble square root filter (EnSRF; Whitaker and Hamill, 2002; Zhou et al., 2022), and hybrid perturbations combining SV perturbations with ensemble data assimilation (EDA; Isaksen et al., 2010) have been developed. In terms of model perturbation, proposed methods include the multi-physics scheme (Chen et al., 2003), the stochastically perturbed parameterization tendencies (SPPT; Buizza et al., 1999; Tan et al., 2013), the stochastic kinetic energy backscatter scheme (SKEB; Shutts, 2005; Peng et al., 2019), and the stochastically perturbed parameterization (SPP; Jankov et al., 2019; Xu et al., 2019). These studies have significantly contributed to the continuous improvement of operational EPSSs in terms of overall performance and reliability.

Since the 1990s, significant progress has been made in operational ensemble forecasting, driven by advancements in ensemble forecasting theory and technology. In 1992, the ECMWF and the NCEP established the first global medium-range EPS, marking a major milestone in

operational ensemble forecasting (Toth and Kalnay, 1993; Molteni et al., 1996; Buizza, 1997). Following remarkable improvements in model technology and computer power, ECMWF took a pioneering step in June 2023 by upgrading its operational NWP system. The traditional EPS combined a high-resolution deterministic model with a lower-resolution ensemble forecast model. The new system, however, employs high-resolution models for both deterministic and ensemble forecasts, offering a more accurate representation of forecast uncertainty at higher resolutions. In 2006, the World Meteorological Organization (WMO) launched the Observing System Research and Predictability Experiment (THORPEX) Interactive Grand Global Ensemble (TIGGE) as an important component of the THORPEX initiative (Jiao, 2010). In recent years, ensemble forecasting has expanded to fields such as data assimilation, physical process parameterization, flood forecasting, severe weather prediction, and typhoon track forecasting. The ensemble forecasting approach has become an integral part of the entire operational NWP process (Buizza et al., 2018).

Chinese scientists have also been actively engaged in ensemble forecasting research and operations since the early stages (Li et al., 1997; Chen et al., 2002; Li and Chen, 2002). In terms of initial condition perturbation, methods such as conditional nonlinear optimal perturbation (CNOP; Mu et al., 2003; Duan and Huo, 2016), the heterogeneous physical mode method (Chen et al., 2005), nonlinear local Lyapunov vectors (NLLV; Feng et al., 2014), multi-scale hybrid initial perturbation (Zhang et al., 2015), and multi-scale SV (Ye et al., 2020) have been proposed. Regarding model perturbation, studies have focused on methods such as the multi-physics scheme (Chen et al., 2003), SPPT (Tan et al., 2013), nonlinear forcing SV (NFSV; Duan et al., 2013), SPP (Xu et al., 2019), and a model tendency perturbation method combining systematic and random errors (Han et al., 2023). China's operational EPS has evolved from initially relying on imported spectral models to the self-developed GRAPES-EPS (Chen and Li, 2020). The initial perturbation methods have advanced from the ETKF (Ma X. L. et al., 2008; Long et al., 2011; Zhang et al., 2017) to the multi-scale SV initial perturbation method (Liu et al., 2011; Ye et al., 2020). Furthermore, a combination of SPPT and SKEB has been implemented to generate model perturbations. These advancements collectively represent a milestone in the development of China's EPSSs (Chen and Li, 2020; Shen et al., 2020).

This paper first reviews the ensemble weather forecasting methods and the primary techniques employed in

the main EPSs designed by China and other countries, emphasizing the development of China Meteorological Administration (CMA) global EPS (CMA-GEPS) and regional EPS (CMA-REPS), including operational technologies, initial and model perturbation methods, and the applications of ensemble forecast. Besides, a comparison of the EPS settings and verification results between the CMA-GEPS and other major international NWP centers is presented. Furthermore, the paper addresses some of the key challenges in ensemble forecast technologies from the aspects of operation, science, integration with artificial intelligence (AI), merging of weather and climate models, and challenging user requirements. The summaries and future perspectives of ensemble forecast are concluded at the end.

2. Review of ensemble forecast methods

2.1 Sources of initial errors and advances in initial perturbation methods

Initial errors play a critical role in numerical simulations, primarily arising from inaccuracies in observation, as well as errors introduced during the processes of data assimilation. The initial condition perturbation (ICP) method, which aims to describe errors in model initial conditions, has been developed within global medium-range EPSs. The development of ICP methods has gone through several stages, from the initial human-generated random perturbations to statistically-based perturbation techniques, such as the Monte Carlo, BGM, SV, ETKF, and CNOP. These methods have been widely applied in the construction of various EPSs, providing effective information about forecasting uncertainty and enhancing prediction skill.

2.1.1 Initial perturbation for estimating assimilation analysis errors

The primary function of these methods is to estimate the error probability distribution in assimilated analyses. For instance, the Monte Carlo random perturbation method generates initial conditions by superimposing random noise or perturbations on control fields (Hollingsworth, 1979), while the time-lagged average method utilizes analyses from different past times as perturbed initial conditions (Hoffman and Kalnay, 1983). As an example, the initial version of the GEPS developed by the Canadian Meteorological Centre (CMC) employed Monte Carlo-based random perturbations of observations to create ensemble initial conditions (Houtekamer et al., 1996). With the development of data assimilation systems and EPSs, ensemble Kalman filter (EnKF) initial perturba-

tion method, which randomly perturbed observation data, is used by Canada's EPS (Houtekamer and Mitchell, 2005) and the NCEP (Zhou et al., 2016). While these perturbation methods demonstrate advantages in effectively representing uncertainties in initial analysis errors and enabling generation of substantial ensemble members, they exhibit limitations including constrained growth of initial perturbation energy, insufficient ensemble spread, and inadequate characterization of forecast uncertainties. Recent innovations by Pan et al. (2021) and Wang et al. (2023) introduced an analysis-constrained method that effectively incorporates characteristics of assimilation analysis increments. This method improved initial perturbation quality and the consistency between ensemble spread and forecast error.

2.1.2 Estimation of initial perturbations with the fastest error growth

These types of initial perturbations are based on research findings in atmospheric predictability. By analyzing the direction and speed of numerical prediction error growth in phase space, perturbations are introduced along the most unstable directions in initial conditions. Representative methods include the SV developed based on nonlinear dynamical finite-time instability theory (Molteni et al., 1996; Yang et al., 2002; Liu et al., 2011) and the BGM employed by NCEP, in which random perturbations are cycled through a nonlinear model and scaled to generate the final fields (Toth and Kalnay, 1993). Both the two methods aim to capture initial errors in atmospheric baroclinic instability regions, representing the forecast uncertainty of large-scale unstable waves. The SV, often developed in conjunction with a 4D variational system, has clear physical significance and can achieve favorable results. It has been successfully applied in medium-range weather ensemble forecasting by ECMWF, Japan, China, and others. Moreover, Zhang et al. (2020) proposed an initial perturbation method combining ensemble sensitivity analysis (ESA) and BGM to represent the characteristics of initial analysis errors. The distribution of perturbations, which is derived from sensitivity modes and rapidly growing perturbations calculated by BGM, adapts to changes in weather conditions and provides accurate simulations of the location and intensity of convective systems. The BGM method is favored by operational forecasting centers due to its effectiveness, simplicity, and low computational cost. Additionally, the NCEP developed the ETR (Wei et al., 2008), a method similar to the BGM but with perturbation structures that are nearly orthogonal and may reduce the correlation among ensemble members.

The SV method has achieved significant success in

operational ensemble forecasting. However, since it relies on a linear approximation of the nonlinear model, it fails to fully capture the impact of nonlinear physical processes, which limits further improvements in ensemble forecast skill. Mu et al. (2003) proposed the CNOP method, which comprehensively considers the effects of nonlinear processes and identifies the fastest-growing initial perturbations in nonlinear models, thereby addressing the limitations of the SV. Duan and Huo (2016) extended the CNOP to different perturbation phase spaces, developing an orthogonal scheme for ensemble forecast initial perturbations. This method has been successfully applied in the study of typhoon track ensemble forecasting, demonstrating that the initial perturbations generated by orthogonal CNOP significantly improve the typhoon track forecasting skill, especially for abnormal tracks, compared to those generated by the SV and the BGM methods (Zhang H. et al., 2023).

To address the rapid growth of errors caused by convective instability in meso- and small-scale systems, Chen et al. (2005) proposed a new method, known as the Heterogeneous Physical Mode Method, for constructing initial perturbations in ensemble forecasting that targets convective instability and exhibits mesoscale motion characteristics. By identifying forecast deviations from different cumulus convective parameterization schemes, this method identifies regions sensitive to convection, extracts perturbation variables, structures, and magnitudes, and constructs perturbed initial conditions. Unlike traditional initial perturbation methods such as the BGM or SV, this approach perturbs the initial values in convectively unstable regions, thereby promoting rapid growth of the unstable perturbations related to convection.

2.1.3 Estimating initial perturbations for multi-scale system errors

Despite the continuous improvement in numerical weather prediction resolution, the rapid growth of multi-scale initial errors continues to significantly impact forecast skill. Limited-area models face dual challenges, i.e., insufficient large-scale initial error information and inadequate characterization of small-scale errors from dynamical downscaling. In response to these issues, Chinese scientists developed the hybrid initial condition perturbation method for enhanced ensemble perturbation generation (Wang Y. et al., 2014; Zhang et al., 2015; Zhuang et al., 2017; Ma et al., 2018). This method uses filtering and spectral analysis techniques to extract small-scale perturbations from regional ensemble forecasts and large-scale perturbations from global ensemble forecasts, and combines the two to generate the growth ICPs. The hybrid ICPs effectively represent both large- and small-

scale uncertainties in the analysis, aligning more closely with the lateral boundary perturbations provided by global EPSs, thereby significantly improving the forecast skill of regional EPSs. Furthermore, the CMA (Ye et al., 2020; Liu et al., 2024) and the Japan Meteorological Agency (JMA; Ono et al., 2021) have conducted research on multiscale SV ICP methods based on regional and global ensemble prediction models. These studies have generated ICPs containing multiscale initial uncertainty, which more effectively reflect the multiscale variations in initial errors and improve both regional and global ensemble prediction performance.

In general, ensemble forecasts based on ICPs formed by analysis errors tend to show slightly higher forecast accuracy but a lower spread, while ensemble forecasts based on initial perturbations derived from dynamical instability growth theory generally exhibit an inverse pattern, i.e., a slightly higher spread but marginally lower accuracy (Pauluis and Schumacher, 2013).

2.2 Model error sources and development of model perturbation methods

Early ensemble forecasting studies primarily focused on initial perturbations. However, as research advanced, it became evident that addressing initial condition uncertainty alone could lead to drawbacks such as insufficient ensemble spread, which further results in systematically biased forecasts (Palmer et al., 2009). Therefore, it is necessary to consider the uncertainty caused by model errors or defects in ensemble forecasting. The model errors primarily originate from inadequate representations of both physical and dynamical processes in the numerical model (Mu et al., 2011). Currently, the predominant model errors considered by ensemble forecasting systems stem from imperfect representations of subgrid-scale parameterization processes.

2.2.1 Estimation of model uncertainty

The uncertainty of the physical process of the model is described by a combination of different physical parameterization schemes within the framework of a single model (Houtekamer et al., 1996), which is called a multi-physics scheme. This scheme was applied to the global EPS of Environment Canada in the early days. It has also been commonly used in regional ensemble forecasting in recent years (Stensrud et al., 2000; Chen et al., 2003; Zhi et al., 2013) to improve the probabilistic forecasting skill. Another model perturbation method is the multi-model method (Krishnamurti et al., 1999), which can represent the systematic bias in the forecasts of different models, so that the ensemble members can produce a larger ensemble spread, which greatly improves the accuracy of

medium-term forecasts and small- and medium-scale ensemble forecasts in limited areas (Zhang et al., 2017). However, since the ensemble members in multi-model or multi-physics schemes employ different models or physical processes, their physical consistency is compromised. This makes it challenging to satisfy the equal-likelihood requirement for members and increases the complexity of determining member weights for probability calculations (Berner et al., 2015). Consequently, the model perturbation method has been increasingly developed in recent years to estimate the uncertainty of the parameterization scheme of the model physical process.

2.2.2 Estimation of physical parameterization scheme uncertainty

Stochastic physical perturbation is now a standard technique for assessing physical parameterization uncertainty in modeling. The theoretical basis is to introduce a random process or factor to perturb some parameter values or related terms of a model (such as tendency and diffusion) to characterize the uncertainty of the model. This type of perturbation scheme mainly includes methods such as the stochastically perturbed physics parameterization tendency term (SPPT; Buizza et al., 1999; Yuan et al., 2016), stochastic kinetic energy backscattering (SKEB; Shutts, 2005), random parameters (RP; Bowler et al., 2008), and stochastically perturbed parameterization (SPP; Chen et al., 2003; Tan et al., 2013; Jankov et al., 2017).

The SPPT scheme employs spatiotemporally continuous random numbers conforming to a uniform distribution to multiplicatively perturb parameterized subgrid physical tendencies. By applying distinct perturbation values to different tendency terms, this method effectively enhances ensemble spread and improves probabilistic forecast reliability in ensemble forecasting systems (Li et al., 2008). The SPPT is currently the most widely adopted method among global NWP centers (Charron et al., 2010; Sanchez et al., 2016; Peng et al., 2020), yet it also has some drawbacks. For instance, perturbation amplitude may cause discontinuity and energy non-conservation in model top and near-surface fluxes, and the long-term integration will produce systematic bias (Leutbecher et al., 2017). Yuan et al. (2016) introduces SPPT technology to the CMA-REPS, which significantly improves the precipitation forecast of heavy rainfall in the late medium-range forecast. Qiao et al. (2017) developed a novel parameterization scheme for stochastic perturbation dissipation based on the uncertainty of the model dissipation term. It follows the same procedure as the SPPT method, but uses a recursive filter to generate smooth perturbations. It also uses horizontal and vertical

localization to maintain the influence of perturbation in regions with strong wind shear, which effectively improves the systematically weak vortex intensity. Similar results have been demonstrated in perturbation studies of microphysical processes (Qiao et al., 2018).

The SKEB scheme targets excessive momentum dissipation errors in the numerical model's dynamic framework at truncated scales, compensating for dissipated energy effects on resolvable scales through stochastic perturbations. It has been applied in global medium-range ensemble forecasting (Berner et al., 2009; Zhou et al., 2017) and is being progressively implemented in regional convective ensemble forecasting (Cai et al., 2017; Yang et al., 2024). Peng et al. (2020) found that the combined application of the SKEB and the SPPT in the GRAPES global ensemble forecast can effectively improve the ensemble spread in the tropics. The SKEB scheme can excite possible instabilities in atmospheric motion (Berner et al., 2015), thereby producing larger ensemble spread and capturing low-probability events. However, due to the higher computational cost of the SKEB compared to the SPPT and the unsatisfactory balance between spread and root mean square error (RMSE), the usefulness of the SKEB scheme has gradually decreased. It is reported that the ECMWF has discontinued the operational implementation of this scheme (Chen C. H. et al., 2021).

The RP/SPP method characterizes model uncertainty by performing random perturbations of subjective empirical and semi-empirical parameters within physical parameterization schemes (Bowler et al., 2008), which represents the predicted uncertainty of small-scale system changes. Although this method improves the ensemble forecasting, it has certain irrationalities. For instance, the parameter values will change suddenly at a certain time (Jankov et al., 2019). In addition, it is found that the SPP method has a more significant effect on the parameters in cumulus convection and planetary boundary layer schemes (Xu et al., 2019). However, the RP/SPP method yields inadequate ensemble spread for medium- and extended-range forecasts (Leutbecher et al., 2017).

Beyond these three prevalent stochastic physical perturbation schemes, researchers have developed various approaches to address uncertainty sources in other model components, such as the stochastic trigger of convection (STC; Li et al., 2015), stochastic boundary-layer humidity (SHUM; Tompkins and Berner, 2008), vorticity confinement (VC; Steinhoff and Underhill, 1994), meso- and medium-scale terrain perturbation schemes (Li et al., 2017), and model tendency perturbation methods combining model systematic bias and random errors (Han et

al., 2023). Zhao and Torn (2022) applied an independent stochastic perturbation parameterization scheme to a single parametric perturbation, and the results showed that turbulent mixed stochastic perturbation can increase the ensemble standard deviation of typhoon intensity, while the impact of stochastic perturbation on microphysics, radiation, and cumulus tendencies is negligible. The SHUM scheme improves the consistency between ensemble spread and forecast error in the tropical regions, and reduces the ensemble mean forecast error. The SHUM is applied to the ensemble data assimilation module of the NCEP global forecast system EnKF/3DVar. The VC method is applied to the NCEP GEPS, which increases the subtropical ensemble spread. The model tendency perturbation method combining model systematic bias and random error reduces the systematic bias of the model, which is applied to the CMA-GEPS.

2.2.3 Estimation of nonlinear characteristics of model error growth

Accurately characterizing the nonlinear nature of model error growth remains a critical challenge for cutting-edge high-resolution ensemble forecasting systems. Wang et al. (2020) implemented the CNOP perturbation (CNOP-P) method in the GRAPES model for convection-scale ensemble forecasting. The experiments demonstrated enhanced ensemble spread in tropospheric humidity and temperature predictions, along with improved forecast reliability for near-surface variables and precipitation. Xu et al. (2022a) applied a nonlinear forcing SV method (NFSV), also known as CNOP forcing (CNOP-F). Duan and Zhou (2013) described the combined effects of model errors from different sources by superimposing NFSV perturbations with specific structures on top of SPPT perturbations, which better characterizes the impact of model uncertainty on convection-scale ensemble forecasting (Xu et al., 2022b). In addition, Zhang Y. C. et al. (2023) extended the NFSV to the orthogonal subspace of inclined perturbations and developed an orthogonal NFSV-based ensemble prediction model perturbation method. When applied to typhoon intensity ensemble prediction, this method demonstrated significantly higher forecasting skill than the SPPT and the SKEB, particularly in providing earlier warning information for predicting typhoon rapid intensification processes.

In summary, the multi-model and multi-physics schemes have been gradually eliminated due to the average error of members, the inequality of meteorological significance, and the complexity of probability calculation, while the stochastic physics perturbation method has received more and more attention. However, whether the representativeness and flow dependence of reasonable

stochastic functions and their spatiotemporal correlation scales of random noises truly reflect the uncertainty of physical processes still need to be studied.

2.3 Post-processing and verification assessment of ensemble forecast uncertainty information

The EPS generates a vast array of data products. In contrast to a single deterministic forecast, ensemble forecasting can provide a probability distribution of possible future atmospheric states and characteristics of forecast uncertainty, demonstrating superior advantages in the early warning of extreme weather events (Du and Chen, 2010; Gao et al., 2019). The critical task in the post-processing of ensemble forecasts is to design appropriate post-processing techniques that provide meteorologically meaningful interpretations of the distribution of ensemble forecast products, as well as to extract similarities, discrepancies, and extreme information among the ensemble forecast members (Williams et al., 2014). The primary techniques currently employed in ensemble forecast post-processing include obtaining probabilistic forecast information to describe forecast uncertainty, correcting systematic model errors, and extracting extreme weather forecast information from the tails of the ensemble forecast probability distribution.

2.3.1 Probability forecast information based on ensemble prediction

The ensemble mean and ensemble spread are two fundamental products of ensemble forecasting. The ensemble mean represents the primary level of information in ensemble forecast products. It filters out unpredictable elements from the ensemble members, providing an overall forecast trend. However, due to its inherent smoothing effect, the ensemble mean might not completely retain extreme values in weather events. The ensemble spread is a measure of the uncertainty in ensemble forecasting or the amplitude of variation of ensemble members relative to the ensemble mean. It can be measured by the standard deviation of the ensemble members relative to the ensemble mean forecast, or by the average anomaly correlation coefficient (ACC) of each member relative to the overall mean field. The spread, to some extent, can represent the skill of ensemble forecasting. Generally speaking, a smaller spread indicates higher forecast confidence with potentially higher forecast skill; however, a larger spread indicates lower forecast confidence but does not necessarily imply lower forecast skill.

Probabilistic forecast of weather elements is a crucial ensemble forecast product. By calculating the forecast probabilities for different thresholds of weather elements such as precipitation, temperature, and wind, it repres-

ents the occurrence likelihood of specific weather conditions. For instance, probability of precipitation exceeding 1, 5, 10, and 20 mm day⁻¹, or that of a specific weather element (e.g., snow) to occur at certain stations can be provided. Another probabilistic forecast product is the spaghetti plot, which involves selecting a specific characteristic contour line and plotting forecasted contours from all ensemble members on the same chart. Generally, the degree of divergence among contour lines roughly indicates forecast confidence, and more concentrated contours correspond to higher confidence.

Ensemble forecast clustering products utilize clustering analysis to classify ensemble member forecasts into distinct clusters, offering various possible outputs. Common clustering algorithms include hierarchical, non-hierarchical, and tubing clustering approaches. Through clustering classification, the number of members in each cluster can be compared, allowing for determination of the most probable weather scenarios. Additionally, two commonly used visualization tools, i.e., plume chart and spaghetti chart, apply clustering analysis principles to classify 500-hPa geopotential height patterns, providing forecasters with operationally actionable guidance.

2.3.2 Post-processing methods for systematic error correction in ensemble forecasting

Common post-processing methods for systematic bias in ensemble forecasting can be broadly categorized into parametric and non-parametric approaches (Mylne et al., 2022). Parametric post-processing methods generally assume that the target variable follows a specific probability distribution and conduct regression on its parameters. Examples include ensemble model output statistics (EMOS), Bayesian model averaging (BMA), and Bayesian joint probability (BJP). Additionally, correction methods based on different probability distributions have been developed for various physical quantities. For instance, Gaussian distribution is used for temperature and pressure, truncated normal distribution and lognormal distribution for wind, and Gamma (and truncated Gamma) distribution for precipitation. Furthermore, ensemble forecast correction based on precipitation categories has been proposed and applied to improve the forecast skill for different levels of precipitation, particularly extreme precipitation (Ji et al., 2023).

Non-parametric post-processing methods eliminate the need for predefined probability distributions of target variables, thereby providing enhanced flexibility. Common non-parametric methods include quantile mapping (QM), frequency matching (FM), probability matching (PM), member by member (MBM), the optimal percentile method, and isotonic distributional regression (IDR).

Meanwhile, by combining effective forecast information obtained from multiple models for multi-model ensemble forecasting, non-equal-weight ensemble methods based on error analysis, such as super ensemble and Kalman filtering, have been developed (Zhi et al., 2013; Zhu et al., 2021). The concept of a sliding training period (Zhi et al., 2012) has been widely applied in the correction of forecasts for tropical cyclones, temperature, precipitation, wind, and other variables. To reduce the positional and structural biases in the target system forecasts, various post-processing methods based on spatial features have been proposed, such as neighborhood methods, model projection methods, and object-based correction techniques (Ji et al., 2020).

In recent years, machine learning techniques have been widely adopted due to their high nonlinearity and strong robustness. Deep learning algorithms, such as random forests, support vector machines, convolutional neural networks, long short-term memory neural networks, and U-Net neural networks, have also been extensively applied to the post-processing of EPSs, achieving remarkable results (Zhi et al., 2020; Yang et al., 2022; Lyu et al., 2023). At the same time, by integrating machine learning algorithms with traditional statistical methods, hybrid machine learning and statistical approaches have been designed to further enhance the reliability of ensemble forecasting. Detailed descriptions of the aforementioned systematic error post-processing methods can be found in the ensemble forecast guidance manual published by the WMO (Mylne et al., 2022).

2.3.3 Extreme forecast information extraction based on ensembles

To extract early warning information for extreme weather from ensemble forecasts, the Extreme Forecast Index (EFI) developed through ECMWF ensemble forecast results was established (Lalaurette, 2003). This index characterizes the continuous difference between the cumulative probability distribution of the ensemble forecast results and the model climate cumulative probability distribution. The larger the value, the greater the deviation of the forecast from the model climate. Such deviations indicate a higher probability of extreme forecasts, and consequently, an increased probability of extreme weather occurring in reality. The NCEP has also conducted studies on extreme weather forecasting based on EFI. Guan and Zhu (2017) pointed out that the EFI index based on the NCEP global ensemble forecasts has certain medium-range forecasting ability for extreme precipitation and extreme low-temperature weather in the U.S. during the winter of 2013/14. In addition to conventional weather elements such as temperature, precipitation, and

10-m wind, the EFI index can also be applied to diagnostic quantities such as convective parameters (e.g., convective available potential energy) and total column water vapor flux (Tsonevsky et al., 2018). The application range of the EFI index has gradually expanded from large-scale medium-range ensemble forecasts to convective-scale and sub-seasonal to seasonal ensemble forecasts (Dutra et al., 2013; Raynaud et al., 2018). Many domestic scholars have also studied the EFI index (Xia and Chen, 2012; Liu et al., 2018; Peng et al., 2024). Research on extreme weather forecasting methods based on the EFI using the CMA-GEPS has shown that the EFI has certain recognition ability for extreme low temperatures and extreme heavy precipitation in China, with good medium-range forecasting skill. Another similar method is the Ensemble Anomaly Forecasting by ranking ensemble forecasts relative to observed climatology rather than model climatology (Du et al., 2014).

2.3.4 Verification and evaluation metrics

The accuracy and reliability of ensemble forecasting are crucial in meteorological prediction, driving the continuous development and refinement of verification and evaluation methods for ensemble forecasts. The works of Wilks (2011) and Jolliffe and Stephenson (2012) comprehensively summarize the verification methods for meteorological forecasts, covering deterministic, probabilistic, and qualitative forecasts, along with their applications. For deterministic forecasts based on ensemble means or individual ensemble members, commonly used verification metrics include the RMSE and spatial correlation coefficients. In the realm of probabilistic forecast verification, primary evaluation methods encompass the Brier skill score (BSS), continuous ranked probability skill score (CRPSS; Hersbach, 2000), reliability diagrams, relative operating characteristics (ROC) curve, and potential economic value (Zhu et al., 2002).

To quantify uncertainty, ensemble spread and the Talagrand distribution are key metrics. The spread–error relationship, commonly referred to as the spread–skill relationship, is often employed to evaluate the predictability of EPSs. However, real-world systems often exhibit under-dispersion in ensemble spread, particularly in quantitative precipitation forecasting (Li, 2001; Li et al., 2009; Su et al., 2014). Uncertainty quantification typically relies on the bootstrapping method (Hamill, 1999), with error bars representing the confidence intervals of the scores. Additionally, Torn and Hakim (2008) introduced the ESA method to assess the sensitivity of forecast outcomes to minor changes in initial conditions or other input parameters.

Su et al. (2014) proposed an area-weighted verifica-

tion method, applying spatial weighting to probabilistic precipitation forecast scores. In the field of precipitation forecasting, some researchers have integrated these spatial verification methods into ensemble forecast evaluation. For instance, Ji et al. (2020) explored the integration of the Method for Object-based Diagnostic Evaluation (MODE) with ensemble forecasting techniques, finding significant advantages in evaluating and improving the spatial structural characteristics of precipitation forecasts. Chen et al. (2018) investigated the spatial relationship between ensemble spread and forecast error for Meiyu precipitation forecasts in the Yangtze–Huaihe region using the precipitation consistency scale method, revealing the impact of positional errors on forecast skill under different precipitation thresholds. Furthermore, Du and Deng (2020) introduced two new forecast evaluation metrics, i.e., the measure of forecast challenge (MFC) and the predictability horizon diagram index (PHDX). The MFC, which may comprehensively consider forecast error and uncertainty, serves as a novel metric for assessing forecast difficulty. The PHDX further investigates how the temporal evolution of forecasts influences decision-making processes, aiming to better represent the dynamic nature of ensemble forecast information.

It is important to note that the accuracy of scores in ensemble forecast verification may be influenced by reference values. For example, the calculation of BSS by using high-frequency grid-based climatological samples could potentially lead to score underestimation (Hamill and Juras, 2006). Additionally, the uncertainty and quality of verification data can significantly impact evaluation outcomes, particularly in precipitation forecast verification (Yuan et al., 2005). With the ongoing advancement of high-resolution ensemble forecasting and observational data, ensemble forecast verification methods are expected to become more refined and scientifically robust in the future.

3. Chinese operational EPS and its applications

3.1 Development history of GEPS and REPS

3.1.1 GEPS of spectral model

In the 1990s, the National Meteorological Centre (NMC) of CMA developed the GEPS (Li et al., 1997). In 1996, a T63L16 EPS had been implemented using the time-lagged method, consisting of 12 ensemble members with a 10-day forecast lead time. In 1999, a T106L19 EPS was developed by the SV method (Li and Chen, 2002), which includes 32 ensemble members with a 10-day forecast lead time. In 2007, based on the BGM

method (Tian et al., 2007), a T213 EPS was established, which includes 15 ensemble members with a 10-day forecast lead time. The year 2014 saw the upgrade of CMA's operational EPS from T213 to T639, which for the first time incorporated SPPT to generate model perturbations (Tan et al., 2013).

3.1.2 REPS

Exploratory efforts toward REPS commenced in the early 21st century. In 2005, based on the mesoscale model MM5, a heterogeneous physical mode method was proposed to generate initial perturbation (Chen et al., 2005), alongside experimental studies incorporating multi-physics configurations and stochastic physical parameter perturbations (Chen et al., 2003; Feng et al., 2006; Wang et al., 2007). By 2006, the NMC developed a regional mesoscale EPS over China using the Weather Research and Forecasting (WRF) model, integrating the BGM method and multi-physics ensemble method (Deng et al., 2010). This system became operational EPS on 15 November 2010, operating at a horizontal resolution of 15 km with 15 ensemble members and running 4 daily forecast cycles (Ma Q. et al., 2008).

3.1.3 GRAPES-based GEPS and REPS

The GRAPES model is a new-generation numerical weather prediction system developed through self-innovation by the CMA (Chen et al., 2008). Since 2005, research and development efforts have been dedicated to advancing ensemble forecast techniques within the GRAPES framework. The Chinese Academy of Meteorological Sciences has successively implemented both the BGM and ETKF methods to investigate GRAPES regional ensemble prediction techniques and systems (Tan and Chen, 2007; Tian and Zhuang, 2008; Ji et al., 2011), ultimately developing its REPS (GRAPES-REPS). The model's version 1.0 became operational in 2014 at NMC, utilizing ETKF for initial condition perturbations and superseding the WRF REPS. This system featured a 15-km horizontal resolution with 15 ensemble members. In 2015, the initial condition perturbation scheme was upgraded to a multi-scale blending (MSB) method (Zhang et al., 2015) that combined large-scale initial perturbations derived from the T639 EPS by using the BGM method with meso-scale initial perturbations obtained through the ETKF method within the GRAPES-REPS. This upgrade culminated in the release of version 2.0. In September 2019, the lateral boundary perturbations for the new GRAPES-REPS were sourced from the GRAPES-GEPS. The parallel enhancement of initial condition perturbation and model perturbation methods led to the operational system's evolution to version 3.0 (Chen and Li, 2020).

Development of the GRAPES-GEPS began in 2008, with initial research focused on the global ETKF method (Ma X. L. et al., 2008). With the advancement of the GRAPES global 4DVar data assimilation technology, the GRAPES SV ICP technique was developed (Li and Liu, 2019). In December 2018, the GRAPES global medium-range EPS, based on the SV ICP, became operational and replaced the T639 EPS. This achievement represents China's first successful implementation of both global and regional operational EPSs utilizing fully independent and self-controlled technologies (Chen and Li, 2020).

To characterize uncertainties in the GRAPES model, researchers have investigated multiple model perturbation methods through GRAPES's REPS and GEPS, building upon prior studies of the NMC's T213 and T639 global medium-range ensemble models. The methods used include the multi-physics (MP; Feng et al., 2006), SPPT (Yuan et al., 2016), SKEB (Peng et al., 2019), and SPP (Xu et al., 2019). The SPPT scheme became operational in the GRAPES-REPS in 2015, followed by implementation of both the SPPT and the SKEB in the GEPS in 2018 (Chen and Li, 2020).

Since 2018, significant methodological improvements have been achieved in the tropical cyclone SV technique, blended SV-EDA perturbation scheme, and sea surface temperature perturbation scheme (Huo et al., 2020; Qi et al., 2022). Based on the GRAPES-REPS forecast model, extensive advancements in initial condition and model perturbations include stochastic parameter perturbation (Xu et al., 2019), conditional typhoon vortex relocation (Wu et al., 2020), radar reflectivity algorithm (Chen Y. X. et al., 2021), and model tendency perturbation combined with systematic bias and random errors (Han et al., 2023).

Since 2020, active research on convective-scale ensemble forecasting has been conducted using the GRAPES-Meso 3-km model. Researchers have performed various experiments involving both model and initial condition perturbations, including multi-scale initial condition perturbation experiments (Ma et al., 2023) and the construction of model perturbations through conditional non-linear optimal perturbations combined with stochastic physical perturbations (Xu et al., 2022a, b). The established 3-km regional ensemble forecast experimental system has been applied in major events such as the Beijing Winter Olympics and the Hangzhou Asian Games. Under the WMO Research Demonstration Project for the Hangzhou Asian Games, a fully independent 3-km convective-scale ensemble prediction system was successfully developed and implemented in Hangzhou, providing ensemble forecast services for the broader East China

region. This system commenced operational service on 21 May 2023, delivering comprehensive probabilistic forecasts and multifaceted decision-support capabilities to forecasters.

3.2 Technology of CMA EPSSs

3.2.1 Global/regional integrated multi-scale SV initial perturbation technique

The SV initial perturbations effectively capture both the primary error information and the sensitive features of analysis errors that grow and intensify within the basic flow. The development of the Tangent Linear Model (TLM) and Adjoint Model (ADM) within the CMA global 4DVar system has provided the necessary foundation for the research and advancement of SV perturbation techniques in the CMA-GEPS (Liu et al., 2017). The SV initial perturbation calculation scheme for the CMA-GEPS is composed of two parts, i.e., calculations of SVs for the extratropical regions and for tropical cyclones, respectively. The specific mathematical treatment scheme is outlined as follows.

3.2.1.1 SV calculation scheme for extratropical regions

In the CMA-GEPS, the SV calculation can be mathematically formulated as maximizing the ratio of the evolved perturbation vector norm to the initial perturbation vector norm, as shown in Eq. (1):

$$(\mathbf{E}^{-1/2} \mathbf{L}^T \mathbf{P}^T \mathbf{E} \mathbf{P} \mathbf{L} \mathbf{E}^{-1/2}) \mathbf{x}(t_0) = \lambda^2 \mathbf{x}(t_0), \quad (1)$$

where \mathbf{L} represents TLM in the CMA global 4DVar assimilation system, \mathbf{L}^T indicates the corresponding adjoint model, \mathbf{E} is the weight matrix that measures the perturbation magnitude, \mathbf{x} denotes the SV, λ is the corresponding singular value, and \mathbf{P} stands for a projection operator that sets the SV perturbations outside the target region to zero. From Eq. (1), it is evident that the SV structure is influenced by three key factors.

(1) Definition of the weight matrix \mathbf{E} . This matrix determines the relative importance of different components of the perturbation vector.

(2) Characteristics of the TLM (\mathbf{L}) and ADM (\mathbf{L}^T). These models primarily involve the use of linearized physical process parameterization schemes.

(3) Integration of time length for the TLM (\mathbf{L}) forward and the ADM (\mathbf{L}^T) backward, referred to as the optimal time interval.

The CMA global SVs employ the dry total energy norm to define the weight matrix \mathbf{E} (Liu et al., 2013). Let the prognostic variables of the CMA global TLM and ADM include the horizontal wind components (i.e., u and v), perturbation potential temperature (θ'), and perturbation dimensionless pressure (Π'). The correspond-

ing perturbation quantities can be expressed as u' , v' , $(\theta')'$, and $(\Pi')'$. The total energy norm \mathbf{E} is calculated as follows:

$$\mathbf{E} = \iiint_V \left\{ \frac{\rho_r \cos \phi}{2} (u')^2 + \frac{\rho_r \cos \phi}{2} (v')^2 + \frac{\rho_r \cos \phi C_p T_r}{(\theta_r)^2} [(\theta')']^2 + \frac{\rho_r \cos \phi C_p T_r}{(\Pi_r)^2} [(\Pi')']^2 \right\} dV. \quad (2)$$

In Eq. (2), the sums of the first two and last two terms represent the perturbation kinetic energy (KE) and the perturbation potential energy (PE) norms, respectively. The third and fourth terms correspond to the contributions of perturbation potential temperature and perturbation dimensionless pressure to the PE norm. Meanwhile, $dV = d\lambda d\varphi d\hat{z}$, among which \hat{z} denotes the terrain-following coordinate, and λ and φ represent longitude and latitude in the spherical coordinate system of the model. In addition, C_p is the specific heat capacity of dry air at constant pressure; and T_r , θ_r , Π_r , and ρ_r denote the reference temperature, reference potential temperature, reference dimensionless pressure, and reference density, respectively.

Based on the aforementioned CMA global SV calculation technique, three types of SVs with different scales are defined by varying resolutions and optimal time intervals. Large-scale SV (LSV) has a horizontal resolution of 2.5° and an optimal time interval of 48 h. It is computed by using dry linearized physical processes and is primarily used to capture uncertainty information at the synoptic scale. Mesoscale SV (MSV), with a horizontal resolution of 1.5° and an optimal time interval of 24 h, also employs dry linearized physical processes and is designed to capture uncertainty at the meso- α scale. Small-scale SV (SSV), which has a horizontal resolution of 0.5° and an optimal time interval of 6 h, incorporates moist linearized physical processes in its computation compared to LSV and MSV, and is mainly used to capture uncertainty at the meso- β scale. The specific configurations of the CMA global/regional integrated multi-scale SVs are summarized in Table 1.

3.2.1.2 Tropical cyclone SV calculation scheme

Considering the unique characteristics of tropical cyclones, a specialized tropical cyclone SV calculation scheme has been developed for the CMA-GEPS (Huo et al., 2020). Specifically, for tropical cyclone cases, the target region for SV computation is defined as a $10^\circ \times 10^\circ$ (latitude \times longitude) domain centered on the cyclone position. A maximum of six tropical cyclone SVs can be calculated.

As a result, the target regions for SV computation in the CMA-GEPS consist of the extratropical regions of

Table 1. The settings of calculation for multi-scale SVs

	LSV	MSV	SSV
Target region	30°–80°N; 30°–80°S	30°–80°N; 30°–80°S	20°–50°N, 105°–125°E
TLM resolution/optimization time	2.5°/48 h	1.5°/24 h	0.5°/6 h
Energy norm	Dry energy	Dry energy	Dry energy
Linearized physical process	Subgrid topographic drag, vertical diffusion	Subgrid topographic drag, vertical diffusion	Subgrid topographic drag, vertical diffusion, large-scale condensation
Number of SVs	10	10	10

the Northern and Southern Hemispheres and the tropical cyclone region. This approach not only captures the development of baroclinically unstable perturbations in the mid-to-high latitudes of both hemispheres but also accounts for the evolution of tropical cyclones when they are present.

Based on the initial SV computed for the Northern and Southern Hemispheres as well as tropical cyclones, Gaussian sampling techniques are employed to construct the initial perturbation fields for the CMA-GEPS. Additionally, a scaling amplification factor σ is applied to normalize the amplitudes of all initial perturbations, ensuring that their magnitudes are consistent with the actual analysis error levels. Finally, by adding and subtracting the initial perturbations to/from the assimilation analysis initial conditions in pairs, the perturbed initial conditions for the CMA-GEPS are generated.

3.2.2 MSB initial perturbation techniques in short-range regional ensembles

Due to the resolution limitations of GEPSSs, the dynamical downscaling method fails to fully capture uncertainties at smaller scales that can be resolved by regional models. However, meso- and small-scale initial perturbations are essential for accurately capturing uncertainties in local severe weather. Therefore, the CMA-REPS has developed an MSB initial perturbation technique based on the ETKF scheme.

3.2.2.1 ETKF initial perturbation method

The ETKF method is an initial perturbation scheme developed based on Kalman filter theory. It rapidly estimates analysis errors from ensemble forecast perturbations and observational error variances (Long et al., 2011), thereby effectively capturing uncertainties in meso- and small-scale initial conditions.

In the CMA-REPS, the forecast perturbation vector X^f is transformed into the analysis perturbation vector X^a through the transformation matrix T , as shown in Eq. (3). Specifically, the analysis perturbations at the current time are derived by linearly combining the forecast perturbation vector using the transformation matrix T .

$$X^a = X^f T. \quad (3)$$

The core of the ETKF initial perturbation method lies

in obtaining the transformation matrix T . In the ETKF scheme of CMA-REPS, the transformation matrix T is defined using the method proposed by Wang and Bishop (2003). To centralize the perturbed members relative to the ensemble mean, we employ a spherical simplex centralization scheme. In the operational version of the CMA-REPS, simulated observations are derived by interpolating the analysis fields from the GEPS to the observational space (Wang et al., 2018). The variables used to compute the transformation matrix T include meridional wind v and zonal wind u , and the variables for calculating the amplification factor are u , v , and specific humidity q . The system applies a 6-h cycling perturbation scheme to provide forecast perturbation fields for the ETKF cycle. The model runs 4 times per day at 0000, 0600, 1200, and 1800 UTC, among which the 0000 and 1200 UTC cycles are integrated for 84-h forecasts, and the other two (i.e., 0600 and 1800 UTC cycles) for 6-h forecasts.

3.2.2.2 MSB initial perturbation method

Practical experience has shown that the initial perturbation fields in regional ensemble prediction, derived from the dynamical downscaling of global ensemble prediction, contain more large-scale perturbation information. Since 2016, the CMA-REPS has adopted an MSB initial perturbation technique (Zhang et al., 2015; Xia et al., 2019). This method employs the two-dimensional discrete cosine transform (2D-DCT; Denis et al., 2002) to filter the initial perturbations from both the ETKF method in regional ensemble prediction and the dynamical downscaling of the T639 global ensemble prediction. Subsequently, through the 2D-DCT inverse transformation, the initial perturbation fields are reconstructed by integrating large-scale perturbations from the global ensemble prediction with meso- and small-scale perturbations from the regional ensemble prediction, with more multiscale initial condition uncertainties.

3.2.2.3 Conditional typhoon vortex relocation technique in ensemble forecasting

To more reasonably describe the uncertainty in the positioning of typhoon vortex centers in ensemble forecasting, Wu et al. (2020) applied the best track data of typhoons from the CMA and the JMA during 2009 and 2018. They analyzed the uncertainties of typhoon vortex

center positioning in these best tracks and developed a conditional typhoon vortex relocation method. Additionally, they established procedures for threshold discrimination of typhoon vortex relocation for ensemble members, as well as mathematical treatment processes for typhoon vortex separation and vortex relocation.

By analyzing and comparing the best typhoon vortex center locations from the CMA and the JMA during the study period, it was found that the maximum annual average positioning error was 17.18 km (in 2009), while the minimum was 11.98 km (in 2015). Given the 10-km horizontal resolution of the CMA-REPS, a critical threshold of 15 km was established to determine whether conditional typhoon vortex relocation should be performed for ensemble members. Specifically, if the distance between the typhoon vortex center in an ensemble member's initial field and the best track exceeds 15 km, the typhoon vortex positioning in that member's initial condition is deemed beyond the range of analyzed uncertainties, requiring conditional vortex relocation. Otherwise, the error is considered within the range of analyzed uncertainties, making relocation unnecessary. The relocation method utilizes the typhoon vortex separation technique developed by the U.S. Geophysical Fluid Dynamics Laboratory to isolate the typhoon vortex from the initial fields of ensemble members requiring relocation. The separated typhoon is then shifted to the observed position through interpolation. This scheme was implemented in the CMA-REPS v3.0 in 2019.

3.2.3 Stochastic model perturbation technique for subgrid physical processes

The EPS relying solely on initial perturbations suffers from insufficient divergence among ensemble members and inadequate reliability. In a complete system, incorporating model perturbation techniques is essential to effectively characterize model uncertainty and account for model errors (Palmer et al., 2009). Therefore, CMA's GEPS and REPS should introduce such techniques to represent uncertainties arising from model imperfections. At present, the SPPT and the SKEB model perturbation techniques are mainly used in the CMA's forecast system, which are briefly introduced below.

3.2.3.1 SPPT scheme

The SPPT scheme in CMA's GEPS and REPS shares the same fundamental design principles as ECMWF's implementation (Buizza et al., 1999). Specifically, during numerical integration, the model integration term can be decomposed into the integral term of non-parametric (dynamic) processes, and the integral tendency term of parametric physical processes. The model integral tendency term perturbed by the SPPT scheme can then be

expressed as:

$$e_j(t_e) = \int_0^{t_e} [A(e_j, t) + \Psi(\lambda, \phi, t)P(e_j, t)] dt, \quad (4)$$

where $\Psi(\lambda, \phi, t)$ is a three-dimensional random field with spatiotemporal correlation features. A stretch function is introduced to implement a range of custom perturbations (Li et al., 2008), which is defined as:

$$\Psi(\lambda, \phi, t) = \mu + \left\{ 2 - \frac{1 - \exp\left[\beta\left(\frac{\psi - \mu}{\Psi_{\max} - \mu}\right)\right]^2}{1 - \exp(\beta)} \right\} (\psi - \mu), \quad (5)$$

where β is a constant with a value of -1.27 and $\mu = (\Psi_{\max} + \Psi_{\min})/2$; Ψ_{\max} and Ψ_{\min} represent the upper and lower bounds of the random field $\Psi(\lambda, \phi, t)$, respectively; and ψ is a three-dimensional random field with spatiotemporal correlated characteristics, defined as:

$$\psi(\lambda, \phi, t) = \mu + \sum_{l=1}^L \sum_{m=-l}^l \alpha_{l,m}(t) Y_{l,m}(\lambda, \phi), \quad (6)$$

where λ , ϕ , and t represent the longitude, latitude, and time, respectively; and $Y_{l,m}$ is the spherical harmonic function, L is the horizontal truncation scale of the random field, l and m are the total wavenumber and latitude wavenumber in the horizontal direction, and $\alpha_{l,m}(t)$ is the spectral coefficient of the random field. The correlation of the stochastic spectral coefficient $\alpha_{l,m}(t)$ in the time dimension is realized by the first-order Markov chain stochastic process (also known as the first-order autoregressive stochastic process), as shown in the following formula:

$$\alpha_{l,m}(t + \Delta t) = e^{-\Delta t/\tau} \alpha_{l,m}(t) + \sqrt{\frac{4\pi\sigma^2(1 - e^{-2\Delta t/\tau})}{L(L+2)}} R_{l,m}(t), \quad (7)$$

where Δt is the specific time interval that can correspond to the integration step of the model in the CMA's EPS; τ is the timescale of the random field decorrelation; and $R_{l,m}(t)$ is a Gaussian distribution stochastic process that obeys a variance of 1 and a mean of 0. The stochastic pattern constructed by the above formula features time-space scale correlation and controllable perturbation amount. In the CMA's GEPS and REPS, perturbations are applied only to the net tendency terms of potential temperature, horizontal wind components, and humidity variables.

Considering the different integration times and forecast objects of global and regional models, the parameter settings of the SPPT scheme in the CMA's GEPS and REPS are different. In the CMA-REPS, the SPPT perturbation amplitude ranges from 0.2 to 1.8 with a mean

of 1.0. The scheme uses a decorrelation time scale of 6 h, a maximum wavenumber of 24, and a standard deviation σ of 0.27. In the CMA-GEPS, to ensure operational stability, the stochastic function's perturbation amplitude is constrained to 0.5–1.5. The vertical profile of tendency perturbations is introduced at both the near-surface layer and atmospheric top, where physical tendencies either remain unperturbed or receive only minimal-amplitude perturbations. The SPPT scheme can significantly improve the ensemble spread and missing rate of temperature and wind speed forecasts, improve the probabilistic forecasting skills of heavy precipitation, and significantly increase the ensemble spread of various elements (altitude field, temperature field, and wind speed) forecasts in tropical areas, which makes up for the lack of perturbation in tropical areas by SV initial perturbation technique (Peng et al., 2020).

3.2.3.2 SKEB scheme

The main purpose of the SKEB scheme is to represent the stochastic process and uncertainty of subgrid-scale energy upscaling transitions in the model (Shutts, 2005). The SKEB scheme in the CMA-GEPS addresses excessive energy dissipation by employing a random flow function forcing with specific spatiotemporal correlations and local dynamic dissipation rates (Peng et al., 2019). The random stream function forcing F_ψ is defined as follows.

$$F_\psi = \frac{\alpha \Delta x}{\Delta t} \Psi(\lambda, \varphi, t) \sqrt{\Delta t D(\lambda, \varphi, \eta, t)}. \quad (8)$$

In Eq. (8), $\Psi(\lambda, \varphi, t)$ is a random type, and its generation method is the same as the definition of the random type of SPPT, as shown in Eq. (5); and $D(\lambda, \varphi, \eta, t)$ is the dissipation rate of local dynamic energy. At present, the SKEB scheme in the CMA-GEPS constructs the local dynamic energy dissipation rate based on the explicit horizontal scheme.

$$D(\lambda, \varphi, \eta, t) = -k \times \mathbf{u} \times \mathbf{u}', \quad (9)$$

where k is a constant factor greater than 1, \mathbf{u} is the horizontal wind speed, and \mathbf{u}' is the change in the horizontal wind speed before and after the application of the horizontal diffusion scheme. Based on the relationship between the stream function and the rotational component of the horizontal wind field, the perturbation forcing terms of the CMA-GEPS in horizontal wind field are constructed as S_u and S_v . These stochastic forcing terms are then added to the tendency terms of the model's horizontal wind field.

$$S_u = -\frac{1}{\alpha} \frac{\partial F_\psi}{\partial \phi}, \quad (10)$$

$$S_v = \frac{1}{\alpha \cos \phi} \frac{\partial F_\psi}{\partial \lambda}. \quad (11)$$

In the CMA-GEPS, the minimum and maximum truncation wavenumbers L_{\min} and L_{\max} of the Ψ random pattern of the SKEB scheme are 10 and 80, the decorrelation time scale τ is 6 h, the mean μ is 0, the standard deviation σ is 0.27, and the maximum Ψ_{\max} (minimum Ψ_{\min}) is 0.8 (-0.8). The application of the SKEB scheme can improve the simulation ability of the CMA-GEPS on the atmospheric kinetic energy spectrum, and significantly enhance the relationship between the ensemble mean error and the ensemble spread of the wind field forecast in the tropics.

3.2.4 Systematic bias and extreme information extraction

In addition to the aforementioned initial perturbation and model perturbation techniques, the CMA's GEPS and REPS have innovatively developed a suite of ensemble post-processing techniques and extreme weather forecasting applications. These include an ensemble dynamical method for correcting bias (Chen et al., 2020), the radar reflectivity factor for subgrid-scale precipitation (Chen Y. X. et al., 2021), and the model of extreme weather index. These techniques addressed the challenges of extracting extreme forecast signals from the tails of probabilistic forecast distributions and correcting systematic model biases, leading to notable improvements in probability density distributions and enhanced extreme weather prediction capabilities in CMA's GEPS and REPS. A brief overview of these technologies is as follows.

3.2.4.1 Ensemble forecasting dynamic correction method based on subtracting model systematic bias in tendency term

Current ensemble prediction techniques primarily address model stochastic errors. However, under pronounced model systematic biases, sole dependence on either initial perturbation or model perturbation methods is inadequate to optimize the error–spread relationship in ensemble forecasts. Following the research on the impacts of CMA global/regional model systematic biases on ensemble probability density distributions, an ensemble forecasting method was developed, applying dynamic correction through subtraction of model systematic bias in the tendency term. This approach removes bias tendencies from both dynamical and physical tendency terms during model integration, as shown in Eq. (12) (Chen et al., 2020; Han et al., 2023).

$$e_j(t_e) = \int_0^{t_e} \left[A(e_j, t) + P(e_j, t) - \hat{B}_l(e_0) \right] dt. \quad (12)$$

In the equation, $e_j(t)$ is the total tendency, $A(e_j, t)$ is

the dynamic tendency, $P(e_j, t)$ is a physical tendency, and $\hat{B}_l(e_0)$ is the subtracted bias. A linear bias coefficient is derived by linear regression and systematically removed from tendency terms at each model integration step. This method significantly enhances both the first- (ensemble mean) and second-order moment (ensemble spread) of the ensemble forecast probability density distribution, thereby improving the probabilistic prediction skill for surface meteorological elements.

3.2.4.2 A new calculated method of radar reflectivity factor for subgrid-scale precipitation

Since radar reflectivity can reveal the characteristics of strong convective weather events, forecasters pay close attention to uncertainty in model forecasts. To better quantify radar reflectivity uncertainties, we developed and implemented a new radar reflectivity calculation method for subgrid-scale precipitation in the CMA-REPS, as presented in Eq. (13). This addresses the current limitation where simulated radar reflectivity fails to represent precipitation from the Kain–Fritsch cumulus parameterization scheme (Chen Y. X. et al., 2021).

$$Z_{\text{total}} = Z_{\text{micro}} + AR_{\text{cu}}^b \quad (13)$$

In the equation, Z_{total} is the new radar reflectivity, Z_{micro} is the radar reflectivity from the cloud microphysics parameterization scheme, R_{cu} is the subgrid precipitation, and A and b are empirical parameters of the Z – R relationship for the radar quantitative precipitation estimation. The methodology operates by subtracting the downdraft evaporation rate from the subgrid precipitation rate, followed by layer-wise estimation of radar reflectivity using the radar-derived reflectivity–rainfall (i.e., Z – R) relationship. This new reflectivity is then integrated with microphysics-simulated radar echoes to generate a novel three-dimensional radar reflectivity field. The new reflectivity fields improved diagnostic capability for precipitation generated by the cumulus parameterization scheme, particularly under conditions dominated by subgrid convective processes. It can better simulate the radar reflectivity related to subgrid precipitation of a single ensemble member and then improve the probabilistic prediction technique of radar reflectivity in CMA-REPS.

3.2.4.3 Extreme weather prediction product

Based on the forecast, model climate, and historical climate data of 31 ensemble members of the CMA-GEPS, extreme weather prediction products are developed, including EFI of ground elements, 2-m temperature anomaly probability forecast product, and medium anomaly probability forecast product based on the Kalman filter bias correction technology.

(1) Ground elements EFI product based on climate

probability of CMA-GEPS. With the development of deterministic and ensemble forecasting for each forecasting element (precipitation, wind, temperature, and cloud cover), forecasters aim to utilize EPS for generating early warning signals of extreme events. However, it is difficult to directly compare the difference between the observed meteorological elements and the forecast output of the model. Therefore, based on the CMA-GEPS, a CMA ensemble forecast method of extreme weather called EFI was developed. The principle is to calculate the difference between the cumulative model climate probability distribution and the ensemble forecast probability distribution. The calculation of the model climate percentile is a crucial step in calculation of the EFI, but the CMA-GEPS has less historical data. Thus, the model climate percentile was calculated by using a 15-day sliding time window (± 7 days centered on the target date) and a spatial window comprising the target grid point and its 9 nearest neighbors. The daily model climate sequence was constructed from this spatiotemporal sampling, and percentiles (1, 2, ..., 99, 100) of the forecast field were extracted to form the model's climate percentile distribution (Wang J. Y. et al., 2014; Peng et al., 2024).

(2) Medium anomaly probability forecast product based on Kalman filter bias correction technology. The medium anomaly probability forecast products based on Kalman filter bias correction technology are developed, including atmospheric circulation anomaly probability products of 500-hPa geopotential height (Z500), 850-hPa temperature, and 850-hPa wind. These products help to better grasp the characteristics of extreme disaster weather and improve the probability forecasting skills of extreme disaster weather.

3.3 China's GEPS and REPS and their comparison with international systems

3.3.1 China's GEPS and REPS

In response to the characteristics of initial and model errors in CMA's GEPS and REPS, Chinese technicians have independently developed key technologies for ensemble forecast, thereby overcoming the bottlenecks in initial perturbation, model perturbation, and ensemble forecast application technologies. They have developed an independent and controllable GEPS with a horizontal resolution of 50 km for 0–15-day forecasts, along with a REPS featuring 10-km resolution and a convective-scale EPS with 3-km resolution, both for 0–3-day forecasts. These systems provide robust scientific and technological support for global ensemble prediction, propelling China's numerical prediction operations to achieve remarkable progress and reach an advanced level. The operational implementation of the CMA's GEPS and REPS

marks a significant milestone in advancing China's NWP capabilities. Beyond daily forecasting services and major meteorological support, the products from the CMA's GEPS and REPS continue to supply a wide range of forecast data to the TIGGE Center. They also provide core support for the World Meteorological Centre (WMC) Beijing and the WMO Severe Weather Forecasting Demonstration Programme for Southeast Asia (SWFDP-SeA) website, further elevating China's role in global meteorological science. **Table 2** illustrates the parameter settings of the CMA's GEPS, REPS, and convective-scale ensemble forecast (CAEF) in detail.

The CMA-GEPS integrates SV initial perturbation, model random physical perturbation, and generation technologies for extreme medium-range probabilistic forecast products. It features a horizontal resolution of 0.5° , a forecast range of 0–15 days, and 31 ensemble members. Leveraging ecFlow technology (a client/server workflow package of ECMWF that enables users to run a large number of programs in a controlled environment), the operational process of the CMA-GEPS has been established, enabling daily operations at 0000 and 1200 UTC. The system provides 29 types of products, including conventional forecasts and probabilistic forecasts for extreme weather events over the 0–15-day period.

The CMA-REPS v3.0 integrates a suite of advanced techniques, including the ETKF, multi-scale blended initial perturbation, SPPT, blended dynamic perturbations of lateral boundary conditions (LBCs), Conditional Typhoon Vortex Relocation (CTVR), and a new sub-grid scale precipitation radar echo reflectivity algorithm. The

system operates with 15 ensemble members and runs 4 times daily, with a forecast duration of 84 h for the 0000 and 1200 UTC cycles, and 6 h for the 0600 and 1800 UTC cycles. The system primarily supports the ETKF forecast cycle, preparing data for ETKF calculations and generating forecast perturbations. The system offers 49 types of ensemble products, with a focus on probabilistic forecasts for localized severe weather and tropical cyclones.

3.3.2 Analysis of international major numerical prediction centers

3.3.2.1 Comparison on key techniques and systems of ensemble prediction

Table 3 is a comprehensive comparison of major international EPSs. The main technical characteristics of the CMA are demonstrated as follows.

(1) The SV initial perturbation technique, a core component of the CMA-GEPS, has been innovatively enhanced with advanced algorithms for perturbation energy modulus, parallel computing, scaling, and linear sampling of initial perturbations based on a Gaussian distribution in a three-dimensional structure. This technique effectively captures the growth of large-scale baroclinic instability perturbation energy, significantly improving the accuracy and reliability of global ensemble forecasts. Given its high technical complexity, the SV initial perturbation technique is employed by only a few leading GEPSSs, including those of ECMWF, the French Meteorological Service, and the JMA. The successful implementation of this technique marks CMA's entry into the global forefront of meteorological forecasting, with its

Table 2. The parameter configuration for CMA's GEPS, REPS, and CAEF

	CMA-GEPS	CMA-REPS	CMA-CAEF
Model version	CMA-GFS v3.1	CMA-MESO v4.3	CMA-MESO v5.1
Horizontal resolution	0.5°	0.1°	0.03°
Vertical level	87	50	50
Domain	90°S – 90°N , 180°W – 180°E	10° – 60°N , 70° – 145°E	10° – 60°N , 70° – 145°E
Initial perturbation of control member	Upscaled from analysis field via CMA-GFS 4DVAR	Downscaled from CMA-GEPS	Generated from analysis field via 3DVAR
Initial perturbation scheme	SVs	ETKF	Multi-scale blending based on SVs
Model perturbation scheme	SPPT, SKEB	SPPT	SPPT
Lateral boundary perturbation scheme	/	From CMA-GEPS	Blended dynamic lateral boundary perturbations
Number of members	31	15	15
Forecast length	360 h (0000/1200 UTC)	84 h (0000/1200 UTC) 6 h (0600/1800 UTC)	72 h (0600/1800 UTC)
Output interval	0–84 h (3 h), 84–360 h (6 h)	1 h	1 h
Postprocessing product	GRIB2 data, ensemble prediction products for normal, extreme weather and typhoon	GRIB2 data, ensemble prediction products for normal, emergency service and typhoon	GRIB2 data, ensemble prediction products for normal, emergency service and typhoon

Note: GRIB2 refers to the Gridded Binary Format Edition 2, which is a standard encoding format sponsored by the WMO for the transmission of gridded data between the national meteorological centers.

technical level matching that of the world's most advanced NWP centers.

(2) The CMA-REPS incorporates several cutting-edge techniques, including the ETKF for initial perturbations, the MSB, and the CTVR. These advancements have significantly improved the accuracy of probabilistic precipitation forecasts in China. The ETKF method is based on the same principles and framework as the local ensemble transform Kalman filter (LETKF), which is operationally used by both the Meteorological Service of Canada and the JMA. Meanwhile, the MSB and CTVR techniques are originally developed by the CMA, showcasing the agency's innovation in REPS.

(3) The stochastic perturbation methods for physical processes, such as the SPPT and the SKEB, utilize a first-order autoregressive stochastic process and spherical harmonic function expansion. These methods produce random functions and key parameters that are temporally and spatially correlated and normally distributed. By introducing stochastic perturbations to the tendencies of physical processes and the dissipation of small-scale kinetic energy, these techniques provide a realistic characterization of the stochastic error growth in sub-grid physical process parameterizations. This approach is consistent with the stochastic perturbation schemes employed by leading meteorological centers in Europe (e.g., ECMWF and the UK) and North America (e.g., the U.S. and Canada).

(4) Aiming at ensemble forecast applications, we have designed techniques for extracting extreme information, dynamically correcting systematic biases, and developing algorithms that consider radar reflectivity for model convective precipitation. Additionally, we have developed technologies for generating probabilistic forecast products such as extreme weather forecast indices. These

advancements provide valuable reference information for extreme weather forecasts.

(5) An independent and proprietary operational system for ensemble forecast has been established in China, encompassing global medium-range forecasts (0–15 days), regional short-range forecasts (0–72 h), and convective-scale EPS. These advancements have significantly propelled progress in China's numerical weather prediction capabilities. A comparison of ensemble forecast techniques and system parameters across various international NWP centers reveals that global ensemble prediction resolutions typically range from 9 to 50 km, with ensemble members varying from 18 to 51. China's GEPS aligns closely with these international standards, featuring a horizontal resolution of 50 km and 31 ensemble members. In the realm of REPS, countries with smaller land areas, such as the UK, France, and Germany, have achieved resolutions of approximately 2.5 km, supported by 10–20 ensemble members. In contrast, countries with vast territories, like Canada, the U.S., and China, have generally established both regional meso- and convective-scale EPSs. China's REPS has a horizontal resolution of 10 km, while its convective-scale ensemble forecasting system achieves an even finer resolution of 3 km. Such a dual-system approach highlights China's commitments to enhancing forecast accuracy and resolution.

In summary, unique technical methods and products with independent intellectual property rights have been developed in China in the aspects of system parameters and construction schemes for CMA's GEPS and REPS.

3.3.2.2 Forecast capability comparison between CMA's GEPS and REPS and major international NWP centers

Based on data provided by the JMA from the Interna-

Table 3. Technical status of international major operational EPSs

Country	GEPS/horizontal resolution (number of members)/technology	REPS/horizontal resolution (number of members)/technology
USA	Spectral model/25 km (31)/ETKF, SPPT, SKEB	WRF, NMMB/16 km (26, North America)/3 km (11, key areas)/hybrid ensemble Kalman filter with BGM, dynamical downscaling, multi-physics
UK	Grid-point model MOGREPS-G/20 km (18)/EDA, SPPT, SKEB	Grid-point model MOGREPS-UK/2.2 km (3 perturbations, 18 lagged ensemble members)/Dynamical Downscaling
France	Spectral model/7.5–37 km (35)/SVs, EDA, MP	Spectral model AROME-EPS/2.5 km (16)/EDA, SPP
Germany	Grid-point model GEPS/40 km (40)/LETKF, SPP	Grid-point model COSMO-DE/2.2 km (20)/LETKF, SPP
Canada	Grid-point model/39 km (21)/LETKF, SPP, SKEB	Limited area model HREPS/15 km (21)/dynamical downscaling, physics perturbation
Japan	Spectral model/27–40 km (27)/SVs, LETKF, SPPT	Grid-point model/5 km (21)/multi-scale SVs
EU	Spectral model/9 km (51)/SVs, EDA, SPPT	/
China	Grid-point model GRAPES/50 km (31)/SVs, SPPT, SKEB	Grid-point model GRAPES/10 (15, China) and 3 km (15, China)/ETKF, MSB, multi-scale SVs, SPPT

Note: MOGREPS refers to the Met Office Global and Regional Ensemble Prediction System, and “G” in MOGREPS-G stands for “global”; NMMB indicates the Nonhydrostatic Mesoscale Model on the B-grid developed by NOAA; AROME denotes the Application of Research to Operations at Mesoscale; and COSMO-DE is a short for the Consortium for Small-scale Modeling Deutschland (“Germany” in English) Edition operated by the Deutscher Wetterdienst (DWD) of Germany.

tional Global Ensemble Forecast Verification Center, Fig. 1 presents a comparison of the continuous ranked probability score (CRPS) for the day 6 of ensemble forecast and the average RMSE for the day 10 of ensemble forecast of the 500-hPa geopotential height field from multiple centers over the period of 2012–2024. The verification data mainly include the GEPSs of several numerical centers, such as the ECMWF, the NCEP, the United Kingdom Met Office (UKMO), the CMC, the Bureau of Meteorology, Australia (BOM), the JMA, the Korea Meteorological Administration (KMA), and the CMA. It can be seen that since the operational implementation of the CMA-GEPS in December 2018, its forecast skill for medium-range forecasts has been comparable to that of the GEPSs of Europe and the U.S.

4. Challenges and prospects in the development of ensemble forecast

4.1 Scientific challenges: Spatiotemporal variability of predictability in convective-scale models

The predictability of convective-scale models exhibits significant spatiotemporal variability. First, the predictability of convective-scale models is influenced by the interactions of circulations across different spatiotemporal scales. Even if the magnitude of initial errors is sufficiently small, these errors can rapidly saturate and undergo upscale growth, thereby affecting the predictability of large-scale weather systems (Zhang et al., 2007). The growth of initial errors in convective-scale models has the characteristics of strong nonlinearity, with the error growth rate being approximately 10 times that of synoptic-scale models (Hohenegger and Schar, 2007). Second, convective-scale models typically employ more sophisticated and complex parameterization schemes such as cloud microphysics and turbulent diffusion. However, these approaches are subject to numerous assumptions and empirical parameters, as well as errors arising from finite-difference methods and computational truncation, all of which contribute to the strongly nonlinear growth of initial errors in convective-scale models (Mu et al., 2011; Yano et al., 2018). Third, the growth of initial errors at different scales and their impact on precipitation are closely related to environmental forcing conditions (Johnson, 2014; Zhuang et al., 2021). The relative importance of initial errors at different scales depends on the quantity and type of moist convection, with more moist convection leading to greater growth of small-scale error perturbation energy (Nielsen, 2016). Consequently, convective-scale ensemble forecast techniques are facing significant challenges.

4.2 Frontier exploration of a combination of ensemble forecast and AI

In recent years, AI has achieved groundbreaking progress in meteorological forecast applications. The computational cost of deep learning-based large-scale meteorological models is only 1×10^{-4} of that of traditional numerical models, offering a new potential pathway for the development of large-sample ensemble forecast. However, this also introduces new challenges. First, the sources of uncertainty in AI models differ from those in traditional numerical models, and the impact of these uncertainties on the forecast skill and performance of AI models requires further in-depth investigation. Second, the uncertainty characteristics of deep learning models differ from those of traditional numerical forecast, and how to appropriately describe the uncertainty features of deep learning models remains an urgent issue to be addressed. Third, the physical consistency and interpretability of AI-based ensemble forecast models remain debated, necessitating corresponding research and improvements alongside the development of AI models. Furthermore, the advancement of AI has brought new challenges and opportunities to ensemble forecast. The framework and concepts of ensemble forecast may also need to undergo transformations in the future. For instance, (1) ensemble forecast may revert to probabilistic forecast, with the core objective of obtaining the PDF of variables, which conceptually derives from seeking the probability distribution of forecast errors. (2) The significance of representative perturbations may decrease. As machine learning models can rapidly generate tens of thousands of ensemble forecasts, a broader spectrum of forecast possibilities can be explored, which provides better probabilistic and statistical characteristics. (3) Partial replacement of dynamical models may occur, as smaller-scale structures in the atmosphere require high-resolution dynamical models for identification. Here, dynamical models essentially act as effective but time-consuming random perturbation generators. Whether machine learning models can learn and generate such perturbations is a topic worthy of further research.

4.3 Operational challenges: Inapplicability of traditional ensemble forecast techniques

In recent years, operational ensemble forecasting development has encountered three fundamental limitations. First, global and regional operational ensemble forecast is moving toward convective-scale resolution. Under the convective-scale modeling framework, the multi-scale structures of initial condition and model errors, along with their nonlinear evolution and interac-

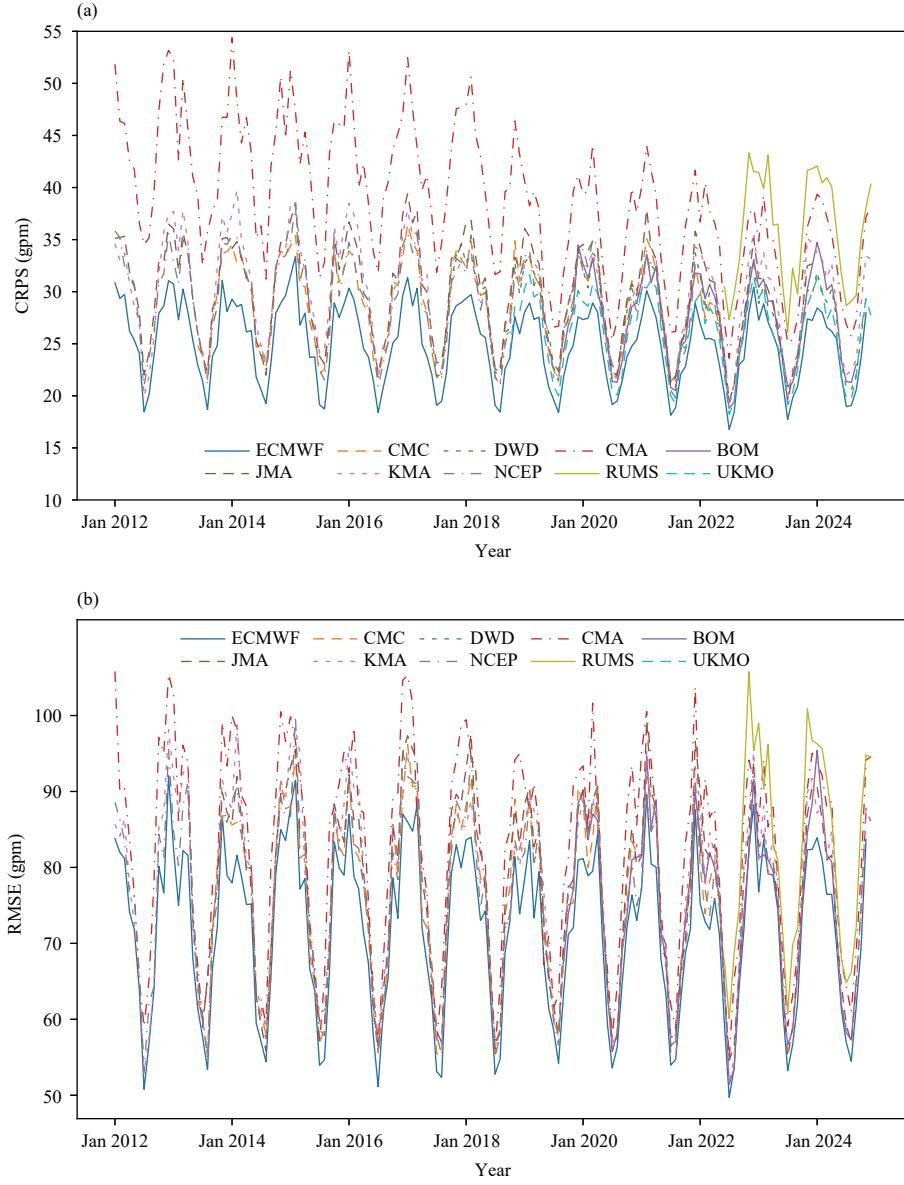


Fig. 1. The ensemble mean (a) continuous ranked probability score (CRPS) and (b) RMSE of the 500-hPa geopotential height (Z500) forecasts at 144 (day 6) and 240 h (day 10) from major international global EPSs, averaged monthly from January 2012 to December 2024. Different colors represent different NWP centers. RUMS refers to the regional unified model system.

tions, remain key scientific challenges in understanding the mechanisms of forecast error growth. How to construct ensemble perturbation methods is a technical bottleneck in the development of convective-scale ensemble forecast. Second, earth system models are evolving toward seamless weather–climate integrated EPSs, while traditional ensemble forecast techniques have been developed separately for temporal and spatial scales, making them inadequate for seamless spatiotemporal ensemble forecasting requirements. Third, ensemble forecast is gradually transitioning toward earth system ensemble forecast. Yet, traditional ensemble forecast techniques primarily address the uncertainty in forecasting derived from the error growth of dynamically unstable

atmospheric systems. The understanding of error characteristics from coupled component models and coupled data assimilation remains insufficient, requiring the development of ensemble forecast techniques suitable for multi-component coupled models.

5. Conclusions

In recent years, ensemble forecasting in China has achieved remarkable development and progress, from theoretical methods to operational applications, making China a world leader in operational ensemble forecasting. However, there is still a certain gap compared to the performance of EPSs at the most advanced operational

centers such as the ECMWF. Currently, ensemble forecasting has been intentionally developed to gradually replace deterministic forecasting, emerging as the primary form and important trend in operational forecasting. The concept of probabilistic forecasting is also gradually being popularized and accepted among researchers, forecasters, and the public. Below, we briefly summarize and discuss the challenges facing the future development of ensemble forecasting, which are also key issues that need to be addressed in the scientific research and operational development of ensemble forecasting in China.

(1) The ensemble members of global ensemble forecasting systems will advance towards convective-scale resolution in the future. This will facilitate the sampling of fine-scale features from mesoscale to convective scale and the estimation of uncertainties in ensemble forecasting on a global scale. ECMWF is at the forefront of this development, targeting a global ensemble forecast at a resolution of 5 km by 2025. Higher resolution not only poses a significant challenge to the high-performance computing resources of major operational centers but also raises highly challenging scientific and engineering questions regarding how to effectively generate multi-scale initial ensemble perturbations and model physics perturbations, while ensuring the physical consistency of these perturbations.

(2) It is important to develop integrated ensemble forecasting for global medium-range to subseasonal-seasonal (S2S) predictions. As the sources of uncertainty in subseasonal and seasonal predictions become increasingly diverse and complex, conducting ensemble forecasting for S2S is an international frontier issue. The current trend among major international operational centers is to use integrated atmosphere–land–ocean coupled models for ensemble forecasting of short- to medium-range and subseasonal predictions. The main frontier scientific questions in this area are how to construct initial ensemble perturbations for the atmosphere–ocean coupled system and how to account for the impact of external forcing uncertainties on S2S ensemble forecasts.

(3) With the increasing resolution of global ensemble forecasting, it is necessary to reconsider the developmental positioning and key scientific issues of REPSs. Regional ensemble forecast requires enhanced spatial resolutions than global ensemble forecast, potentially advancing towards sub-kilometer resolution in the future. Its forecasting targets may also gradually encompass smaller urban-scale meteorological elements. Moreover, how to more effectively consider the coordination and nonlinear interactions of initial, lateral boundary, and model process multi-source perturbations remains a crit-

ical scientific issue in regional ensemble forecasting.

(4) The rise of AI technology presents unprecedented challenges and opportunities for ensemble forecasting. The essence of ensemble forecasting lies in statistical sampling, while AI tools excel in constructing statistical characteristics of variables and their nonlinear relationships. How to effectively apply AI technology to ensemble forecasting is a vast and cutting-edge scientific and engineering issue. Future developments in this area hold the promise of significantly advancing the level of ensemble forecasting.

REFERENCES

- Arakawa, A., 2004: The cumulus parameterization problem: Past, present, and future. *J. Climate*, **17**, 2493–2525, [https://doi.org/10.1175/1520-0442\(2004\)017<2493:RATCPP>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<2493:RATCPP>2.0.CO;2).
- Bauer, P., A. Thorpe, and G. Brunet, 2015: The quiet revolution of numerical weather prediction. *Nature*, **525**, 47–55, <https://doi.org/10.1038/nature14956>.
- 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>.
- Berner, J., K. R. Fossell, S. Y. Ha, et al., 2015: Increasing the skill of probabilistic forecasts: Understanding performance improvements from model-error representations. *Mon. Wea. Rev.*, **143**, 1295–1320, <https://doi.org/10.1175/MWR-D-14-00091.1>.
- Bowler, N. E., A. Arribas, K. R. Mylne, et al., 2008: The MOGREPS short-range ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **134**, 703–722, <https://doi.org/10.1002/qj.234>.
- 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:PFSEOP>2.0.CO;2](https://doi.org/10.1175/1520-0493(1997)125<0099:PFSEOP>2.0.CO;2).
- Buizza, R., M. Miller, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887–2908, <https://doi.org/10.1002/qj.49712556006>.
- Buizza, R., J. Du, Z. Toth, et al., 2018: Major operational ensemble prediction systems (EPS) and the future of EPS. *Handbook of Hydrometeorological Ensemble Forecasting*, Q. Y. Duan, F. Pappenberger, A. Wood, et al., Eds., Springer, Berlin and Heidelberg, 1–43, https://doi.org/10.1007/978-3-642-40457-3_14-1.
- Cai, W. C., J. Z. Min, and X. R. Zhuang, 2017: Comparison of different stochastic physics perturbation schemes on a storm-scale ensemble forecast in a heavy rain event. *Plateau Meteor.*, **36**, 407–423, <https://doi.org/10.7522/j.issn.1000-0534.2016.00024>. (in Chinese)
- Charron, M., G. Pellerin, L. Spacek, et al., 2010: Toward random sampling of model error in the Canadian ensemble prediction system. *Mon. Wea. Rev.*, **138**, 1877–1901, <https://doi.org/10.1175/2009MWR3187.1>.

- Chen, C. H., Y. Wang, H. R. He, et al., 2021: Review of the ensemble prediction using stochastic physics. *Adv. Meteor. Sci. Technol.*, **11**, 48–57, <https://doi.org/10.3969/j.issn.2095-1973.2021.03.007>. (in Chinese)
- Chen, D. H., J. S. Xue, X. S. Yang, et al., 2008: New generation of multi-scale NWP system (GRAPES): General scientific design. *Chinese Sci. Bull.*, **53**, 3433–3445, <https://doi.org/10.1007/s11434-008-0494-z>.
- Chen, J., and X. L. Li, 2020: The review of 10 years development of the GRAPES global/regional ensemble prediction. *Adv. Meteor. Sci. Technol.*, **10**, 9–18, 29, <https://doi.org/10.3969/j.issn.2095-1973.2020.02.003>. (in Chinese)
- Chen, J., D. H. Chen, and H. Yan, 2002: A brief review on the development of ensemble prediction system. *J. Appl. Meteor. Sci.*, **13**, 497–507, <https://doi.org/10.3969/j.issn.1001-7313.2002.04.013>. (in Chinese)
- Chen, J., J. S. Xue, and H. Yan, 2003: The uncertainty of mesoscale numerical prediction of South China heavy rain and the ensemble simulations. *Acta Meteor. Sinica*, **61**, 432–446, <https://doi.org/10.11676/qxxb2003.042>. (in Chinese)
- Chen, J., J. S. Xue, and H. Yan, 2004: Impacts of diabatic physics parameterization schemes on mesoscale heavy rainfall short-range simulation. *Acta Meteor. Sinica*, **18**, 51–72. Available online at <http://jmr.cmsjournal.net/en/article/id/943>. Accessed on 6 May 2025.
- Chen, J., J. S. Xue, and H. Yan, 2005: A new initial perturbation method of ensemble mesoscale heavy rain prediction. *Chinese J. Atmos. Sci.*, **29**, 717–726, <https://doi.org/10.3878/j.issn.1006-9895.2005.05.05>. (in Chinese)
- Chen, J., J. Z. Wang, J. Du, et al., 2020: Forecast bias correction through model integration: A dynamical wholesale approach. *Quart. J. Roy. Meteor. Soc.*, **146**, 1149–1168, <https://doi.org/10.1002/qj.3730>.
- Chen, X., H. L. Yuan, and M. Xue, 2018: Spatial spread-skill relationship in terms of agreement scales for precipitation forecasts in a convection-allowing ensemble. *Quart. J. Roy. Meteor. Soc.*, **144**, 85–98, <https://doi.org/10.1002/qj.3186>.
- Chen, Y. X., J. Chen, D. H. Chen, et al., 2021: A simulated radar reflectivity calculation method in numerical weather prediction models. *Wea. Forecasting*, **36**, 341–359, <https://doi.org/10.1175/WAF-D-20-0030.1>.
- Chou, J. F., 1986: Some general properties of the atmospheric model in H space, R space, point mapping, cell mapping. Proceedings of International Summer Colloquium on Nonlinear Dynamics of the Atmosphere, 10–20 August, Beijing. Science Press, Beijing, 187–189.
- Deng, G., J. D. Gong, L. T. Deng, et al., 2010: Development of mesoscale ensemble prediction system at national meteorological center. *J. Appl. Meteor. Sci.*, **21**, 513–523, <https://doi.org/10.3969/j.issn.1001-7313.2010.05.001>. (in Chinese)
- Denis, B., J. Côté, and R. Laprise, 2002: Spectral decomposition of two-dimensional atmospheric fields on limited-area domains using the discrete cosine transform (DCT). *Mon. Wea. Rev.*, **130**, 1812–1829, [https://doi.org/10.1175/1520-0493\(2002\)130<1812:SDOTDA>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<1812:SDOTDA>2.0.CO;2).
- Du, J., and J. Chen, 2010: The corner stone in facilitating the transition from deterministic to probabilistic forecasts-ensemble forecasting and its impact on numerical weather prediction. *Meteor. Mon.*, **36**, 1–11, <https://doi.org/10.7519/j.issn.1000-0526.2010.11.001>. (in Chinese)
- Du, J., and G. Deng, 2020: An introduction to “measure of forecast challenge” and “predictability horizon diagram index”. *Adv. Meteor. Sci. Technol.*, **10**, 75–77, <https://doi.org/10.3969/j.issn.2095-1973.2020.02.009>. (in Chinese)
- Du, J., R. H. Grumm, and G. Deng, 2014: Ensemble anomaly forecasting approach to predicting extreme weather demonstrated by extremely heavy rain event in Beijing. *Chinese J. Atmos. Sci.*, **38**, 685–699, <https://doi.org/10.3878/j.issn.1006-9895.2013.13218>. (in Chinese)
- Du, J., J. Berner, R. Buizza, et al., 2018: Ensemble methods for meteorological predictions. *Handbook of Hydrometeorological Ensemble Forecasting*, Q. Y. Duan, F. Pappenberger, A. Wood, et al., Eds., Springer, Berlin and Heidelberg, 1–52, https://doi.org/10.1007/978-3-642-40457-3_13-1.
- Duan, W. S., and F. F. Zhou, 2013: Non-linear forcing singular vector of a two-dimensional quasi-geostrophic model. *Tellus A*, **65**, 18452, <https://doi.org/10.3402/tellusa.v65i0.18452>.
- 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., R. Q. Ding, and F. F. Zhou, 2013: Several dynamical methods used in predictability studies for numerical weather forecasts and climate prediction. *Climatic Environ. Res.*, **18**, 524–538, <https://doi.org/10.3878/j.issn.1006-9585.2012.12009>. (in Chinese)
- Dutra, E., L. Magnusson, F. Wetterhall, et al., 2013: The 2010–2011 drought in the Horn of Africa in ECMWF reanalysis and seasonal forecast products. *Int. J. Climatol.*, **33**, 1720–1729, <https://doi.org/10.1002/joc.3545>.
- Epstein, E. S., 1969: Stochastic dynamic prediction. *Tellus A*, **21**, 739–759, <https://doi.org/10.3402/tellusa.v21i6.10143>.
- Feng, H. Z., J. Chen, G. B. He, et al., 2006: Simulation and test of short-range ensemble prediction system for heavy rainfall in the upper reach of Changjiang River. *Meteor. Mon.*, **32**, 12–16, <https://doi.org/10.3969/j.issn.1000-0526.2006.08.002>. (in Chinese)
- Feng, J., R. Q. Ding, D. Q. Liu, et al., 2014: The application of nonlinear local Lyapunov vectors to ensemble predictions in Lorenz systems. *J. Atmos. Sci.*, **71**, 3554–3567, <https://doi.org/10.1175/JAS-D-13-0270.1>.
- Gao, L., J. Chen, J. W. Zheng, et al., 2019: Progress in researches on ensemble forecasting of extreme weather based on numerical models. *Adv. Earth Sci.*, **34**, 706–716, <https://doi.org/10.11867/j.issn.1001-8166.2019.07.0706>. (in Chinese)
- Guan, H., and Y. J. Zhu, 2017: Development of verification methodology for extreme weather forecasts. *Wea. Forecasting*, **32**, 479–491, <https://doi.org/10.1175/WAF-D-16-0123.1>.
- Hamill, T. M., 1999: Hypothesis tests for evaluating numerical precipitation forecasts. *Wea. Forecasting*, **14**, 155–167, [https://doi.org/10.1175/1520-0434\(1999\)014<0155:HTFENP>2.0.CO;2](https://doi.org/10.1175/1520-0434(1999)014<0155:HTFENP>2.0.CO;2).
- Hamill, T. M., and J. Juras, 2006: Measuring forecast skill: Is it real skill or is it the varying climatology? *Quart. J. Roy. Meteor. Soc.*, **132**, 2905–2923, <https://doi.org/10.1256/qj.06.25>.
- Han, Y. M., J. Chen, F. Peng, et al., 2023: A model tendency per-

- turbation method that combines systematic bias of potential temperature and random errors in global ensemble prediction. *Acta Meteor. Sinica*, **81**, 592–604, <https://doi.org/10.11676/qxb2023.20220203>. (in Chinese)
- Hersbach, H., R. Mureau, J. D. Opsteegh, et al., 2000: A short-range to early-medium-range ensemble prediction system for the European area. *Mon. Wea. Rev.*, **128**, 3501–3519, [https://doi.org/10.1175/1520-0493\(2000\)128<3501:ASRTEM>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<3501:ASRTEM>2.0.CO;2).
- Hoffman, R. N., and E. Kalnay, 1983: Lagged average forecasting, an alternative to Monte Carlo forecasting. *Tellus A*, **35A**, 100–118. Available online at <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1600-0870.1983.tb00189.x>. Accessed on 6 May 2025.
- Hohenegger, C., and C. Schar, 2007: Atmospheric predictability at synoptic versus cloud-resolving scales. *Bull. Amer. Meteor. Soc.*, **88**, 1783–1794, <https://doi.org/10.1175/BAMS-88-11-783>.
- Hollingsworth, A., 1979: An experiment in Monte Carlo forecasting. Proceedings of ECMWF Workshop on Stochastic Dynamic Forecasting, ECMWF, Reading, 65–85. Available online at <https://www.ecmwf.int/en/elibrary/74849-experiment-monte-carlo-forecasting>. Accessed on 6 May 2025.
- 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](https://doi.org/10.1175/1520-0493(1995)123<2181:MFEP>2.0.CO;2).
- Houtekamer, P. L., and H. L. Mitchell, 2005: Ensemble Kalman filtering. *Quart. J. Roy. Meteor. Soc.*, **131**, 3269–3289, <https://doi.org/10.1256/qj.05.135>.
- Houtekamer, P. L., L. Lefavre, J. Derome, et al., 1996: A system simulation approach to ensemble prediction. *Mon. Wea. Rev.*, **124**, 1225–1242, [https://doi.org/10.1175/1520-0493\(1996\)124<1225:ASSATE>2.0.CO;2](https://doi.org/10.1175/1520-0493(1996)124<1225:ASSATE>2.0.CO;2).
- Huo, Z. H., Y. Z. Liu, J. Chen, et al., 2020: The preliminary application of tropical cyclone targeted singular vectors in the GRAPES global ensemble forecasts. *Acta Meteor. Sinica*, **78**, 48–59, <https://doi.org/10.11676/qxb2020.006>. (in Chinese)
- Isaksen, L., M. Bonavita, R. Buizza, et al., 2010: Ensemble of Data Assimilations at ECMWF. Technical Memorandum No.636, ECMWF, 42 pp. Available online at <https://www.ecmwf.int/en/elibrary/74969-ensemble-data-assimilations-ecmwf>. Accessed on 6 May 2025.
- Jankov, I., J. Berner, J. Beck, et al., 2017: A performance comparison between multiphysics and stochastic approaches within a North American RAP ensemble. *Mon. Wea. Rev.*, **145**, 1161–1179, <https://doi.org/10.1175/MWR-D-16-0160.1>.
- Jankov, I., J. Beck, J. Wolff, et al., 2019: Stochastically perturbed parameterizations in an HRRR-based ensemble. *Mon. Wea. Rev.*, **147**, 153–173, <https://doi.org/10.1175/MWR-D-18-0092.1>.
- Ji, L. Y., X. F. Zhi, C. Simmer, et al., 2020: Multimodel ensemble forecasts of precipitation based on an object-based diagnostic evaluation. *Mon. Wea. Rev.*, **148**, 2591–2606, <https://doi.org/10.1175/MWR-D-19-0266.1>.
- Ji, Y., X. F. Zhi, L. Y. Ji, et al., 2023: Conditional ensemble model output statistics for postprocessing of ensemble precipitation forecasting. *Wea. Forecasting*, **38**, 1707–1718, <https://doi.org/10.1175/WAF-D-22-0190.1>.
- Jiang, Y. M., J. Chen, M. Y. Jiao, et al., 2011: The preliminary experiment of GRAPES-MESO ensemble prediction based on TIGGE data. *Meteor. Mon.*, **37**, 392–402. (in Chinese)
- Jiao, M. Y., 2010: Progress on the key technology development in application of ensemble prediction products associated with TIGGE. *J. Meteor. Res.*, **24**, 136. Available online at <http://jmr.cmsjournal.net/en/article/id/1240>. Accessed on 6 May 2025.
- Johnson, A., 2014: Optimal design of a multi-scale ensemble system for convective scale probabilistic forecasts: Data assimilation and initial condition perturbation methods. Ph.D. dissertation, University of Oklahoma, USA, 156 pp.
- Jolliffe, I. T., and D. B. Stephenson, 2012: *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. 2nd ed. Wiley-Blackwell, Chichester, 274 pp.
- Krishnamurti, T. N., C. M. Kishtawal, T. E. LaRow, et al., 1999: Improved weather and seasonal climate forecasts from multimodel superensemble. *Science*, **285**, 1548–1550, <https://doi.org/10.1126/science.285.5433.1548>.
- Lalaurette, F., 2003: Early detection of abnormal weather conditions using a probabilistic extreme forecast index. *Quart. J. Roy. Meteor. Soc.*, **129**, 3037–3057, <https://doi.org/10.1256/qj.02.152>.
- 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<409:TSOMCF>2.0.CO;2](https://doi.org/10.1175/1520-0493(1974)102<409:TSOMCF>2.0.CO;2).
- Leutbecher, M., S.-J. Lock, P. Ollinaho, et al., 2017: Stochastic representations of model uncertainties at ECMWF: State of the art and future vision. *Quart. J. Roy. Meteor. Soc.*, **143**, 2315–2339, <https://doi.org/10.1002/qj.3094>.
- Li, C., 2001: An integrated decision-making model of ecological and environmental management for sustainable development in the Yangtze River Basin. Ph.D. dissertation, China Institute of Water Resources and Hydropower Research, China. (in Chinese)
- Li, J., J. Du, M. H. Wang, et al., 2009: Experiments of perturbing initial conditions in the development of mesoscale ensemble prediction system for heavy rainstorm forecasting. *Plateau Meteor.*, **28**, 1365–1375. (in Chinese)
- Li, J., J. Du, and Y. Liu, 2015: A comparison of initial condition-, multi-physics- and stochastic physics-based ensembles in predicting Beijing “7.21” excessive storm rain event. *Acta Meteor. Sinica*, **73**, 50–71, <https://doi.org/10.11676/qxb2015.008>.
- Li, J., J. Du, Y. Liu, et al., 2017: Similarities and differences in the evolution of ensemble spread using various ensemble perturbation methods including topography perturbation. *Acta Meteor. Sinica*, **75**, 123–146, <https://doi.org/10.11676/qxb2017.011>. (in Chinese)
- Li, J. P., Q. C. Zeng, and J. F. Chou, 2000: Computational uncertainty principle in nonlinear ordinary differential equations. I. Numerical results. *Sci. China Ser. E*, **43**, 449–460. Available online at <https://www.science.org/Sci%20China%20Tech%20Sci%20E/doi/10.1360/ye2000-43-5-449>. Accessed on 6 May 2025.
- Li, X. L., and Y. Z. Liu, 2019: The improvement of GRAPES global extratropical singular vectors and experimental study. *Acta Meteor. Sinica*, **77**, 552–562, <https://doi.org/10.11676/qxb2019.020>. (in Chinese)

- Li, X. L., M. Charron, L. Spacek, et al., 2008: A regional ensemble prediction system based on moist targeted singular vectors and stochastic parameter perturbations. *Mon. Wea. Rev.*, **136**, 443–462, <https://doi.org/10.1175/2007MWR2109.1>.
- Li, X. Q., J. D. Liu, and Y. H. Wang, 1997: The ensemble prediction and its application in medium range weather forecast. *Meteor. Mon.*, **23**, 3–9. (in Chinese)
- Li, Z. C., and D. H. Chen, 2002: The development and application of the operational ensemble prediction system at National Meteorological Center. *J. Appl. Meteor. Sci.*, **13**, 1–15, <https://doi.org/10.3969/j.issn.1001-7313.2002.01.001>. (in Chinese)
- Liu, L., J. Chen, and J. Y. Wang, 2018: A study on medium-range objective weather forecast technology for persistent heavy rainfall events based on T639 ensemble forecast. *Acta Meteor. Sinica*, **76**, 228–240, <https://doi.org/10.11676/qxb2018.002>. (in Chinese)
- Liu, X., J. Chen, Y. Z. Liu, et al., 2024: An initial perturbation method for the multiscale singular vector in global ensemble prediction. *Adv. Atmos. Sci.*, **41**, 545–563, <https://doi.org/10.1007/s00376-023-3035-4>.
- Liu, Y. Z., X. S. Yang, and H. Q. Wang, 2011: Research on GRAPES singular vectors and application to heavy rain ensemble prediction. *Acta Sci. Nat. Univ. Pekinensis*, **47**, 271–277, <https://doi.org/10.13209/j.0479-8023.2011.039>. (in Chinese)
- Liu, Y. Z., X. S. Shen, and X. L. Li, 2013: Research on the singular vector perturbation of the GRAPES global model based on the total energy norm. *Acta Meteor. Sinica*, **71**, 517–526, <https://doi.org/10.11676/qxb2013.043>. (in Chinese)
- Liu, Y. Z., L. Zhang, and Z. Y. Jin, 2017: The optimization of GRAPES global tangent linear model and adjoint model. *J. Appl. Meteor. Sci.*, **28**, 62–71, <https://doi.org/10.1198/1001-7313.20170106>. (in Chinese)
- Long, K. J., J. Chen, X. L. Ma, et al., 2011: The preliminary study on ensemble prediction of GRAPES-meso based on ETKF. *J. Chengdu Univ. Inf. Technol.*, **26**, 37–46, <https://doi.org/10.3969/j.issn.1671-1742.2011.01.008>. (in Chinese)
- Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141, [https://doi.org/10.1175/1520-0469\(1963\)020<0130:DNF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2).
- Lyu, Y., S. P. Zhu, X. F. Zhi, et al., 2023: Improving subseasonal-to-seasonal prediction of summer extreme precipitation over southern China based on a deep learning method. *Geophys. Res. Lett.*, **50**, e2023GL106245, <https://doi.org/10.1029/2023GL106245>.
- Ma, Q., J. D. Gong, L. Li, et al., 2008: Study on the 2nd moment spread-correction of mesoscale ensemble forecast system. *Meteor. Mon.*, **34**, 15–21, <https://doi.org/10.7519/j.issn.1000-0526.2008.11.003>. (in Chinese)
- Ma, X. L., J. S. Xue, and W. S. Lu, 2008: Preliminary study on ensemble transform Kalman filter-based initial perturbation scheme in GRAPES global ensemble prediction. *Acta Meteor. Sinica*, **66**, 526–536, <https://doi.org/10.11676/qxb2008.050>. (in Chinese)
- Ma, X. L., Y. X. Ji, B. Y. Zhou, et al., 2018: A new scheme of blending initial perturbation of the GRAPES regional ensemble prediction system. *Trans. Atmos. Sci.*, **41**, 248–257, <https://doi.org/10.13878/j.cnki.dqkxb.20160104001>. (in Chinese)
- Ma, Y. N., J. Chen, Z. Z. Xu, et al., 2023: Evolution characteristics of initial perturbation energy at different scales in convection-permitting ensemble prediction of GRAPES. *Chinese J. Atmos. Sci.*, **47**, 1541–1556, <https://doi.org/10.3878/j.issn.1006-9895.2202.21242>. (in Chinese)
- 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. *Nonlinear Proc. Geophys.*, **10**, 493–501, <https://doi.org/10.5194/npg-10-493-2003>.
- Mu, M., B. Y. Chen, F. F. Zhou, et al., 2011: Methods and uncertainties of meteorological forecast. *Meteor. Mon.*, **37**, 1–13. Available online at http://qxqk.nmc.cn/qx/ch/reader/view_abstract.aspx?file_no=20110101&flag=1. Accessed on 6 May 2025. (in Chinese)
- Mylne, K., J. Chen, A. Erfani, et al., 2022: Guidelines for Ensemble Prediction System. WMO-1254, World Meteorological Organization, 40 pp.
- Nielsen, E. R., 2016: Using convection-allowing ensembles to understand the predictability of extreme rainfall. Master dissertation, Colorado State University, USA, 178 pp.
- Ono, K., M. Kunii, and Y. Honda, 2021: The regional model-based mesoscale ensemble prediction system, MEPS, at the Japan Meteorological Agency. *Quart. J. Roy. Meteor. Soc.*, **147**, 465–484, <https://doi.org/10.1002/qj.3928>.
- Palmer, T. N., R. Buizza, F. Doblas-Reyes, et al., 2009: Stochastic Parametrization and Model Uncertainty. ECMWF Technical Memoranda 598, ECMWF, 42 pp, <https://doi.org/10.21957/ps8gbwbdv>.
- Pan, X., Q. P. Wang, Y. Zhang, et al., 2021: Analysis constraints scheme of initial perturbation of ensemble prediction. *Chinese J. Atmos. Sci.*, **45**, 1327–1344. Available online at <http://www.iapjournals.ac.cn/dqkx/article/doi/10.3878/j.issn.1006-9895.2103.21029>. Accessed on 6 May 2025. (in Chinese)
- Pauluis, O., and J. Schumacher, 2013: Radiation impacts on conditionally unstable moist convection. *J. Atmos. Sci.*, **70**, 1187–1203, <https://doi.org/10.1175/JAS-D-12-0127.1>.
- Peng, F., X. L. Li, J. Chen, et al., 2019: A stochastic kinetic energy backscatter scheme for model perturbations in the GRAPES global ensemble prediction system. *Acta Meteor. Sinica*, **77**, 180–195, <https://doi.org/10.11676/qxb2019.009>. (in Chinese)
- Peng, F., X. L. Li, and J. Chen, 2020: Impacts of different stochastic physics perturbation schemes on the GRAPES global ensemble prediction system. *Acta Meteor. Sinica*, **78**, 972–987, <https://doi.org/10.11676/qxb2020.074>. (in Chinese)
- Peng, F., J. Chen, X. L. Li, et al., 2024: Development of the CMA-GEPS extreme forecast index and its application to verification of summer 2022 extreme high temperature forecasts. *Acta Meteor. Sinica*, **82**, 190–207, <https://doi.org/10.11676/qxb2024.20230017>. (in Chinese)
- Qi, Q. Q., Y. J. Zhu, J. Chen, et al., 2022: Error diagnosis and assessment of sub-seasonal forecast using GRAPES-GFS model. *Chinese J. Atmos. Sci.*, **46**, 327–345. Available online at <http://www.iapjournals.ac.cn/dqkx/article/doi/10.3878/j.issn.1006-9895.2202.21242>.

- 006-9895.2008.20157. Accessed on 6 May 2025. (in Chinese)
- Qiao, X. S., S. Z. Wang, and J. Z. Min, 2017: A stochastic perturbed parameterization tendency scheme for diffusion (SP PTD) and its application to an idealized supercell simulation. *Mon. Wea. Rev.*, **145**, 2119–2139, <https://doi.org/10.1175/MWR-D-16-0307.1>.
- Qiao, X. S., S. Z. Wang, and J. Z. Min, 2018: The impact of a stochastically perturbing microphysics scheme on an idealized supercell storm. *Mon. Wea. Rev.*, **146**, 95–118, <https://doi.org/10.1175/MWR-D-17-0064.1>.
- Raynaud, L., B. Touzé, and P. Arbogast, 2018: Detection of severe weather events in a high-resolution ensemble prediction system using the extreme forecast index (EFI) and shift of tails (SOT). *Wea. Forecasting*, **33**, 901–908, <https://doi.org/10.1175/WAF-D-17-0183.1>.
- Sanchez, C., K. D. Williams, and M. Collins, 2016: Improved stochastic physics schemes for global weather and climate models. *Quart. J. Roy. Meteor. Soc.*, **142**, 147–159, <https://doi.org/10.1002/qj.2640>.
- Shen, X. S., J. J. Wang, Z. C. Li, et al., 2020: China's independent and innovative development of numerical weather prediction. *Acta Meteor. Sinica*, **78**, 451–476, <https://doi.org/10.11676/qxb2020.030>. (in Chinese)
- Shutts, G., 2005: A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quart. J. Roy. Meteor. Soc.*, **131**, 3079–3102, <https://doi.org/10.1256/qj.04.106>.
- Steinhoff, J., and D. Underhill, 1994: Modification of the Euler equations for “vorticity confinement”: Application to the computation of interacting vortex rings. *Phys. Fluids*, **6**, 2738–2744, <https://doi.org/10.1063/1.868164>.
- Stensrud, D. J., J. W. Bao, and T. T. Warner, 2000: Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Mon. Wea. Rev.*, **128**, 2077–2107, [https://doi.org/10.1175/1520-0493\(2000\)128<2077:UICAMP>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<2077:UICAMP>2.0.CO;2).
- Su, X., H. L. Yuan, Y. J. Zhu, et al., 2014: Evaluation of TIGGE ensemble predictions of Northern Hemisphere summer precipitation during 2008–2012. *J. Geophys. Res. Atmos.*, **119**, 7292–7310, <https://doi.org/10.1002/2014JD021733>.
- Tan, N., J. Chen, and H. Tian, 2013: Comparison between two global model stochastic perturbation schemes and analysis of perturbation propagation. *Meteor. Mon.*, **39**, 543–555. Available online at http://qxqk.nmc.cn/qx/ch/reader/view_abstract.aspx?flag=1&file_no=20130502&journal_id=qx. Accessed on 7 May 2025. (in Chinese)
- Tan, Y., and D. H. Chen, 2007: Meso-scale ensemble forecasts on physical perturbation using a non-hydrostatic model. *J. Appl. Meteor. Sci.*, **18**, 396–406, <https://doi.org/10.3969/j.issn.1001-7313.2007.03.017>. (in Chinese)
- Tian, H., G. Deng, J. K. Hu, et al., 2007: Introduction to the global T213 numerical ensemble forecast service system. *Meteor. Mon.*, **34**, 2658–2662. (in Chinese)
- Tian, W. H., and S. Y. Zhuang, 2008: Application of ETKF method to regional ensemble forecasts. *Meteor. Mon.*, **34**, 35–39, <https://doi.org/10.7519/j.issn.1000-0526.2008.08.005>. (in Chinese)
- Tompkins, A. M., and J. Berner, 2008: A stochastic convective approach to account for model uncertainty due to unresolved humidity variability. *J. Geophys. Res.*, **113**, D18101, <https://doi.org/10.1029/2007JD009284>.
- Torn, R. D., and G. J. Hakim, 2008: Performance characteristics of a pseudo-operational ensemble Kalman filter. *Mon. Wea. Rev.*, **136**, 3947–3963, <https://doi.org/10.1175/2008MWR2443.1>.
- 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](https://doi.org/10.1175/1520-0477(1993)074<2317:EFANTG>2.0.CO;2).
- Tsonevsky, I., C. A. Doswell III, and H. E. Brooks, 2018: Early warnings of severe convection using the ECMWF extreme forecast index. *Wea. Forecasting*, **33**, 857–871, <https://doi.org/10.1175/WAF-D-18-0030.1>.
- Wang, C. X., J. Q. Yao, and X. D. Liang, 2007: The comparing experiment of improving the operational ensemble prediction system for Shanghai regional precipitation. *J. Meteor. Sci.*, **27**, 481–487, <https://doi.org/10.3969/j.issn.1009-0827.2007.05.002>. (in Chinese)
- Wang, J. Y., J. Chen, L. Liu, et al., 2014: The sensitivity of the extreme precipitation forecast index on climatological cumulative probability distribution. *Torrential Rain Disaster*, **33**, 313–319, <https://doi.org/10.3969/j.issn.1004-9045.2014.04.002>. (in Chinese)
- Wang, J. Z., J. Chen, Z. R. Zhuang, et al., 2018: Characteristics of initial perturbation growth rate in the regional ensemble prediction system of GRAPES. *Chinese J. Atmos. Sci.*, **42**, 367–382, <https://doi.org/10.3878/j.issn.1006-9895.1708.17141>. (in Chinese)
- Wang, L., X. S. Shen, J. J. Liu, et al., 2020: Model uncertainty representation for a convection-allowing ensemble prediction system based on CNOP-P. *Adv. Atmos. Sci.*, **37**, 817–831, <https://doi.org/10.1007/s00376-020-9262-z>.
- Wang, Q. P., X. Pan, B. Y. Zhou, et al., 2023: Cosine analysis constraint scheme based on ETKF initial perturbations in the GRAPES regional ensemble prediction system. *Chinese J. Atmos. Sci.*, **47**, 1731–1745. Available online at <http://www.iap-journals.ac.cn/dqkx/en/article/doi/10.3878/j.issn.1006-9895.2210.22062>. Accessed on 7 May 2025. (in Chinese)
- 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](https://doi.org/10.1175/1520-0469(2003)060<1140:ACOBAE>2.0.CO;2).
- Wang, Y., M. Bellus, J. F. Geleyn, et al., 2014: A new method for generating initial condition perturbations in a regional ensemble prediction system: Blending. *Mon. Wea. Rev.*, **142**, 2043–2059, <https://doi.org/10.1175/MWR-D-12-00354.1>.
- 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*, **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*, **60**, 62–79, <https://doi.org/10.1111/j.1600-0870.2007.00273.x>.
- Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924, [https://doi.org/10.1175/1520-0493\(2002\)130<1913:EDAWPO>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2).

- Wilks, D. S., 2011: Chapter 8: Forecast verification. *Int. J. Geophys.*, **100**, 301–394, <https://doi.org/10.1016/B978-0-12-385022-5.00008-7>.
- Williams, R. M., C. A. T. Ferro, and F. Kwasniok, 2014: A comparison of ensemble post-processing methods for extreme events. *Quart. J. Roy. Meteor. Soc.*, **140**, 1112–1120, <https://doi.org/10.1002/qj.2198>.
- Wu, Z. Q., J. Zhang, J. Chen, et al., 2020: The study on the method of conditional typhoon vortex relocation for GRAPES regional ensemble prediction. *Acta Meteor. Sinica*, **78**, 163–176, <https://doi.org/10.11676/qxxb2020.027>. (in Chinese)
- Xia, F., and J. Chen, 2012: The research of extreme forecast index based on the T213 ensemble forecast and the experiment in predicting temperature. *Meteor. Mon.*, **38**, 1492–1501. Available online at <http://qxqk.nmc.cn/qxen/article/abstract/20121206>. Accessed on 7 May 2025. (in Chinese)
- Xia, Y., J. Chen, J. Du, et al., 2019: A unified scheme of stochastic physics and bias correction in an ensemble model to reduce both random and systematic errors. *Wea. Forecasting*, **34**, 1675–1691, <https://doi.org/10.1175/WAF-D-19-0032.1>.
- Xu, Z. Z., J. Chen, Y. Wang, et al., 2019: Sensitivity experiments of a stochastically perturbed parameterizations (SPP) scheme for mesoscale precipitation ensemble prediction. *Acta Meteor. Sinica*, **77**, 849–868, <https://doi.org/10.11676/qxxb2019.039>. (in Chinese)
- Xu, Z. Z., J. Chen, M. Mu, et al., 2022a: A nonlinear representation of model uncertainty in a convective-scale ensemble prediction system. *Adv. Atmos. Sci.*, **39**, 1432–1450, <https://doi.org/10.1007/s00376-022-1341-x>.
- Xu, Z. Z., J. Chen, M. Mu, et al., 2022b: A stochastic and non-linear representation of model uncertainty in a convective-scale ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **148**, 2507–2531, <https://doi.org/10.1002/qj.4322>.
- Yang, M., P. L. Yu, L. F. Zhang, et al., 2024: Predictability of the 7·20 extreme rainstorm in Zhengzhou in stochastic kinetic-energy backscatter ensembles. *Sci. China Earth Sci.*, **67**, 2226–2241, <https://doi.org/10.1007/s11430-023-1357-1>.
- Yang, X., K. Dai, and Y. J. Zhu, 2022: Progress and challenges of deep learning techniques in intelligent grid weather forecasting. *Acta Meteor. Sinica*, **80**, 649–667, <https://doi.org/10.11676/qxxb2022.051>. (in Chinese)
- Yang, X. S., D. H. Chen, T. B. Leng, et al., 2002: The comparison experiments of SV and LAF initial perturbation techniques used at the NMC ensemble prediction system. *J. Appl. Meteor. Sci.*, **13**, 62–66, <https://doi.org/10.3969/j.issn.1001-7313.2002.01.007>. (in Chinese)
- Yano, J.-I., M. Z. Ziemiański, M. Cullen, et al., 2018: Scientific challenges of convective-scale numerical weather prediction. *Bull. Amer. Meteor. Soc.*, **99**, 699–710, <https://doi.org/10.1175/BAMS-D-17-0125.1>.
- Ye, L., Y. Z. Liu, J. Chen, et al., 2020: A study on multi-scale singular vector initial perturbation method for ensemble prediction. *Acta Meteor. Sinica*, **78**, 648–664, <https://doi.org/10.11676/qxxb2020.042>. (in Chinese)
- Yuan, H. L., S. L. Mullen, X. G. Gao, et al., 2005: Verification of probabilistic quantitative precipitation forecasts over the southwest United States during winter 2002/03 by the RSM ensemble system. *Mon. Wea. Rev.*, **133**, 279–294, <https://doi.org/10.1175/MWR-2858.1>.
- Yuan, Y., X. L. Li, J. Chen, et al., 2016: Stochastic parameterization toward model uncertainty for the GRAPES mesoscale ensemble prediction system. *Meteor. Mon.*, **42**, 1161–1175. Available online at <http://qxqk.nmc.cn/html/2016/10/20161001.html>. Accessed on 7 May 2025. (in Chinese)
- Zhang, F. Q., N. F. Bei, R. Rotunno, et al., 2007: Mesoscale predictability of moist baroclinic waves: Convection-permitting experiments and multistage error growth dynamics. *J. Atmos. Sci.*, **64**, 3579–3594, <https://doi.org/10.1175/JAS4028.1>.
- Zhang, H., W. S. Duan, and Y. C. Zhang, 2023: 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, 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>.
- Zhang, H. B., X. F. Zhi, J. Chen, et al., 2017: Achievement of perturbation methods for regional ensemble forecast. *Trans. Atmos. Sci.*, **40**, 145–157, <https://doi.org/10.13878/j.cnki.dqkxb.20160405001>. (in Chinese)
- Zhang, X. Y., J. Z. Min, and T. J. Wu, 2020: A study of ensemble-sensitivity-based initial condition perturbation methods for convection-permitting ensemble forecasts. *Atmos. Res.*, **234**, 104741, <https://doi.org/10.1016/j.atmosres.2019.104741>.
- Zhang, Y. C., W. S. Duan, S. Vannitsem, et al., 2023: 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>.
- Zhao, X. H., and R. D. Torn, 2022: Evaluation of independent stochastically perturbed parameterization tendency (iSPPT) scheme on HWRF-based ensemble tropical cyclone intensity forecasts. *Mon. Wea. Rev.*, **150**, 2659–2674, <https://doi.org/10.1175/MWR-D-21-0303.1>.
- Zhi, X. F., H. X. Qi, Y. Q. Bai, et al., 2012: A comparison of three kinds of multimodel ensemble forecast techniques based on the TIGGE data. *Acta Meteor. Sinica*, **26**, 41–51, <https://doi.org/10.1007/S13351-012-0104-5>.
- Zhi, X. F., X. D. Ji, J. Zhang, et al., 2013: Multimodel ensemble forecasts of surface air temperature and precipitation using TIGGE datasets. *Trans. Atmos. Sci.*, **36**, 257–266, <https://doi.org/10.3969/j.issn.1674-7097.2013.03.001>. (in Chinese)
- Zhi, X. F., T. Wang, and Y. Ji, 2020: Multimodel ensemble forecasts of surface air temperature over China based on deep learning approach. *Trans. Atmos. Sci.*, **43**, 435–446, <https://doi.org/10.13878/j.cnki.dqkxxb.20200219003>. (in Chinese)
- Zhou, X. Q., Y. J. Zhu, D. C. Hou, et al., 2016: A comparison of perturbations from an ensemble transform and an ensemble Kalman filter for the NCEP global ensemble forecast system. *Wea. Forecasting*, **31**, 2057–2074, <https://doi.org/10.1175/WAF-D-16-0109.1>.
- Zhou, X. Q., Y. J. Zhu, D. C. Hou, et al., 2017: Performance of the new NCEP global ensemble forecast system in a parallel experiment. *Wea. Forecasting*, **32**, 1989–2004, <https://doi.org/10.1175/WAF-D-17-0023.1>.

- 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, S. P., X. F. Zhi, F. Ge, et al., 2021: Subseasonal forecast of surface air temperature using superensemble approaches: Experiments over Northeast Asia for 2018. *Wea. Forecasting*, **36**, 39–51, <https://doi.org/10.1175/WAF-D-20-0096.1>.
- Zhu, Y. J., Z. Toth, R. Wobus, et al., 2002: The economic value of ensemble-based weather forecasts. *Bull. Amer. Meteor. Soc.*, **83**, 73–84, [https://doi.org/10.1175/1520-0477\(2002\)083<007](https://doi.org/10.1175/1520-0477(2002)083<007)
- 3:TEVOEB>2.3.CO;2.
- Zhuang, X. R., J. Z. Min, T. J. Wu, et al., 2017: Development mechanism of multi-scale perturbation based on different perturbation methods in convection-allowing ensemble prediction. *Plateau Meteor.*, **36**, 811–825. (in Chinese)
- Zhuang, X. R., M. Xue, J. Z. Min, et al., 2021: Error growth dynamics within convection-allowing ensemble forecasts over central U.S. regions for days of active convection. *Mon. Wea. Rev.*, **149**, 959–977, <https://doi.org/10.1175/MWR-D-20-0329.1>.