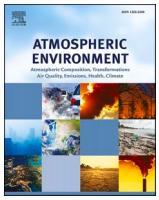
Atmospheric Environment 306 (2023) 119831

Contents lists available at ScienceDirect

Atmospheric Environment

journal homepage: www.elsevier.com/locate/atmosenv



Dust storm forecasting through coupling LOTOS-EUROS with localized ensemble Kalman filter

Mijie Pang [a], Jianbing Jin [a,*], Arjo Segers [b], Huiya Jiang [c], Li Fang [a], Hai Xiang Lin [d,e], Hong Liao [a]

[a] Joint International Research Laboratory of Climate and Environment Change, Jiangsu Key Laboratory of Atmospheric Environment Monitoring and Pollution Control, Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology, School of Environmental Science and Engineering, Nanjing University of Information Science and Technology, Nanjing, Jiangsu, China
[b] TNO, Department of Climate, Air and Sustainability, the Netherlands
[c] College of Environment and Resources, Nanjing Agricultural University, China
[d] Institute of Environmental Sciences, Leiden University, the Netherlands
[e] Delft Institute of Applied Mathematics, Delft University of Technology, Delft, the Netherlands

HIGHLIGHTS

- A dust storm forecasting system is developed by coupling LOTOS-EUROS and EnKF.
- Dust forecast is improved when data assimilation is applied.
- LEnKF is proved to be superior than EnKF in dust storm forecasting.
- Sensitivities of dust forecast to the localization distance in LEnKF are exploited.

ARTICLE INFO

Keywords:
Dust aerosol
Data assimilation
EnKF

ABSTRACT

Super dust storms re-occurred over East Asia in 2021 spring and casted great health damages and property losses. It is essential to achieve an accurate dust forecast to reduce the damage for early warning. The forecasting system fundamentally relies on a numerical model which can forecast the full evolution of dust storms. However, large uncertainties exist in model forecasts. Meanwhile, various near-real-time observations are available that contain valuable dust information. A dust storm forecasting system is here developed through coupling a chemical transport model, LOTOS-EUROS, and Localized EnKF (LEnKF) assimilation approach. The assimilations are carried out via an interface of our self-designed assimilation toolbox, PyFilter v1.0. Ground-based PM_{10} measurements from air quality monitoring network are assimilated. Sequential assimilation tests are carried out over the 2021 spring super dust storms. The results show that the assimilation-based forecasting system produces a promising dust forecast than model-only forecast, and the improvements is also validated through comparing to the independent MODIS aerosol optical depth (AOD). Superior performance is obtained when LEnKF is implemented, as the localization helps EnKF in resolving the PM_{10} measurements that have a large spatial variability with limited ensemble members. In addition, sensitivity experiments are conducted to exploit the distance-dependent localization for the LEnKF. Considering both cases, the optimal choice of the distance is tested to be around 500 km: the larger distance is less effective in removing the spurious correction, while the smaller one easily falls into the local optimum and the model would become divergent rapidly.

1. Introduction

Dust storms are defined as severe meteorological disasters by the World Meteorological Organization. These events are prevalent in arid and semi-arid land (Wang et al., 2017), and are caused by wind erosion that liberates particles from exposed, dry surfaces (Joshi, 2021). Dust aerosols can be lifted over several miles high depending on the intensity of the wind and transport over long distances, even in a global scale, e.g., from Africa to the Americas crossing the pacific ocean (Zhang et al., 2018). Together with the substantial amount of dust aerosols, it has also

* Corresponding author.
E-mail address: jianbing.jin@nuist.edu.cn (J. Jin).

https://doi.org/10.1016/j.atmosenv.2023.119831
Received 13 December 2022; Received in revised form 13 April 2023; Accepted 5 May 2023
Available online 6 May 2023

been proved that long-distance transport of dust in the atmosphere is in favour of environmental microorganisms which consist of bacteria, virus and spores (Yuan et al., 2017). They pose great threats to human health by causing various illness, e.g., breathlessness, cardiovascular disorders and allergies (WMO, 2018). In addition, the reduction in visibility would cause severe disorders of the transportation and aviation systems. Taking the 2021 super dust storm in East Asia for example, 10 deaths were reported in Mongolia (Chen and Walsh, 2021) and 12 provinces in the northern China were affected with thousands of flights grounded and public transportation systems halted (Jin et al., 2022).

Dry soil, rare vegetation, instability in the lower troposphere and strong cyclone are the primary factors contributing to dust storms (Zou and Zhai, 2004). Dust storms frequently took place in the Arabian Peninsula, Southwest Asia, Africa and East Asia (Tanaka and Chiba, 2006). For century, East Asia has been afflicted by the dust storms, of which the source are mainly attributed to the Mongolian and Chinese Gobi Desert (Qian et al., 2002). The peak time of the dust storms in Gobi Desert is in March, April and May (Kurosaki and Mikami, 2003). When it comes to interannual variations, the dust storms occurred frequently from the 1950s–1970s, and become less frequent during 1980s and 1990s (Yin et al., 2021). In the recent two decades, the dust frequency stays at a low level thanks to a series of large scale projects initiated by Chinese government, such as 'Three-North' Shelter Forest Program (Guo et al., 2018). On the other hand, the ongoing desertification is reported to aggravate the dust hazards (Han et al., 2021). In spring 2021, however, there were several super dust storms re-occurred in East Asia (Liu et al., 2021), which brought huge property losses and health damages (Filonchyk, 2022; Jin et al., 2022; Yin et al., 2021).

Numerical model is an effective tool for implementing dust storm forecast, and efforts have been made to develop varies of dust models. Typically, a dust model consists of a chemical transport model and modules for dust emission and deposition (Shao and Dong, 2006). Since the early 1990s, several dust forecast system have been developed, e.g., ECMWF's Integrated Forecast System (IFS) (Morcrette et al., 2008, 2009), LOTOS-EUROS (Manders et al., 2017; Timmermans et al., 2017) and an improved dust emission model implemented in Kok et al. (2014). However, numerical prediction of dust storms suffers from uncertainties in various aspects, especially in the emission uncertainties (Kontos et al., 2021). The lack of knowledge about the soil characteristics and wind fields that vary in fine-scale, which is difficult for model to resolve, give rise to the high levels of uncertainty in dust storm forecast (Jin et al., 2019a; Jin et al., 2020, 2021). It was found that even from models with similar transport patterns, the predicted concentrations on the level of surface are different by over two orders of magnitude (Lin et al., 2008).

In recent decades, there has been an increased focus on the atmospheric environment, resulting in the promotion of the construction of ground-based observation and remote sensing technology. Ground-based observations, as a fundamental method wildly used by government and scientist, can provide data with high accuracy and temporal resolution. For example, Ministry of Ecology and Environment has been publishing hourly-averaged concentrations of $PM_{2.5}$, PM_{10}, SO_2, NO_2, CO, and O_3 from over 1600 ground stations online since January 2013 (Guo et al., 2019). However, due to inconsistent spatial distribution of stations, leading to spatial sparseness in the ground-based observations. Moreover, these stations are located in the densely populated regions for the air quality index monitoring, and therefore are far away from the dust source region. Remote sensing technology, owning to its broad coverage both in spatial and temporal extent (only for geostationary onboard instruments), is considered to be a promising method to identify and monitor the dust storms (Rayegani et al., 2020). There are various kinds of remote sensing measurements that are available for dust storm monitoring, e.g., aerosol optical depth (AOD) from ground-based sun photometers, polar-orbiting satellite and geostationary meteorological satellite. However, there are larger uncertainties existing while AODs are used for representing dust loading, e.g., satellite instrument usually measure the column-integrated aerosol extinctions while contains no information of the vertical profile (Lorente et al., 2017); large uncertainties were found in the Himawari-8 product due to uncertainty in assumptions regarding aerosol models and surface reflectance estimation in the retrieval algorithm (Zhang et al., 2019).

As aforementioned, there are deficiencies in using observation data or model solely. To make the best of the observation data and model, a method called 'Data Assimilation' was introduced. Data assimilation can intercorporate the observations with models through estimating the uncertain initial conditions or parameters, thereby improving forecasts (Bertino et al., 2003). It has been wildly applied to many fields, e.g., meteorology, oceanography and geography. Data assimilation methodologies can be mainly categorized as variational approaches and filtering approaches (Bocquet et al., 2015). The variational approaches such as 4DVar aim to obtain the optimal analysis by minimizing the cost function constructed in time and space in assimilation window (Jin et al., 2018). Based on variational approaches, many inverse models have been developed and applied in many fields such as atmospheric pollutant emission inversion (Corazza et al., 2011; Bergamaschi et al., 2010; Jin et al., 2018, 2019a). The filtering approaches, e.g., Particle filter (Leeuwen et al., 2019), Extended Kalman Filter (EKF) which is a nonlinear solution extended from linear system filter: Kalman Filter (KF) (Judd, 2003) (these two can hardly be used in the large scale model) and Ensemble Kalman Filter (EnKF) (Evensen, 1994) using the background error covariance matrix calculated from ensemble statistics and observation state vector to produce optimal posteriori, which used to be computational unaffordable when applied to high dimensional models (Houtekamer and Zhang, 2016). Recently, with the continuous increase of computing resource, EnKF has been more popular in high dimensional models such as global atmospheric models. It can result in effective improvement in the forecast with affordable extra numerical cost (Burgers et al., 1998). Despite these advantages, there are drawbacks in EnKF such as its dependence on the relative small ensemble number compared to high model dimensions to estimate background error covariances and spurious correlations associated with observations that are spatially remote (Hamill et al., 2001; Houtekamer and Mitchell, 2001). In order to 'cut off' the spurious correlations existing in background error covariance matrix, various localization methods were introduced. The most representative one is the distance-dependent scheme, in which the correlation between any two grid cells decreases from 1.0 to 0.0 as the spatial distance increases (Lei and Anderson, 2014).

Many assimilation systems have been developed to improve predictability of atmospheric aerosols. For instance, the China Unified Atmospheric Chemistry Environment/Dust (CUACE/Dust) forecast system (Gong and Zhang, 2008; Niu et al., 2008; Wang et al., 2008; Zhou et al., 2008) used a 3DVar method to reconstruct the initial conditions. Observations including Aerosol Optical Depth (AOD) data from the Chinese geostationary satellite FY-2C and surface visibility (phenomena) from the China Meteorological Agency (CMA) ground monitoring system are incorporated. Kong et al. (2021) developed a $PM_{2.5}$ forecast system coupling 4D-LETKF and WRF-Chem model and results shows that the initial concentrations were optimized, RMSE were decreased and correlation coefficients were increased in a large extent. Osores et al. (2020) developed a volcanic ash forecast system which coupled ETKF and FALL3D, and it turned out that both estimation of ash concentration and time-dependent eruption source parameters were optimized. Recently, we have developed an emission inversion system by combining LOTOS-EUROS and 4DVar assimilation algorithm (Jin et al., 2018), which effectively assimilated the aforementioned valuable dust measurements including the ground-based PM_{10} from the air quality monitoring network, Himawari-8 and MODIS AOD. Both data quality controls for AOD (Jin et al., 2019a; 2022) and PM_{10} measurements (Jin et al., 2019b) were proposed to accurately represent dust loading using those observations. An adjoint method was proposed to backtrack the potential dust source regions for better constructing the emission background covariance matrix (Jin et al., 2020). Dust plume position errors in the

long-distance transport were solved using grid distorted data assimilation (Jin et al., 2021). However, those studies only focused on calculating the optimal emission field and all attributed the observation-minus-simulation to errors in the emission. Model driven by the posterior dust emission field could be used to forecast the dust concentration, however, an accurate forecast of the dust plume especially in fine scale is not guaranteed. This is because the deposition and transport errors accumulate along the long-distance movement from the Mongolia and Gobi Desert source. Uncertainty of the dust plume simulation in fine scale was also determined by those errors next to the uncertainty in the emission. When the dust plume already moves to the downwind region, it is difficult to reproduce the actual dust field in fine scale through the emission inversion by only nudging the emission intensity pattern. Sequential data assimilation systems are indeed in demand for short-term dust plume prediction over East Asia.

We develop a dust storm forecasting system through coupling the regional chemical transport model LOTOS-EUROS and EnKF and Localized EnKF (LEnKF) that are embedded in a self-developed assimilation toolbox, PyFilter v1.0. It is capable of providing a high-quality dust storm forecast over East Asia. The system is here tested to forecast the 2021 spring super dust storms by assimilating PM_{10} concentration measurements. The forecasting skill is validated by comparing to the independent MODIS Deep Blue AOD measurements. Meanwhile, the role of localization on the assimilation analysis is exploited, and the optimal localization scheme is determined.

The paper is organized as follows. Section 2 introduces the framework of our self-developed dust forecasting system. It is composed of a CTM, LOTOS-EUROS model, and the Ensemble Kalman Filter (EnKF) and the localized Ensemble Kalman Filter (LEnKF) assimilation algorithm which are embedded in the PyFilter toolbox. In Section 3, two super dust storm events occurred in the east Asia in 2021 spring are described as well as the ground-based PM_{10} observation data for assimilation and MODIS AOD for independent evaluation. The CAMS reanalysis data is also used to evaluate the performance. Section 4 evaluates and discusses the assimilation analysis and forecasts. The benefit of localization is emphasized. Section 5 further concludes the highlight of our dust forecasting system and discuss the future work.

2. Dust forecasting system

2.1. System framework

The system is mainly constructed by coupling the LOTOS-EUROS model with the Ensemble Kalman Filter (EnKF) and Localized Ensemble Kalman Filter (LEnKF) assimilation algorithm. The filter algorithms are embedded in our self-developed Python-based toolbox, PyFilter v1.0, which is designed with various of sequential data assimilation algorithms, e.g., EnKF, particle filter and three-dimensional (3DVar) variational method. An interface is designed in PyFilter for data transmission between the model and the assimilation, as well as the model initialization. The feature of ensemble model parallelization largely accelerates the computing speed. Different from some data-assimilation built-in systems, our toolbox is highly extendable and not bound with a specific numerical model or assimilation algorithm. Any other numerical model can be adapted into the PyFilter, and filter algorithms can be updated conveniently.

As shown in Fig. 1 is the diagram of the processes involved in the dust forecasting system. Initially, ensemble members (*Ne*) of model realization forward with perturbed inputs, which are here the ensemble emissions as described in Section 2.2. When observations are available, assimilation analysis based on the EnKF and LEnKF algorithm will be performed to update the ensemble posteriori dust fields. The ensemble posteriors are then written back as the model restarting files for new forecasts. The assimilation analysis is repeated until the end of the simulation.

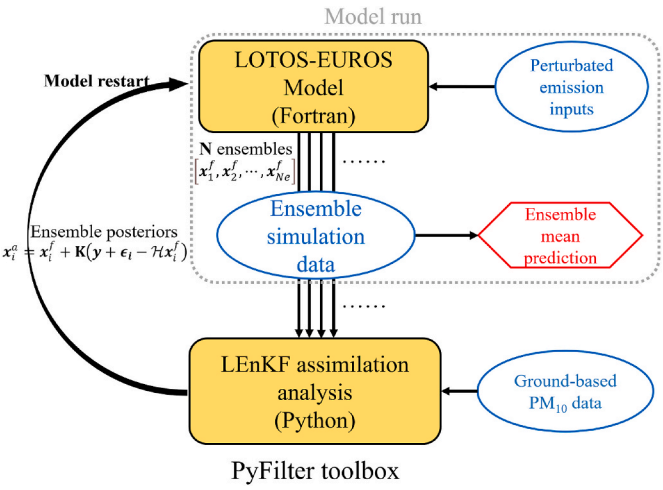


Fig. 1. Flowchart of the data assimilation system scheme for dust storm forecast. The blue oval means the input. The red hexagon denotes the output. The yellow rounded rectangle is the two main parts of the system. Parts in gray rounded rectangle are typical ensemble model forecast, while outside of them are the additional parts of ensemble data assimilation.

2.2. LOTOS-EUROS forecast

The LOTOS-EUROS model is a regional chemical transport model (CTM) originating from Long-Term Ozone Simulation (LOTOS) and European Operational Smog model (EUROS) in the Netherlands in the 1980s which are developed by Netherlands Organization for Applied Scientific Research (TNO) and National Institute for Public Health and the Environment (RIVM) respectively and it is under continuous development by a consortium of institutes in the Netherlands during the past 10 years. LOTOS-EUROS has been widely used in many applications such as air pollution forecast including ozone (Curier et al., 2012), NO_2 (Skoulidou et al., 2021), dust storm emission inversion (Jin et al., 2018, 2019a, 2019b, 2020, 2021, 2022) and source apportionment (Kranenburg et al., 2013; Timmermans et al., 2017).

In this study, LOTOS-EUROS is utilized to simulate dust storms occurred in East Asia. The pure LOTOS-EUROS dust storm forecast/simulation is referred to as the priori or control model run in this study. The simulation is carried out over a domain from 15°N to 50°N and 70°E to 140°E, with a resolution of 0.25° × 0.25°. The model consists of eight vertical layers, with a top at 10 km. It has zero boundary conditions on the assumption that all dust aerosols are emitted in regional area and the external dust flows can be ignored. Dust aerosols are described using five bins with the geometric mean radius spanning from 0.01 to 10 μm. Physical processes included are the dust emission, deposition, advection, diffusion, and sedimentation. The dust forecast is driven by European Center for Medium-Ranged Weather Forecast (ECMWF) operational forecasts over 3–12 h, retrieved at a regular longitude/latitude grid resolution of about 7 km. An interface to the ECMWF output set is designed, which not only interpolates the default 3-h ECMWF short-term forecast meteorology to hour values, but also averages the forecast to fit the LOTOS-EUROS spatial resolutions. The model forecast is output at the interval of 1 h.

The goal of the work is to estimate the dust concentration field with the available PM_{10} measurements, which will then lead to more accurate dust forecast. It is then necessary to define the uncertainty in the dust storm simulation. In this study, the main model error is attributed to the uncertainty in the dust emission parameterization. While other model processes like transport, deposition are uncertain as well. They are assumed to be less important than the emission parameterization. The priori dust aerosol emission scheme adopted in the LOTOS-EUROS is the windblown parameterization proposed by Zender et al. (2003). The ensemble emissions for initializing the ensemble model forecasts are

randomly generated following the a priori emission and an error covariance matrix describing the uncertainty of the emission parameterization scheme. Details concerning the ensemble dust emission generation could be found in the Supplement.

2.3. Assimilation algorithm

The Ensemble Kalman Filter (EnKF) provides a practical method to cope with high-dimensional atmospheric model dynamic by approximating the low-dimensional background error covariance matrix. It is adopted as the assimilation algorithm in this study.

To begin with, ensemble individual x_i^f which represents the dust concentration is generated by the LOTOS-EUROS model. The model operator $\mathcal{M}$ integrates from the analysis vector $x_{i,t-1}^a$ at the previous time step t-1.

$$x_i^f = \mathcal{M}\left(x_{i,t-1}^a\right) \tag{1}$$

$$\mathbf{X}^f = \left[x_1^f, x_2^f, \cdots, x_{Ne}^f\right] \tag{2}$$

$\bar{x}^f$ is the average of the ensemble model forecast stored in $\mathbf{X}^f$, and $\mathbf{X}^{f'}$ is the ensemble perturbation matrix. *Ne* is the total ensemble number. In this paper, it is assigned to be 32 for it gives a high freedom for representing the uncertainty in the dust field and is computational affordable. A larger ensemble member Ne > 32 is tested but the improvement is limited. Experiments about the sensitivity of assimilation analysis to the ensemble number is depicted in Fig. 2(a). The ensemble-based background covariance matrix $\mathbf{P}^f$ is calculated based on $\mathbf{X}^{f'}$.

$$\bar{x}^f = \frac{1}{Ne}\sum_{i=1}^{Ne} x_i^f \tag{3}$$

$$\mathbf{X}^{f'} = \left[x_1^f - \bar{x}^f, x_2^f - \bar{x}^f, \cdots, x_{Ne}^f - \bar{x}^f\right] \tag{4}$$

$$\mathbf{P}^f = \frac{1}{Ne-1}\sum_{i=1}^{Ne}\left(x_i^f - \bar{x}^f\right)\left(x_i^f - \bar{x}^f\right)^{\mathrm{T}} = \frac{1}{Ne-1}\mathbf{X}^{f'}\,\mathbf{X}^{f'\mathrm{T}} \tag{5}$$

The factor $\mathbf{K}$ is usually referred to as the Kalman gain. $\mathcal{H}$ is the observation operator which maps model state variables into observation space. $\boldsymbol{y}$ is the observation vector storing the PM_{10} measurements, and $\mathbf{O}$ is the observation error covariance matrix that quantify the penalty of the simulation mismatching the observations $\boldsymbol{y}$. However, PM_{10} concentration observations consists of not only dust aerosol, but also non-dust particles in practice. The non-dust fraction is removed before these PM_{10} data are assimilated. Description about the non-dust bias correction and definition of $\mathbf{O}$ are presented in Section 3.2.

$$\mathbf{K} = \mathbf{P}^f\mathcal{H}^{\mathrm{T}}\left(\mathcal{H}\mathbf{P}^f\mathcal{H}^{\mathrm{T}} + \mathbf{O}\right)^{-1} \tag{6}$$

With all these, a posteriori estimation individual x_i^a can be updated via:

$$x_i^a = x_i^f + \mathbf{K}\left(\boldsymbol{y} + \boldsymbol{\varepsilon}_i - \mathcal{H}x_i^f\right) \tag{7}$$

$\boldsymbol{\varepsilon}_i$ refers to the sampling error. It is a random vector that subjects to the normal distribution. Its mean is zero and the covariance is adapted by the root of diagonal from $\mathbf{O}$. Furthermore, in order to 'cut-off' the spurious correlation in the background covariance matrix $\mathbf{P}^f$, a correlation matrix $\mathbf{L}$ filled with local supports is introduced. Local support is a term meaning that the function is only non-zero in a local region and is zero elsewhere. The local support function is taken from set proposed in (Gaspari and Cohn, 1999) as:

$$\mathbf{S}_{i,j} = \frac{\mathbf{D}_{i,j}}{L_{thres}} \tag{8}$$

$$\mathbf{L}_{i,j} = \begin{cases} 1 - \frac{5}{3}\mathbf{S}_{i,j}^2 + \frac{5}{8}\mathbf{S}_{i,j}^3 + \frac{1}{2}\mathbf{S}_{i,j}^4 - \frac{1}{4}\mathbf{S}_{i,j}^5, & \mathbf{S}_{i,j} < 1 \\ -\frac{2}{3}\mathbf{S}_{i,j}^{-1} + 4 - 5\mathbf{S}_{i,j} + \frac{5}{3}\mathbf{S}_{i,j}^2 + \frac{5}{8}\mathbf{S}_{i,j}^3 - \frac{1}{2}\mathbf{S}_{i,j}^4 + \frac{1}{12}\mathbf{S}_{i,j}^5, & 1 \le \mathbf{S}_{i,j} < 2 \\ 0, & \mathbf{S}_{i,j} \ge 2 \end{cases} \tag{9}$$

where $\mathbf{D}_{i,j}$ refers to the distance between the two grid cells *i* and *j*, while L_{thres} represents the localization distance threshold. The individual element $\mathbf{L}_{i,j}$ in localization matrix is constructed as Eq. (10). As Fig. 2(b) shows, the correlation declines as the distance increases and shorter distance threshold brings greater descent rate. It should be noted that when distance arrives at the distance threshold, the correlation between two grid cell *i* and *j* is not reduced to zero but keeps declining until the distance comes to the twice the distance threshold.

With the localized correction matrix $\mathbf{L}$, the localized $\mathbf{P}^{f,local}$ is calculated through a Schur product:

$$\mathbf{P}^{f,local} = \mathbf{P}^f \circ \mathbf{L} \tag{10}$$

Subsequently, an updated localized Kalman gain is generated. Apply the localized Kalman gain $\mathbf{K}^{local}$ to Eq. (6) and the localized posteriori can be retrieved.

3. Dust storm events and observation data

The test cases of these super dust storms are illustrated in Section 3.1, while PM_{10} concentration measurements for assimilation is described in Section 3.2. MODIS AOD product that is used to evaluate our dust storm forecast independently is presented in Section 3.3, and the operational CAMS reanalysis to evaluate our dust field analysis is introduced in Section 3.4. The timeline, the choice of ensemble number, localization threshold concerning the assimilation experiments were illustrated in Section 3.5.

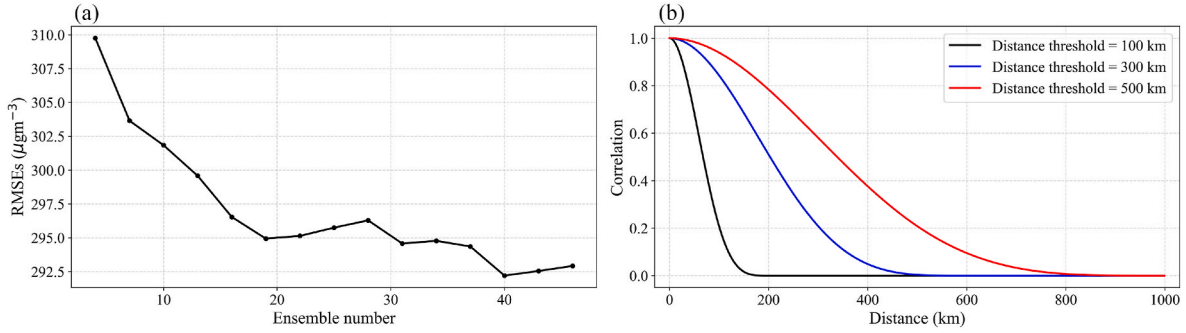


Fig. 2. Trend of average RMSEs from EnKF analysis with ensemble number ranging from 4 to 46 ensembles from 8:00 to 20:00 (at the interval of 1 h) 28th 2021 **(a)** and correlation distribution for different localization distance threshold L_{thres} **(b)**.

3.1. Dust storm events

Several super dust storms took place in March and April of 2021 in East Asia after an absence of two decades. This study focused on two of them which occurred around March 28 and April 15, respectively. The dust plumes were clearly observed by the ground-based air quality monitoring network. The maximum PM_{10} concentration detected by ground monitoring network has exceeded 9000 μg/m^3 on 28th March and 4000 μg/m^3 on 15th April. Fig. 3 shows the frequency density of ground-based hourly mean PM_{10} concentration in North Chinese Plain (NCP) and Fenwei Plain (FWP) from 27th March to 18th April when two dust storms occurred. The duration of two dust storms sustains 2–3 days. Both of them have a great impact on these regions (Xu et al., 2022). In this study, two dust storms events from March and April of 2021 will be investigated using the system we designed, and they will be referred to as DSE1 and DSE2 respectively in the following section.

3.2. p.m.$_{10}$ concentration observation for assimilation

At the present, there is an air quality monitoring network consisting of more than 1600 ground-based stations covering all over China, part of which in the northern regions could be seen in Fig. 3(a). The real-time hourly PM_{10} concentration data is made public. These ground-based PM_{10} observation data, owing to its high temporary resolution and less uncertainty, has become an essential source of measurements on dust storms.

However, the PM_{10} observation data can't be utilized in data assimilation directly since it comprises not only dust itself, but also some other particles. Those particles are released from anthropogenic sources including vehicles and industry as well as natural sources including sea spray and volcanic eruption. One of the basic assumption of data assimilation is that the observation is unbiased, while in reality the biases in observation are often inevitable. The assimilation system will diverge rapidly if the PM_{10} concentration is assimilated directly into the system, especially when the majority of PM_{10} is non-dust particle. In our recent 4DVar-based dust emission inversion (Jin et al., 2022), the non-dust fraction of the total PM_{10} is simulated using a separate model, and the baseline-removed PM_{10} measurements is calculated. Details concerning the non-dust baseline removal could be found in Jin et al. (2022). Snapshots of the original PM_{10} and baseline-removed (BR) dust observations in DSE1 and DSE2 can be seen in Fig. 5(a and b) and Fig. 6 (a and b), respectively. The baseline-removed measurements captured a

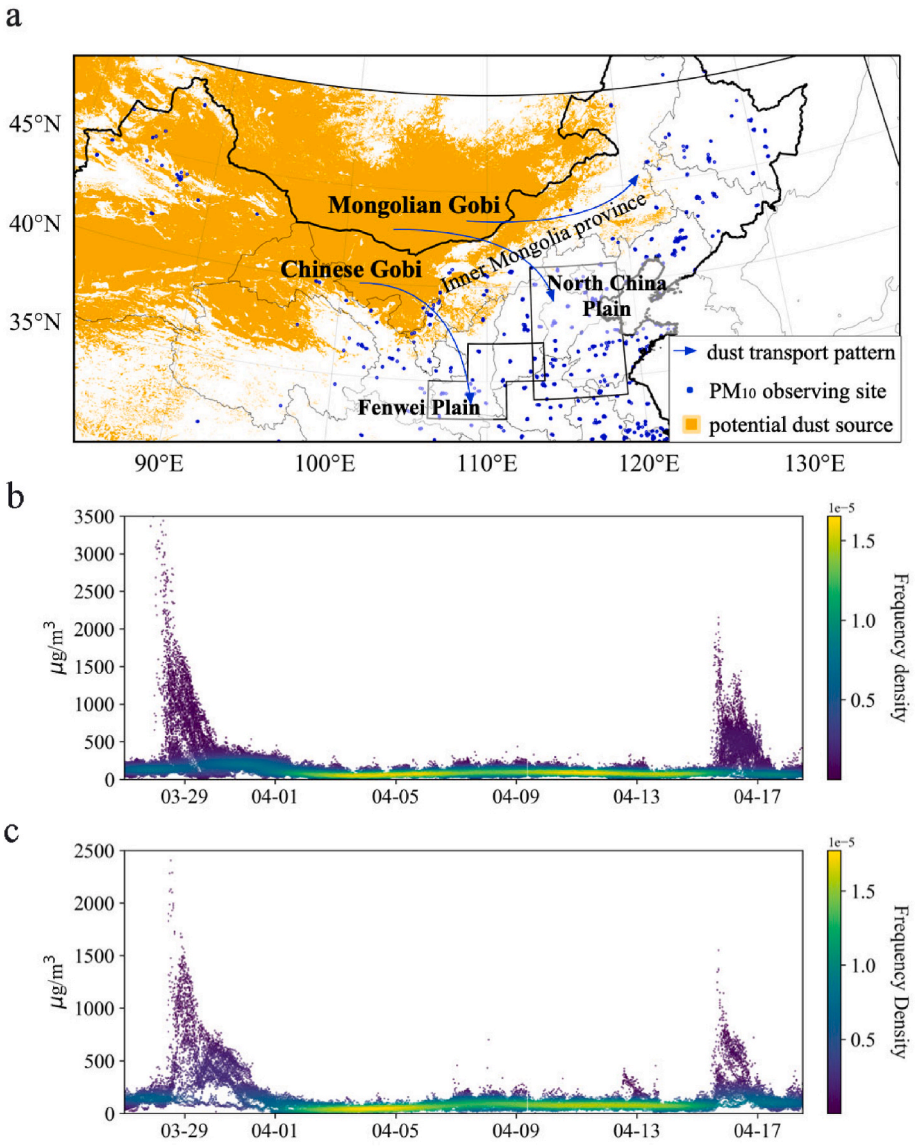


Fig. 3. Distribution of the potential dust source over east Asia and ground-based air quality observing met work over northern China (**a**). Frequency density of ground-based hourly mean PM10 concentration in NCP (**b**) and FWP (**c**) during two dust storms.

clearer shape of the dust plume. For instance, the original PM_{10} observations in Fig. 5(a) indicated that the dust plume already arrived in Shandong Province (yellow scatters in red box ranging from 100 to 300 μg/m^3). They were identified as the non-dust aerosol, and Shandong Province is almost free of dust as shown in Fig. 5(b) which is consistent to our priori dust simulation plotted in Fig. 5(d).

Concerning the observations error covariance **O** that required by the Kalman gain calculation in Eq. (6), it is defined to be a diagonal matrix with all observations are assumed to be independent. Meanwhile, the uncertainty (square root of the individual diagonal element in **O**) of the pre-processed PM_{10} measurements for assimilation is assumed to be joint due to uncertainty in the PM_{10} data and the non-dust PM_{10} bias correction. We used the empirical Eq. (11) to quantify the representation errors $\boldsymbol{\sigma}^{PM}$ in PM_{10} following our previous choice in Jin et al. (2022).

$$\boldsymbol{\sigma}_i^{PM} = \begin{cases} 200, & \boldsymbol{y}_i \le 200 \\ (\boldsymbol{y}_i - 200) \times 0.1 + 200, & \boldsymbol{y}_i > 200 \end{cases} \tag{11}$$

We chose an empirical value of 200 μg/m^3 as the minimum of the $\boldsymbol{\sigma}^{PM}$. This is mainly to prevent the posteriori from getting too close to the low-value PM_{10} observations and hence being model divergent. Details about the errors in observation error covariance matrix **O** can be found in Supporting Information.

3.3. MODIS AOD for independent evaluation

The Moderate Resolution Imaging Spectroradiometer (MODIS), launched onboard the Terra and the Aqua satellite, is one of the fundamental instruments of NASA Earth Observing System (EOS) and produce reflectance at multiple wavelengths (Xie et al., 2017). Satellite measurements have advantage in coverage over ground-based measurements while it also has greater uncertainties. In this study, the MODIS Deep Blue AOD products are used to compare AOD forecast from the priori and posteriori simulation. Note that these AOD could be used as part of the observations for assimilation to have the prediction as accurate as possible in practice. But they are used as the independent measurements for the validation of our dust forecasting system. Similarly, instead of using the original MODIS AOD that contains both dust and non-dust contribution, it is the baseline-removed AOD used for accurately representing the dust load, and the pre-processing method is described in Supplementary Material Text S3.

3.4. CAMS reanalysis for comparison

The Copernicus Atmosphere Monitoring Service (CAMS) is a global reanalysis dataset produced by European Center for Medium-Range Weather Forecasts (ECMWF) (Inness et al., 2019). It provides three-dimensional simulations of the atmospheric composition obtained by combining a global atmospheric chemistry model and observations. Data assimilation is also used as the principle to produce the best estimate of the state of the atmosphere. It has global coverage, horizontal resolution of 0.75° × 0.75°, 60 model levels and temporal resolution of 3-hourly. In this paper, PM_{10} on surface level from the CAMS product is used as a benchmark to evaluate the assimilation analysis performance of our PyFilter system. Note that CAMS provides the full coarse aerosol simulation instead of the dust product solely, therefore, it is directly compared to the raw PM_{10} observation records.

3.5. Assimilation experiments

To evaluate the capability of the dust forecasting system, data assimilation experiments on the DSE1 and DSE2 are carried out. Fig. 4(a) shows the timeline of the sequential assimilation designed for DSE1 in this study. Dust emissions occurred at the start of the simulation window, 26th March, when the plumes were lifted high to be carried southward to the northern China. The assimilation is not performed until 28th March, as the dust plume is not observed by ground-based monitoring network before. The assimilation analysis is then performed at an interval of 3 h starting from 8:00 to 20:00 on 28th of March. The posteriori dust field would be used as the initial condition for the new forecast. The assimilation analysis is terminated when the average dust load from BR-PM_{10} observations is reduced to a low level around 150 μg/m^3. The baseline-removed PM_{10} concentration observations described in Section 3.2 are assimilated for the dust state field estimation in each of the analysis. The updated posteriors are then used to forward the model for short-term forecast with a horizon of 1–6 h. For an operational dust forecast system in practise, the assimilation could be conducted every hour, therefore all the PM_{10} data would be assimilated. Similar timeline of assimilation sets for DSE2 could be found in Fig. 4(b).

To further explore the effectiveness of localization on the analysis and forecasts, a set of experiments with different localization distance threshold are performed. Note that the control model here represents the average of ensemble simulation without assimilating any observation. It is severed as a benchmark of the pure model-based forecast for the comparison against the assimilation-based forecast without localization or using distance threshold L_{thres} = 100, 300, 500, 700 and 1000 km (referred to as EnKF, L100, L300, L500, L700, L1000 respectively). The metric, Root-Mean Square Error (RMSE) and Normalized Mean Error (NME), is introduced here for evaluating the accuracy of these simulations against the baseline-removed dust observation (PM_{10} and AOD). In the calculation of RMSE and NME, the mode simulation is compared to the observation in the observational space using the nearest search

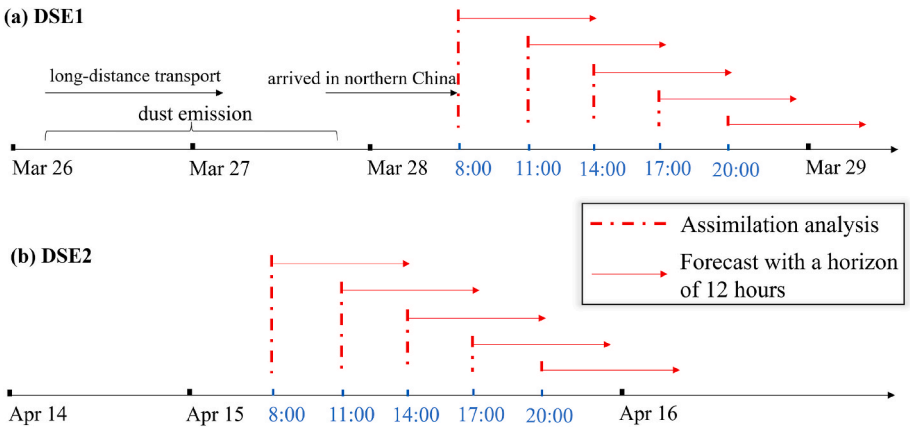


Fig. 4. Sequential assimilation set for DSE1 (**a**) and DSE2 (**b**). The assimilation analysis (vertical dash line) is performed at the interval of 3 h from 8:00 to 20:00 on March 28th in DSE1, and from 8:00 to 20:00 on April 15th in DSE2. The rolling model forecast (red solid line) is made with a horizon of 12 h based on the each of the assimilation analysis.

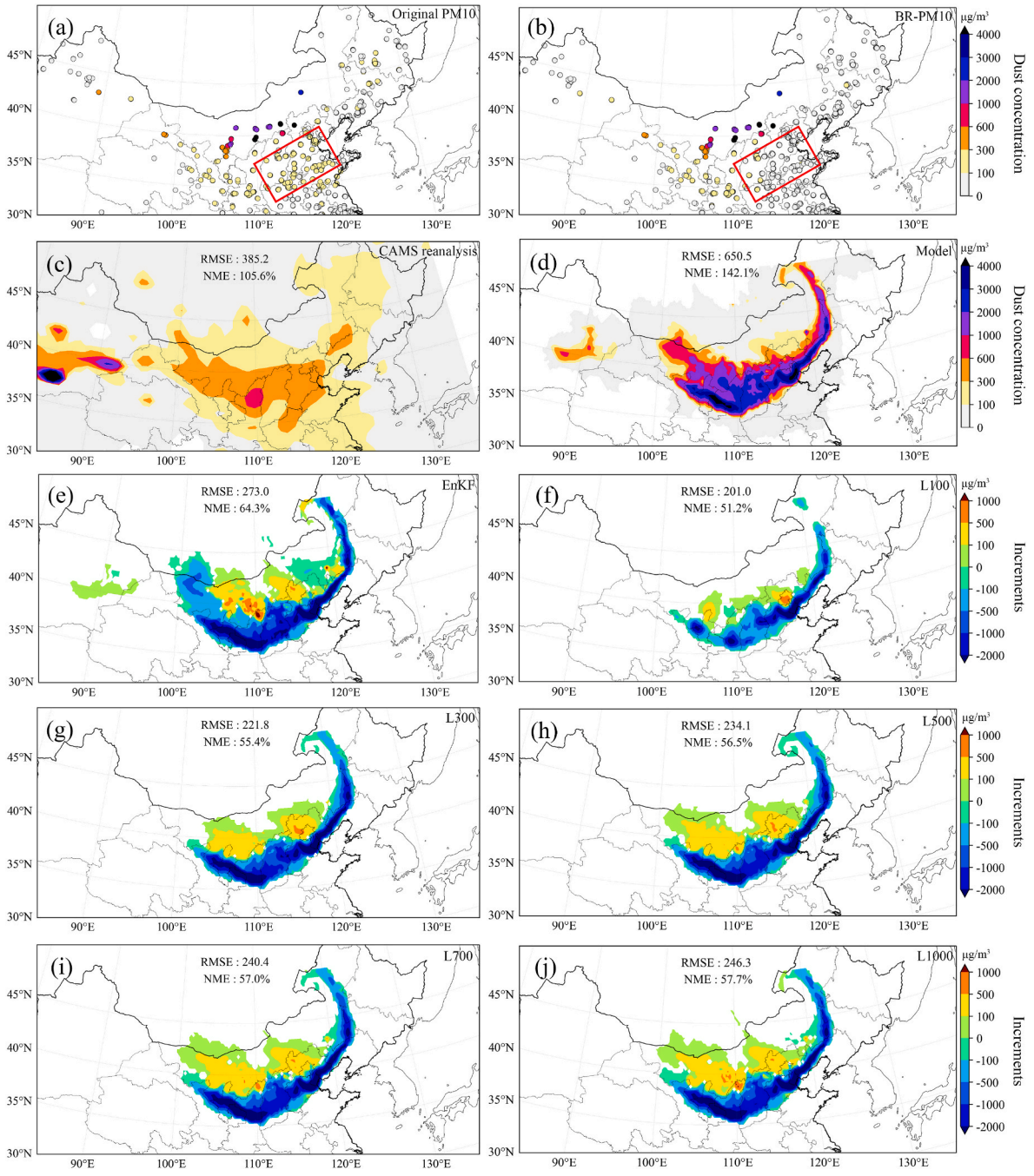


Fig. 5. Spatial distribution of ground-based raw PM_{10} (**a**), BR-PM_{10} concentration observations (**b**), reanalysis data of PM_{10} from CAMS at the surface level (**c**), the surface dust concentration (SDC) simulation from control model (**d**), increments from the posteriori SDC updated using EnKF (assimilation analysis field minus control model run, figures below are the same) (**e**), using LEnKF with L_{thres} = 100 km (**f**), using LEnKF with L_{thres} = 300 km (**g**), using LEnKF with L_{thres} = 500 km (**h**), using LEnKF with L_{thres} = 700 km (**i**) and using LEnKF with L_{thres} = 1000 km (**j**) at 08:00, 28th March 2021 (CST). BR-PM_{10}: baseline-removed PM_{10}.

method.

4. Results and discussions

Section 4.1 presented the evaluation of our assimilation analysis, as well as the comparison against the reanalysis product from CAMS. Section 4.2 evaluated the performance of our dust storm forecast and effectiveness of the localization. In Section 4.3, the dust storm forecasts are further evaluated using the independent MODIS AOD measurements.

4.1. Assimilation analysis

Fig. 5 is the original PM_{10} concentration observations (panel a), baseline-removed PM_{10} concentration observation (panel b), surface PM_{10} concentration simulation from CAMS reanalysis dataset (panel c), surface dust concentration forecast from the pure model (panel d), increments from the EnKF analysis (panel e) and EnKF with different localization schemes (panel f–j) at 8:00 on 28th of March, China standard time (CST). Note that increments from assimilation analysis is obtained by calculating the difference between assimilation analysis and pure model simulation. The first analysis for DSE1 is implemented at 8:00. At this instance, the model simulated a clear dust plume and it is

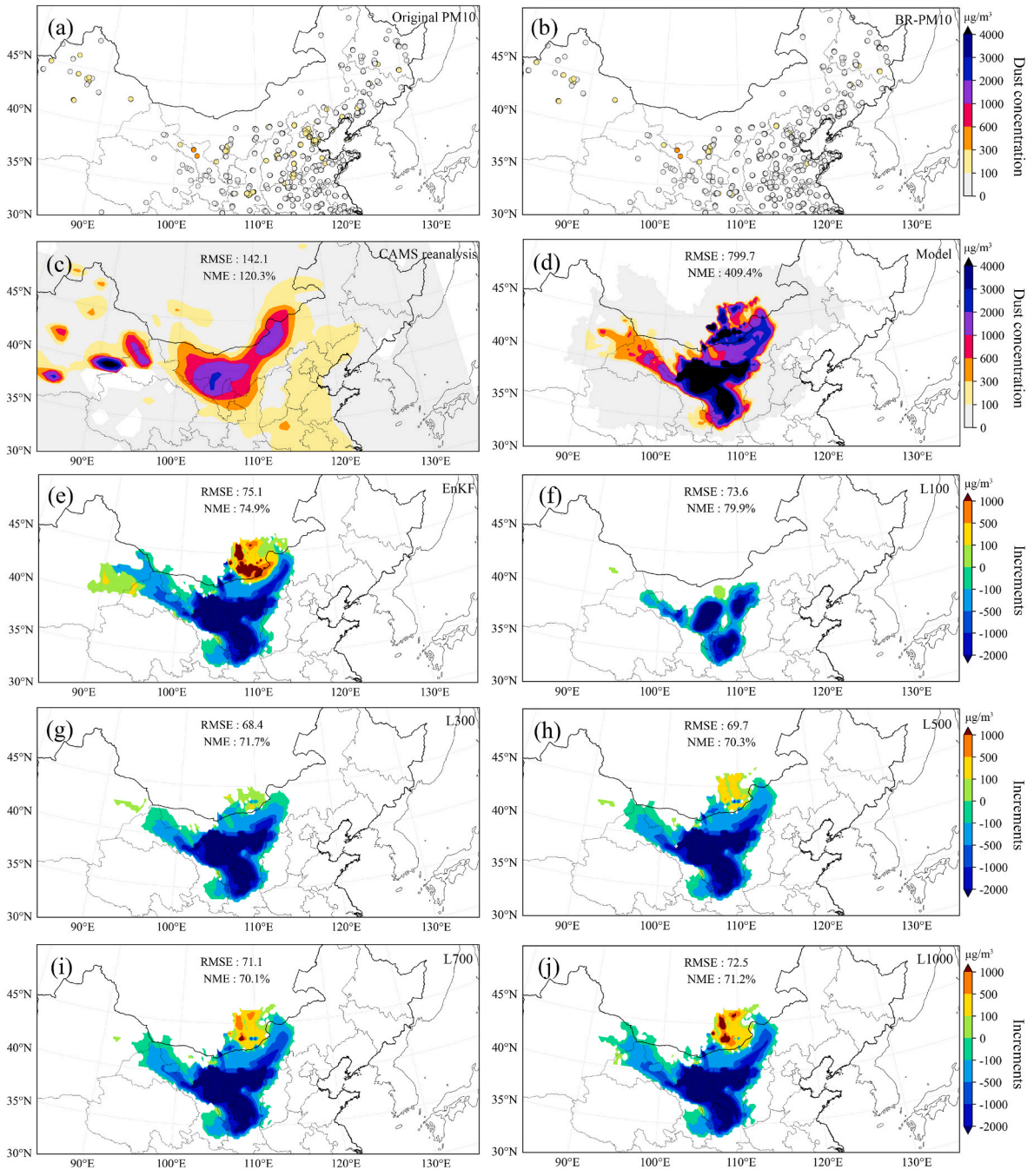


Fig. 6. Spatial distribution of ground-based raw PM_{10} (**a**), BR-PM_{10} concentration observations (**b**), reanalysis data of PM_{10} from CAMS at the surface level (**c**), the surface dust concentration (SDC) simulation from control model (**d**), increments from the posteriori SDC updated using EnKF (assimilation analysis field minus control model run, figures below are the same) (**e**), using LEnKF with L_{thres} = 100 km (**f**), using LEnKF with L_{thres} = 300 km (**g**), using LEnKF with L_{thres} = 500 km (**h**), using LEnKF with L_{thres} = 700 km (**i**) and using LEnKF with L_{thres} = 1000 km (**j**) at 08:00, 15th April 2021 (CST). BR-PM_{10}: baseline-removed PM_{10}.

consistent with the observations to some extent. However, there is plenty of overestimation from model in south-east front of the dust plume. It results in an RMSE as high as 676 μg/m^3 and NME of 142.1%. By assimilating BR-PM_{10} data (in panel b) using the EnKF (without localization), the dust plume simulations are corrected to a great extent. The overestimation in north-east front is greatly reduced with increments greater than −1000 μg/m^3 and part of dust load in north-west of the front is increased which better fits the observations with the RMSE drops to less than 300 μg/m^3 and NME to 64.3%. It indicates that our assimilation analysis is effective in dust field estimation. The posteriori dust field at the moment from the assimilation analysis can be seen in Fig. 5(e–j). The dust cloud in the southern Inner Mongolia, FWP and NCP is attenuated in all posteriors. The west part of the plume in Xinjiang remains consistent due to the lack of measurements available there.

In terms of RMSE and NME, localization is found to be effective in improving the assimilation analysis. The RMSE is 273 μg/m^3 and NME is 64.3% when the pure EnKF is employed, and it is further reduced to 201, 221.8, 234.1, 240.4, 246.3 μg/m^3 and 51.2%, 55.4%, 56.5%, 57.0%, 57.7% when the localization distance threshold L_{thres} is set to as 100, 300, 500, 700 and 1000 km respectively. Among them, L100 seems to have the largest capacity of resolving the measurements. However, there is the danger of breaking the model consistence of using such a low threshold, and hence bring worse prediction as will be discussed in Section 4.2. Meanwhile, it is also found that with longer distance

threshold, increments from assimilation analysis have a greater coverage and is closer to the pure EnKF analysis. For example, through comparison between EnKF (panel e) and L1000 (panel j), we can see a quite similar increment pattern. It is in accord with the localization principle, which is a long distance threshold can reserve more information about correlation in ensembles and thus resolve more distant observations.

Compared to dust simulation shown in Fig. 5(d–j), the CAMS reanalysis reproduced a broader coverage of high-polluted full-aerosol PM_{10} field. It resulted in better performance with RMSE of 385.7 μg/m^3 and NME of 105.5% than the pure LOTOS-EUROS model prediction. However, it didn't capture the main dust plume and underestimated the intensity of the dust as can be seen in comparison between panel (a) and panel (c). Compared to the CAMS reanalysis, all of assimilation analysis show a lower RMSE and NME.

Snapshots of original PM_{10}, BR-PM_{10}, CAMS reanalysis of PM_{10}, the control model and increments from posteriori dust load simulation for DSE2 are shown in Fig. 6. Only a small part of dust plume can be observed by the ground-based monitoring network at 8:00 15th April as shown in panel (b). While the control model run in panel (d) indicated that a larger piece of dust cloud has already arrived in the FWP in China. After first assimilation analysis at 8:00 15th April, an improved dust field is obtained. The west stretch of the dust plume remains almost the same owing to the scarcity of observations nearby. The north of the plume is strengthened. The south plume is greatly reduced to better fit the observations as there is an increment over −2000 μg/m^3 in south of the dust plume. The EnKF analysis effectively resolve the few PM_{10} concentration measurements that contains valuable dust information, and therefore the RMSE is reduced sharply from over 700 μg/m^3 down to less than 100 μg/m^3 and NME is reduced from over 400% to less than 80% in all assimilation sets. Differences between localization (panel f–j) and non-localization (panel d) are also quite small considering the RMSE and NME. However, in areas where observations are unavailable, there are differences clearly depicted. For example, distribution of increments from L100 in panel (f) is greatly different from the EnKF in panel (e). Much too short distance threshold limits the capability of assimilation algorithm to resolve the observations far away. So the north dust plume remains still. Similar for DSE2, analysis with longer distance threshold produces a comparable posteriori to EnKF especially for L1000. Impact on the following forecast will be examined. In addition, CAMS reanalysis showed a relatively low RMSE (142.1 μg/m^3) and NME (120.3%) compared to the priori simulation. But it outperforms far worse than our assimilation analysis.

4.2. Forecasting skill

The experiments discussed so far focus on the evaluation of the posteriori using the EnKF analysis, which would be used as the initial condition for dust forecasts. An overall validation of the short-term dust forecast for these two events is then carried out here.

As shown in Fig. 7 is the time series of RMSE and NME calculated from the dust forecast restarted by EnKF and LEnKF analysis during 8:00 to 20:00 in DSE1 and DSE2. In DSE1, as shown in panel (a) and panel (c), the forecast from the EnKF (blue dash line) is steadily slightly higher than the forecast where localization is applied. It indicates that the localization can improve the prediction quality effectively. Meanwhile, the optimal choice for improving the prediction is obtained when L_{thres} is set from 300 to 500 km. It is because that a longer localization distance threshold would be less efficient in removing the spurious spatial correction. While localized EnKF (L_{thres} = 100 km) did result in the best performance in the posteriori analysis, it only estimated the dust field within 200 km ($r = 2\ L_{thres}$) and would not nudge the dust states beyond this range. It can be best seen from the green line in panel (a) and panel (c). Although the L100 reaches the lowest RMSE (201 μg/m^3) and NME (51.2%) at 8:00, but the forecast error rapidly grows and is even worse than the forecast from the pure EnKF in the following instances. Similar pattern can also be seen in Fig. S1 which is the time series of the dust forecast starting from subsequent assimilation analysis.

In DSE2, the diversion of L100 is much more significant than that of the DSE1 as shown in Fig. 7(b and d) and Fig. S2. This is mainly because that only a small part of the whole dust plume was observed by the PM_{10} monitoring network at 8:00 and 11:00. The insufficient observations with L100 only estimated the dust field with a limited range, while the main dust plume keeps unchanged. Different from DSE1, diversion, though not that significant, also occurs in L300 at 8:00 and the RMSEs of L300 forecasts are slightly higher than other sets (L500, L700) in the following time window. Our forecasting system still produces stable prediction using EnKF with L_{thres} = 500 and 700 km as shown in Fig. 7.

Fig. 8 is the time series of hourly RMSE of the pure model forecast (light blue dash line) and improvement on 3-h RMSE using the assimilation-based forecast (histogram) starting from 5 assimilation time points (8:00, 11:00, 14:00, 17:00 and 20:00) and from 8:00 to 20:00 in DSE1 and DSE2. Note that the left axis represents the relative RMSE improvements with respect to the pure model run. The higher value represents the better performance. Evaluation results using NME are also shown in Fig. S3. This figure clearly shows the added value of assimilation compared to pure model run and impact of L_{thres} on LEnKF. In general, our assimilation-based forecasting system outperforms the pure model forecast in varying degrees over the whole tested window.

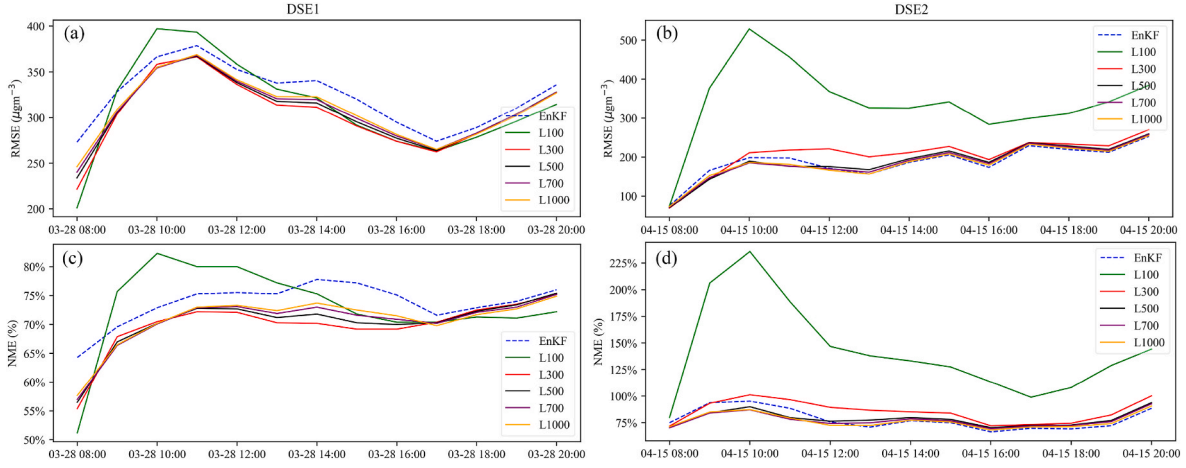


Fig. 7. Time series of RMSE and NME calculated from the dust forecast restarted by EnKF and LEnKF (L_{thres} = 100, 300, 500, 700, 1000 km) analysis during 8:00 to 20:00 in DSE1 (**a** and **c**) and DSE2 (**b** and **d**).

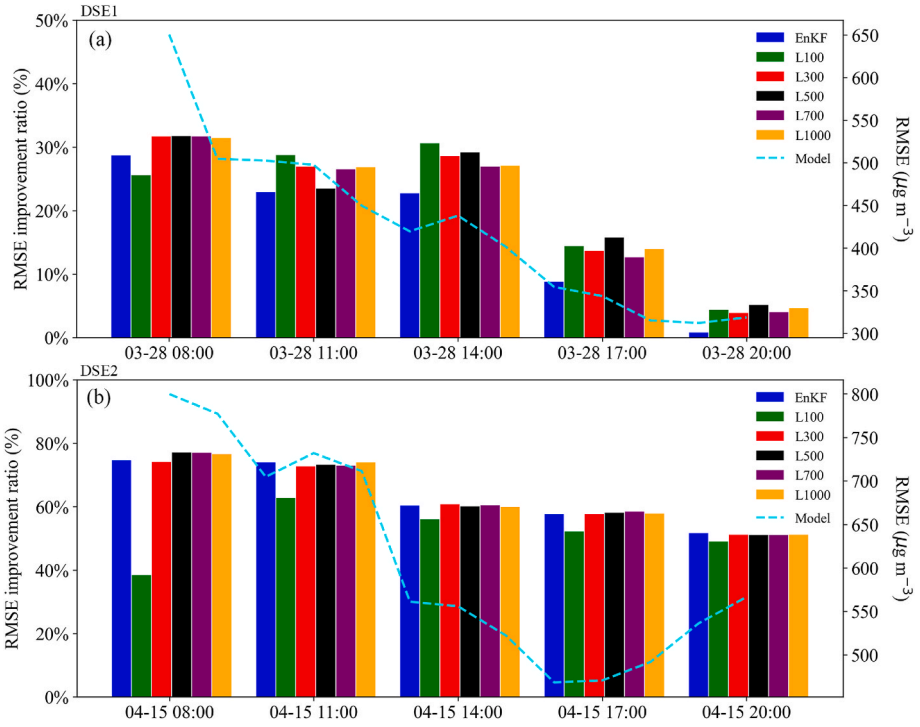


Fig. 8. Relative RMSE improvement of the 3-h assimilation-based forecast with respect to pure model forecast starting from 5 assimilation time points (8:00, 11:00, 14:00, 17:00 and 20:00) and time series of hourly RMSE from pure model forecast (light blue dash line) during 8:00 to 20:00 in DSE1 (**a**) and DSE2 (**b**). The left axis calculates the improvement ratio and the right represents the RMSE of the pure model forecast.

In DSE1, the RMSE of the LOTOS-EUROS forecast (light blue dash line) stayed at a relative high level constantly, RMSE ranging from 300 to over 600 μg/m^3 and NME ranging from 90% to 140%. Forwarded with the assimilation analysis, the RMSE is reduced remarkably and remains less than 300 μg/m^3 and NME is reduced to around 70%. Relative improvements using assimilation are around 30%. Most of the LEnKF-based forecast outperforms the EnKF-based forecast. The greatest reduction is in initial assimilation analysis at 8:00 given the maximum model RMSE. The L100 helps the least in the forecast from the first assimilation which can also be seen in Fig. 7(a). While in the following time points, L100 produces relatively high contribution to the forecast. It is caused by the high-frequency assimilation that corrects the cumulative errors. Among all experiment sets, L500 obtains the best performance in 3 out of 5 time points.

In DSE2, our forecasting system outperforms the pure model forecast to a considerable extent over the time window. The RMSE drops sharply from over 700 μg/m^3 (control model) to less than 100 μg/m^3 and remains stable around 200 μg/m^3. The NME drops from over 400% to less than 100%. Similar to DSE1, L100 in DSE2 achieves the worst performance on the first assimilation-based forecast. In fact, it is the worst one throughout the five time points. In contrast, all experiment sets reach comparable levels considering the RMSE and NME, which indicates that the sensitivity of forecast in DSE2 to L_{thres} is not very high. The localization only helps a little in this case. In general, if localization is enabled, L_{thres} between 500 and 700 km can be the best choice that wouldn't degrade the forecast. In both cases, localized EnKF with $L_{thres} = 100$ km would lead to a divergent model though it results in the best assimilation analysis. It indicates that it reaches the local optimum instead of the global one. L300 also has the risk of falling into a divergent model at 8:00 which can also be seen in Fig. 7(b). In fact, the localization aims to reduce the spurious correlation generated by the limited ensemble model simulation. If the localization distance threshold is too short, observations would only be used to tune the model simulation at the very local grids around the observations, while the whole simulated dust plume could not be fully estimated. It will break model consistency and thus the forecast goes divergent rapidly.

Conclusively, both of the EnKF and LEnKF would help in improving the forecast skill and the latter is steadily better than the former in DSE1. In DSE2, assimilation-based forecast is not improved a lot when longer L_{thres} is applied. Hence, the choice of L_{thres} can be case-independent. Meanwhile, with a high-frequency data assimilation and rolling forecast, the possible divergence of model can be alleviated. For our LEnKF for dust storm forecasting, L_{thres} around 500 km is shown to be the optimal choice that can improve the forecast to some degree and avoid the model divergence.

4.3. Forecast evaluation using AOD

To further evaluate our dust forecast system, the prediction is also compared to the MODIS Deep Blue AOD product described in Section 3.3. Fig. 9 shows snapshots of the baseline-removed MODIS AOD and the AOD prediction from the control model, EnKF and LEnKF ($L_{thres} = 500$ km) at 11:00 28th March. As is clearly shown in MODIS AOD in panel (a), the dust load is concentrated in the northeast China and it stretches to west. The priori model approximately matches the shape of MODIS AOD but overestimates the AOD in the front of the dust plume. Meanwhile, less-severe dust plume was observed in the Heilongjiang and east China (in blue box) from the MODIS AOD. However, the priori model doesn't reproduce this pattern as shown in Fig. 9(b), and the plume there is neither captured by the EnKF and L300 simulation. This inconsistence may arise from the MODIS AOD retrieval error, and this part of data is not successfully excluded by our data quality control described in Jin et al. (2022). The strength of AOD from EnKF and LEnKF are reduced after assimilation at 8:00 which is in better harmony with the MODIS AOD compared to the priori model, and the RMSE drops from 0.938 to 0.752 and 0.713, respectively.

Another snapshot of MODIS AOD, the a priori and the posteriori AOD simulation during the DSE2 is shown Fig. 10. The AOD measurements in panel (a) indicated that the dust plume is concentrated in the Inner Mongolia of China. The AOD simulation of the control model in panel

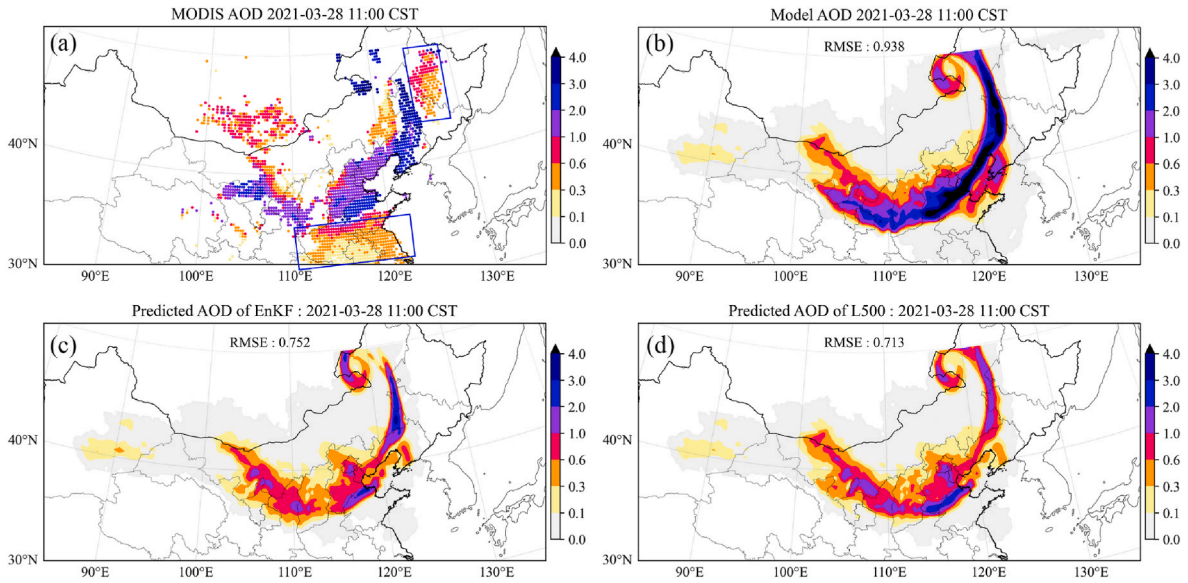


Fig. 9. MODIS Deep Blue AOD at 550 nm (**a**), control model AOD (**b**), predicted AOD of EnKF (**c**) and L500 (**d**) (model restarted and run from 8:00 after the initial assimilation at 8:00) on 11:00 28th March 2021 (CST).

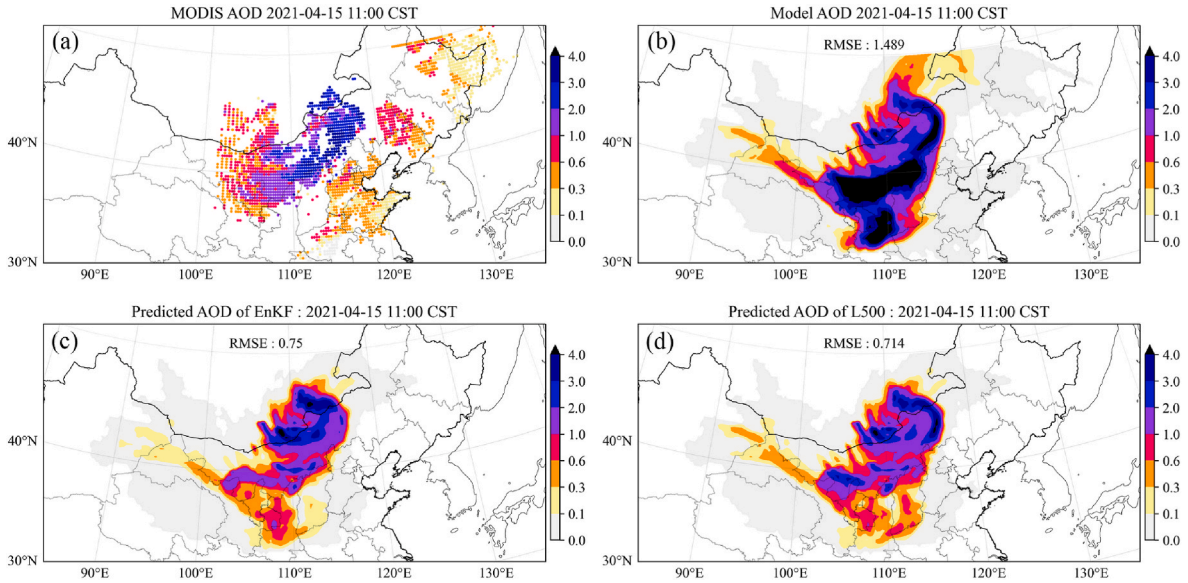


Fig. 10. MODIS Deep Blue AOD at 550 nm (**a**), control model AOD (**b**), predicted AOD of EnKF (**c**) and L500 (**d**) (model restarted and run from 8:00 after the initial assimilation at 8:00) on 11:00 15th April 2021 (CST).

(b) has a more extent coverage and greater intensity than the MODIS AOD. The assimilation-based prediction in panel (c-d) presents less-severe AODs that stretches to the south, and the RMSE drops from 1.489 to 0.75 and 0.714 respectively.

5. Conclusions

Super dust storms reoccurred in East Asia in 2021 spring after an absence of two decades, and casted great health damages and property losses. High-quality dust storm forecasts are therefore in demand for reducing the adverse impacts. A dust forecasting system has been developed in this study through combining a chemical transport model LOTOS-EUROS and EnKF/LEnKF data assimilation algorithms. Measurements are assimilated when they are available to estimate the dust state, which are then used as the initial condition for new forecast. The system was tested in the 2021 super dust storms, and the ground-based PM_{10} concentration measurements covering all over the China were assimilated. Sequential assimilation tests show that our system can produce a promising dust forecast than pure model-based forecast. Main features of the dust plume were reproduced, and the forecast agreed with the PM_{10} concentration data well. Parts of dust simulation error remained, which is because our ensemble members cannot fully represent the position uncertain in the dust transport, and therefore is unable to resolve all the observations. Independent evaluations are carried out, our assimilation analysis of dust concentration field is validated to outperform the operational CAMS product. Besides, the improvements on dust prediction were also validated through a comparison against the independent MODIS Deep Blue AODs.

Experiments were performed to compare EnKF and LEnKF, and to explore the sensitivities of dust forecasting to the localization distance threshold in LEnKF. In terms of the assimilation analysis, LEnKF achieved lower RMSE than the EnKF steadily in DSE1. The best performance

was obtained when a shortest localization distance threshold L_{thres} = 100 km was used. It is because the LEnKF with the shortest-distance localization has the largest freedom to resolve the measurements. It is also found that with a longer distance threshold, LEnKF can produce a similar result compared to EnKF. When we come to the forecasting skill, however, it was found that even though L100 achieves best performance in assimilation analysis, it would break the model consistency. The forecast onward from the estimated field diverges rapidly. Meanwhile, with a high-frequency assimilation, the errors arise from much too short distance threshold can be alleviated to some degree. On the other hand, forecast with a long localization distance is also unsatisfying as it is less efficient in removing the spurious spatial correction especially in DSE1. Sensitivity of forecast in DSE2 to distance threshold is not as high as in DSE1. While much too short L_{thres} can still degrade the forecast. If localization is applied, considering both cases, distance threshold of around 500 km is proved to be the optimal choice that can resolve the observations and avoid the model divergence.

Credit author statement

JJ: designed PyFilter and dust forecasting system. **MP**: wrote the PyFilter and carried out the data assimilation experiments. **AS**: The manuscript is prepared with contributions from. **HJ**: The manuscript is prepared with contributions from. **LF**: The manuscript is prepared with contributions from. **HXL**: The manuscript is prepared with contributions from. **HL**: The manuscript is prepared with contributions from.

Declaration of competing interest

There is no conflict of interest.

Data availability

Data will be made available on request.

Acknowledgments

This work is supported by the National Natural Science Foundation of China [grant No. 42105109] and Natural Science Foundation of Jiangsu Province (NO. BK20210664).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.atmosenv.2023.119831.

References

Bergamaschi, P., Krol, M., Meirink, J.F., Dentener, F., Segers, A., van Aardenne, J., et al., 2010. Inverse modeling of European CH_4 emissions 2001–2006. J. Geophys. Res. 115 (D22), D22309 https://doi.org/10.1029/2010JD014180.

Bertino, L., Evensen, G., Wackernagel, H., 2003. Sequential data assimilation techniques in oceanography. Int. Stat. Rev. 71 (2), 223–241. https://doi.org/10.1111/j.1751-5823.2003.tb00194.x.

Bocquet, M., Elbern, H., Eskes, H., Hirtl, M., Žabkar, R., Carmichael, G.R., et al., 2015. Data assimilation in atmospheric chemistry models: current status and future prospects for coupled chemistry meteorology models. Atmos. Chem. Phys. 15 (10), 5325–5358. https://doi.org/10.5194/acp-15-5325-2015.

Burgers, G., Jan van Leeuwen, P., Evensen, G., 1998. Analysis scheme in the ensemble Kalman filter. Mon. Weather Rev. 126 (6), 1719–1724. https://doi.org/10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2.

Chen, L., Walsh, M., 2021. Vast Sandstorms Expose Mongolia's Long-Ignored Ecological Crisis. Retrieved from. https://asia.nikkei.com/Spotlight/Caixin/Vast-sandstorms-expose-Mongolia-s-long-ignored-ecological-crisis.

Corazza, M., Bergamaschi, P., Vermeulen, A.T., Aalto, T., Haszpra, L., Meinhardt, F., et al., 2011. Inverse modelling of European N_2O emissions: assimilating observations from different networks. Atmos. Chem. Phys. 11 (5), 2381–2398. https://doi.org/10.5194/acp-11-2381-2011.

Curier, R.L., Timmermans, R., Calabretta-Jongen, S., Eskes, H., Segers, A., Swart, D., Schaap, M., 2012. Improving ozone forecasts over Europe by synergistic use of the LOTOS-EUROS chemical transport model and in-situ measurements. Atmos. Environ. 60, 217–226. https://doi.org/10.1016/j.atmosenv.2012.06.017.

Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J. Geophys. Res. 99 (C5), 10143 https://doi.org/10.1029/94JC00572.

Filonchyk, M., 2022. Characteristics of the severe March 2021 Gobi Desert dust storm and its impact on air pollution in China. Chemosphere 287, 132219. https://doi.org/10.1016/j.chemosphere.2021.132219.

Gaspari, G., Cohn, S.E., 1999. Construction of correlation functions in two and three dimensions. Q. J. R. Meteorol. Soc. 125 (554), 723–757. https://doi.org/10.1002/qj.49712555417.

Gong, S.L., Zhang, X.Y., 2008. CUACE/Dust – an integrated system of observation and modeling systems for operational dust forecasting in Asia. Atmos. Chem. Phys. 8 (9), 2333–2340. https://doi.org/10.5194/acp-8-2333-2008.

Guo, H., Gu, X., Ma, G., Shi, S., Wang, W., Zuo, X., Zhang, X., 2019. Spatial and temporal variations of air quality and six air pollutants in China during 2015–2017. Sci. Rep. 9 (1), 15201 https://doi.org/10.1038/s41598-019-50655-6.

Guo, L., Fan, B., Zhang, F., Jin, Z., Lin, H., 2018. The clustering of severe dust storm occurrence in China from 1958 to 2007. J. Geophys. Res. Atmos. 123 (15), 8035–8046. https://doi.org/10.1029/2018JD029042.

Hamill, T.M., Whitaker, J.S., Snyder, C., 2001. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. Mon. Weather Rev. 129 (11), 2776–2790. https://doi.org/10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2.

Han, J., Dai, H., Gu, Z., 2021. Sandstorms and desertification in Mongolia, an example of future climate events: a review. Environ. Chem. Lett. 19 (6), 4063–4073. https://doi.org/10.1007/s10311-021-01285-w.

Houtekamer, P.L., Mitchell, H.L., 2001. A sequential ensemble Kalman filter for atmospheric data assimilation. Mon. Weather Rev. 129 (1), 123–137. https://doi.org/10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2.

Houtekamer, P.L., Zhang, F., 2016. Review of the ensemble Kalman filter for atmospheric data assimilation. Mon. Weather Rev. 144 (12), 4489–4532. https://doi.org/10.1175/MWR-D-15-0440.1.

Inness, A., Ades, M., Agustí-Panareda, A., Barré, J., Benedictow, A., Blechschmidt, A.-M., et al., 2019. The CAMS reanalysis of atmospheric composition. Atmos. Chem. Phys. 19 (6), 3515–3556. https://doi.org/10.5194/acp-19-3515-2019.

Jin, J., Lin, H.X., Heemink, A., Segers, A., 2018. Spatially varying parameter estimation for dust emissions using reduced-tangent-linearization 4DVar. Atmos. Environ. 187, 358–373. https://doi.org/10.1016/j.atmosenv.2018.05.060.

Jin, J., Segers, A., Heemink, A., Yoshida, M., Han, W., Lin, H.-X., 2019a. Dust emission inversion using Himawari-8 AODs Over East Asia: an extreme dust event in may 2017. J. Adv. Model. Earth Syst. 11 (2), 446–467. https://doi.org/10.1029/2018MS001491.

Jin, J., Lin, H.X., Segers, A., Xie, Y., Heemink, A., 2019b. Machine learning for observation bias correction with application to dust storm data assimilation. Atmos. Chem. Phys. 19 (15), 10009–10026. https://doi.org/10.5194/acp-19-10009-2019.

Jin, J., Segers, A., Liao, H., Heemink, A., Kranenburg, R., Lin, H.X., 2020. Source backtracking for dust storm emission inversion using an adjoint method: case study of Northeast China. Atmos. Chem. Phys. 20 (23), 15207–15225. https://doi.org/10.5194/acp-20-15207-2020.

Jin, J., Segers, A., Lin, H.X., Henzing, B., Wang, X., Heemink, A., Liao, H., 2021. Position correction in dust storm forecasting using LOTOS-EUROS v2.1: grid-distorted data assimilation v1.0. Geosci. Model Dev. 14 (9), 5607–5622. https://doi.org/10.5194/gmd-14-5607-2021.

Jin, J., Pang, M., Segers, A., Han, W., Fang, L., Li, B., et al., 2022. Inverse modeling of the 2021 spring super dust storms in East Asia. Atmos. Chem. Phys. 22 (10), 6393–6410. https://doi.org/10.5194/acp-22-6393-2022.

Joshi, J.R., 2021. Quantifying the impact of cropland wind erosion on air quality: a high-resolution modeling case study of an Arizona dust storm. Atmos. Environ. 263, 118658 https://doi.org/10.1016/j.atmosenv.2021.118658.

Judd, K., 2003. Nonlinear state estimation, indistinguishable states, and the extended Kalman filter. Phys. Nonlinear Phenom. 183 (3–4), 273–281. https://doi.org/10.1016/S0167-2789(03)00180-5.

Kok, J.F., Mahowald, N.M., Fratini, G., Gillies, J.A., Ishizuka, M., Leys, J.F., et al., 2014. An improved dust emission model – Part 1: model description and comparison against measurements. Atmos. Chem. Phys. 14 (23), 13023–13041. https://doi.org/10.5194/acp-14-13023-2014.

Kong, Y., Sheng, L., Li, Y., Zhang, W., Zhou, Y., Wang, W., Zhao, Y., 2021. 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.

Kontos, S., Kakosimos, K., Liora, N., Poupkou, A., Melas, D., 2021. Towards a regional dust modeling system in the central Middle East: evaluation, uncertainties and recommendations. Atmos. Environ. 246, 118160 https://doi.org/10.1016/j.atmosenv.2020.118160.

Kranenburg, R., Segers, A.J., Hendriks, C., Schaap, M., 2013. Source apportionment using LOTOS-EUROS: module description and evaluation. Geosci. Model Dev. 6 (3), 721–733. https://doi.org/10.5194/gmd-6-721-2013.

Kurosaki, Y., Mikami, M., 2003. Recent frequent dust events and their relation to surface wind in East Asia. Geophys. Res. Lett. 30 (14) https://doi.org/10.1029/2003GL017261.

Leeuwen, P.J., Künsch, H.R., Nerger, L., Potthast, R., Reich, S., 2019. Particle filters for high-dimensional geoscience applications: a review. Q. J. R. Meteorol. Soc. 145 (723), 2335–2365. https://doi.org/10.1002/qj.3551.

Lei, L., Anderson, J.L., 2014. Comparisons of empirical localization techniques for serial ensemble Kalman filters in a simple atmospheric general circulation model. Mon. Weather Rev. 142 (2), 739–754. https://doi.org/10.1175/MWR-D-13-00152.1.

Lin, C., Zhu, J., Wang, Z., 2008. Model bias correction for dust storm forecast using ensemble Kalman filter. J. Geophys. Res. 113 (D14), D14306 https://doi.org/10.1029/2007JD009498.

Liu, S., Xing, J., Sahu, S.K., Liu, X., Liu, S., Jiang, Y., et al., 2021. Wind-blown dust and its impacts on particulate matter pollution in Northern China: current and future scenarios. Environ. Res. Lett. 16 (11), 114041 https://doi.org/10.1088/1748-9326/ac31ec.

Lorente, A., Folkert Boersma, K., Yu, H., Dörner, S., Hilboll, A., Richter, A., et al., 2017. Structural uncertainty in air mass factor calculation for NO_2 and HCHO satellite retrievals. Atmos. Meas. Tech. 10 (3), 759–782. https://doi.org/10.5194/amt-10-759-2017.

Manders, A.M.M., Builtjes, P.J.H., Curier, L., Denier van der Gon, H.A.C., Hendriks, C., Jonkers, S., et al., 2017. Curriculum vitae of the LOTOS–EUROS (v2.0) chemistry transport model. Geosci. Model Dev. 10 (11), 4145–4173. https://doi.org/10.5194/gmd-10-4145-2017.

Morcrette, J.-J., Beljaars, A., Benedetti, A., Jones, L., Boucher, O., 2008. Sea-salt and dust aerosols in the ECMWF IFS model. Geophys. Res. Lett. 35 (24), L24813 https://doi.org/10.1029/2008GL036041.

Morcrette, J.-J., Boucher, O., Jones, L., Salmond, D., Bechtold, P., Beljaars, A., et al., 2009. Aerosol analysis and forecast in the European centre for medium-range weather forecasts integrated forecast system: forward modeling. J. Geophys. Res. 114 (D6), D06206 https://doi.org/10.1029/2008JD011235.

Niu, T., Gong, S.L., Zhu, G.F., Liu, H.L., Hu, X.Q., Zhou, C.H., Wang, Y.Q., 2008. Data assimilation of dust aerosol observations for the CUACE/dust forecasting system. Atmos. Chem. Phys. 8 (13), 3473–3482. https://doi.org/10.5194/acp-8-3473-2008.

Osores, S., Ruiz, J., Folch, A., Collini, E., 2020. Volcanic ash forecast using ensemble-based data assimilation: an ensemble transform Kalman filter coupled with the FALL3D-7.2 model (ETKF–FALL3D version 1.0). Geosci. Model Dev. 13 (1), 1–22. https://doi.org/10.5194/gmd-13-1-2020.

Qian, W., Quan, L., Shi, S., 2002. Variations of the dust storm in China and its climatic control. J. Clim. 15 (10), 1216–1229. https://doi.org/10.1175/1520-0442(2002)015<1216:VOTDSI>2.0.CO;2.

Rayegani, B., Barati, S., Goshtasb, H., Gachpaz, S., Ramezani, J., Sarkheil, H., 2020. Sand and dust storm sources identification: a remote sensing approach. Ecol. Indicat. 112, 106099 https://doi.org/10.1016/j.ecolind.2020.106099.

Shao, Y., Dong, C.H., 2006. A review on East Asian dust storm climate, modelling and monitoring. Global Planet. Change 52 (1–4), 1–22. https://doi.org/10.1016/j.gloplacha.2006.02.011.

Skoulidou, I., Koukouli, M.-E., Manders, A., Segers, A., Karagkiozidis, D., Gratsea, M., et al., 2021. Evaluation of the LOTOS-EUROS NO_2 simulations using ground-based measurements and S5P/TROPOMI observations over Greece. Atmos. Chem. Phys. 21 (7), 5269–5288. https://doi.org/10.5194/acp-21-5269-2021.

Tanaka, T.Y., Chiba, M., 2006. A numerical study of the contributions of dust source regions to the global dust budget. Global Planet. Change 52 (1–4), 88–104. https://doi.org/10.1016/j.gloplacha.2006.02.002.

Timmermans, R., Kranenburg, R., Manders, A., Hendriks, C., Segers, A., Dammers, E., et al., 2017. Source apportionment of $PM_{2.5}$ across China using LOTOS-EUROS. Atmos. Environ. 164, 370–386. https://doi.org/10.1016/j.atmosenv.2017.06.003.

Wang, R., Liu, B., Li, H., Zou, X., Wang, J., Liu, W., et al., 2017. Variation of strong dust storm events in Northern China during 1978–2007. Atmos. Res. 183, 166–172. https://doi.org/10.1016/j.atmosres.2016.09.002.

Wang, Y.Q., Zhang, X.Y., Gong, S.L., Zhou, C.H., Hu, X.Q., Liu, H.L., et al., 2008. Surface observation of sand and dust storm in East Asia and its application in CUACE/Dust. Atmos. Chem. Phys. 8 (3), 545–553. https://doi.org/10.5194/acp-8-545-2008.

World Meteorological Organization, 2018. WMO Airborne Dust Bulletin: Sand and Dust Storm Warning Advisory and Assessment System. Retrieved from. https://library.wmo.int/index.php?lvl=bulletin_display&id=3948.

Xie, Y., Zhang, W., Qu, J., 2017. Detection of Asian dust storm using MODIS measurements. Rem. Sens. 9 (8), 869. https://doi.org/10.3390/rs9080869.

Xu, X., Zhao, P., Yin, Y., Cheng, W., Wang, J., Li, P., et al., 2022. Dust particles transport during the rare strong sandstorm process in Northern China in early year 2021. Air Qual. Atmos. Health. https://doi.org/10.1007/s11869-022-01159-2.

Yin, Z., Wan, Y., Zhang, Y., Wang, H., 2021. Why super sandstorm 2021 in North China? Natl. Sci. Rev. https://doi.org/10.1093/nsr/nwab165 nwab165.

Yuan, H., Zhang, D., Shi, Y., Li, B., Yang, J., Yu, X., et al., 2017. Cell concentration, viability and culture composition of airborne bacteria during a dust event in Beijing. J. Environ. Sci. 55, 33–40. https://doi.org/10.1016/j.jes.2016.03.033.

Zender, C.S., Bian, H., Newman, D., 2003. Mineral dust entrainment and deposition (DEAD) model: description and 1990s dust climatology. J. Geophys. Res. Atmos. 108 (D14) https://doi.org/10.1029/2002JD002775.

Zhang, X., Sharratt, B., Liu, L., Wang, Z., Pan, X., Lei, J., et al., 2018. East Asian dust storm in May 2017: observations, modelling, and its influence on the Asia-Pacific region. Atmos. Chem. Phys. 18 (11), 8353–8371. https://doi.org/10.5194/acp-18-8353-2018.

Zhang, Z., Wu, W., Fan, M., Tao, M., Wei, J., Jin, J., et al., 2019. Validation of Himawari-8 aerosol optical depth retrievals over China. Atmos. Environ. 199, 32–44. https://doi.org/10.1016/j.atmosenv.2018.11.024.

Zhou, C.H., Gong, S.L., Zhang, X.Y., Wang, Y.Q., Niu, T., Liu, H.L., et al., 2008. Development and evaluation of an operational SDS forecasting system for East Asia: CUACE/Dust. Atmos. Chem. Phys. 8 (4), 787–798. https://doi.org/10.5194/acp-8-787-2008.

Zou, X.K., Zhai, P.M., 2004. Relationship between vegetation coverage and spring dust storms over northern China: relationship between NDVI and dust storm. J. Geophys. Res. 109 (D3) https://doi.org/10.1029/2003JD003913.