Geoscientific
Model Development

Sensitivity studies of a four-dimensional local ensemble transform Kalman filter coupled with WRF-Chem version 3.9.1 for improving particulate matter simulation accuracy

Jianyu Lin[1,2], Tie Dai[2], Lifang Sheng[1], Weihang Zhang[1], Shangfei Hai[3], and Yawen Kong[4]

[1]College of Oceanic and Atmospheric Sciences, Ocean University of China, Qingdao, 266100, China
[2]State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics,
Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China
[3]CMA Earth System Modeling and Prediction Centre, China Meteorological Administration (CMA), Beijing 100081, China
[4]State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute,
Chinese Academy of Sciences, Beijing 100101, China

Correspondence: Tie Dai (daitie@mail.iap.ac.cn)

Abstract. Accurately simulating severe haze events through numerical models remains a challenge because of uncertainty in anthropogenic emissions and physical parameterizations of particulate matter ($PM_{2.5}$ and PM_{10}). In this study, a coupled Weather Research and Forecasting with Chemistry (WRF-Chem)–four-dimensional local ensemble transform Kalman filter (4D-LETKF) data assimilation system has been successfully developed to optimize particulate matter concentration by assimilating hourly ground-based observations in winter over the Beijing–Tianjin–Hebei (BTH) region and surrounding provinces. The effectiveness of the 4D-LETKF system and its sensitivity to the ensemble member size and length of the assimilation window are investigated. The promising results show that significant improvements have been made by analysis in the simulation of particulate matter during a severe haze event. The assimilation reduces root mean square errors in $PM_{2.5}$ from 69.93 to $31.19\,\mu g\,m^{-3}$ and of PM_{10} from 106.88 to $76.83\,\mu g\,m^{-3}$. Smaller root mean square errors and larger correlation coefficients in the analysis of $PM_{2.5}$ and PM_{10} are observed across nearly all verification stations, indicating that the 4D-LETKF assimilation optimizes the simulation of $PM_{2.5}$ and PM_{10} concentration efficiently. Sensitivity experiments reveal that the combination of 48 h of assimilation window length and 40 ensemble members shows the best performance for reproducing the severe haze event. In view of the performance

of ensemble members, an increasing ensemble member size improves ensemble spread among each forecasting member, facilitates the spread of state vectors about $PM_{2.5}$ and PM_{10} information in the first guess, favors the variances between each initial condition in the next assimilation cycle, and leads to better simulation performance in both severe and moderate haze events. This study advances our understanding of the selection of basic parameters in the 4D-LETKF assimilation system and the performance of ensemble simulations in a particulate-matter-polluted environment.

1 Introduction

Although great progress regarding air pollution control has been made during recent years, China is facing the highest levels of particulate matter in the world (van Donkelaar et al., 2016). Particulate matter consists of $PM_{2.5}$ and PM_{10}, referring to particles with aerodynamic diameters of less than 2.5 and $10\,\mu m$, respectively. A high concentration of particulate matter is a major factor for severe haze events (air quality index larger than 200) in the Beijing–Tianjin–Hebei (BTH) region of China, especially during winter (Yan et al., 2016; Zhang et al., 2018). Numerical models are considered to be useful tools for simulating haze events as they take complex physical and chemical mechanisms into account, but the un-

certainty in emissions and physical parameterizations still remains a significant barrier in improving the simulation accuracy (Gao et al., 2017; Feng et al., 2018).

As an effective statistical approach, data assimilation is capable of improving the accuracy of pollution simulations by limiting the performance of models. A lot of data assimilation approaches have been applied to atmospheric science, including three-dimensional variation (3D-Var) (Lorenc, 1986; Parrish and Derber, 1992; Sun et al., 2020), four-dimensional variation (4D-Var) (Huang et al., 2009; Benedetti et al., 2009), and ensemble Kalman filter algorithms and their variants (Evensen, 1994; Whitaker and Hamill, 2002; Miyazaki et al., 2012a). Among them, four-dimensional local ensemble transform Kalman filter (4D-LETKF) has shown unique characteristics in numerical simulation (Evensen, 2003; Kong et al., 2021). Firstly, derived from finite-forecasting members, the background error covariance matrix of 4D-LETKF features flow-dependent characteristics, and the linear combinations of ensemble members produce a global analysis (Hunt et al., 2007). Secondly, the computational time for 4D-LETKF remains robust as the observation numbers increase, exhibiting strong computational ability in the parallel architecture when assimilating various measurements (Miyoshi et al., 2007; Hunt et al., 2007; Dai et al., 2021). Lastly, 4D-LETKF can assimilate time slots of asynchronous observations to optimize the current state within the assimilation window, which efficiently improves the quality of pollution prediction (Evensen, 2003; Ott et al., 2004; Dai et al., 2019; Cheng et al., 2019).

The characteristics of 4D-LETKF underscore the importance of the ensemble member size and length of the assimilation window for its effectiveness. The background error covariance matrix, which represents the uncertainty in the ensemble simulations, is determined by the spread of the ensemble members (Peng et al., 2017). In general, 4D-LETKF considers approximate model trajectories by observing linear combinations of the background ensemble trajectories. However, limited numbers of ensemble members may bring about insufficient dispersion of ensemble systems (Hunt et al., 2004). Overall, the 4D-LETKF system can greatly improve the utilization rate of observations by constraining the state variables in an asynchronous hourly slot within the assimilation window. A longer assimilation window efficiently reduces the computational load by avoiding frequent switches between state and forecast variables. But the trajectories in a long assimilation window may diverge enough that linear combinations will not approximate the model trajectories. Moreover, the model ensemble trajectory may not fit the observations well over the entire interval with the presence of model errors (Dai et al., 2019). Many studies have discussed the choice of these two parameters for ensemble Kalman filter algorithms and their variants. When optimizing hourly aerosol fields by satellite observations, Cheng et al. (2019) revealed that the forecast with a 24 h assimilation window was comparable to those with 1 h; the root mean square error for aerosol optical depth (AOD) is 0.091 and 0.110, respectively, indicating the weights determined at the end of the 24 h assimilation window are valid to optimize the ensemble trajectories. However, Dai et al. (2019) proposed that over 80 % of the hourly assimilation efficiencies for the 1 h assimilation window are higher than those with 6 or 24 h in 4D-LETKF experiments, suggesting that assimilation efficiency decreases with the increase in the assimilation window interval. These different opinions reveal that there is still a large uncertainty in the selection of parameters in the 4D-LETKF assimilation system.

The accuracy simulation of severe haze events with air quality index (AQI) larger than 200 has been a challenging problem for a long time, posing severe threats to human daily life and public health (Wang et al., 2014; Kong et al., 2021; Gao et al., 2017). Although 4D-LETKF has unique advantages in computational efficiency and analysis, there is little research that investigates the impacts of 4D-LETKF assimilation on pollutant simulations, especially in severe haze events; in addition, it is also imperative to explore the basic optimal combination of assimilation parameters and its explanation in this method. Our major objectives are to not only evaluate the performance of 4D-LETKF in reproducing particulate matter concentration during a severe haze event, but also summarize the influence rules of ensemble size and assimilation window length on particulate matter simulation and explore whether these rules are applicable to a moderate haze event (air quality index smaller than 200) as well. The results have great significance for verifying and quantifying the effect of 4D-LETKF assimilation on numerical simulations of $PM_{2.5}$ and PM_{10}, subsequently providing a general rule for parameter selection in the 4D-LETKF during severe haze events. Herein, we utilize the 4D-LETKF system, which is coupled with the Weather Research and Forecasting with Chemistry (WRF-Chem) model to improve simulative skill of particulate matter among northern China during the winter of 2020. Section 2 briefly introduces detail setting of WRF-Chem model, 4D-LETKF, observations and numerical experiment designs. Section 3 compares the assimilation with those in the prior simulation, summarizes and explains sensitivity rules for parametric selection, and followed by a conclusion in Sect. 4 lastly.

2 Methodology

2.1 Configuration of the forecast model

In our implementation, the fully coupled online WRF-Chem version 3.9.1 is employed as a numeral forward model to describe the meteorological and chemical conditions simultaneously, in which it fully considers extensive chemical transport processes including advection, convection and sedimentation processes (Grell et al., 2005). The WRF-Chem model is configured with two domains (d01 and d02), both using

100 (west–east) $\times$ 100 (south–north) grid points but with horizontal resolutions of 30 and 10 km, respectively. As shown in Fig. 1a, the d01 domain covers most of east Asia, and the area under the blue shadow is the d02 domain. The vertical grid contains 40 full sigma levels, extending from the surface to 50 hPa.

The initial and lateral boundary conditions of meteorological fields are derived from the National Centers for Environmental Prediction Final (NCEP FNL) analysis data with a spatial resolution of $1° \times 1°$ and temporal interval of 6 h. A state-of-the-art and highly non-linear gas-phase chemical mechanism Regional Atmospheric Chemistry Mechanism (RACM) (Stockwell et al., 1997) is selected as the gas-phase mechanism, and Goddard Chemistry Aerosol Radiation and Transport (GOCART) (Schwartz et al., 2012) is adopted as the aerosol mechanism. The parameterization scheme used in research is shown in Table 1.

The anthropogenic emissions are obtained from the Multi-resolution Emission Inventory for China (MEIC, 2024) compiled by Tsinghua University. The inventory includes anthropogenic emissions from agriculture, industry, power, residential, and transportation sectors (Zheng et al., 2021). The inventory has a spatial resolution of $0.25° \times 0.25°$ and has been interpolated to match the simulation resolution. The biogenic emissions are calculated online by Guenther scheme (Guenther et al., 1995). The $PM_{2.5}$, PM_{10} concentrations output from WRF-Chem are linearly interpolated to site observations. The evaluation of uncertainty in the emission inventory has been shown in previous research (Zhang et al., 2009).

2.2 The 4D-LETKF algorithm and the state variables

The 4D-LETKF coupled with the WRF-Chem model is implemented to investigate the influence of assimilation on particulate matter simulation in this research. In this section, we introduce the 4D-LETKF algorithm and corresponding state variables briefly, a more detailed information can be found in Hunt et al. (2007). The LETKF features a flow-dependent covariance matrix from ensemble simulation and determines the analysis ensemble mean, $\overline{x}^a$ (a posteriori), according to the following formula:

$$\overline{x}^a = \overline{x}^b + \mathbf{X}^b \overline{w}^a, \tag{1}$$

where $\overline{x}^b$ and $\mathbf{X}^b$ denote the ensemble mean of the first guess and background ensemble perturbations, respectively. The ensemble perturbation matrix ($\mathbf{X}^b$) is calculated by $x^b(i) - \overline{x}^b \{i = 1, 2, \ldots, k\}$, where k represents the ensemble member size. The perturbation weight matrix $\overline{w}^a$ is the Kalman gain which linearly determines the increment between the analysis and the first guess, and can be calculated as follows:

$$\overline{w}^a = \tilde{P}^a \left(Y^b\right)^T \mathbf{R}^{-1} (y^0 - \overline{y}^b), \tag{2}$$

where $\tilde{P}^a$ is the analysis error covariance in ensemble space. y^0 and $\overline{y}^b$ denote the observations vector and en-

semble mean background observations, respectively. Ensemble mean background observations derived from applying the observation vector to the ensemble member state vector $\mathbf{H}$ ($\overline{x}^b$). The matrix $\mathbf{R}$ is the observation error covariance matrix. The matrix $\mathbf{Y}^b$ represents ensemble background observation perturbations, whose ith column is $y^b(i) - \overline{y}^b$, $\{i = 1, 2, \ldots, k\}$. $\tilde{P}^a$ can be obtained as follows:

$$\tilde{P}^a = [(k-1)\mathbf{I}/\rho + (\mathbf{Y}^b)^T \mathbf{R}^{-1} \mathbf{Y}^b]^{-1}, \tag{3}$$

where $\mathbf{I}$ denotes the identity matrix and k is ensemble member size. The multiplicative inflation factor ρ is necessary to avoid the filter divergence (Anderson and Anderson, 1999; Sekiyama et al., 2010), which is set to 1.1 to inflate the analysis covariance (Dai et al., 2019; Anderson, 2007). Analysis ensemble perturbation (X^a) is calculated by

$$X^a = \mathbf{X}^b \left[(k-1)\tilde{P}^a\right]^{1/2} = \mathbf{X}^b W^a. \tag{4}$$

Calculated by the sum of the $\overline{x}^a$ and each of the columns of X^a, the ensemble analyses are served as optimal initial conditions in each ensemble member to generate the first guess in the next cycle.

Figure 2 is the flow chart of the WRF-Chem–4D-LETKF assimilation system applied in our implementation. The system conducts these processes within each assimilation cycle. The 4D-LETKF generates a flow-dependent background error covariance matrix by ensemble member. Given that the emissions inventory is an important source of uncertainty in simulation (Pagowski and Grell, 2012), the research randomly perturbs anthropogenic emissions of PM, black carbon (BC), and organic carbon (OC) in January for each member to create the ensemble members, and the perturbation follows a log-normal distribution in the k-dimensional space. The mean values of perturbations of $PM_{2.5}$, PM_{10}, BC, and OC emissions are equal to 1, and the variances of these emissions are set according to corresponding uncertainty in MEIC inventory (130 %, 132 %, 208 %, and 258 % for $PM_{2.5}$, PM_{10}, BC, and OC, respectively) (Luo et al., 2023). Such ensemble anthropogenic emissions are perfectly correlated in spatial and temporal dimension and should not be regarded as overly restrictive (Schutgens et al., 2010; Dai et al., 2021). This study only adds perturbations one time into emissions at the first cycle of assimilation to provide the information spread of particulate matter. It is enough to make the most out of the site information affecting the first-guess field. The WRF-Chem–4D-LETKF system propagates the ensemble forward simulation for the entire assimilation window time and outputs the first-guess fields at each hourly time slot. The ensemble mean of the first guess ($\overline{x}^b$) and background ensemble perturbations ($\mathbf{X}^b$) can be obtained from the ensemble member here. Combining the observation and the observation operator, the innovation ($y^0 - \overline{y}^b$) and $\mathbf{Y}^b$ can be obtained in each time slot. The perturbation weight matrix ($\overline{w}^a$) is valid within a relative short assimilation window (e.g., 24 or 48 h) (Hunt et al., 2004; Cheng et al., 2019).

Table 1. WRF-Chem parameterization scheme in this study. RRTMG stands for Rapid Radiative Transfer Model for General Circulation Models, and YSU stands for Yonsei University.

Parameterization	WRF-Chem option
Microphysics	Morrison two-moment scheme (Morrison et al., 2009)
Longwave radiation	RRTMG longwave scheme (Iacono et al., 2008)
Shortwave radiation	RRTMG shortwave scheme (Iacono et al., 2008)
Planetary boundary layer	YSU scheme (Hong et al., 2006)
Cumulus parameterization	Grell 3D ensemble scheme (Grell, 1993)
Land surface model	Noah (Tewari et al., 2004)
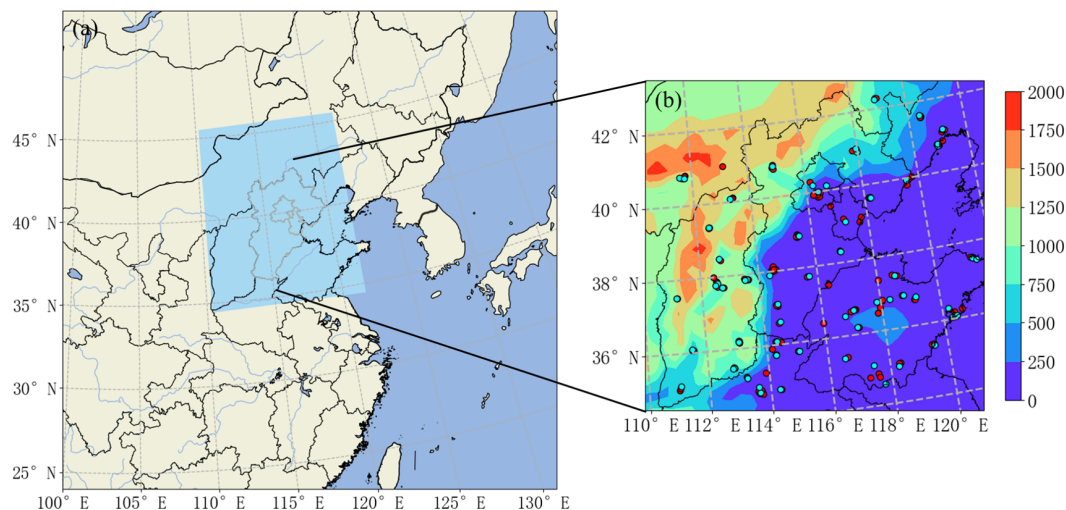
Figure 1. (a) WRF-Chem model domains. The grey border implies the BTH region. **(b)** Location of assimilated and independent verification observation sites with topography (m). The red and blue dots imply the assimilated and independent verification observation site, respectively. Publisher's remark: please note that the above figure contains disputed territories.

The analysis ensemble derived from $\overline{w}^a$ at the end of time slots will serve as chemical initial conditions for the next assimilation window. As the cycle of assimilation proceeds, a linear combination of the analysis ensemble is continuously obtained.

The ensemble Kalman filter generally encounters a spurious long-distance correlation problem because of the limited numbers of ensemble members (Miyazaki et al., 2012a). To avoid the problem above, it is necessary to apply observation localizations to filter observations from a long distance. Overall, 4D-LETKF offers a flexible choice of observation localizations in horizontal, vertical, and temporal dimensions for each grid point (Cheng et al., 2019). In this study, the horizontal localization factor is calculated as the Gaussian function (Miyoshi et al., 2007), which gradually reduces the effect of observations as the increasing departure from the analysis grid:

$$f(r) = \exp(-r^2/2\sigma^2). \tag{5}$$

Here, r represents the physical distance from the observation to the analysis grid, and σ represents localization length. We limit the localization factor from 0 to 3.65 times the localization length (Zhao et al., 2015), ignoring the observation beyond 3.65 times the localization length to the analysis grid.

The selection of the state variables depends on the generative mechanism of aerosol. As a result, 16 kinds of WRF-Chem–GOCART aerosol variables are treated as state variables. For the $PM_{2.5}$ observations, the observation operator is described as follows:

$$
\begin{aligned}
y^b_{PM_{2.5}} = \rho_d[&P_{2.5} + 1.375S + 1.8(OC_1 + OC_2) \\
&+ BC_1 + BC_2 + D_1 + 0.286D_2 + S_1 + 0.942S_2], \quad (6)
\end{aligned}
$$

where ρ_d represents the dry-air density; $P_{2.5}$ is the fine unspecified aerosol contribution; S represents sulfate; and OC_1 and OC_2 are hydrophobic and hydrophilic organic carbon, respectively.

BC_1 and BC_2 are hydrophobic and hydrophilic black carbon, D_1 and D_2 are dusts with effective radii of 0.5 and 1.4 μm, and S_1 and S_2 are sea salts with effective radii of 0.3 and 1.0 μm, respectively (Peng et al., 2018).
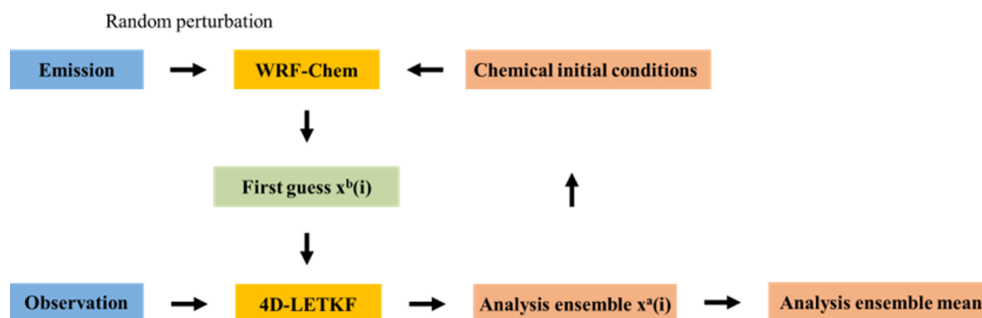
Figure 2. Flow chart of the WRF-Chem–4D-LETKF assimilation system for particulate matter.

Similarly, the observation operator for PM_{10} is shown as follows:

$$y^b_{PM_{10}} = \rho_d[P_{10} + P_{2.5} + 1.375S + 1.8(OC_1 + OC_2) \\ + BC_1 + BC_2 + D_1 + 0.286D_2 + D_3 + 0.87D_4 \\ + S_1 + 0.942S_2 + S_3], \quad (7)$$

where P_{10} is the coarse unspecified aerosol contribution and D_3 and D_4 are dusts with effective radii of 2.4 and 4.5 μm. S_3 is sea salt with an effective radius of 3.2 μm. Therefore, the simulated $PM_{10–2.5}$ is

$$y^b_{PM_{10–2.5}} = \rho_d[P_{10} + D_3 + 0.87D_4 + S_3]. \quad (8)$$

In this research, $y^o_{PM_{10–2.5}}$ calculated using $y^o_{PM_{10}} - y^o_{PM_{2.5}}$ is used to analyze state variables including D_5 and S_4, which are dust with effective radii of 8 μm and sea salt with effective radii of 7.5 μm, respectively.

2.3 Site observation data and errors

Ground-based observations feature a high temporal resolution, which can capture variation in pollution concentration on an hourly scale at the bottom of the troposphere, providing continuous and reliable observations. The quality-assured and quality-controlled hourly observation data of $PM_{2.5}$ and PM_{10} are used to explore the influence of 4D-LETKF assimilation in this research. The pollution data were obtained from the China National Environmental Monitoring Center (CNEMC, 2024). As the research primarily focuses on the BTH region, the assimilation and verification sites are mainly located in the BTH region and neighboring provinces, primarily in urban and suburban areas. In order to obtain more reliable observation data, the quality control of observation data in this study includes hourly observations of default value and extreme value detection. First, during the haze period, if the number of missing values for either type of pollutant at one site exceeds 24 h, this site is considered to have a certain uncertainty in observation quality, and data will not be assimilated. Second, for each kind of observation in different station, the hourly observations outside the range of $m \pm 3\sigma$, where the m and σ are the mean value and standard deviation of daily concentration, respectively, will not be assimilated. When selecting assimilation and verification sites, spatial distribution uniformity is ensured for better assimilation performance; consequently, those sites are randomly selected. Finally, 127 assimilation sites and 69 verification sites in the BTH region and surrounding provinces are selected (Fig. 1b). It can be seen that the assimilation and verification sites have a relatively uniform spatial distribution.

The observation error covariance matrix (**R**) is assumed to be diagonal, implying that observational errors among each pollution species are uncorrelated. The observation error (r) consists of the measurement error (ε_0) and the representation error (ε_r):

$$r = \sqrt{\varepsilon_0^2 + \varepsilon_r^2}. \quad (9)$$

The measurement error (ε_0) is defined as

$$\varepsilon_0 = \text{ermax} + 0.0075 \times \Pi_0, \quad (10)$$

where ermax is the base error, which is set to 1 for $PM_{2.5}$ and PM_{10} (Chen et al., 2019a), and Π_0 denotes the observation of concentration. Produced by the observation operator, the representation error can be calculated by the following formula (Elbern et al., 2007):

$$\varepsilon_r = \gamma \varepsilon_0 \sqrt{\Delta l/L}, \quad (11)$$

where γ is the tunable scaling factor set to 0.5, Δl is the spatial resolution of gridding (30 and 10 km for d01 and d02, respectively), and L depends on station location, which denotes the range that an observation can reflect; here, L is 2 km for this calculation.

Meteorological data were collected from the National Climatic Data Center (NCDC, 2024), which provides the hourly air temperature, dew point, and wind speed data. The observational meteorological data are used to validate the performance of the simulations in this study.

2.4 Experiment design

A series of control and data assimilation experiments during severe and moderate haze events, as listed in Table 2, have

been carried out to achieve our major objective. The control experiments refer to numerical experiments without data assimilation. The Severe-FR experiment with a 48 h spin-up time is performed firstly to quantify the necessity of adjusting particulate matter concentration during a severe haze event. Severe-FR-24h, Severe-FR-48h, and Severe-FR-72h are accompanied by a restart every 24, 48, and 72 h, respectively, and update meteorological boundary conditions. Except Severe-FR, the rest of the experiments all have 24 h of spin-up time at the beginning of each restart or assimilation cycle. The settings of the assimilation experiment cycle time are the same as in corresponding control experiments. The detailed descriptions of experiment cycle time are shown in Fig. S1 in the Supplement. Since the effectiveness of 4D-LETKF is highly related to the ensemble member size and length of the assimilation window (Rubin et al., 2016), the sensitivity analysis is employed to investigate the influence from two parameters on assimilation effect (Kong et al., 2023). The selection of assimilation parameters for the sensitivity experiments includes 20, 40, and 60 for ensemble members and 24, 48, and 72 h for the length of the assimilation window empirically (Kong et al., 2021; Dai et al., 2021). All sensitivity experiments use identical WRF-Chem physical parameterizations, anthropogenic emissions, and random perturbations. Through the comparison between all assimilation experiments, the influence rules of 4D-LETKF assimilation have on the simulation of particulate matter in severe haze can be retrieved. Lastly, aiming to determine the applicable range of obtained influence rules above, two assimilation experiments in a moderate haze event are performed to validate whether the rules are also suitable for a less polluted environment. The reasons for the selection of parameters are described in detail in the next section.

The root mean square error (RMSE), mean error (bias), mean absolute error (MAE), and correlation coefficient are calculated in this study to evaluate the performance of each numerical experiment. The assimilation efficiency (AE) for estimating the data assimilation performance is also calculated by the following formulation (Yumimoto and Takemura, 2011):

$$AE = \frac{RMSE^f - RMSE^a}{RMSE^f} \times 100\%, \tag{12}$$

where $RMSE^a$ and $RMSE^f$ are RMSE with and without assimilation, respectively. According to the definition, if AE is positive, it means that RMSE has decreased due to the assimilation effect. When AE is equal to 1, RMSE in analysis completely disappears, and the analysis is equal to observations. The formulations of the correlation coefficient and root mean square error are shown in the Supplement.

3 Results

3.1 Comparison of the analysis with the control experiment

3.1.1 The reproduction of a severe haze event in BTH

It is essential to discuss the basic evolution of pollutants and the necessity of pollutant data assimilation in a severe haze event before conducting the assimilation experiments. The severe haze event selected in this study occurred from 00:00 15 January 2020 to 00:00 21 January 2020 (UTC). The AQI is a comprehensive indicator of overall air pollution and a criterion for severe haze events (Zhan et al., 2018; Bao et al., 2015). As a result, Fig. 3a shows the temporal variation in AQI at the six sites in the BTH region during the investigated period. The peak value of AQI mainly appeared on 18 January and then rapidly decreased on 19 and 20 January. The temporal averages of AQI have exceeded 200, with particulate matter identified as the primary pollutant. Figure 3b provides the correlation coefficients and standardized standard deviations of five parameters from Severe-FR shown against observations. Meteorological variables, including air temperature, dew point temperature, and wind speed, are well simulated when compared with $PM_{2.5}$ and PM_{10}. The correlation coefficients of meteorological factors are all larger than 0.6, while those of pollutant concentrations are all below 0.4. Therefore, when the meteorological conditions can be retrieved relatively accurately, particulate matter assimilation is the key to improving the simulative skill of pollutants.

3.1.2 The improvement in the reproduction of severe haze achieved by 4D-LETKF

The divergence between the assimilation and control experiments reflects the contribution from 4D-LETKF initial condition adjustment. Consequently, the study takes an ensemble member size of 40 and assimilation window length of 48 h to conduct the sensitivity experiment and compare it with Severe-FR-48h, which has the same integration time in each cycle to validate the effectiveness of the 4D-LETKF assimilation system (the analysis from the selection of 40 ensemble members and 48 h of the assimilation window length is presented here because it shows the best performance among sensitivity experiments in the next section). Figure 4 reveals the performance of control and assimilation experiments in the severe haze event. The RMSE values of $PM_{2.5}$ and PM_{10} in Severe-FR-48h are 69.93 and 106.88 µg m^{-3}, and both have a scattered distribution, indicating substantial uncertainty exists in reproducing this severe haze event. In Severe-40m-48h, the RMSE values of $PM_{2.5}$ and PM_{10} are 31.19 and 76.83 µg m^{-3}, decreasing by 55.40 % and 28.12 %, respectively, in a high-particulate-matter-concentration environment. The decreased RMSE values also imply that the assimilation system has reached a well-calibrated stage. Not

Table 2. Design of the numerical experiments in this research.

|	Experiment	Design of simulation
Control experiments	Severe-FR	Free-run experiment in a severe haze event and without restart in integration process.
	Severe-FR-24h Severe-FR-48h Severe-FR-72h	Free-run experiment in a severe haze event and with a restart every 24, 48 and 72 h, providing deterministic simulation corresponding to data assimilation experiment.
	Moderate-FR-48h	Free-run experiment in moderate haze events and with a restart every 48 h, providing deterministic simulation corresponding to data assimilation experiment.
Data assimilation experiments in a severe haze event	Severe-20m-24h Severe-20m-48h Severe-20m-72h	Assimilation experiment in a severe haze event with 20 ensemble members and 24, 48, and 72 h assimilation window length, respectively.
	Severe-40m-24h Severe-40m-48h Severe-40m-72h	Assimilation experiment in a severe haze event with 40 ensemble members and 24, 48, and 72 h assimilation window length, respectively.
	Severe-60m-24h Severe-60m-48h Severe-60m-72h	Assimilation experiment in a severe haze event with 60 ensemble members and 24, 48, and 72 h assimilation window length, respectively.
Data assimilation experiments in a moderate haze event	Moderate-20m-48h Moderate-40m-48h	Assimilation experiment in a moderate haze event with 20 and 40 ensemble members combine with 48 h assimilation window length.
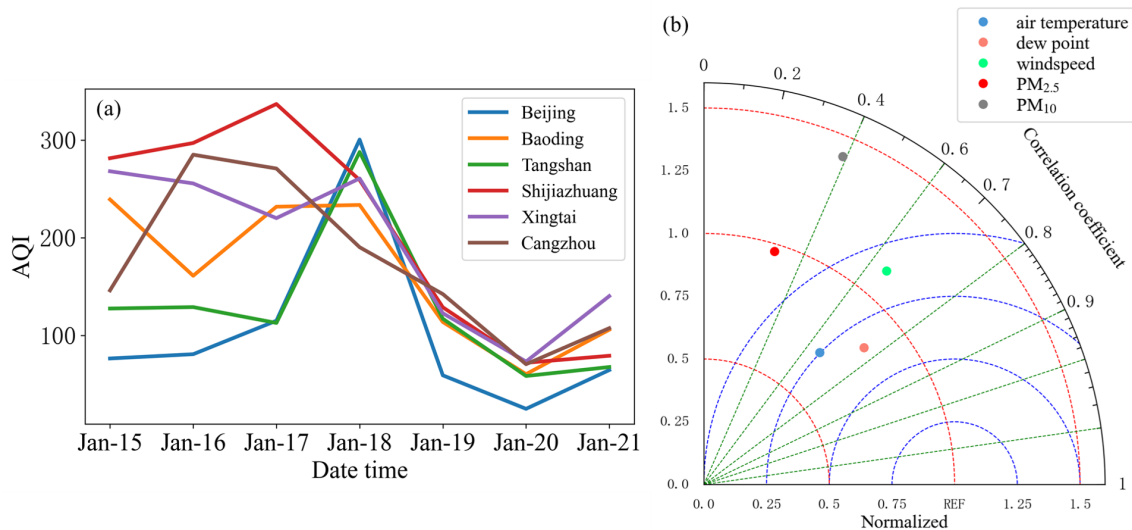
Figure 3. (a) Temporal variation about AQI at six sites in the severe haze event. **(b)** A Taylor graph describing the simulation from Severe-FR with five kinds of parameters compared with the observed ones in the BTH region.

only are more points grouped together, but smaller simulation errors for $PM_{2.5}$ and PM_{10} also imply that Severe-40m-48h outperforms Severe-FR-48h in this severe haze event.

In order to acquire the base distribution of simulation errors for particulate matter, Fig. 5 presents the frequency distribution of deviations between observed and simulated particulate matter concentrations in Severe-FR-48 and Severe-

40m-48h experiments. It is obvious that Severe-40m-48h increases the frequency of low deviations and decreases those of high deviations in the simulation of $PM_{2.5}$. The deviation pattern of $PM_{2.5}$ in Severe-40m-48h is generally more squeezed together, with higher peaks, and symmetrical to the value of 0 than Severe-FR-48h. For the deviation distribution pattern of PM_{10}, it shows a high frequency of negative devia-
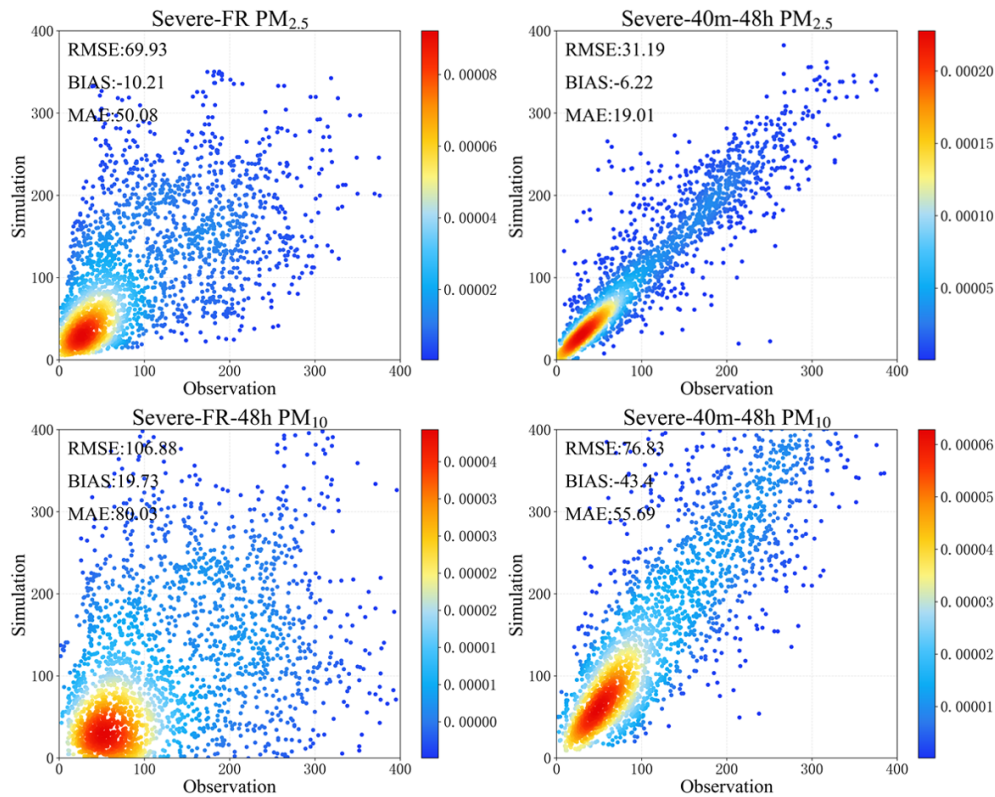
Figure 4. Scatterplot and density plot of $PM_{2.5}$ and PM_{10} in Severe-FR-48h and Severe-40m-48h versus observations from verification stations ($\mu g\,m^{-3}$). The color bar represents the Gaussian kernel density estimation.

tions and great underestimation in the Severe-FR-48h, and this underestimation has been effectively corrected by the adjustment of initial conditions and step analysis in Severe-40m-48h. In particular, the proportion of deviation within $20\,\mu g\,m^{-3}$ in the Severe-40m-48h is 69.98 % for $PM_{2.5}$ and 31.90 % for PM_{10}.

Figure 6 exhibits the spatial distribution of four statistical parameters about RMSE for particulate matter in the BTH region. By comparing from the Severe-FR-48h and Severe-40m-48h, it is clear there is a significant RMSE reduction for $PM_{2.5}$ after assimilation, implying that the actual evolution of $PM_{2.5}$ can be better represented by Severe-40m-48h. For instance, the RMSE values of $PM_{2.5}$ in Baoding, Hengshui, and Cangzhou have significantly decreased to 29.85, 18.98, and $19.06\,\mu g\,m^{-3}$, respectively, compared to 80.55, 55.22, and $76.32\,\mu g\,m^{-3}$ in Severe-FR-48h. AE in most verification stations exceeding 50 % also suggests the high efficiency of 4D-LETKF assimilation for the simulation of $PM_{2.5}$. Although the performance of the assimilation experiment in Shijiazhuang does not have a good agreement with observation and shows a positive difference, high values of AE in most verification stations also prove the validation of the assimilating effect for PM_{10}. Compared to Severe-FR-48h, Severe-40m-48h productively reduces the RMSE of PM_{10}, accompanied by high values of 61.18 %, 59.17 %, and

52.18 % of AE on Zhangjiakou, Tangshan, and Hengshui, respectively.

The spatial distribution of correlation coefficients from Severe-40m-48h, Severe-FR-48h, and their difference for $PM_{2.5}$ and PM_{10} is also illustrated in Fig. 7. The assimilation experiment increases the correlation coefficients to more than 0.6 at all sites in the simulations of $PM_{2.5}$ and exceeds 0.7 in the southern BTH region in the simulations of PM_{10}. Severe-40m-48h also reverses the opposite trend of $PM_{2.5}$ and PM_{10} series in Severe-FR-48h versus observations; for example, the correlation coefficients in Severe-FR-48h at Chengde and Zhangjiakou are −0.42 and −0.53, but they increase to 0.52 and 0.69 after assimilation in the simulations of PM_{10}. Incorporating more assimilable observations may further increase the correlation coefficient in the simulation of particulate matter (Kong et al., 2021). Data assimilation by multiple observations from diverse platforms is necessary because it can integrate and coordinate observational information into aerosol forecasts them well and then improve air pollutant forecast accuracy (Barbu et al., 2009; Ma et al., 2020).

The temporal variations in particulate matter from Severe-40m-48h and Severe-FR-48h and observations from six independent verification stations are shown in Figs. S2 and S3. The six independent verification stations have experienced different levels of air pollution and distributed uniformly over
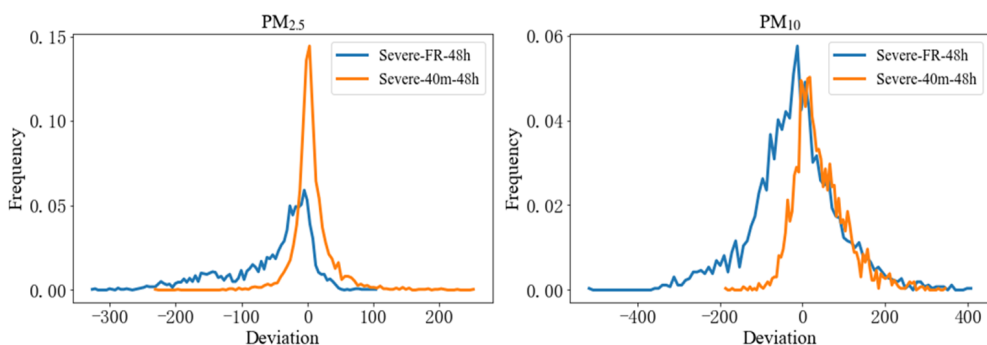
Figure 5. Frequency distribution of the deviations about the simulated $PM_{2.5}$ and PM_{10} concentrations in Severe-40m-48h and Severe-FR-48h with the observed ones subtracted (deviation is in $\mu g\,m^{-3}$).
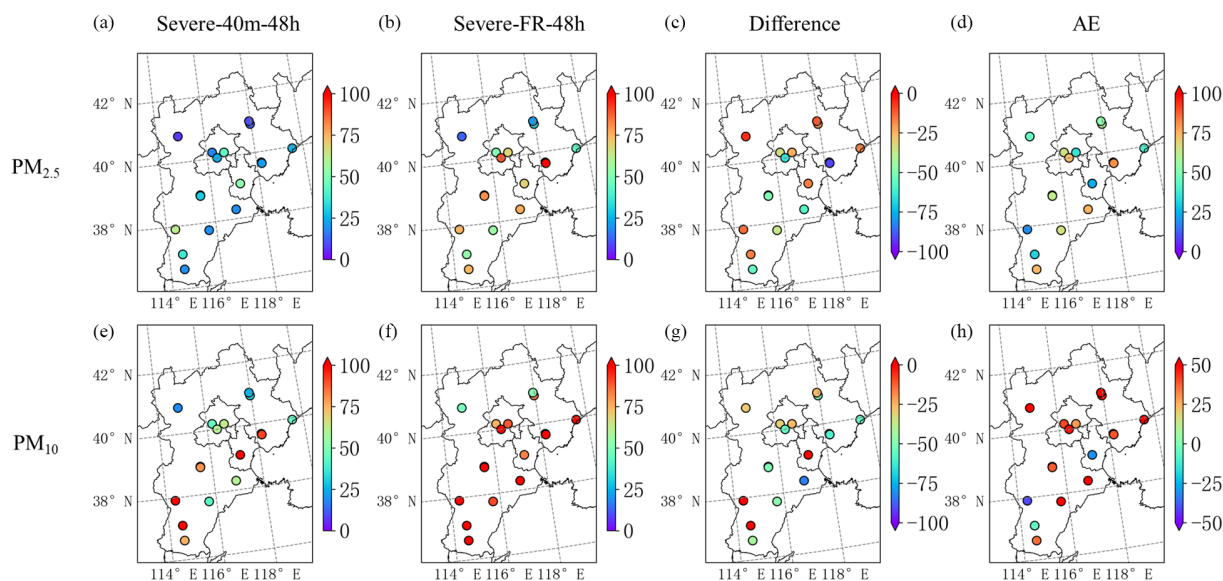

Figure 6. Spatial distribution of RMSE values from Severe-40m-48h **(a, e)**, Severe-FR-48h **(b, f)**, and their difference **(c, g)** and AE **(d, h)** for $PM_{2.5}$ **(a–d)** and PM_{10} **(e–h)** from 15 to 21 January among verification stations in the BTH region. The difference implies the RMSE in Severe-40m-48h without those in Severe-FR-48h ($\mu g\,m^{-3}$). AE is assimilation efficiency and has been described in the Methodology section. The grey border denotes the BTH region.

the BTH region. It is apparent that the analysis at six stations have good agreement with observations for both $PM_{2.5}$ and PM_{10}, which can better characterize the peaks and valleys of particulate matter concentration over investigated period.

Table 3 lists the ΔRMSE, ΔCORR, and AE in the simulations of particulate matter at independent stations outside the BTH region. The RMSEs and correlation coefficients have decreased and increased, respectively, after assimilating ground-based observations, suggesting that the uncertainty in Severe-FR-48h has been well optimized in not only the BTH region, but also the whole simulation domain. The main source of these gains is generated from local initial field assimilation. Compared to Severe-FR-48h, the analysis in Yuncheng shows that the RMSE values of $PM_{2.5}$ and PM_{10} have decreased remarkably by 98.26 and 144.56 $\mu g\,m^{-3}$, and such a great improvement may be related to the enhanced es-

timation capability about state variables of particulate matter. The high values of AE also suggest that verification observation sites outside the BTH region have achieved a good Kalman gain. In previous research, predicting heavy haze events in northern China, especially over the BTH Region, remained a challenge when compared to other regions like Pearl River Delta and Yangtze River Delta in China (Feng et al., 2018; Gao et al., 2017). In this research, the analysis is propagated by meteorological elements, including temperature, air pressure, and wind fields, which come from NCEP Final analysis data and may provide an optimal meteorological boundary conditions for the assimilation of pollutant concentrations.
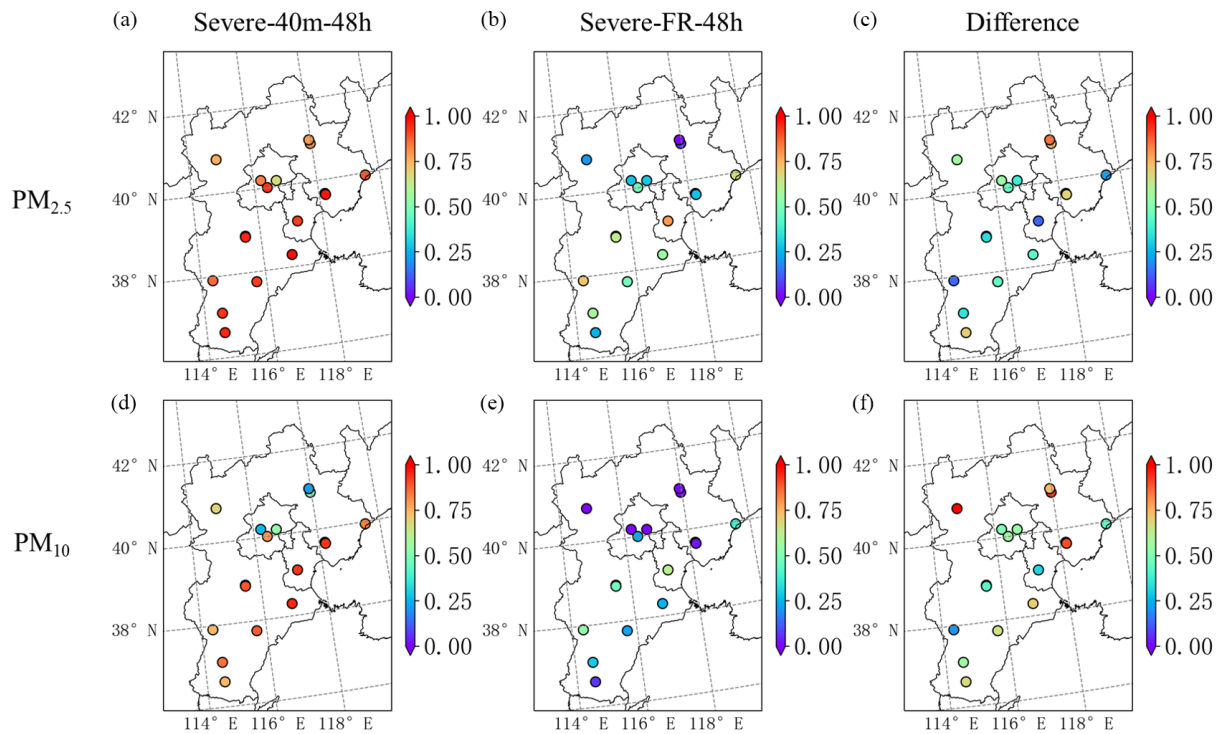
Figure 7. Spatial distribution of correlation coefficients from Severe-40m-48h **(a, d)**, Severe-FR-48h **(b, e)**, and their difference **(c, f)** for $PM_{2.5}$ (first row) and PM_{10} (second row) from 15 to 21 January among verification stations in the BTH region. The difference implies the correlation coefficient in Severe-40m-48h without those in Severe-FR-48h. The grey border denotes the BTH region.

Table 3. Statistics about $PM_{2.5}$ and PM_{10} from analyses in the cities among neighboring provinces of the BTH region. ΔRMSE (ΔCORR) represents the RMSE (correlation coefficient) from the analysis with those from Severe-FR-48h subtracted (ΔRMSE is in $\mu g\,m^{-3}$).

City/statistical variable	$PM_{2.5}$			PM_{10}		
	ΔRMSE	ΔCORR	AE	ΔRMSE	ΔCORR	AE
Taiyuan	−21.30	+0.41	23.93 %	−54.8	+0.71	39.05 %
Changzhi	−38.39	+0.38	65.93 %	−63.06	+0.68	63.22 %
Jincheng	−37.94	+0.37	66.85 %	−94.32	+0.89	72.45 %
Shuozhou	−27.07	+0.31	58.96 %	−100.08	+0.84	69.96 %
Yuncheng	−98.26	+0.67	77.85 %	−144.56	+1.25	80.64 %
Hohhot	−92.30	+0.67	74.53 %	−121.79	+1.41	68.92 %
Chifeng	−16.90	+0.55	64.95 %	−38.18	+0.95	60.85 %
Huludao	−38.56	+0.20	59.11 %	−95.95	+1.04	65.76 %
Jinzhou	−42.97	+0.21	61.17 %	−46.26	+0.83	45.83 %
Chaoyang	−39.37	+0.37	51.14 %	−83.04	+1.23	61.86 %
Jinan	−44.90	+0.63	69.71 %	−71.22	+0.59	62.93 %
Qingdao	−23.99	+0.27	37.98 %	−72.05	+0.44	72.21 %
Shouguang	−28.70	+0.21	58.03 %	−48.92	+0.39	47.57 %
Anyang	−35.87	+0.41	62.53 %	−32.08	+0.60	33.15 %
Zhengzhou	−26.26	+0.37	37.51 %	−3.64	+0.42	3.73 %

3.2 The sensitivity of 4D-LETKF to ensemble member size and length of the assimilation window

In the previous section, the performance from the assimilation experiment with 40 ensemble members and 48 h of assimilation window length is compared against those which do not integrate hourly pollutant observations. The results fully demonstrate the ability of the 4D-LETKF assimilation method to reproduce severe haze events in spatial and temporal dimensions. However, the 4D-LETKF assimilation effect highly relies on the selection of the ensemble member size and length of the assimilation window; so how is the assimilation approach different from the parameterized selection in the severe haze event? It is of great meaning to conduct sensitivity experiments based on the ensemble member size and length of the assimilation window, compare each of their performances according to statistical metrics, and summarize the general influence rule of the 4D-LETKF parameter selection. Consequently, nine panels of sensitivity experiments are conducted with the selection of the ensemble member size (20, 40, and 60 members) and the length of the assimilation window (24, 48, and 72 h) to maximize the positive innovation in this section.

Figure 8 reveals the heatmap of RMSE in each sensitivity experiment of particulate matter over verification sites among the BTH region. The results of the free-run experiment with different integration times (24, 48, and 72 h) are offered here for comparison with analysis which have the same assimilation cycle time. The RMSEs of $PM_{2.5}$ and PM_{10} in each free-run experiment exceed 60 and $100\,\mu g\,m^{-3}$, respectively. It is apparent that 4D-LETKF performs better than the FR experiment in the simulation focusing on $PM_{2.5}$ and PM_{10} over a wide range of ensemble member sizes and assimilation window lengths, illustrating the broad applicability of 4D-LETKF data assimilation to these parameters. However, it can be found that the analysis of $PM_{2.5}$ and PM_{10} is dependent on the length of the assimilation window and dramatically related to ensemble member size in all sensitivity experiments. Unlike the short-lived and chemical reactive species (such as SO_2 and NO_2) which easily undergo complex and non-linear photochemical reactions, a relatively longer assimilation window length seems more suitable for assimilating ground-based particulate matter observations (Peng et al., 2017; Kong et al., 2021). A longer assimilation window length could also avoid the underestimation of model spread, which implies overconfidence in the first-guess state estimate by enough integration time for each member (Schutgens et al., 2010; Miyazaki et al., 2012a; Hunt et al., 2007). Hence, 48 or 72 h assimilation windows are advised to optimize the ensemble concentration trajectories. On the other hand, increasing ensemble member size efficiently reduces uncertainty in $PM_{2.5}$ and PM_{10}, as evidenced by the decrease in RMSEs from free-run to assimilation experiments with 20 and 40 members. However, when compared with the results from 40 ensemble members, the accuracy of numerical simulations does not significantly improve for $PM_{2.5}$ and PM_{10} with 60 ensemble members either, indicating that 40 members are sufficient and feasible to provide a reliable estimation of the background error and analysis rather than more numerical source consumption. Considering numerical source consumption and RMSE values in the simulations of $PM_{2.5}$ and PM_{10}, Severe-40m-48h is more comparable to the observations when compared with the other eight panels of sensitivity experiments.

3.3 The influence of ensemble member size on the ensemble spread

In order to explore why increasing the ensemble member size can efficiently reduce the uncertainty in the analysis of $PM_{2.5}$ and PM_{10}, as revealed in Fig. 8, this study investigates the spatial distribution of the standard deviations of $PM_{2.5}$ and $PM_{10-2.5}$ of the first guess ($x^b(i)$) and analysis field ($x^a(i)$) in terms of ensemble members. The standard deviations of ensemble members describe how the emission perturbation propagates among the forward model, and this perturbation is driven by the underlying surface pollution emission inputs and meteorological conditions. Therefore, the standard deviation in the first-guess fields quantifies the dispersion degree of the ensemble background; substantially impacts the calculation of assimilation parameters, such as ensemble state vector perturbations; and further affects the performance of particulate matter predictions.

Since the RMSE decreases with increasing ensemble member size when 20 and 40 members are in the setting, and a 48 h assimilation window corresponds to a smaller RMSE, this study compares the spatial distribution of ensemble standard deviations from Severe-20m-48h and Severe-40m-48h to explain the relationship between the ensemble member size and simulation errors in the analysis result. Figure 9 depicts contour maps of the spatial distribution of temporal averaged standard deviations in the first guess and analysis of Severe-40m-48h, Severe-20m-48h, and their difference for $PM_{2.5}$ and $PM_{10-2.5}$ during the severe haze event. The first guess in Severe-40m-48h and Severe-20m-48h shows that the relatively high standard deviations are generally observed in the southern BTH region, while those in the northern areas are close to zero for both $PM_{2.5}$ and $PM_{10-2.5}$. High-value centers are distributed in densely populated areas and urban centers including Shijiazhuang, Xingtai, Tianjin, and Tangshan, where the standard deviations generally exceed $30\,\mu g\,m^{-3}$. Combined with Fig. S4, it can be seen that the areas with large concentration standard deviations correspond well to the spatial distribution of anthropogenic emissions and the areas with large standard deviations of emission sources. The standard deviations of concentrations of $PM_{2.5}$ and $PM_{10-2.5}$ have a close relationship with the allocation and configuration of anthropogenic emission sources because disturbances are only added to emission sources for each ensemble member without disturbing
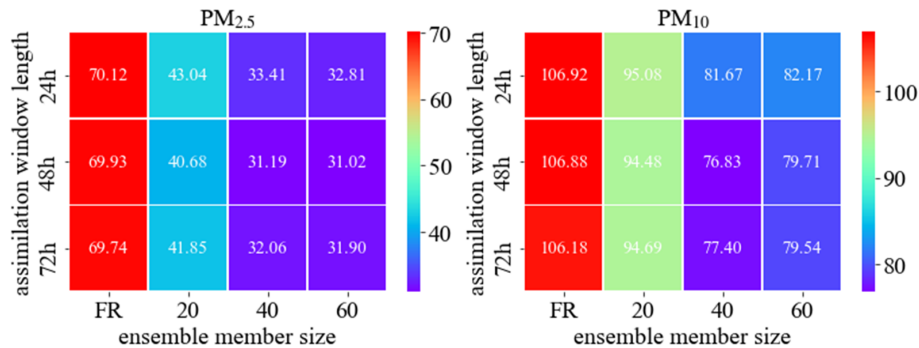
Figure 8. Heatmap showing the RMSE in each sensitivity experiment of particulate matter over verification sites ($\mu g\,m^{-3}$). The number in each small square represents the RMSE between the observation and simulation for each combination of ensemble member sizes and the lengths of assimilation window methods.

the meteorological field in this haze event. The variation in the difference in the third column entirely comes from increasing the ensemble member size. The positive difference between Severe-40m-48h and Severe-20m-48h in the first guess suggests that increasing the ensemble member size leads to greater differences among each ensemble for both $PM_{2.5}$ and $PM_{10-2.5}$ over BTH areas. The high efficiency of 4D-LETKF is strongly influenced by sufficient information spread among ensemble members, which integrate spreading out observational information to produce the analysis from the first guess (Rubin et al., 2016). As a result, the increasing ensemble member size improves divergence for each member and facilitates the state vectors of $PM_{2.5}$ and $PM_{10-2.5}$ information spread in the first guess, which means that Severe-40m-48h performs better than Severe-20m-48h in this severe haze event. The standard deviations of $PM_{2.5}$ in the analysis are generally lower than those in the first guess. Due to the localization of 4D-LETKF, that is, the ground-based observation data only optimized for the simulation grid within a certain range, square-like areas of low standard deviations appear in the analysis of $PM_{2.5}$ for both 40 and 20 ensemble members. Nearly all assimilated stations are located at the center of low-value square areas, suggesting that 4D-LETKF tunes all $PM_{2.5}$ trajectories to a small range with a low standard deviation at each slot of analysis by the assimilation of ground-based observations. For $PM_{10-2.5}$, there are no square-like areas of low standard deviations in the analysis for both 40 and 20 ensemble members, indicating that 4D-LETKF does not have an obvious limitation for $PM_{10-2.5}$ trajectories; however, the decreased standard deviations effect from 4D-LETKF is still distinct for the particulate matter because PM_{10} consists of $PM_{2.5}$ and $PM_{10-2.5}$ in the simulation. Enlarging the ensemble member size helps improve the standard deviations of $PM_{2.5}$ and $PM_{10-2.5}$ in the analysis, while the improving magnitude of $PM_{2.5}$ is obviously smaller than $PM_{10-2.5}$. The assimilation results are not directly influenced by the increased standard deviations in analysis. Such a low increase in standard deviations (generally below $3\,\mu g\,m^{-3}$)

is unlikely to induce uncertainty in the fitting and averaging process but facilitates divergence in initial conditions between forecasting members in the next assimilation cycle. In addition, Fig. S5 depicts the spatial distribution of the standard deviation from Severe-60m-48h, Severe-20m-48h, and their difference in the first-guess and analysis fields. It can be seen that increasing the number of ensemble members generally also improves the standard deviation in the first guess and analysis over the BTH region for both $PM_{2.5}$ and $PM_{10-2.5}$. Overall, the increase in standard deviations generated by the increasing ensemble member size directly improves the information spread of ensemble members in the first-guess field and the assimilation effect of 4D-LETKF, while the positive difference in the standard deviation in the analysis favors the variances between each initial condition in the next assimilation window during the severe haze event.

On the other hand, for both the first guess from Severe-40m-48h and Severe-20m-48h, the high standard deviations are found near the Shijiazhuang region in Fig. 9, but Shijiazhuang station ($37.91°$ N, $114.35°$ E) still has a larger RMSE and smaller AE, as shown in Fig. 6. This seems contrary to the opinion that increasing standard deviations in the first-guess field is beneficial to raising the accuracy of pollutant simulations. Therefore, Shijiazhuang station and the stations with high values of AE (exceed 50) and differences in the standard deviation in the first guess (exceed $1\,\mu g\,m^{-3}$), including Beijing, Tangshan, Handan, Baoding, Cangzhou, and Hengshui regions, are selected to explore the temporal distribution of standard deviation differences between 40 and 20 ensemble members so as to further advance our understanding of the relationship between the ensemble member size and simulation uncertainty in the 4D-LETKF system. The geographical locations of observations are shown in Fig. S6 of the Supplement. Figure 10 examines the temporal distribution of the standard deviation difference for $PM_{2.5}$ and $PM_{10-2.5}$ during the investigated period at Shijiazhuang station and results averaged from the selected stations. From 17 to 18 January, the standard deviation difference in the first
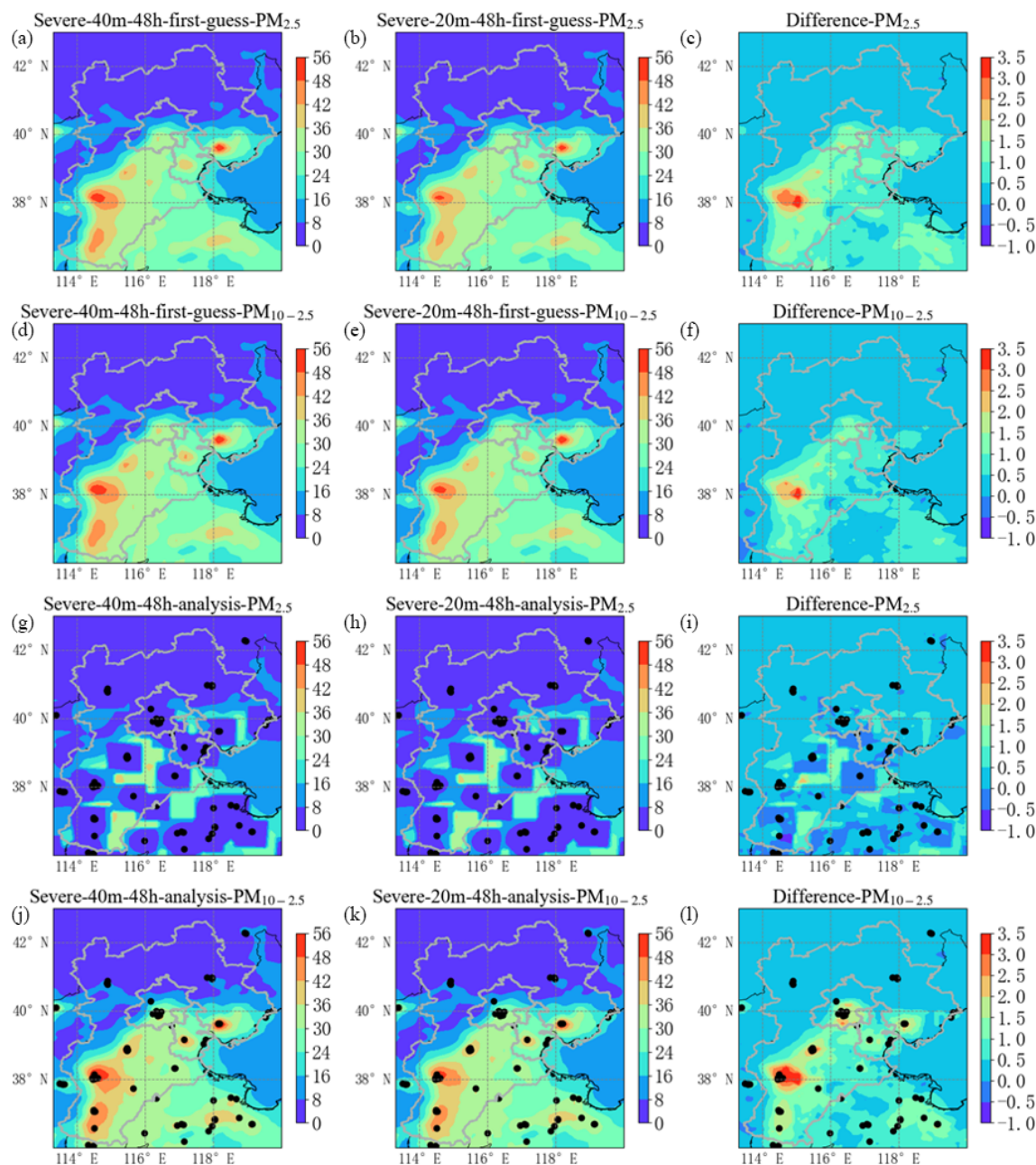
Figure 9. Contour maps of spatial distributions of temporal averaged $PM_{2.5}$ and $PM_{10-2.5}$ standard deviations in the first guess **(a–f)** and analysis **(g–l)** of Severe-40m-48h, Severe-20m-48h, and their difference (Severe-40m-48h without Severe-20m-48h) within the simulation period ($\mu g\,m^{-3}$). The black dots in the analysis of $PM_{2.5}$ and $PM_{10-2.5}$ show the location of assimilated stations. The grey border denotes the BTH region.

guess at Shijiazhuang station increased drastically and exceeded $10\,\mu g\,m^{-3}$ for both $PM_{2.5}$ and $PM_{10-2.5}$. This uneven temporal distribution results in a large standard deviation difference of the first guess in Fig. 9. This huge divergency between ensemble members may be attributed to the peak pollutant levels with AQI exceeding 300 at Shijiazhuang station occurring on 17 January, as shown in Fig. 3. In highly polluted environments, 40 forecasting members with differ-

ent perturbations in emission sources are more likely to have different concentrations of particulate matter in first-guess fields. An excessively large spread of $PM_{2.5}$ and $PM_{10-2.5}$ for ensemble members may cause an overly high estimation of background error variance and obtain a poor Kalman gain. Moreover, it can be found that the standard deviation difference in $PM_{2.5}$ and $PM_{10-2.5}$ at Shijiazhuang station is obviously lower than the average from selected stations, except

for the high dispersion time, suggesting the increasing number of ensemble members has limited impact on the spread between each ensemble member at Shijiazhuang during these dates. Standard deviations that are too low imply filter divergence near Shijiazhuang station, which may induce the underestimation of model spread, reduce the effect of observation information, and make the system more certain of the state estimate of particulate matter concentrations in the first guess (Hunt et al., 2007). In addition, reducing uncertainty in the mixed anthropogenic emission inventory may be an important approach to avoid filter convergence near the Shijiazhuang region. Generally edited by empirical and statistical data such as anthropogenic emission factors and activity dataset, the anthropogenic emissions based on the bottom-up method can hardly capture the real spatiotemporal distribution of anthropogenic emissions over China as there are frequent variations in energy consumption, even in the latest version. In the southern BTH region, great positive innovations in particulate matter emissions in posterior estimation have been introduced in previous research, implying that the update of underestimated emissions in this region may enlarge ensemble spread since a large quantity of emissions corresponds to a higher degree of perturbation (Peng et al., 2017; Feng et al., 2023). In a word, the perturbations added to emissions and meteorological fields need to be executed carefully in the 4D-LETKF system to avoid too high or too low of an ensemble spread degree which determines how analysis results perform against observation information and first-guess fields (Dai et al., 2021).

The results above suggest that the increasing ensemble member size enlarges ensemble spread, benefits the information spread in the first guess, and finally improves the simulation skill in the severe haze event. However, it has not been determined whether these influence rules are also practical for a more common and less polluted condition. Therefore, two assimilation experiments in a moderate haze event, Moderate-20m-48h and Moderate-40m-48h, are performed to examine the applicable range. As shown in Fig. S7, the moderate haze event spans from 00:00 15 January 2019 to 00:00 21 January 2021 (UTC). This moderate event began on 15 January, with AQI increasing until 18 January, reaching a moderate level but not lasting for a long time, and then decreasing on 19 and 20 January. Most areas experienced mild or moderate air pollution, with AQI generally below 200; the primary pollutant was particulate matter after calculation. The simulations of the moderate haze event utilize the same anthropogenic emission inventory as used in the severe haze event since the two events both happen in January, thereby avoiding the additional influence introduction from emission source variation and the perturbations to the information spread and assimilation effect.

Figure S8 shows the simulated concentrations of $PM_{2.5}$ and PM_{10} against ground-based observations during the moderate air pollution event. The RMSEs of $PM_{2.5}$ in Moderate-FR-48h, Moderate-20m-48h, and Moderate-40m-48h are 40.40, 24.12, and 18.52 $\mu g\,m^{-3}$, respectively, and the RMSEs of PM_{10} are 73.47, 67.81, and 57.04 $\mu g\,m^{-3}$, respectively. The concentrations of $PM_{2.5}$ and PM_{10} in assimilation experiments are more in agreement with observations, suggesting the validation of 4D-LETKF initial condition adjustment in the moderate haze event. The phenomenon that the simulation error in $PM_{2.5}$ and PM_{10} decreases with an increasing ensemble member size is the same as those characteristics have shown in the severe haze event before.

Similar to Fig. 9, Fig. S9 presents the spatial distributions of standard deviations of $PM_{2.5}$ and PM_{10} in the first guess of Moderate-40m-48h, Moderate-20m-48h, and their difference. The relatively smaller magnitude of the standard deviation difference in the first guess may relate to relatively low $PM_{2.5}$ and PM_{10} concentrations in the moderate haze event. A positive difference in the first guess and analysis for particulate matter implies the Moderate-40m-48h obtains a higher diversity of ensemble members than Moderate-20m-48h which are also similar with those happening in the severe haze event.

4 Summary

The numerical reproduction of severe haze events with an AQI larger than 200 has been a challenging problem in the field of atmospheric pollution for a long time. In this research, a WRF-Chem–4D-LETKF coupled data assimilation system has been successfully developed by an ensemble member with perturbed anthropogenic emissions to improve the simulative skill of particulate matter in the severe haze event during the winter of 2020. The research validates the effectiveness of 4D-LETKF data assimilation, discusses the optimal parameter combination of the ensemble member size and length of the assimilation window for the 4D-LETKF assimilation system, and summarizes and explains the influence rules from parametric selection to the 4D-LETKF assimilation effect during the severe and moderate haze events.

It is concluded that the Severe-40m-48h experiment shows the best performance in the simulations of $PM_{2.5}$ and PM_{10} after comparing the statistical errors and computing resource consumption across multiple sensitivity analyses, with RMSEs of 31.19 and 76.83 $\mu g\,m^{-3}$ for $PM_{2.5}$ and PM_{10} in the severe haze event. Severe-40m-48h optimizes the underestimation of particulate matter concentrations in Severe-FR-48h and remarkably improves the simulation accuracy in the entire BTH region and neighboring provinces. For example, the RMSEs of $PM_{2.5}$ in Baoding, Hengshui, and Cangzhou decrease to 29.85, 18.98, and 19.06 $\mu g\,m^{-3}$, respectively, from 80.55, 55.22, and 76.32 $\mu g\,m^{-3}$ in Severe-FR-48h. Severe-40m-48h is also capable of retrieving the peaks and valleys of particulate matter concentration over the investigated period. To examine the dependence of the assimilation effect of 4D-LETKF, nine panels of sensitivity tests were conducted according to the ensemble member size and length of the as-
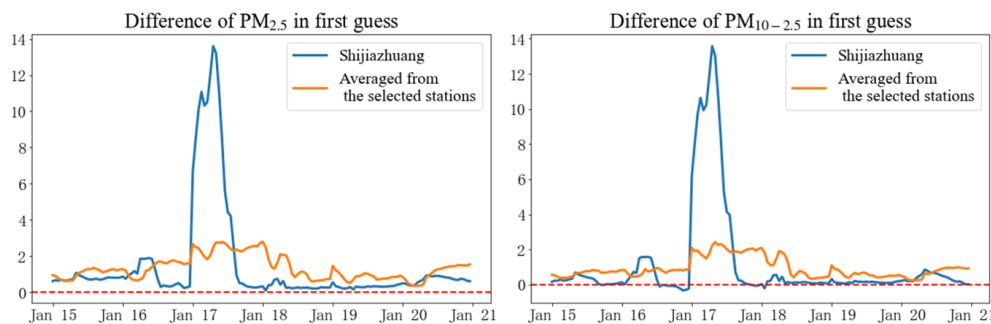
Figure 10. Temporal distribution of standard deviation difference (Severe-40m-48h without Severe-20m-48h) in the first guess for $PM_{2.5}$ and $PM_{10-2.5}$ at Shijiazhuang station, averaged from the selected stations (μg m^{-3}). The dashed red line represents a value of zero.

simulation window. The findings suggest that the simulation accuracy of $PM_{2.5}$ and PM_{10} can be strongly improved by increasing the ensemble member size from 20 to 40. A relatively longer assimilation window length such as 48 or 72 h combined with an ensemble member size of 40 is advised in the 4D-LETKF assimilation system. In view of performance of the ensemble members, increasing ensemble member size improves ensemble spread among each forecasting member, facilitates the spread of state vectors of $PM_{2.5}$ and PM_{10} information in the first guess, favors the variances between each initial condition in the next assimilation window, and leads to better performance in the simulation of the severe haze event. A similar conclusion can also be drawn from the moderate haze event, suggesting that this influence rule is applicable in both severe and moderate haze conditions.

There are still some deficiencies in this research. Although we have performed data quality control in this study, we do not use approaches such as super observations to improve the correspondence between grid points and observations (Jin et al., 2022; Miyazaki et al., 2012a), which may increase the representational error and result in the possibility of two stations with different concentrations interpolating on the same grid. Improving the spatial resolution of the forward model or introducing super observations may mitigate this problem (Miyazaki et al., 2012b; Feng et al., 2020b). Furthermore, the concentration of state variables of particulate matters in initial conditions is optimized in this study, but there still remain large uncertainties in anthropogenic emission data, which is important chemical boundary input for pollutant simulations. These uncertainty sources may play a significant role in the over- or underestimation of pollutant ensemble modeling. The anthropogenic emission inversion based on an ensemble Kalman filter and their variants is recognized as an effective approach to reducing uncertainty in anthropogenic emission sources (Peng et al., 2018; Feng et al., 2020a; Chen et al., 2019b). The jointly adjusted initial conditions and emission source with 4D-LETKF should be the focus of future work to further improve the forecast skills regarding air pollutants during heavy-pollution events.

References

Anderson, J. L.: An adaptive covariance inflation error correction algorithm for ensemble filters, Tellus A, 59, 210–224, https://doi.org/10.1111/j.1600-0870.2006.00216.x, 2007.

Anderson, J. L. and Anderson, S. L.: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts, Mon. Weather Rev., 127, 2741–2758, 1999.

Bao, J., Yang, X., Zhao, Z., Wang, Z., Yu, C., and Li, X.: The spatial-temporal characteristics of air pollution in China from 2001–2014, Int. J. Environ. Res. Publ. Health, 12, 15875e15887, https://doi.org/10.3390/ijerph121215029, 2015.

Barbu, A. L., Segers, A. J., Schaap, M., Heemink, A. W., and Builtjes, P. J. H.: A multi-component data assimilation experiment directed to sulphur dioxide and sulphate over Europe, Atmos. Environ., 43, 1622–1631, https://doi.org/10.1016/j.atmosenv.2008.12.005, 2009.

Benedetti, A., Morcrette, J., Boucher, O., Dethof, A., Engelen, R. J., Fisher, M., Flentje, H., Huneeus, N., Jones, L., Kaiser, J. W., Kinne, S., Mangold, A., Razinger, M., Simmons, A. J., and Suttie, M.: Aerosol analysis and forecast in the European Centre for Medium-Range Weather Forecasts Integrated Forecast System: 2. Data assimilation, J. Geophys. Res., 114, D13205, https://doi.org/10.1029/2008JD011115, 2009.

Cheng, Y., Dai, T., Goto, D., Schutgens, N. A. J., Shi, G., and Nakajima, T.: Investigating the assimilation of CALIPSO global aerosol vertical observations using a four-dimensional ensemble Kalman filter, Atmos. Chem. Phys., 19, 13445–13467, https://doi.org/10.5194/acp-19-13445-2019, 2019.

Chen, D., Liu, Z., Ban, J., and Chen, M.: The 2015 and 2016 wintertime air pollution in China: SO$_2$ emission changes derived from a WRF-Chem/EnKF coupled data assimilation system , Atmos. Chem. Phys., 19, 8619–8650, https://doi.org/10.5194/acp-19-8619-2019, 2019a.

Chen, D., Liu, Z., Ban, J., Zhao, P., and Chen, M.: Retrospective analysis of 2015–2017 wintertime PM$_{2.5}$ in China: response to emission regulations and the role of meteorology, Atmos. Chem. Phys., 19, 7409–7427, https://doi.org/10.5194/acp-19-7409-2019, 2019b.

CNEMC: China National Environmental Monitoring Center, https://www.cnemc.cn (last access: 31 December 2024), 2024.

Dai, T., Cheng, Y., Suzuki, K., Goto, D., Kikuchi, M., Schutgens, N. A. J., Yoshida, M., Zhang, P., Husi, L., Shi, G., and Nakajima, T.: Hourly Aerosol Assimilation of Himawari-8 AOT Using the Four-Dimensional Local Ensemble Transform Kalman Filter, J. Adv. Model. Earth Sy., 11, 680–711, https://doi.org/10.1029/2018MS001475, 2019.

Dai, T., Cheng, Y., Goto, D., Li, Y., Tang, X., Shi, G., and Nakajima, T.: Revealing the sulfur dioxide emission reductions in China by assimilating surface observations in WRF-Chem, Atmos. Chem. Phys., 21, 4357–4379, https://doi.org/10.5194/acp-21-4357-2021, 2021.

Elbern, H., Strunk, A., Schmidt, H., and Talagrand, O.: Emission rate and chemical state estimation by 4-dimensional variational inversion, Atmos. Chem. Phys., 7, 3749–3769, https://doi.org/10.5194/acp-7-3749-2007, 2007.

Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, J. Geophys. Res.-Oceans, 99, 10143–10162, https://doi.org/10.1029/94JC00572, 1994.

Evensen, G.: The Ensemble Kalman Filter: theoretical formulation and practical implementation, Ocean Dynam., 53, 343–367, https://doi.org/10.1007/s10236-003-0036-9, 2003.

Feng, S., Jiang, F., Jiang, Z., Wang, H., Cai, Z., and Zhang, L.: Impact of 3DVAR assimilation of surface PM$_{2.5}$ observations on PM$_{2.5}$ forecasts over China during wintertime, Atmos. Environ., 187, 34–49, https://doi.org/10.1016/j.atmosenv.2018.05.049, 2018.

Feng, S., Jiang, F., Wang, H., Wang, H., Ju, W., Shen, Y., Zheng, Y., Wu, Z., and Ding, A.: NO$_x$ emission changes over China during the COVID-19 epidemic inferred from surface NO$_2$ observations, Geophys. Res. Lett., 47, e2020GL090080, https://doi.org/10.1029/2020GL090080, 2020a.

Feng, S., Jiang, F., Wu, Z., Wang, H., Ju, W., and Wang, H.: CO emissions inferred from surface CO observations over China in December 2013 and 2017, J. Geophys. Res.-Atmos., 125, e2019JD031808, https://doi.org/10.1029/2019JD031808, 2020b.

Feng, S., Jiang, F., Wu, Z., Wang, H., He, W., Shen, Y., Zhang, L., Zheng, Y., Lou, C., Jiang, Z., and Ju, W.: A Regional multi-Air Pollutant Assimilation System (RAPAS v1.0) for emission estimates: system development and application, Geosci. Model Dev., 16, 5949–5977, https://doi.org/10.5194/gmd-16-5949-2023, 2023.

Gao, M., Saide, P. E., Xin, J., Wang, Y., Liu, Z., Wang, Y., Wang, Z., Pagowski, M., Guttikunda, S. K., and Carmichael, G. R.: Estimates of Health Impacts and Radiative Forcing in Winter Haze in Eastern China through Constraints of Surface PM$_{2.5}$ Predictions, Environ. Sci. Technol., 51, 2178–2185, https://doi.org/10.1021/acs.est.6b03745, 2017.

Grell, G., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock, W. C., and Eder, B.: Fully coupled "online" chemistry within the WRF model, Atmos. Environ, 39, 6957–6975, https://doi.org/10.1016/j.atmosenv.2005.04.027, 2005.

Grell, G. A.: Prognostic Evaluation of Assumptions Used by Cumulus Parameterizations, Mon. Weather Rev., 121, 764–787, https://doi.org/10.1175/1520-0493(1993)121<0764:PEOAUB>2.0.CO;2, 1993.

Guenther, A., Hewitt, C. N., Erickson, D., Fall, R., Geron, C., Graedel, T., Harley, P., Klinger, L., Lerdau, M., Mckay, W. A., Pierce, T., Scholes, B., Steinbrecher, R., Tallamraju, R., Taylor, J., and Zimmerman, P.: A global model of natural volatile organic compound emissions, J. Geophys. Res.-Atmos., 100, 8873–8892, https://doi.org/10.1029/94JD02950, 1995.

Hong, S., Noh, Y., and Dudhia, J.: A new vertical diffusion package with an explicit treatment of entrainment processes, Mon. Weather Rev., 134, 2318–2341, https://doi.org/10.1175/MWR3199.1, 2006.

Huang, X. Y., Xiao, Q., Barker, D. M., Zhang, X., Michalakes, J., Huang, W., Henderson, T., Bray, J., Chen, Y., Ma, Z., Dudhia, J., Guo, Y., Zhang, X., Won, D., Lin, H., and Kuo, Y.: Four-dimensional variational data assimilation for WRF: Formulation and preliminary results, Mon. Weather Rev., 137, 299–314, https://doi.org/10.1175/2008MWR2577.1, 2009.

Hunt, B. R., Kalnay, E., Kostelich, E. J., Ott, E., Patil, D. J., Sauer, T., Szunyogh, I., Yorke, J. A., and Zimin, A. V.: Four-dimensional ensemble Kalman filtering, Tellus A, 56, 273–277, https://doi.org/10.3402/tellusa.v56i4.14424, 2004.

Hunt, B. R., Kostelich, E. J., and Szunyogh, I.: Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter, Physica D, 230, 112–126, https://doi.org/10.1016/j.physd.2006.11.008, 2007.

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., and Collins, W. D.: Radiative forcing by long–lived greenhouse gases: Calculations with the AER radiative transfer models, J. Geophys. Res., 113, D13103, https://doi.org/10.1029/2008JD009944, 2008.

Jin, J., Pang, M., Segers, A., Han, W., Fang, L., Li, B., Feng, H., Lin, H. X., and Liao, H.: Inverse modeling of the 2021 spring super dust storms in East Asia, Atmos. Chem. Phys., 22, 6393–6410, https://doi.org/10.5194/acp-22-6393-2022, 2022.

Kong, L., Tang, X., Zhu, J., Wang, Z., Sun, Y., Fu, P., Gao, M., Wu, H., Lu, M., Wu, Q., Huang, S., Sui, W., Li, J., Pan, X., Wu, L., Akimoto, H., and Carmichael, G. R.: Unbalanced emission reductions of different species and sectors in China during COVID-19 lockdown derived by multi-species surface observation assimilation, Atmos. Chem. Phys., 23, 6217–6240, https://doi.org/10.5194/acp-23-6217-2023, 2023.

Kong, Y., Sheng, L., Li, Y., Zhang, W., Zhou, Y., Wang, W., and Zhao, Y.: Improving $PM_{2.5}$ forecast during haze episodes over China based on a coupled 4D-LETKF and WRF-Chem system, Atmos. Res., 249, 105366, https://doi.org/10.1016/j.atmosres.2020.105366, 2021.

Lin, J.: LETKF-WRF-Chem, Zenodo [code, data set], https://doi.org/10.5281/zenodo.14010521, 2024.

Lin, J.: Observation, Zenodo [data set], https://doi.org/10.5281/zenodo.14835237, 2025.

Lorenc, A. C.: Analysis methods for numerical weather prediction, Q. J. Roy. Meteor. Soc., 112, 1177–1194, https://doi.org/10.1002/qj.49711247414, 1986.

Luo, X., Tang, X., Wang, H., Kong, L., Wu, H., Wang, W., SONG, Y., Luo, H., Wang, Y., Zhu, J., and Wang, Z.: Investigating the Changes in Air Pollutant Emissions over the Beijing-Tianjin-Hebei Region in February from 2014 to 2019 through an Inverse Emission Method, Adv. Atmos. Sci. 40, 601–618, https://doi.org/10.1007/s00376-022-2039-9, 2023.

Ma, C., Wang, T., Jiang, Z., Wu, H., Zhao, M., Zhuang, B., Li, S., Xie, M., Li, M., Liu, J., and Wu, R.: Importance of bias correction in data assimilation of multiple observations over eastern China using WRF-Chem/DART, J. Geophys. Res.-Atmos., 125, e2019JD031465, https://doi.org/10.1029/2019JD031465, 2020.

MEIC: Multi-resolution Emission Inventory for China, http://www.meicmodel.org/ (last access: 31 December 2024), 2024.

Miyoshi, T., Yamane, S., and Enomoto, T.: Localizing the Error Covariance by Physical Distances within a Local Ensemble Transform Kalman Filter (LETKF), Scient. Online Lett. Atmos., 3, 89–92, https://doi.org/10.2151/sola.2007-023, 2007.

Morrison, H., Thompson, G., and Tatarskii, V.: Impact of Cloud Microphysics on the Development of Trailing Stratiform Precipitation in a Simulated Squall Line: Comparison of One– and Two–Moment Schemes, Mon. Weather Rev., 137, 991–1007, https://doi.org/10.1175/2008MWR2556.1, 2009.

Miyazaki, K., Eskes, H. J., and Sudo, K.: Global NO_x emission estimates derived from an assimilation of OMI tropospheric NO_2 columns, Atmos. Chem. Phys., 12, 2263–2288, https://doi.org/10.5194/acp-12-2263-2012, 2012a.

Miyazaki, K., Eskes, H. J., Sudo, K., Takigawa, M., van Weele, M., and Boersma, K. F.: Simultaneous assimilation of satellite NO_2, O_3, CO, and HNO_3 data for the analysis of tropospheric chemical composition and emissions, Atmos. Chem. Phys., 12, 9545–9579, https://doi.org/10.5194/acp-12-9545-2012, 2012b.

Miyoshi, T.: LETKF source codes, GitHub, https://github.com/takemasa-miyoshi/letkf, last access: 1 January 2024.

NCDC: National Climatic Data Center, https://www.ncei.noaa.gov (last access: 31 December 2024), 2024.

Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E. J., Corazza, M., Kalnay, E., Patil, D. J., and Yorke, J. A.: A local ensemble Kalman filter for atmospheric data assimilation, Tellus A, 56, 415–428, https://doi.org/10.1111/j.1600-0870.2004.00076.x, 2004.

Pagowski, M. and Grell, G. A.: Experiments with the assimilation of fine aerosols using an ensemble Kalman filter, J. Geophys. Res., 117, D21302, https://doi.org/10.1029/2012JD018333, 2012.

Parrish, D. F. and Derber, J. C.: The National Meteorological Center's Spectral Statistical-Interpolation Analysis System, Mon. Weather Rev., 120, 1747–1763, https://doi.org/10.1175/1520-0493(1992)120<1747:TNMCSS>2.0.CO;2, 1992.

Peng, Z., Liu, Z., Chen, D., and Ban, J.: Improving $PM_{2.5}$ forecast over China by the joint adjustment of initial conditions and source emissions with an ensemble Kalman filter, Atmos. Chem. Phys., 17, 4837–4855, https://doi.org/10.5194/acp-17-4837-2017, 2017.

Peng, Z., Lei, L., Liu, Z., Sun, J., Ding, A., Ban, J., Chen, D., Kou, X., and Chu, K.: The impact of multi-species surface chemical observation assimilation on air quality forecasts in China, Atmos. Chem. Phys., 18, 17387–17404, https://doi.org/10.5194/acp-18-17387-2018, 2018.

Rubin, J. I., Reid, J. S., Hansen, J. A., Anderson, J. L., Collins, N., Hoar, T. J., Hogan, T., Lynch, P., McLay, J., Reynolds, C. A., Sessions, W. R., Westphal, D. L., and Zhang, J.: Development of the Ensemble Navy Aerosol Analysis Prediction System (ENAAPS) and its application of the Data Assimilation Research Testbed (DART) in support of aerosol forecasting, Atmos. Chem. Phys., 16, 3927–3951, https://doi.org/10.5194/acp-16-3927-2016, 2016.

Schutgens, N. A. J., Miyoshi, T., Takemura, T., and Nakajima, T.: Sensitivity tests for an ensemble Kalman filter for aerosol assimilation, Atmos. Chem. Phys., 10, 6583–6600, https://doi.org/10.5194/acp-10-6583-2010, 2010.

Sekiyama, T. T., Tanaka, T. Y., Shimizu, A., and Miyoshi, T.: Data assimilation of CALIPSO aerosol observations, Atmos. Chem. Phys., 10, 39–49, https://doi.org/10.5194/acp-10-39-2010, 2010.

Sun, W., Liu, Z., Chen, D., Zhao, P., and Chen, M.: Development and application of the WRFDA-Chem three-dimensional variational (3DVAR) system: aiming to improve air quality forecasting and diagnose model deficiencies, Atmos. Chem. Phys., 20, 9311–9329, https://doi.org/10.5194/acp-20-9311-2020, 2020.

Schwartz, C. S., Liu, Z., Lin, H. C., and McKeen, S. A.: Simultaneous three-dimensional variational assimilation of surface fine particulate matter and MODIS aerosol optical depth, J. Geophys. Res., 117, D13202, https://doi.org/10.1029/2011JD017383, 2012.

Stockwell, W. R., Kirchner, F., Kuhn, M., and Seefeld, S.: A new mechanism for regional atmospheric chem-

istry modeling, J. Geophys. Res., 102, 25847–25879, https://doi.org/10.1029/97JD00849, 1997.

Tewari, M., Chen, F., Wang, W., Dudhia, J., LeMone, M. A., Mitchell, K., Ek, M., Gayno, G., Wegiel, J., and Cuenca, R. H.: Implementation and verification of the unified NOAH land surface model in the WRF model, 20th conference on weather analysis and forecasting/16th conference on numerical weather prediction, 11–15, 12–16 January 2004, Seattle, WA, USA, https://ams.confex.com/ams/84Annual/techprogram/paper_69061.htm (last access: 4 April 2025), 2004.

van Donkelaar, A., Martin, R. V., Brauer, M., Hsu, N. C., Kahn, R. A., Levy, R. C., Lyapustin, A., Sayer, A. M., and Winker, D. M.: Global estimates of fine particulate matter using a combined geophysical-statistical method with information from satellites, models, and monitors, Environ. Sci. Technol, 50, 3762–3772, https://doi.org/10.1021/acs.est.5b05833, 2016.

Wang, Z., Li, J., Wang, Z., Yang, W., Tang, X., Ge, B., Yan, P., Zhu, L., Chen, X., and Chen, H.: Modeling study of regional severe hazes over mid-eastern China in January 2013 and its implications on pollution prevention and control, Sci. China-Earth Sci., 57, 3–13, https://doi.org/10.1007/s11430-013-4793-0, 2014.

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

Yan, S., Cao, H., Chen, Y., Wu, C., and Fan, H.: Spatial and temporal characteristics of air quality and air pollutants in 2013 in Beijing, Environ. Sci. Pollut Res., 23, 13996–14007, https://doi.org/10.1007/s11356-016-6518-3, 2016.

Yumimoto, K. and Takemura, T.: Direct radiative effect of aerosols estimated using ensemble-based data assimilation in a global aerosol climate model, Geophys. Res. Lett., 38, L21802, https://doi.org/10.1029/2011GL049258, 2011.

Zhan, D., Kwan, M., Zhang, W., Yu, X., Meng, B., and Liu, Q.: The driving factors of air quality index in China, J. Clean. Prod., 197, 1, 1342–1351, https://doi.org/10.1016/j.jclepro.2018.06.108, 2018.

Zhang, Q., Streets, D. G., Carmichael, G. R., He, K. B., Huo, H., Kannari, A., Klimont, Z., Park, I. S., Reddy, S., Fu, J. S., Chen, D., Duan, L., Lei, Y., Wang, L. T., and Yao, Z. L.: Asian emissions in 2006 for the NASA INTEX-B mission, Atmos. Chem. Phys., 9, 5131–5153, https://doi.org/10.5194/acp-9-5131-2009, 2009.

Zhang, Q., Ma, Q., Zhao, B., Liu, X., Wang, Y., Jia, B., and Zhang, X.: Winter haze over North China Plain from 2009 to 2016: Influence of emission and meteorology, Environ. Pollut., 242, 1308–1318, https://doi.org/10.1016/j.envpol.2018.08.019, 2018.

Zhao, Y., Greybush, S. J., Wilson, R. J., Hoffman, R. N., and Kalnay, E.: Impact of assimilation window length on diurnal features in a Mars atmospheric analysis, Tellus A, 67, 26042, https://doi.org/10.3402/tellusa.v67.26042, 2015.

Zheng, B., Cheng, J., Geng, G., Wang, X., Li, M., Shi, Q., Qi, J., Lei, Y., Zhang, Q., and He, K.: Mapping anthropogenic emissions in China at 1 km spatial resolution and its application in air quality modeling, Sci. Bull., 66, 612–620, https://doi.org/10.1016/j.scib.2020.12.008, 2021.