



The sensitivity of aerosol data assimilation to vertical profiles: case study of dust storm assimilation with LOTOS-EUROS v2.2

Mijie Pang^{1,2}, Jianbing Jin¹, Ting Yang³, Xi Chen¹, Arjo Segers⁴, Batjargal Buyantogtokh⁵, Yixuan Gu¹, Jiandong Li¹, Hai Xiang Lin^{2,6}, Hong Liao¹, and Wei Han⁷

¹State Key Laboratory of Climate System Prediction and Risk Management, 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 & Technology, Nanjing, China

²Delft Institute of Applied Mathematics, Delft University of Technology, Delft, the Netherlands

³State Key Laboratory of Atmospheric Boundary Layer Physics and Atmospheric Chemistry (LAPC), Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

⁴Department of Climate, Air and Sustainability, TNO, Utrecht, the Netherlands

⁵Information and Research Institute of Meteorology, Hydrology and Environment, Ulaanbaatar, Mongolia

⁶Institute of Environmental Sciences, Leiden University, Leiden, the Netherlands

⁷Center for Earth System Modeling and Prediction, Chinese Meteorological Administration, Beijing, China

Correspondence: Jianbing Jin (jianbing.jin@nuist.edu.cn) and Wei Han (hanwei@cma.gov.cn)

Received: 14 June 2024 – Discussion started: 19 September 2024

Revised: 13 February 2025 – Accepted: 27 March 2025 – Published: 25 June 2025

Abstract. Modelling and observational techniques are pivotal in aerosol research, yet each approach exhibits inherent limitations. Aerosol observation is constrained by its limited spatial and temporal coverage compared to that of models. On the other hand, models tend to possess higher uncertainties and biases compared to observations. Aerosol data assimilation has gained popularity as it combines the advantages of both methods. Despite numerous studies in this domain, few have addressed the challenges faced in assimilating aerosol data with significant differences in magnitude and degree of freedom between the model state and observations, especially in the vertical direction. These challenges can lead to the preservation – or even the exacerbation – of structural inaccuracies within the assimilation process. This study investigates the sensitivity of dust aerosol data assimilation to the vertical structure of the aerosol profile. We assimilate a variety of dust observations, encompassing ground-based particulate matter (PM₁₀) measurements, and satellite-derived dust optical depth (DOD) data, using the ensemble Kalman filter (EnKF). The assimilation process is elucidated, detailing the assimilation of raw ground-based and satellite-based observations for an optimized three-dimensional (3D)

posterior state. To demonstrate the impact of accurate vs. erroneous prior aerosol vertical profiles on the assimilation result, we select three cases of super dust storms for analysis. Our findings reveal that the assimilation of ground observations would optimize the dust field at the ground in general. However, the vertical structure presents a more complex challenge. When the prior profile accurately reflects the true vertical structure, the assimilation process can successfully preserve this structure. Conversely, if the prior profile introduces an incorrect structure, the assimilation can significantly deteriorate the integrity of the aerosol profile. This is also found in the assimilation of DOD, which exhibits a comparable pattern in its sensitivity to the initial aerosol profile's accuracy.

1 Introduction

In recent decades, atmospheric aerosol has garnered extensive attention due to its impact on the environment (Buseck and Pósfai, 1999; Zhao et al., 2020) and human health (Zhu et al., 2021). Numerous research studies have been conducted

on the characterization (Mhawish et al., 2021), source apportionment (Wu et al., 2018), and model simulation (Lee et al., 2011) of aerosols. One crucial tool for understanding aerosol processes and forecasting is the chemical transport model (CTM), which can generate continuous three-dimensional (3D) aerosol fields on a large scale. However, despite significant efforts to develop and improve CTM parameter schemes, studies have pointed out that significant uncertainties still exist in CTM simulations (Vignati et al., 2010; Stier et al., 2013). These uncertainties primarily arise from emission sources (e.g. intensity, location), meteorological inputs (e.g. wind mixing, precipitation), and incomplete parameter schemes (e.g. chemical reactions) (Benedetti et al., 2018). At the same time, aerosol measurements have advanced, owing to the development of sensors and increased investment. These include chemical kinetics studies in laboratories (Kolb and Worsnop, 2012), a network of loosely distributed ground monitoring stations (Gueymard and Yang, 2020), and regularly scanning satellites providing global coverage (Sogacheva et al., 2020). Although these measurements offer higher accuracy, challenges persist. So far, obtaining continuous 3D aerosol measurements on a large scale remains impractical and costly. Measurements are limited to surface-level concentrations of total particle matter, column-integrated aerosol optical depths (AODs), and vertical profiles of total aerosol extinction. These measurements have fewer degrees of freedom compared to the complexity of reality (Pappalardo et al., 2014; Qin et al., 2021).

Data assimilation incorporates models and measurements to generate a posterior that holds the strengths of both (Bannister, 2017). Various data assimilation techniques have been developed for different applications, including hydrology (Reichle et al., 2002), geology (Peng and Liu, 2022), and atmospheric science (Hamill, 2006). These methods are typically based on either variational or ensemble approaches (Whitaker et al., 2009; Law and Stuart, 2012). The variational method is particularly useful for emission inversion (Bergamaschi et al., 2022; Jin et al., 2023a). Ensemble-based filters have also demonstrated superior performance in sequential forecasting without the adjoint or tangent linear model required by the variational method. By assimilating real-time observations, these filters generate optimized initial fields that enhance forecast accuracy (Houtekamer et al., 2005). Since the introduction of the classic ensemble Kalman filter (EnKF) (Evensen, 1994), a considerable amount of research has focused on its application and improvement. Several variations of the EnKF, such as EnSRF (Whitaker and Hamill, 2002) and LETKF (Hunt et al., 2007), have been proposed to address computational complexity, accuracy, and robustness issues. Although these advancements have been validated, challenges related to the model's background error covariance and observational data quality control can still impact the assimilation results significantly.

In data assimilation methodology, an observation operator is required for deriving the innovation vector that quan-

tifies the difference between the observation field and the simulation field (Ma et al., 2020). These two usually have different degrees of freedom. With that, the optimized state could then be calculated considering the integration of background and observational penalties. From the perspective of aerosol data assimilation, the main objective is to reproduce optimal aerosol states (Liu et al., 2011), considering their spatial, vertical, aerosol species, and size distribution characteristics, while currently available observations commonly measure their mixed state. For example, AOD is the column-integrated optical extinction of all aerosols, and high uncertainties exist in the optical properties of aerosol (Tsikerdekis et al., 2021). Ground PM₁₀ concentration measurement is the additive sum of all particles with a diameter less than 10 µm. The ground PM₁₀ measurements are not directly comparable to the aerosol state. For ground measurements, monitoring stations are sparsely distributed, making high data vacancy in AOD products inevitable as a result of retrieval failures. The number of observations is therefore commonly much lower than the model state, or the degree of freedom is limited compared to that of the model state. Mathematically, the assimilation result strongly depends on the background error covariance for transferring the increments from the low-degree observational space to the full model space. Given a partially erroneous prior, the assumptions about the prior states can be reserved – or even exacerbated – with the lack of observations. Only a few studies have focused on some of these aspects. We recently found that the incorrect assumption of aerosol size distribution from prior models gives rise to a worse posterior (Jin et al., 2023b). In addition, the position error in the dust plume simulation can lead to the failure of assimilating valid observations (Jin et al., 2021). With regard to the vertical structure, most studies have focused on proposing relevant assimilation methods to assimilate observations, such as the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIPSO) (Sekiyama et al., 2010; Cheng et al., 2019; Ye et al., 2021) and light detection and ranging (lidar) (Wang et al., 2022). The sensitivities of assumptions in the prior simulations need to be explored.

For example, in dust aerosol data assimilation, the most representative observation sources are direct observations, like PM₁₀ and PM_{2.5} concentrations, and indirect observations, like dust optical depth (DOD) and extinction coefficient (Jin et al., 2022). Since they are all a combination of different particles, it is necessary to perform bias correction before assimilation (Jin et al., 2019; Ma et al., 2020). Afterwards, through assimilation analysis, few observations can optimize the whole model state space, while in reality, ground stations produce scattered concentrations on the ground, and satellites and lidars receive column-integrated information about the aerosol or a single profile once a day (Sekiyama et al., 2010; Hofer et al., 2017; Cheng et al., 2019; Escribano et al., 2022). None of them can provide continuous vertical information about the aerosol with large spatial coverage. When there is incorrect information about the dust

aerosol structure in the model prior, due to the considerably greater degree of freedom of the model state, data assimilation algorithms may fail to correct it or, even worse, further degrade the vertical dust aerosol loading.

There is limited research addressing the impact of flawed aerosol prior assumptions on the degradation of the assimilated posterior, particularly concerning vertical structures. Gwyther et al. (2023) delved into the implications of 4DVar data assimilation for mesoscale eddy representation in the vertical dimension, highlighting that inadequate prior knowledge of these structures significantly hinders improvement in the assimilated simulations. This issue extends to aerosol data assimilation, where erroneous assumptions regarding aerosol vertical distribution cause assimilation methods to maintain or exacerbate inaccuracies. This study focuses on several major dust storm events that occurred in the spring of 2021, utilizing them as case studies. We conduct separate data assimilation experiments involving ground-level PM₁₀ measurements and satellite-based AOD data to demonstrate explicitly how misleading vertical prior information can distort the posterior results. The LOTOS-EUROS model facilitates the reproduction of dust dynamics and generates the initial dust load priors. For assimilation, we employ an EnKF algorithm, incorporating bias-corrected PM₁₀ data from ground stations and DOD retrievals from the Himawari-8 satellite. For validation, we employ aerosol profile observations from CALIPSO and lidar systems. By presenting both favourable and detrimental scenarios, this work underscores the high sensitivity of data assimilation processes to the accuracy of assumed vertical structures, thereby emphasizing the criticality of precise vertical profiling in aerosol studies.

The remainder is organized as follows. Section 2 describes the model used for the dust aerosol simulation and the dust observations used in this study. Section 3 introduces the assimilation algorithm applied to assimilate the observations and the experiment settings. Section 4 discusses the assimilation results and validates the results with the CALIPSO and lidar observations. Finally, Sect. 5 summarizes the study.

2 Dust aerosol simulation and observations

This section introduces the model used for simulating dust aerosol and describes the observation methods applied for monitoring dust, including a ground monitoring station, Himawari-8, CALIPSO, and lidar. Among these, ground PM₁₀ from ground monitoring stations and DOD from Himawari-8 are used for assimilation. Extinction profiles from CALIPSO and lidar are used to validate the posterior profile.

2.1 Dust simulation

2.2 LOTOS-EUROS

LOTOS-EUROS v2.2 is used to simulate the dust aerosol. The LOTOS-EUROS model is a 3D chemistry transport model used for air quality forecasting (Manders et al., 2017). It has also been applied in source apportionment and emission inversion worldwide. In this study, the modelling domain spans 15 to 50°N and 70 to 140°E with a spatial resolution of $0.25^\circ \times 0.25^\circ$. Vertically, it comprises 21 layers with a top level at 10 km, which is adequate for recognizing the vertical structure. An ECMWF operational forecast of 3 h is used to drive the model. The boundary conditions are set to zero assuming that all the dust aerosols are emitted during the simulation window. Dust aerosol processes, including emission, advection, diffusion, deposition, and sedimentation, are considered in the model.

All simulations commence 1 d prior to the initial assimilation time point, during which no dust emission occurs. The dust emission process is modelled using the Zender03 emission parameterization scheme (Zender et al., 2003). In general, we assign the dust simulation uncertainty to the dust emission. Ensemble emission fields $[f_1, \dots, f_N]$ are generated randomly following the emission uncertainty choice f_{priori} and background error covariance matrices $\mathbf{B}$ in Jin et al. (2022). They are used to forward the LOTOS-EUROS model $\mathcal{M}$ for the ensemble dust simulations $[x_1, \dots, x_N]$ as

$$[x_1, \dots, x_N] = [\mathcal{M}(f_1), \dots, \mathcal{M}(f_N)]. \quad (1)$$

Here, N refers to the total ensemble number.

2.3 Dust observations for assimilation

2.3.1 Ground PM₁₀

There are over 1600 ground air quality monitoring stations located throughout China. Their spatial distribution can be seen in Fig. 1 (blue scatters). They provide hourly air pollutant concentrations, including PM_{2.5}, PM₁₀, O₃, NO₂, SO₂, and CO. PM₁₀ can appropriately represent the dust load as it falls under the size definition of fine and coarse dust (Adebisi et al., 2023). Its uncertainty is also relatively lower than that of some remote sensing instruments. However, PM₁₀ is a combination of various species. Assimilating such observations with inaccurate or incorrect representativeness can diverge the model. It is then necessary to perform bias correction (BC) to remove the non-dust parts before they are assimilated. Details concerning the method can be found in Jin et al. (2022). The bias-corrected observations can be used to correct the model simulation by data assimilation.

2.3.2 Himawari-8 AOD and calculation of DOD

The Himawari-8 geostationary satellite, operated by the Japan Meteorological Agency (JMA), provides aerosol op-

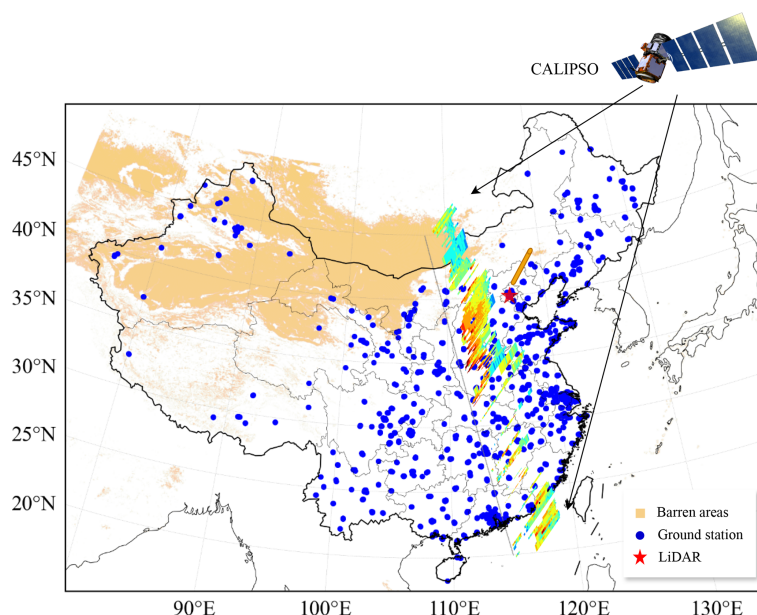


Figure 1. Spatial distribution of ground monitoring stations (blue scatters), location of the lidar (red star) with upward-scanning ray (orange cylinder), and aerosol extinction profile sample scanned by the CALIPSO satellite (trajectory is plotted as the grey line on the ground).

tical information covering East Asia and the western Pacific region at a time resolution of 10 min (Yumimoto et al., 2016). This allows us to monitor the source and movement of dust at high spatial–temporal resolution. Its products, the Ångström exponent ($\text{\AA}$) and aerosol optical thickness (AOT) at 550 nm, are selected to measure the dust optical depth (DOD). These data are merged to model grid ($0.25^\circ \times 0.25^\circ$) and hourly resolution.

To remove the fine-mode non-dust AOD in total AOD, an empirical function concerning $\text{\AA}$ is used to calculate the sub-micron fraction (SMF) (Anderson et al., 2005; Di Tomaso et al., 2022). The dust optical depth (DOD) can then be obtained by the SMF.

$$\text{SMF} = -0.0512 \times \text{\AA}^2 + 0.5089 \times \text{\AA} + 0.02 \quad (2)$$

$$\text{DOD} = \text{AOD} \times (1 - \text{SMF}) \quad (3)$$

Furthermore, the threshold of $\text{\AA} \leq 1$ is set to exclude the fine-mode dominant observations.

2.4 Dust observations for validation

The aerosol extinction coefficient from CALIPSO and lidar is used to validate the vertical structure of dust aerosol.

2.4.1 CALIPSO

The Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP), on board the CALIPSO satellite, provides data on the vertical structures and types of aerosols on a global scale (Winker et al., 2010). It has a spatial resolution of 333 m and a vertical resolution of 30 m. A normal extinction profile

sample can be seen above the dashed black line in Fig. 1a. It has been widely used in investigating aerosol distribution and evolution (Xu et al., 2020; Zhang et al., 2020; Han et al., 2022; Chen et al., 2023) and data assimilation (Sekiyama et al., 2010; Cheng et al., 2019; Escribano et al., 2022). Here, its product, extinction coefficient at 532 nm, is taken as the aerosol profile validation set to identify the vertical distribution of the dust load.

2.4.2 Polarization lidar

Polarization lidar deployed on the ground can provide continuous measurements on the vertical details of the atmospheric environment with high resolution (Hofer et al., 2017). Recently, Wang et al. (2021) collected dust extinction coefficient data through a dual-wavelength aerosol lidar built in the tower of the Institute of Atmospheric Physics ($39^\circ 58' 35''$ N, $116^\circ 22' 41''$ E; red star in Fig. 1a). The device provided dust information with a high vertical resolution of 30 m and temporal resolution of 15 min. The vertical profile development of the several super dust storms that occurred in the spring of 2021 was thoroughly captured by the instrument, thus making it a crucial source of measurement for learning about the vertical structure and its development during dust storms. In this study, these data are employed to compare the vertical structures from the model simulation and assimilation analysis.

3 Assimilation methodology and experiments

This section first illustrates the methodology of the classic EnKF used in Sect. 3.1. The sensitivity of the aerosol posterior to the vertical structure is then emphasized when assimilating the ground PM₁₀ concentrations in Sect. 3.2 and the DOD in Sect. 3.3. Case experiments evaluating this sensitivity are described in Sect. 3.4.

3.1 EnKF

The data assimilation algorithm used in this study is the stochastic EnKF formulated by Burgers et al. (1998). It is implemented in our self-developed PyFilter toolbox (Pang, 2024). Its capability to improve dust forecasting has been proven in our recent research (Pang et al., 2023). We did not adopt any other variant of EnKF because the problem we address is contingent but universal in aerosol data assimilation applications.

EnKF is a Monte Carlo approach based on Kalman filter theory. EnKF maintains a set of model states to approximate the probability distribution of the model state or parameter. It includes the forecast step and the analysis step. In the forecast step, each posterior ensemble member $\mathbf{x}_{t-1}^{a,i}$ at the previous time $t-1$ is integrated forward according to the model dynamics $\mathcal{M}$ to generate a prior forecast $\mathbf{x}_t^{f,i}$ at the next moment t . Here, i refers to the ensemble member.

$$\mathbf{x}_t^{f,i} = \mathcal{M}(\mathbf{x}_{t-1}^{a,i}) \quad (4)$$

In the analysis step, the states of each member are adjusted based on observational data to approach the true state. This adjustment process is implemented based on the error from the model prior and observation. In EnKF, the model prior error covariance matrix $\mathbf{P}^f$ is calculated through the ensemble approximation method following

$$\mathbf{P}^f = \frac{1}{N-1} (\mathbf{x}_i^f - \bar{\mathbf{x}}^f)(\mathbf{x}_i^f - \bar{\mathbf{x}}^f)^T, \quad (5)$$

where N is the number of ensembles, $\mathbf{x}_i^f$ is the i th individual of the ensemble prior states, and $\bar{\mathbf{x}}^f$ is the average of the ensemble members.

Then, a weight matrix $\mathbf{K}$, also referred to as the Kalman gain, can be obtained by

$$\mathbf{K} = \mathbf{P}^f \mathcal{H}^T (\mathcal{H} \mathbf{P}^f \mathcal{H}^T + \mathbf{R})^{-1}. \quad (6)$$

The a posteriori ensemble individual $\mathbf{x}_i^a$ is calculated by

$$\mathbf{x}_i^a = \mathbf{x}_i^f + \mathbf{K}(\mathbf{y} + \boldsymbol{\epsilon}_i - \mathcal{H}\mathbf{x}_i^f), \quad (7)$$

where $\boldsymbol{\epsilon}_i$ is the extra perturbation, and its variance is set according to the diagonal of the observation error covariance $\mathbf{R}$. It serves to maintain the ensemble spread (van Leeuwen, 2020).

Meanwhile, the a posteriori $\mathbf{x}^a$ can be updated via

$$\mathbf{x}^a = \bar{\mathbf{x}}^f + \mathbf{K}(\mathbf{y} - \mathcal{H}\bar{\mathbf{x}}^f). \quad (8)$$

From the equations above, we can tell that the a posteriori is dependent on both the prior and the observations. Within different spaces, these are aligned by the observation operator $\mathcal{H}$.

Figure 2 concisely illustrates the process of EnKF applied in this paper. In the upper row of panel (a.1), we have ensemble AOD priors (hollow grey diamond) and AOD observations (solid blue diamond), and below there are idealized aerosol vertical profiles, including a true profile (red line), ensemble prior profiles (dashed grey line), and an observation at ground level (blue star). It is noticeable that there are great differences between priors and observations both on the ground and in AOD under two conditions. The figures in panel (a) are in a positive condition. Both the AOD and the 3D mass priors underestimate the loading, while the structure is consistent with the true profile. Through assimilating ground-based or AOD observations, both the ground and the vertical dust loading are adjusted to be more consistent with the observations.

3.2 Ground PM₁₀ assimilation

In this section, the impact of ground measurement data assimilation on the aerosol vertical structure is explained. In practice, each of the PM₁₀ observations could be assimilated to optimize the 3D states that are correlated. The increments are transferred from the observation space to the model space via the Kalman gain $\mathbf{K}$ shown previously. To easily evaluate the posterior, we solely perform the assimilation analysis at the pixels where PM₁₀ measurements are available. Thus, here we confine the states only in the vertical direction in one pixel to better identify the sensitivity, and we assume the prior states $\mathbf{x}^f$ with total k levels to be

$$\mathbf{x}^f = [x_1^f, \dots, x_k^f]^T. \quad (9)$$

In this case, the error covariance matrix $\mathbf{P}^f$ is constructed by the variance of ensemble states and covariance in the vertical direction, and the observation operator $\mathcal{H}$ only converts the single observation on the ground level:

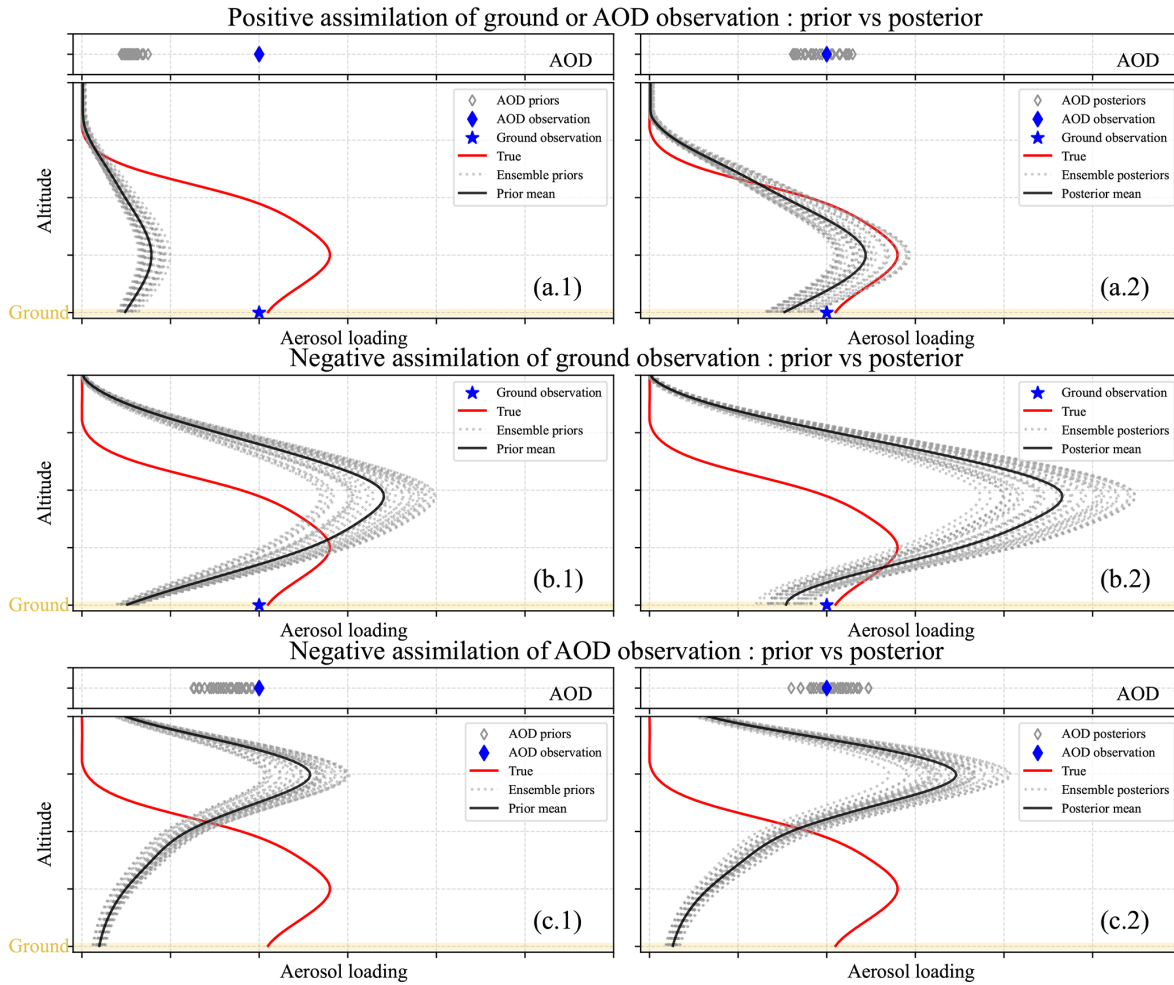


Figure 2. Schematic diagram of the sensitivity of data assimilation to aerosol vertical structure. Assimilation of ground and AOD observations under a positive condition, with AOD priors (hollow grey diamond) and AOD observations (solid blue diamond) in the upper box and true profiles (red line), ensemble priors (dashed grey line), prior means (black line), and ground observations (blue star) in the lower box (a.1). The figures on the right show the posteriors (a.2). The figures in the second row show the prior (b.1) and posterior (b.2) of the ground assimilation under a negative condition. The figures in the third row show the prior (c.1) and posterior (c.2) of the AOD assimilation under a negative condition.

$$\mathbf{P}^f = \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1k} \\ C_{21} & C_{22} & \cdots & C_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ C_{k1} & C_{k2} & \cdots & C_{kk} \end{bmatrix}, \quad (10)$$

$$\mathbf{H} = [1 \quad 0 \quad \cdots \quad 0]. \quad (11)$$

The calculation of $\mathbf{K}$ follows Eq. (6) and substitutes Eq. (10) and Eq. (11) into it:

$$\begin{aligned} \mathbf{K} &= \mathbf{P}^f \mathbf{H}^T (\mathbf{H} \mathbf{P}^f \mathbf{H}^T + \mathbf{R})^{-1} \\ &= [C_{11}, \dots, C_{k1}]^T (C_{11} + \sigma_y^2)^{-1}, \end{aligned} \quad (12)$$

where σ_y^2 is the error in the ground observation.

We substitute $\mathbf{K}$ into the following update function:

$$\begin{aligned} [x_1^a, \dots, x_k^a]^T &= [x_1^f, \dots, x_k^f]^T \\ &+ [C_{11}, \dots, C_{k1}]^T (C_{11} + \sigma_y^2)^{-1} (y - x_1^f). \end{aligned} \quad (13)$$

Here, we have the posterior state x_i^a on each level.

On the ground level, the posterior state is obtained as

$$\mathbf{x}_1^a = \frac{\sigma_y^2}{C_{11} + \sigma_y^2} \mathbf{x}_1^f + \frac{C_{11}}{C_{11} + \sigma_y^2} y. \quad (14)$$

If the observation error is much smaller than the prior error ($\sigma_y^2 \ll C_{11}$), we can have

$$x_0^a \approx y, \quad (15)$$

which means that the posterior on the ground level is close to that of the ground observation. The ground data assimilation has successfully tuned the ground prior state.

On the other hand, for the state on the i th level, the posterior x_i^a is calculated by

$$x_i^a = x_i^f + \frac{C_{k1}}{C_{11} + \sigma_y^2} (y - x_1^f). \quad (16)$$

To better identify the correctness of the vertical structure, we define a vertical ratio v_k here:

$$v_k = \frac{x_i}{x_1}, \quad (17)$$

which is the ratio of the state on the i th level to the ground level.

We divide x_i^f on both sides of Eq. (16):

$$v_i^a = v_i^f + \frac{C_{k1}}{C_{11} + \sigma_y^2} \frac{y - x_1^f}{x_1^f}. \quad (18)$$

When we have a strong vertical correlation and small observation error, we have $\frac{C_{k1}}{C_{11} + \sigma_y^2} \approx 1$. If the prior vertical structure is far from the true structure, $v_i^f \ll v_i^{\text{true}}$ or $v_i^f \gg v_i^{\text{true}}$. Then, the vertical structure of the posterior is strongly dependent on $\frac{y - x_1^f}{x_1^f}$. If x_1^f and y deviate far from each other, the increment can be large, thus further enhancing the incorrect prior structure v_i^f .

Figure 2b illustrates this kind of pattern. In panel (b.1), the priors greatly underestimate the aerosol concentration on the ground level. Meanwhile, over the upper space, the priors overestimate the intensity and altitude of the dust loading to some extent. After assimilating the ground observations, as shown in panel (b.2), the posterior aerosol loading (dashed grey line) is tuned to a large extent and is much closer to the observation on the ground layer, while for the vertical distribution, a greater overestimation than that of the prior can be seen. The original incorrect profile is significantly amplified.

3.3 AOD assimilation

Assimilation of AOD-related observations is carried out by connecting the 3D aerosol priors $\mathbf{x}^f$ and AOD priors $\boldsymbol{\tau}^f$. It is assumed that the vertical structure of the aerosol field remains the same as the AOD lacks vertical information. Then,

a 3D mass concentration field can be calculated by finding the optimal AOD field.

To begin with, the value of the AOD is related to the aerosol mass concentration, aerosol type, and humidity. Here, it is assumed to be linearly related to only the aerosol mass concentration field $\mathbf{x}^f$. The column-integrated AOD is calculated by summing the aerosol mass concentration $\mathbf{x}_i^f$ on each level:

$$\boldsymbol{\tau}^f = \sum_{i=1}^k \mathcal{M} \mathbf{x}_i^f, \quad (19)$$

where $\mathcal{M}$ is the linear model operator that calculates the AOD from the mass concentration.

The calculation of the posterior AOD, $\boldsymbol{\tau}^a$, follows Eq. (8):

$$\boldsymbol{\tau}^a = \bar{\boldsymbol{\tau}}^f + \mathbf{K}(\mathbf{y} - \mathcal{H} \bar{\boldsymbol{\tau}}^f). \quad (20)$$

Next, we substitute Eq. (19) into Eq. (20). We can then bridge the mass concentration field and AOD field:

$$\mathcal{M} \sum_{i=1}^k \mathbf{x}_i^a = \mathcal{M} \sum_{i=1}^k \bar{\mathbf{x}}_i^f + \mathbf{K}(\mathbf{y} - \mathcal{H} \bar{\boldsymbol{\tau}}^f). \quad (21)$$

Similarly for AOD, we confine the states in the vertical direction within one pixel. We have one prior AOD, $\boldsymbol{\tau}^f$, and aerosol mass prior, $\mathbf{x}_i^f$, on a total of k levels. We first define a vertical factor, u_i , which is the ratio of the mass at the i th level to the mass at all levels:

$$u_i = \frac{x_i}{\sum_{i=1}^k x_i}. \quad (22)$$

Next, we substitute it into Eq. (21):

$$\mathcal{M} \frac{\mathbf{x}_i^a}{u_i} = \mathcal{M} \frac{\mathbf{x}_i^f}{u_i} + \mathbf{K}(\mathbf{y} - \mathcal{H} \boldsymbol{\tau}^f). \quad (23)$$

Since we have only one observation and one AOD simulation, the observation operator $\mathcal{H}$, error variance matrix $\mathbf{P}^f$, and Kalman gain $\mathbf{K}$ are

$$\mathcal{H} = [1], \quad \mathbf{P}^f = [\sigma_\tau^2], \quad \mathbf{K} = \left[\frac{\sigma_\tau^2}{\sigma_\tau^2 + \sigma_y^2} \right]. \quad (24)$$

We substitute them into Eq. (23), which gives us the mass posterior x_i^a on each layer:

$$\mathcal{M} \frac{x_i^a}{u_i^f} = \mathcal{M} \frac{x_i^f}{u_i^f} + \frac{\sigma_\tau^2}{\sigma_\tau^2 + \sigma_y^2} (y - \tau^f), \quad (25)$$

$$x_i^a = x_i^f + u_i^f \frac{\sigma_\tau^2}{\sigma_\tau^2 + \sigma_y^2} \frac{y - \tau^f}{\mathcal{M}}. \quad (26)$$

This equation demonstrates that the analysis increment $\frac{\sigma_\tau^2}{\sigma_\tau^2 + \sigma_y^2} \frac{y - \tau^f}{\mathcal{M}}$ from AOD assimilation is allocated to each level

of mass by the vertical factor u_i . When we have an incorrect structure, we note that $u_i \gg u_i^{\text{true}}$ or $u_i \ll u_i^{\text{true}}$. The analysis increment can be reallocated to the wrong level, thus amplifying the incorrect structure.

Figure 2c demonstrates the negative pattern of assimilating AOD. In the upper row of panel (c.1), we have the ensemble AOD priors (hollow grey diamond) and the AOD observation (solid blue diamond) that are combined to optimize the 3D dust mass field (dashed grey line). Similar to panel (b), assume we have an incorrect prior dust mass structure. By assimilating AOD observations, the AOD posteriors can be tuned to better fit the AOD observation, as shown in the upper row in panel (c.2), while for the structure of mass field, this incorrect structure is generally amplified since the prior DOD simulation implies that the model underestimates the total dust column compared to the observation. Another pattern, in which the prior mass is incorrectly situated at the ground level, is also illustrated in Fig. S1 in the Supplement. It is notable that the sensitivity of AOD assimilation to the vertical structure is not as significant as that of the ground assimilation since the AOD contains information on the total columns.

3.4 Case settings

To demonstrate the sensitivity of aerosol data assimilation to vertical structure, five assimilation analysis cases were selected, namely P-Gd-CAL, N-Gd-CAL, N-Gd-Li, N-DOD-CAL, and NP-DOD-CAL. These cases were selected from the ensemble model run starting from 00:00 CST on 14 March to 00:00 CST on 17 March and 00:00 CST on 14 April to 00:00 CST on 17 April 2021. (Hereafter, all times are in CST.) Dust classification is carried out using the CALIPSO depolarization ratio data to ensure that dust is the dominant particle during these time ranges. The procedures and results can be found in the Supplement. The background run consists of 32 ensembles, and each of them are driven by perturbed emission inventories. Details concerning the generation of the ensembles can be found in Jin et al. (2022).

Figure 1 lists information about the experiment cases. P-Gd-CAL is a positive case, which shows how the data assimilation corrects the surface concentrations and maintains the correct aerosol structure. N-Gd-CAL and N-Gd-Li are negative cases, which show how the assimilation deteriorates the profile when assimilating ground observations only. The time of the former two cases is fixed, and their profiles follow the trajectory of CALIPSO. Hence, by comparison, we can observe the impact of assimilation on the vertical structure. The last case focuses on a fixed site. A continuous timeline is used, and assimilation analysis is performed on each time point. The profiles are compared with high-resolution lidar observations. N-DOD-CAL and NP-DOD-CAL are negative cases that focus on assimilating DOD observations. Himawari-8 DOD observations are assimilated. Both of them are validated by CALIPSO.

In addition, to test the impact of vertical localization on the assimilation performance, experiments applying EnKF with vertical localization are also carried out. They are performed on the priors of cases P-Gd-CAL and P-Gd-CAL. These results can be found in the Supplement.

4 Results and discussions

This section delves into the sensitivities of data assimilation from ground and DOD observations, with a particular emphasis on the impact of prior vertical profiles. Regarding the ground observation assimilation, the discussion is enriched by presenting a positive case, alongside two illustrative negative examples. The validity of these scenarios is further reinforced by CALIPSO satellite observations and lidar measurements. This serves to underscore the substantial influence that ground assimilation can exert on the posterior. Then, the examination of DOD assimilation uncovers its intricacies by showcasing two instances, highlighting favourable and unfavourable outcomes. These illustrations collectively emphasize the impact DOD assimilation has on the posterior, thereby underscoring the importance of understanding the vertical profile's role in the assimilation process for both ground and DOD observations.

4.1 Cases on ground assimilation

4.1.1 Positive case

P-Gd-CAL is intended to demonstrate how the assimilation works properly. Figure 3a.1 shows the time series of ground PM_{10} and BC PM_{10} from several ground stations that are close to the CALIPSO trajectory ($< 0.25^\circ$). The assimilation time point, 03:00 on 16 April 2021, is selected in this case (highlighted in green). Although it is during the later stage of the dust storm, the concentrations were as high as approximately $400 \mu\text{g m}^{-3}$. Through comparison between raw and bias-corrected PM_{10} (solid and dashed lines), it is clear that the scanned region was still dominated by coarse-dust aerosol. Figure 3c.1 and d.1 are snapshots of the average of ensemble priors and posteriors. We can see that the priors underestimated the dust load to the south of the domain and overestimated it to the east. After assimilation analysis, the dust load was adjusted to better fit the observations. The ground observations were consistent with the posterior. RMSE was reduced from 254.83 to $118.03 \mu\text{g m}^{-3}$.

Take a closer look at the vertical structure. Figure 3b illustrates the extinction coefficient profile from CALIPSO. From the profile we can tell that the dust was concentrated on the ground over a latitude of $34\text{--}38^\circ$ (circled by the dashed line). Vertically, it extended upwards, accompanied with terrain from 1 km at 34.5° to 2.5 km at 37.5° . Figure 3c.1 shows the prior dust profile following the same trajectory of CALIPSO. The colour bar here is rescaled to show more detail. At around $41\text{--}42^\circ$, a heavy dust load was simulated

Table 1. Experiment configurations.

Name	Time (CST)	Type	Assimilated observations	Validation data
P-Gd-CAL	03:00 on 16 April 2021	Positive	BC PM ₁₀	CALIPSO
N-Gd-CAL	03:00 on 16 March 2021	Negative	BC PM ₁₀	CALIPSO
N-Gd-Li	15 March 2021	Negative	BC PM ₁₀	Lidar
N-DOD-CAL	14:00 on 16 March 2021	Negative	Himawari-8 DOD	CALIPSO
NP-DOD-CAL	14:00 on 29 March 2021	Negative and positive	Himawari-8 DOD	CALIPSO

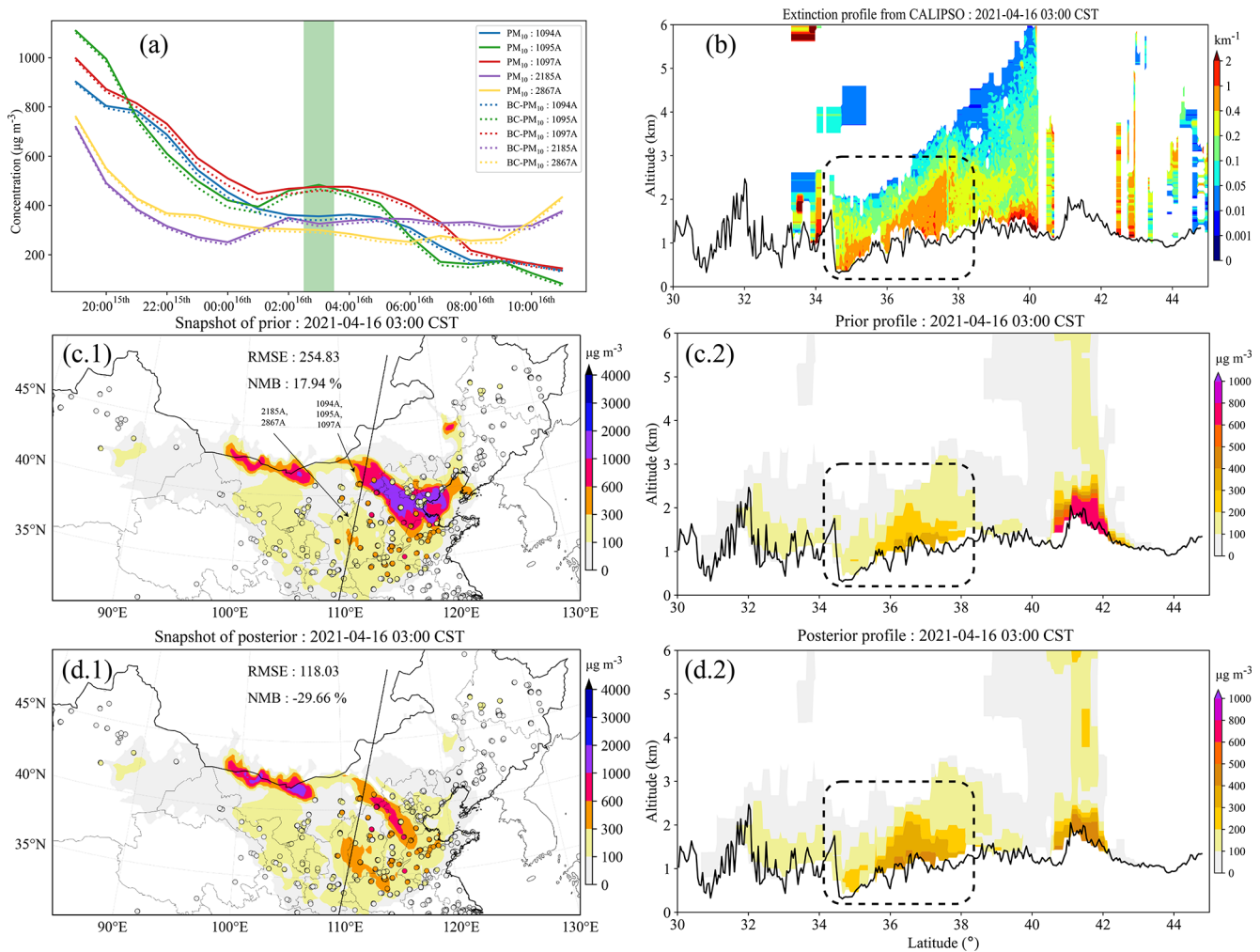


Figure 3. Time series of PM₁₀ and BC PM₁₀ concentrations from several ground stations that are close to the CALIPSO trajectory (a). Spatial distribution of ground dust concentrations from the average of the ensemble priors and the posteriors with scatters of ground BC PM₁₀ observations (c.1 and d.1). The black line inside is where CALIPSO scanned through. The figures in the right column are the extinction profile from CALIPSO (b) and the dust concentration profile following the CALIPSO scanning trajectory from the prior (c.2) and the posterior (d.2). The colour bars in (c.2) and (d.3) are rescaled to show more detail. The black line at the bottom is the terrain altitude. The case time is at 03:00 CST on 16 April 2021.

(600–800 μg m⁻³), while CALIPSO did not capture the data there. Hence, it cannot be validated, and we focus on the dust at 34–38°. At 34–38°, the prior model had successfully reproduced the dust structure indicated by CALIPSO. The concentrations were about 100–300 μg m⁻³, which are underes-

timated, as shown in Fig. 3d.2. After assimilation, the ground concentrations increased to 400–500 μg m⁻³, as shown in (c.1) and (d.1). The vertical structure was also reserved, which is identical to the structure shown by CALIPSO.

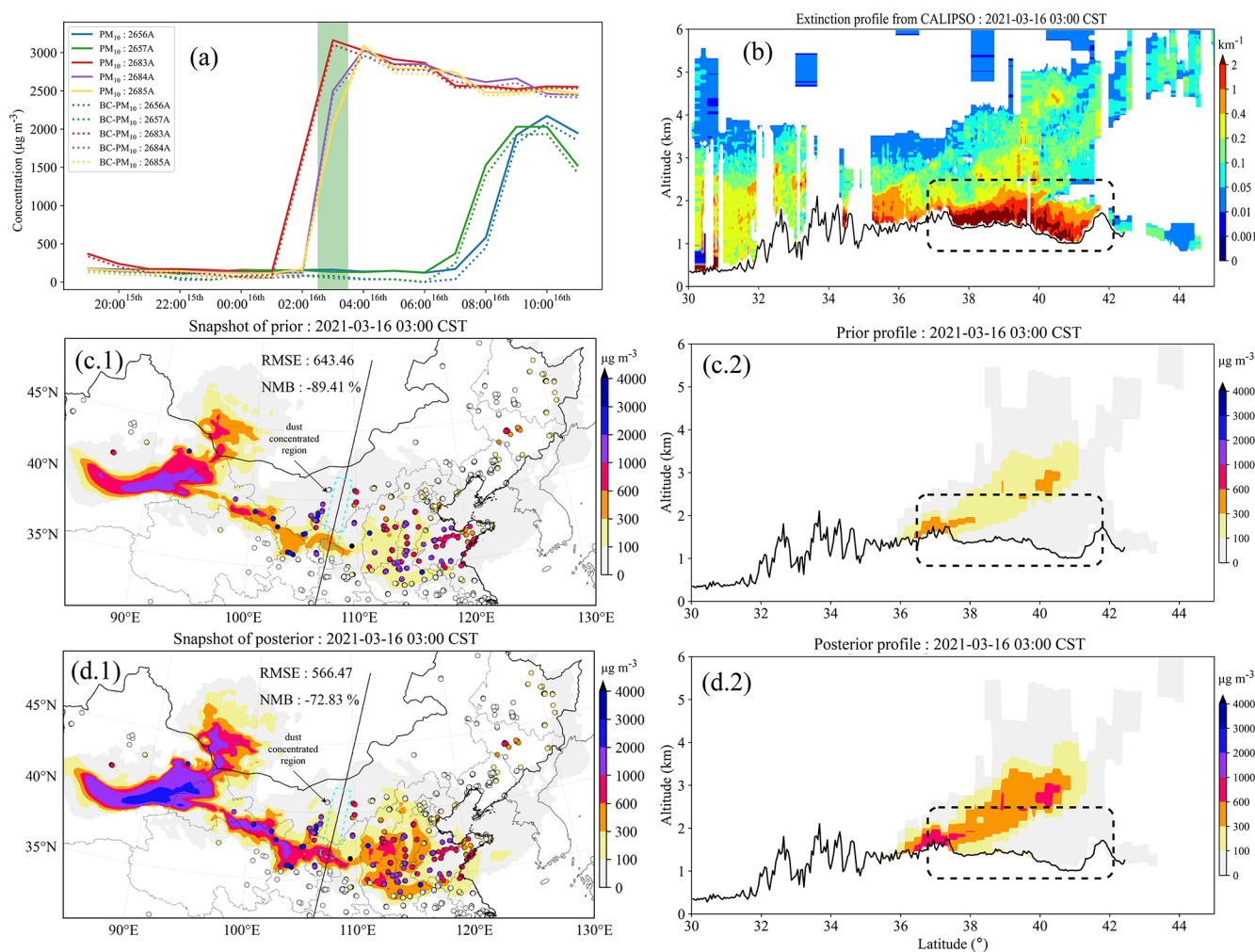


Figure 4. Time series of PM₁₀ and BC PM₁₀ concentrations from several ground stations that are close to the CALIPSO trajectory (a). Spatial distribution of ground dust concentrations from the average of the ensemble priors and the posteriors with scatters of ground BC PM₁₀ observations (c.1 and posterior d.1). The black line inside is where the CALIPSO scanned through. The figures in the right column are the extinction profile from CALIPSO (b) and the dust concentration profile following the CALIPSO scanning trajectory from the prior (c.2) and the posterior (d.2). The black line at the bottom is the terrain altitude. The case time is at 03:00 CST on 16 March 2021.

4.1.2 Negative case validated by CALIPSO

N-Gd-CAL is the negative case that illustrates how data assimilation degrades the aerosol vertical structure. Figure 4a.1 shows the time series of ground PM₁₀ and BC PM₁₀ from several ground stations that are close to the CALIPSO trajectory. As shown in panel (a), the ground dust aerosol loading from ground stations in N-Gd-CAL increased rapidly starting 2 h before assimilation occurred. This means that the region in the lower rectangle in (c.1) was a dust-concentrated region.

Evidence from CALIPSO also indicated that there was heavy dust loading. As shown in Fig. 4b, the extinction coefficient profile from 37 to 42° (circled by the dashed black rectangle) exhibits extremely high values (maximum exceeded 2 km⁻¹). It spread upwards for 1 km at an alti-

tude of 1 km. This region is highlighted by the dashed light blue rectangle in (c.1) and (d.1). As inferred from the farther ground stations (concentrations over 1000 μg m⁻³), this region was dominated by dust aerosols and is situated at ground level.

The dust profile in (c.2) indicates that the ensemble priors agreed that the aerosols were mainly concentrated upwards. The plume floated up to 4 km with latitude. Only a few dust aerosols over 36–37° on the ground (300–600 μg m⁻³) show consistency with the observations (shown in c.1 and c.2). A great discrepancy exists between the prior and independent observations at 37–42°. After assimilating BC PM₁₀ concentrations, the overall loading is increased. The great underestimation in the east is improved. However, the dust-concentrated region circled in (d.1) is only improved to a small extent. This is due to a lack of observations nearby and

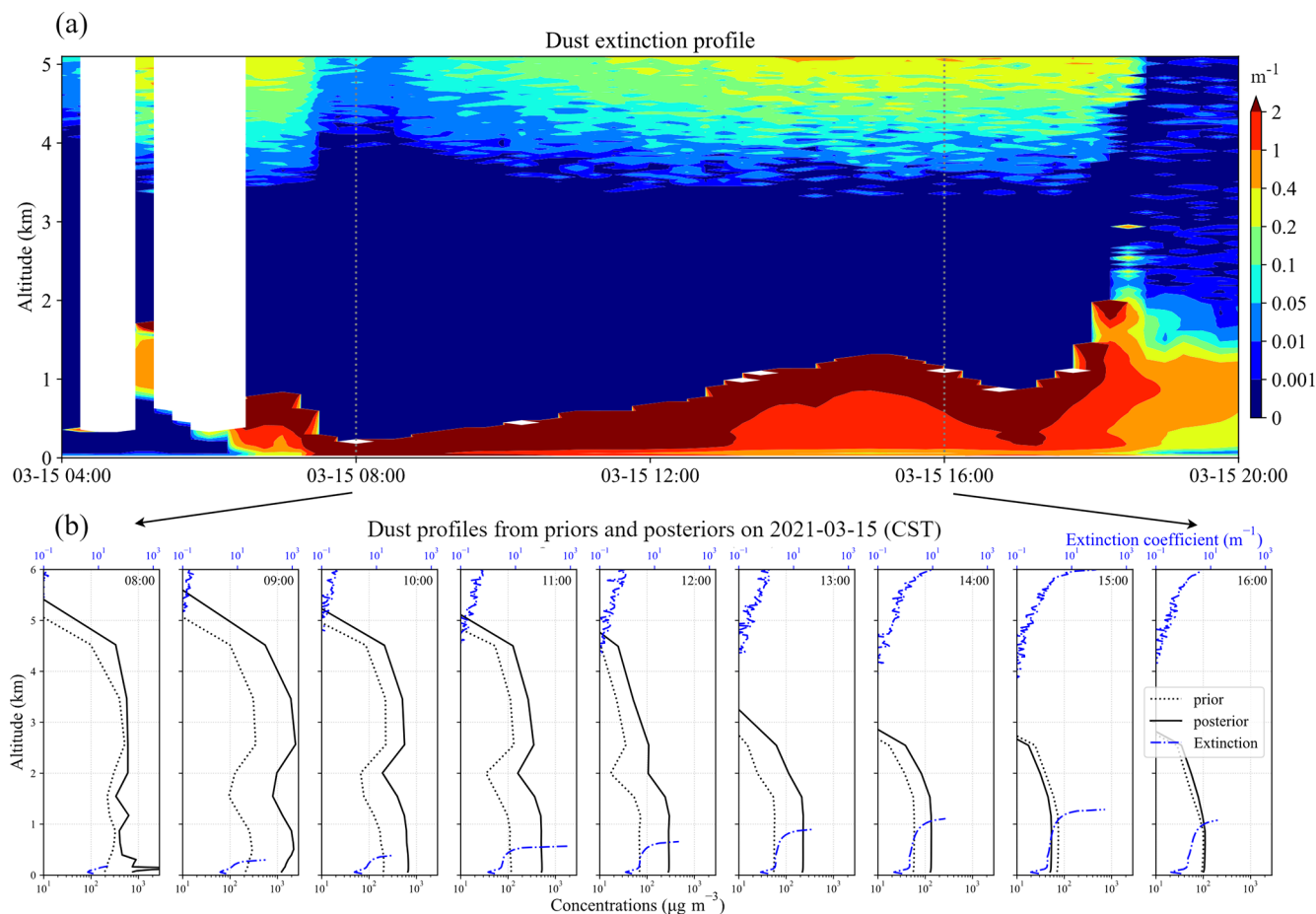


Figure 5. Time series of dust extinction coefficient profile obtained from lidar (a). The assimilation analysis is performed hourly from 08:00 to 16:00. The figure below shows the hourly dust profile line from priors (dashed line) and posteriors (solid line) (b). The extinction coefficient is also plotted (dash-dotted blue line). The profile data are extracted from the closest grid point to the lidar location. The x axis is logarithmically rescaled. Note that instead of using the posterior in the previous time step to propagate the model, analysis here is separately conducted on the static background.

the much lower uncertainties exhibited by ensemble priors. In terms of the vertical structure, a disturbing profile is obtained. As shown in Fig. 4d.2, ground concentrations at 36–37° are aligned to over $600 \mu\text{g m}^{-3}$. This is consistent with surrounding stations. However, for the reason mentioned earlier, dust loads over 37–42° are barely corrected. Yet, these trivial increments significantly enhance the erroneous vertical structure. At ground level, the dust concentrations rise from less than $100 \mu\text{g m}^{-3}$ to nearly $300 \mu\text{g m}^{-3}$ at 37–39°. The vertical loading is almost 3 times greater than the loading from the prior profile. Moreover, the extremely low region at 39–42° represents a more intense amplification effect than that. Since the low values are more sensitive to variations, increments on these values can easily give rise to an increase in the vertical loading. In this case, this phenomenon can be seen clearly through the comparison between priors, posteriors, and measurements for validation.

4.1.3 Negative case validated by lidar

N-Gd-Li is another negative case. In this case, we do not focus on a single time point. A series of assimilation analyses is conducted to investigate the effect. As shown by Fig. 5a, a high-resolution profile of a dust extinction coefficient from lidar is given to demonstrate the changing structure of dust loading on a fixed location. Entering at about 05:00 and then landing at 07:00, this dust storm lasted for about 20 h in the area where the lidar was located. A time range from 08:00 to 16:00 is selected as it comprises most of the dust period. During the whole period, the measured dust loading increased from 0.4 to 1.4 km. Afterwards, the intensity decreased gradually.

Figure 5b illustrates the hourly prior and posterior dust profiles during the selected period. The profiles of the extinction coefficient are also plotted with the dash-dotted blue line. The x axis, which represents concentrations, is logarithmically rescaled here. In terms of prior profiles, dust load-

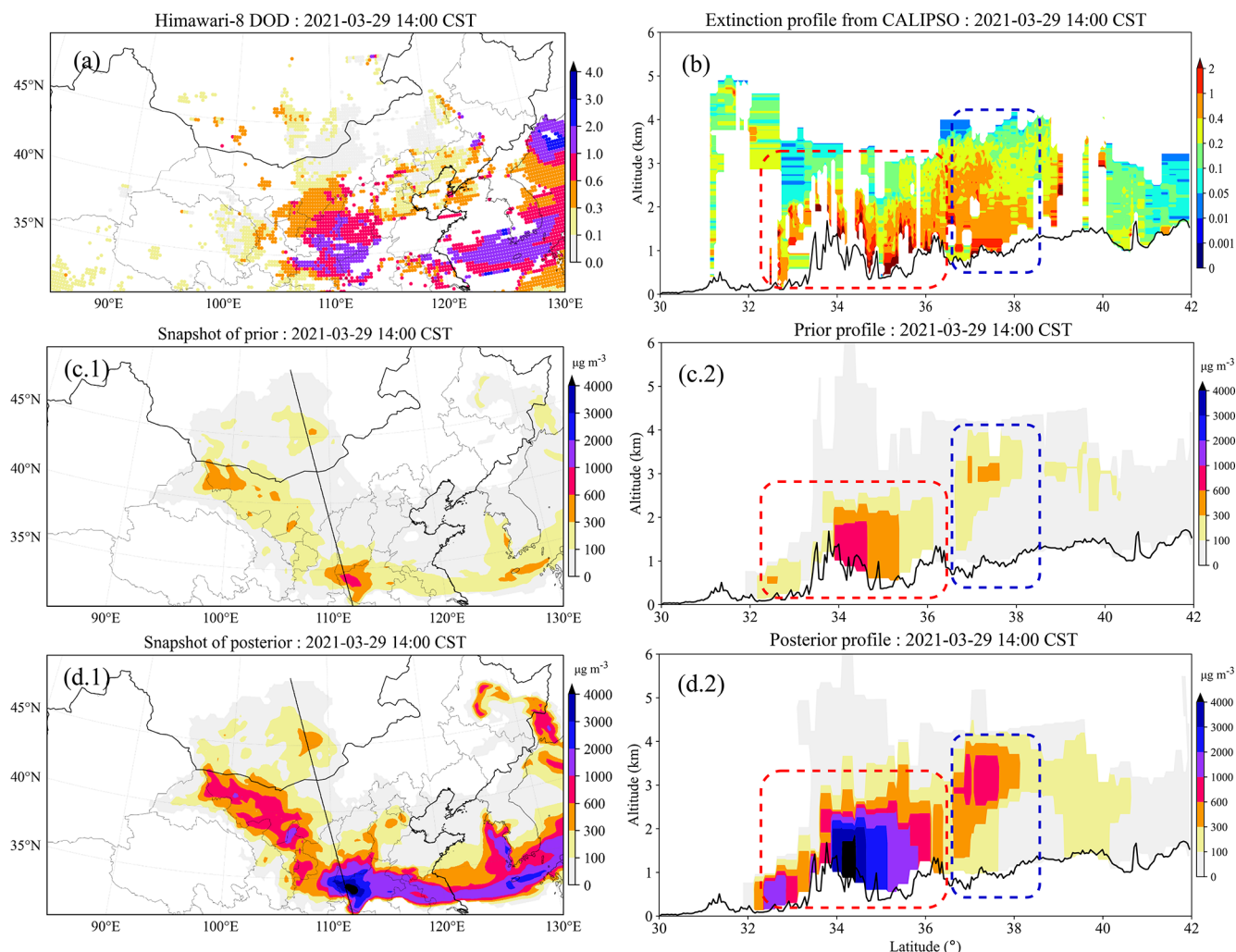


Figure 6. Time series of PM_{10} and BC PM_{10} concentrations from several ground stations that are close to the CALIPSO trajectory (a). Spatial distribution of ground dust concentrations from the average of the ensemble priors and posteriors with scatters of ground BC PM_{10} observations (c.1 and posterior d.1). The black line inside is where CALIPSO scanned through. The figures in the right column are the extinction profile from CALIPSO (b) and dust concentration profile following the CALIPSO scanning trajectory from the prior (c.2) and the posterior (d.2). The black line at the bottom is the terrain altitude. The case time is at 14:00 CST on 29 March 2021.

ing extended up to 5 km at 08:00 and declined to 3 km at 16:00. This structure is inconsistent with the lidar profile, in which only little dust was observed. The dust storm has been overestimated to a large extent in height. The ground aerosol concentrations are lower than $200 \mu\text{g m}^{-3}$ throughout eight time points, which is much lower than those of the observations (over $1000 \mu\text{g m}^{-3}$). After assimilation, this underestimation is mitigated. The ground dust concentrations are amplified several times to better fit the observations. Meanwhile, the erroneous vertical structure is also intensified in the first six moments, as can clearly be seen in the comparison between the concentration and extinction coefficient. At 09:00 in particular, the dust loading above 2 km is amplified to over $1000 \mu\text{g m}^{-3}$, which will provide completely incor-

rect information about the dust storm structure and impact future forecasting.

4.2 Cases on DOD assimilation

In this section, the sensitivity of assimilating DOD to the aerosol vertical profile is presented, starting with the NP-DOD-CAL case. Figure 6a shows the spatial distribution of DOD observed by Himawari-8 at 14:00 on 29 March 2021. The dust plume approached the east of China and reached out to the East China Sea. The prior model reproduces the distribution, while the overall concentration is underestimated (less than $300 \mu\text{g m}^{-3}$), as shown in panel (c.1). By assimilating the Himawari-8 DOD, the overall dust concentration field is increased to over $1000 \mu\text{g m}^{-3}$. The posterior dust plume is consistent with the DOD observations.

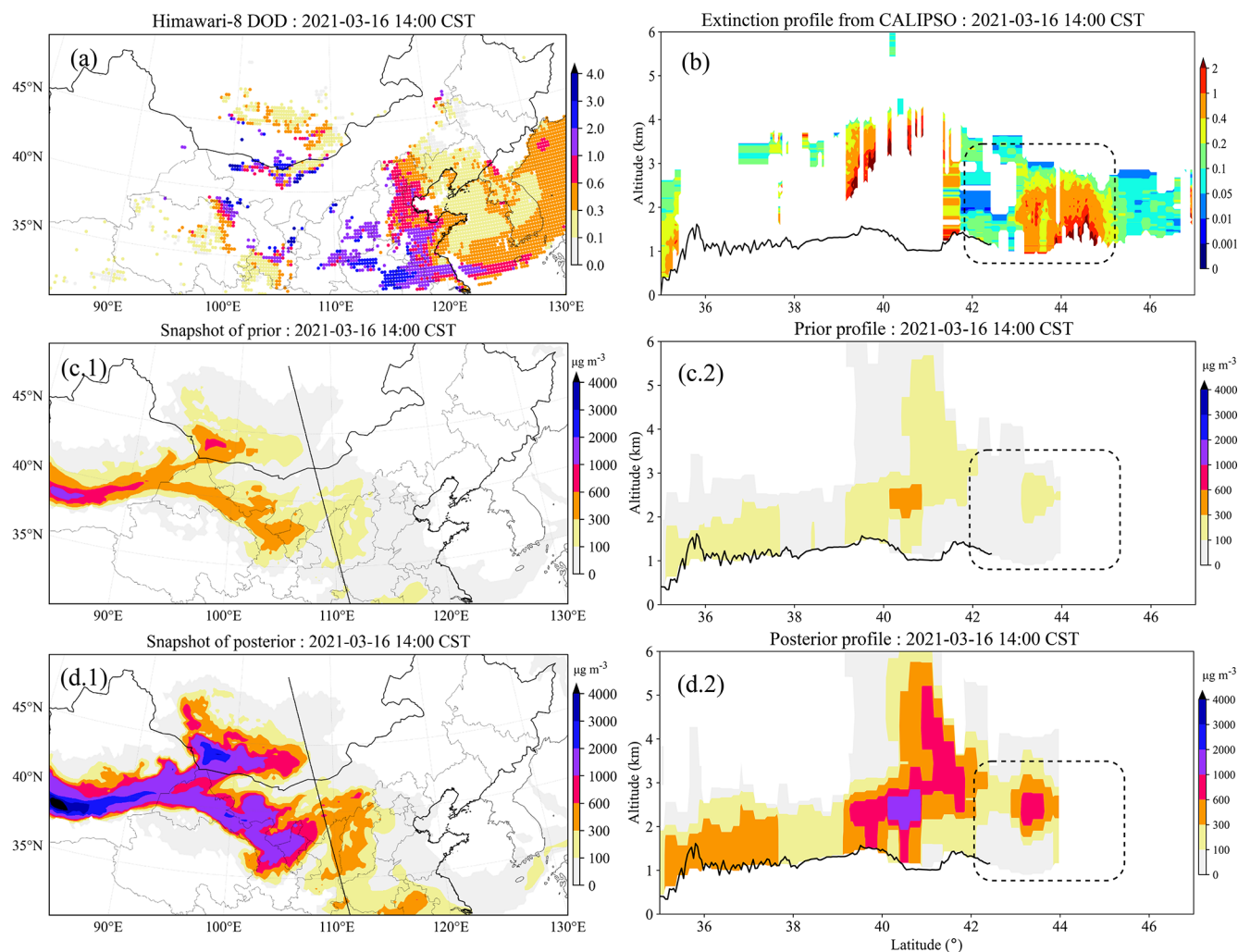


Figure 7. Snapshot of Himawari-8 DOD at the assimilation instance (a). Spatial distribution of ground dust concentrations from the average of the ensemble priors (c.1 and posterior d.1). The black line inside is where CALIPSO scanned through. The figures in the right column are the extinction profile from CALIPSO (b) and the dust concentration profile following the CALIPSO scanning trajectory from the prior (c.2) and the posterior (d.2). The black line at the bottom is the terrain altitude. The case time is at 14:00 CST on 16 March 2021.

As shown by the extinction profile in panel (b), the dust was concentrated around 32–38° and extended upwards to an altitude of 3 km. In terms of the prior, this structure is partly correct. The dust in the red box is located on the lower layer and aligns with the extinction profile, while the dust in the blue box has a higher position. By assimilating DOD, the total column dust field is optimized, and the vertical structure is restored following the prior dust field. As a result, the underestimation is alleviated, and the overall concentrations are increased. The correct dust structure in the red box is maintained, and the intensity is enhanced. However, the incorrect dust structure in the blue box is also amplified, causing an obvious inaccurate dust loading, which should be located on the lower level.

N-DOD-CAL is another case that illustrates the negative aspects of assimilating DOD. Figure 7 comprises the same

assimilation experiment results as in NP-DOD-CAL, except at 14:00 on 16 March 2021. The DOD observations report high values densely distributed in the east and sparsely distributed in the north. The prior model simulates the dust plume that is concentrated in the north-west. To the east, only rare dust is observed. It is caused by the inability of the model to reproduce the long-term transport of dust. If we focus on where CALIPSO scanned through, an increase in dust can be seen.

In terms of CALIPSO, it is certain that the dust is concentrated on the lower level at 43–45°. There is also a dust plume at 39–42°, although it cannot be verified because of the missing observation data. Similarly, the prior dust profiles show an incorrect dust structure, as circled by the black box in panel (c.2). After assimilation, this error is amplified from around $100 \mu\text{g m}^{-3}$ to over $600 \mu\text{g m}^{-3}$.

5 Conclusions

Data assimilation has been widely used in constructing re-analysis datasets and improving the predictability of atmospheric models. The performance of assimilation algorithms relies not only on the methodologies themselves but also on the observations. Nowadays, the complexity and dimensions of atmospheric models have been growing rapidly. For atmospheric aerosol models, high spatial and temporal resolutions are becoming feasible. In terms of observations, they are either outnumbered in space or discontinuous in time compared to models. Particularly in the vertical direction, the available aerosol measurements are still limited both in space and in time. Through assimilating ground or satellite measurements, the aerosol field can be improved, while in 3D space, it can also amplify the erroneous estimation of the aerosol structure.

In this paper, we explore the sensitivity of aerosol data assimilation to the vertical profile by carrying out five assimilation experiments. Dust aerosol, which is simulated by LOTOS-EUROS, is selected as the optimization target. The EnKF data assimilation methodology is adopted. Bias-corrected ground PM₁₀ concentrations obtained from ground monitoring stations and DOD from Himawari-8 are assimilated. An extinction coefficient profile from CALIPSO and polarization lidar are used to validate the aerosol vertical structure. P-Gd-CAL intends to show how ground data assimilation tunes the vertical structure positively. After assimilation, the ground dust field is optimized to better fit the observations, and the vertical structure is enhanced, which is also consistent with the CALIPSO measurements. In contrast, N-Gd-CAL exhibits a drastic negative effect. In this case, even small increments on the ground can greatly enhance the incorrect aerosol structure. N-Gd-Li is another negative case that displays hourly prior and posterior profiles. The continuous dust extinction profile from lidar is utilized to validate the structure. In most of the moments, the assimilation improves the ground aerosol field while significantly degrading the vertical dust loading. N-DOD-CAL and NP-DOD-CAL intend to show how DOD assimilation deteriorates the vertical structure. Different from the ground assimilation, the posterior vertical structure is here reallocated by the vertical ratio. When the vertical ratio is incorrect, assimilating DOD can also pass down this error.

In conclusion, integrating ground- and satellite-derived aerosol observations into models enhances both the analysis and the forecasting accuracy of the model's state. However, challenges persist in reconciling these observations with the model's high-dimensional state. Specifically, the model's initial, potentially flawed vertical aerosol structure could impair assimilation efforts. The underlying reason is that data assimilation relies on the background error covariance to propagate the innovations between states and observations across the entire domain. If the vertical structure is inaccurate, this error may not only persist but could also be am-

plified during the assimilation process. Analytical examples from both ground-based and satellite-based assimilation confirm this effect. This paper has only scratched the surface by presenting a handful of negative instances, yet it is evident that this issue is pervasive in aerosol data assimilation. The path forward entails establishing a complementary network of vertical observations and implementing advanced assimilation methodologies that are sensitive to vertical structures, thereby offering a promising avenue to surmount these challenges.

Appendix A: Conversion of AOD between different wavelengths

One common method for converting AOD between different wavelengths involves the Ångström exponent ($\mathring{A}$) (Jin et al., 2023b). This exponent describes the relationship between AOD and wavelength, defined as the ratio of the logarithm of the AOD ratio at two different wavelengths to the logarithm of the ratio of those wavelengths:

$$\mathring{A} = -\frac{\log(\tau_{\lambda_1}/\tau_{\lambda_2})}{\log(\lambda_1/\lambda_2)}, \quad (\text{A1})$$

where $\mathring{A}$ is the Ångström exponent. τ_{λ_1} and τ_{λ_2} are the AOD values at wavelengths λ_1 and λ_2 , respectively.

Given the AOD at a specific wavelength and the Ångström exponent, the AOD at another wavelength can be estimated using the following formula:

$$\tau_{\lambda} = \tau_{\lambda_0} \left(\frac{\lambda}{\lambda_0} \right)^{-\mathring{A}}. \quad (\text{A2})$$

Here, λ_0 represents the reference wavelength where the AOD is known, and λ is the target wavelength. The wavelengths 550 and 532 nm are very close in the visible spectrum, with a difference of only 18 nm. Dust optical properties, such as extinction coefficient, do not vary significantly over such a small wavelength range. This allows for a reasonable approximation when comparing data at these two wavelengths.

Appendix B: DOD operator

Mie theory is applied to convert the aerosol mass concentration into AOD. It is calculated through the scatter and absorption coefficients of spherical particles with a given radius and refractive index (Gupta et al., 2018) and is defined as

$$\tau = \sum_{k=1}^n \epsilon_d^k z^k, \quad (\text{B1})$$

where τ is the simulated AOD. ϵ_d^k and z^k are the dust extinction coefficient and layer thickness at the k th layer. ϵ_d^k is calculated by the product of extinction efficiency Q_{ext} , the total cross section per unit mass S ($\text{m}^2 \text{g}^{-1}$), and the aerosol mass concentration C (g m^{-3}):

$$\epsilon_d^k = Q_{\text{ext}} S C, \quad (\text{B2})$$

where Q_{ext} is the sum of the scattering and absorption efficiency. It is decided by the ratio of the aerosol radius, the incident wavelength, and the chemical composition (van de Hulst, 1958). S depends on the particle size and aerosol mass density. The dust bins and diameter ranges are shown in Table B1. Detailed descriptions concerning the calculation of Q_{ext} and S can be found in Sect. 2 in Jin et al. (2023b).

Table B1. Dust size bins and diameter ranges.

Bins	dust_ff	dust_f	dust_ccc	dust_cc	dust_c
Diameter range (μm)	0.01–1	1–2.5	2.5–4	4–7	7–10

Code and data availability. PyFilter is archived on Zenodo (<https://doi.org/10.5281/zenodo.14036308>) (Pang, 2024) and also available on GitHub (<https://github.com/xxcvvv/open-PyFilter>, last access: 23 June 2025). The ensemble initial fields and prior and posterior fields used in this paper are archived in <https://doi.org/10.5281/zenodo.14846965> (Pang, 2025). The source code of the LOTOS-EUROS model is available at <https://doi.org/10.5281/zenodo.14039267> (Segers, 2024). The CALIPSO data can be downloaded at https://doi.org/10.5067/CALIOP/CALIPSO/CAL_LID_L2_05km APro-Standard-V4-21 (NASA/LARC/SD/ASDC, 2025). The ground PM_{10} observations can be obtained at <https://quotsoft.net/air/> (last access: 23 June 2025, Wang, 2024). The Himawari-8 aerosol product is available at <https://www.eorc.jaxa.jp/ptree> (JAXA, 2025).

Supplement. The supplement related to this article is available online at <https://doi.org/10.5194/gmd-18-3781-2025-supplement>.

Author contributions. JJ conceived the study. MP carried out the experiments and data analysis. JJ and MP prepared the paper. TY, XC, AS, HXL, HL, and WH provided useful comments on the paper.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

Financial support. This study was supported by the National Key Research and Development Program of China (grant nos. 2024YFE0113700 and 2024YFF0811400) and the National Natural Science Foundation of China (grant no. 42475150). We received technical support from the National large Scientific and Technological Infrastructure “Earth System Numerical Simulation Facility” (<https://cstr.cn/31134.02.EL>, last access: 24 June 2025).

Review statement. This paper was edited by Holger Tost and reviewed by two anonymous referees.

References

- Adebisi, A., Kok, J. F., Murray, B. J., Ryder, C. L., Stuut, J.-B. W., Kahn, R. A., Knippertz, P., Formenti, P., Mahowald, N. M., Pérez García-Pando, C., Klose, M., Ansmann, A., Samset, B. H., Ito, A., Balkanski, Y., Di Biagio, C., Romanias, M. N., Huang, Y., and Meng, J.: A review of coarse mineral dust in the Earth system, *Aeolian Res.*, 60, 100849, <https://doi.org/10.1016/j.aeolia.2022.100849>, 2023.
- Anderson, T. L., Wu, Y., Chu, D. A., Schmid, B., Redemann, J., and Dubovik, O.: Testing the MODIS satellite retrieval of aerosol fine-mode fraction, *J. Geophys. Res.*, 110, D18204, <https://doi.org/10.1029/2005JD005978>, 2005.
- Bannister, R. N.: A review of operational methods of variational and ensemble-variational data assimilation, *Q. J. Roy. Meteor. Soc.*, 143, 607–633, <https://doi.org/10.1002/qj.2982>, 2017.
- Benedetti, A., Reid, J. S., Knippertz, P., Marsham, J. H., Di Giuseppe, F., Rémy, S., Basart, S., Boucher, O., Brooks, I. M., Menut, L., Mona, L., Laj, P., Pappalardo, G., Wiedensohler, A., Baklanov, A., Brooks, M., Colarco, P. R., Cuevas, E., da Silva, A., Escribano, J., Flemming, J., Huneeus, N., Jorba, O., Kazadzis, S., Kinne, S., Popp, T., Quinn, P. K., Sekiyama, T. T., Tanaka, T., and Terradellas, E.: Status and future of numerical atmospheric aerosol prediction with a focus on data requirements, *Atmos. Chem. Phys.*, 18, 10615–10643, <https://doi.org/10.5194/acp-18-10615-2018>, 2018.
- Bergamaschi, P., Segers, A., Brunner, D., Haussaire, J.-M., Henne, S., Ramonet, M., Arnold, T., Biermann, T., Chen, H., Conil, S., Delmotte, M., Forster, G., Frumau, A., Kubistin, D., Lan, X., Leuenberger, M., Lindauer, M., Lopez, M., Manca, G., Müller-Williams, J., O'Doherty, S., Scheeren, B., Steinbacher, M., Trisolino, P., Vítková, G., and Yver Kwok, C.: High-resolution inverse modelling of European CH_4 emissions using the novel FLEXPART-COSMO TM5 4DVAR inverse modelling system, *Atmos. Chem. Phys.*, 22, 13243–13268, <https://doi.org/10.5194/acp-22-13243-2022>, 2022.
- Burgers, G., Jan van Leeuwen, P., and Evensen, G.: Analysis scheme in the ensemble kalman filter, *Mon. Weather*

- Rev., 126, 1719–1724, [https://doi.org/10.1175/1520-0493\(1998\)126<1719:ASITEK>2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2), 1998.
- Buseck, P. R. and Pósfai, M.: Airborne minerals and related aerosol particles: effects on climate and the environment, *P. Natl. Acad. Sci. USA*, 96, 3372–3379, <https://doi.org/10.1073/pnas.96.7.3372>, 1999.
- Chen, B., Dong, L., Huang, J., Wang, Y., Jing, Z., Yan, W., Wang, X., Song, Z., Huang, Z., Guan, X., Dong, X., and Huang, Y.: Analysis of long-term trends in the vertical distribution and transport paths of atmospheric aerosols in typical regions of China using 15 years of CALIOP data, *J. Geophys. Res.-Atmos.*, 128, e2022JD038066, <https://doi.org/10.1029/2022JD038066>, 2023.
- 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.
- Di Tomaso, E., Escribano, J., Basart, S., Ginoux, P., Macchia, F., Barnaba, F., Benincasa, F., Bretonnière, P.-A., Buñuel, A., Castillo, M., Cuevas, E., Formenti, P., Gonçalves, M., Jorba, O., Klose, M., Mona, L., Montané Pinto, G., Mytilinaios, M., Obiso, V., Olid, M., Schutgens, N., Votsis, A., Werner, E., and Pérez García-Pando, C.: The MONARCH high-resolution reanalysis of desert dust aerosol over Northern Africa, the Middle East and Europe (2007–2016), *Earth Syst. Sci. Data*, 14, 2785–2816, <https://doi.org/10.5194/essd-14-2785-2022>, 2022.
- Escribano, J., Di Tomaso, E., Jorba, O., Klose, M., Gonçalves Ageitos, M., Macchia, F., Amiridis, V., Baars, H., Marinou, E., Proestakis, E., Urbanneck, C., Althausen, D., Bühl, J., Mamouri, R.-E., and Pérez García-Pando, C.: Assimilating spaceborne lidar dust extinction can improve dust forecasts, *Atmos. Chem. Phys.*, 22, 535–560, <https://doi.org/10.5194/acp-22-535-2022>, 2022.
- Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, 99, 10143, <https://doi.org/10.1029/94JC00572>, 1994.
- Gueymard, C. A. and Yang, D.: Worldwide validation of CAMS and MERRA-2 reanalysis aerosol optical depth products using 15 years of AERONET observations, *Atmos. Environ.*, 225, 117216, <https://doi.org/10.1016/j.atmosenv.2019.117216>, 2020.
- Gupta, M. C., Ungaro, C., Foley, J. J., and Gray, S. K.: Optical nanostructures design, fabrication, and applications for Solar/thermal energy conversion, *Sol. Energy*, 165, 100–114, <https://doi.org/10.1016/j.solener.2018.01.010>, 2018.
- Gwyther, D. E., Keating, S. R., Kerry, C., and Roughan, M.: How does 4DVar data assimilation affect the vertical representation of mesoscale eddies? A case study with observing system simulation experiments (OSSEs) using ROMS v3.9, *Geosci. Model Dev.*, 16, 157–178, <https://doi.org/10.5194/gmd-16-157-2023>, 2023.
- Hamill, T. M.: Ensemble-based atmospheric data assimilation, in: *Predictability of Weather and Climate*, edited by: Palmer, T. and Hagedorn, R., Cambridge University Press, 1st edn., <https://doi.org/10.1017/CBO9780511617652.007>, 124–156, 2006.
- Han, Y., Wang, T., Tang, J., Wang, C., Jian, B., Huang, Z., and Huang, J.: New insights into the Asian dust cycle derived from CALIPSO lidar measurements, *Remote Sens. Environ.*, 272, 112906, <https://doi.org/10.1016/j.rse.2022.112906>, 2022.
- Hofer, J., Althausen, D., Abdullaev, S. F., Makhmudov, A. N., Nazarov, B. I., Schettler, G., Engelmann, R., Baars, H., Fomba, K. W., Müller, K., Heinold, B., Kandler, K., and Ansmann, A.: Long-term profiling of mineral dust and pollution aerosol with multiwavelength polarization Raman lidar at the Central Asian site of Dushanbe, Tajikistan: case studies, *Atmos. Chem. Phys.*, 17, 14559–14577, <https://doi.org/10.5194/acp-17-14559-2017>, 2017.
- Houtekamer, P. L., Mitchell, H. L., Pellerin, G., Buehner, M., Charron, M., Spacek, L., and Hansen, B.: Atmospheric data assimilation with an ensemble Kalman filter: results with real observations, *Mon. Weather Rev.*, 133, 604–620, <https://doi.org/10.1175/MWR-2864.1>, 2005.
- 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.
- JAXA: JAXA Himawari Monitor (P-Tree System), JAXA [data set], <https://www.eorc.jaxa.jp/ptree>, last access: 23 June 2025.
- Jin, J., Lin, H. X., Segers, A., Xie, Y., and Heemink, A.: Machine learning for observation bias correction with application to dust storm data assimilation, *Atmos. Chem. Phys.*, 19, 10009–10026, <https://doi.org/10.5194/acp-19-10009-2019>, 2019.
- Jin, J., Segers, A., Lin, H. X., Henzing, B., Wang, X., Heemink, A., and Liao, H.: Position correction in dust storm forecasting using LOTOS-EUROS v2.1: grid-distorted data assimilation v1.0, *Geosci. Model Dev.*, 14, 5607–5622, <https://doi.org/10.5194/gmd-14-5607-2021>, 2021.
- 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.
- Jin, J., Fang, L., Li, B., Liao, H., Wang, Y., Han, W., Li, K., Pang, M., Wu, X., and Lin, H. X.: 4DVar-based inversion system for ammonia emission estimation in China through assimilating IASI ammonia retrievals, *Environ. Res. Lett.*, 18, 034005, <https://doi.org/10.1088/1748-9326/acb835>, 2023a.
- Jin, J., Henzing, B., and Segers, A.: How aerosol size matters in aerosol optical depth (AOD) assimilation and the optimization using the Ångström exponent, *Atmos. Chem. Phys.*, 23, 1641–1660, <https://doi.org/10.5194/acp-23-1641-2023>, 2023b.
- Kolb, C. E. and Worsnop, D. R.: Chemistry and composition of atmospheric aerosol particles, *Annu. Rev. Phys. Chem.*, 63, 471–491, <https://doi.org/10.1146/annurev-physchem-032511-143706>, 2012.
- Law, K. J. H. and Stuart, A. M.: Evaluating data assimilation algorithms, *Mon. Weather Rev.*, 140, 3757–3782, <https://doi.org/10.1175/MWR-D-11-00257.1>, 2012.
- Lee, L. A., Carslaw, K. S., Pringle, K. J., Mann, G. W., and Spracklen, D. V.: Emulation of a complex global aerosol model to quantify sensitivity to uncertain parameters, *Atmos. Chem. Phys.*, 11, 12253–12273, <https://doi.org/10.5194/acp-11-12253-2011>, 2011.
- Liu, Z., Liu, Q., Lin, H.-C., Schwartz, C. S., Lee, Y.-H., and Wang, T.: Three-dimensional variational assimilation of MODIS aerosol optical depth: implementation and application to a

- dust storm over East Asia, *J. Geophys. Res.*, 116, D23206, <https://doi.org/10.1029/2011JD016159>, 2011.
- 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.
- Manders, A. M. M., Builtjes, P. J. H., Curier, L., Denier van der Gon, H. A. C., Hendriks, C., Jonkers, S., Kranenburg, R., Kuenen, J. J. P., Segers, A. J., Timmermans, R. M. A., Visschedijk, A. J. H., Wichink Kruit, R. J., van Pul, W. A. J., Sauter, F. J., van der Swaluw, E., Swart, D. P. J., Douros, J., Eskes, H., van Meijgaard, E., van Ulft, B., van Velthoven, P., Banzhaf, S., Mues, A. C., Stern, R., Fu, G., Lu, S., Heemink, A., van Velzen, N., and Schaap, M.: Curriculum vitae of the LOTOS-EUROS (v2.0) chemistry transport model, *Geosci. Model Dev.*, 10, 4145–4173, <https://doi.org/10.5194/gmd-10-4145-2017>, 2017.
- Mhawish, A., Sorek-Hamer, M., Chatfield, R., Banerjee, T., Bilal, M., Kumar, M., Sarangi, C., Franklin, M., Chau, K., Garay, M., and Kalashnikova, O.: Aerosol characteristics from earth observation systems: a comprehensive investigation over South Asia (2000–2019), *Remote Sens. Environ.*, 259, 112410, <https://doi.org/10.1016/j.rse.2021.112410>, 2021.
- NASA/LARC/SD/ASDC: CALIPSO Lidar Level 2 Aerosol Profile, V4-21, NASA Langley Atmospheric Science Data Center DAAC [data set], https://doi.org/10.5067/CALIP/CALIPSO/CAL_LID_L2_05kmAPro-Standard-V4-21, 2025.
- Pang, M.: Xxcvvv/Open-PyFilter: Pyfilter_v1.1, Zenodo [code], <https://doi.org/10.5281/zenodo.14036308>, 2024.
- Pang, M.: Source Code of PyFilter, Zenodo [code], <https://doi.org/10.5281/zenodo.7611976>, 2024.
- Pang, M.: Materials for GMD-2024-113, Zenodo [data set], <https://doi.org/10.5281/zenodo.14846965>, 2025.
- Pang, M., Jin, J., Segers, A., Jiang, H., Fang, L., Lin, H. X., and Liao, H.: Dust storm forecasting through coupling LOTOS-EUROS with localized ensemble kalman filter, *Atmos. Environ.*, 306, 119831, <https://doi.org/10.1016/j.atmosenv.2023.119831>, 2023.
- Pappalardo, G., Amodeo, A., Apituley, A., Comeron, A., Freudenthaler, V., Linné, H., Ansmann, A., Bösenberg, J., D'Amico, G., Mattis, I., Mona, L., Wandinger, U., Amiridis, V., Alados-Arboledas, L., Nicolae, D., and Wiegner, M.: EARLINET: towards an advanced sustainable European aerosol lidar network, *Atmos. Meas. Tech.*, 7, 2389–2409, <https://doi.org/10.5194/amt-7-2389-2014>, 2014.
- Peng, D. and Liu, L.: Quantifying slab sinking rates using global geodynamic models with data-assimilation, *Earth Sci. Rev.*, 230, 104039, <https://doi.org/10.1016/j.earscirev.2022.104039>, 2022.
- Qin, W., Fang, H., Wang, L., Wei, J., Zhang, M., Su, X., Bilal, M., and Liang, X.: MODIS high-resolution MAIAC aerosol product: global validation and analysis, *Atmos. Environ.*, 264, 118684, <https://doi.org/10.1016/j.atmosenv.2021.118684>, 2021.
- Reichle, R. H., McLaughlin, D. B., and Entekhabi, D.: Hydrologic data assimilation with the ensemble Kalman filter, *Mon. Weather Rev.*, 130, 103–114, [https://doi.org/10.1175/1520-0493\(2002\)130<0103:HDAWTE>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<0103:HDAWTE>2.0.CO;2), 2002.
- Segers, A.: LOTOS-EUROS-v2.2 for GMD-2024-113, Zenodo [code], <https://doi.org/10.5281/zenodo.14039267>, 2024.
- 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.
- Sogacheva, L., Popp, T., Sayer, A. M., Dubovik, O., Garay, M. J., Heckel, A., Hsu, N. C., Jethva, H., Kahn, R. A., Kolmonen, P., Kosmale, M., de Leeuw, G., Levy, R. C., Litvinov, P., Lyapustin, A., North, P., Torres, O., and Arola, A.: Merging regional and global aerosol optical depth records from major available satellite products, *Atmos. Chem. Phys.*, 20, 2031–2056, <https://doi.org/10.5194/acp-20-2031-2020>, 2020.
- Stier, P., Schutgens, N. A. J., Bellouin, N., Bian, H., Boucher, O., Chin, M., Ghan, S., Huneus, N., Kinne, S., Lin, G., Ma, X., Myhre, G., Penner, J. E., Randles, C. A., Samset, B., Schulz, M., Takemura, T., Yu, F., Yu, H., and Zhou, C.: Host model uncertainties in aerosol radiative forcing estimates: results from the AeroCom Prescribed intercomparison study, *Atmos. Chem. Phys.*, 13, 3245–3270, <https://doi.org/10.5194/acp-13-3245-2013>, 2013.
- Tsikerdekis, A., Schutgens, N. A. J., and Hasekamp, O. P.: Assimilating aerosol optical properties related to size and absorption from POLDER/PARASOL with an ensemble data assimilation system, *Atmos. Chem. Phys.*, 21, 2637–2674, <https://doi.org/10.5194/acp-21-2637-2021>, 2021.
- van de Hulst, H. C.: Light Scattering by Small Particles. By H. C. van de Hulst. New York (John Wiley and Sons), London (Chapman and Hall), 1957. Pp. Xiii, 470; 103 Figs.; 46 Tables. 96s, Q. J. Roy. Meteor. Soc., 84, 198–199, <https://doi.org/10.1002/qj.49708436025>, 1958.
- van Leeuwen, P. J.: A consistent interpretation of the stochastic version of the ensemble Kalman filter, *Q. J. Roy. Meteor. Soc.*, 146, 2815–2825, <https://doi.org/10.1002/qj.3819>, 2020.
- Vignati, E., Karl, M., Krol, M., Wilson, J., Stier, P., and Cavalli, F.: Sources of uncertainties in modelling black carbon at the global scale, *Atmos. Chem. Phys.*, 10, 2595–2611, <https://doi.org/10.5194/acp-10-2595-2010>, 2010.
- Wang, F., Yang, T., Wang, Z., Cao, J., Liu, B., Liu, J., Chen, S., Liu, S., and Jia, B.: A comparison of the different stages of dust events over Beijing in March 2021: the effects of the vertical structure on near-surface particle concentration, *Remote Sens.-Basel*, 13, 3580, <https://doi.org/10.3390/rs13183580>, 2021.
- Wang, H., Yang, T., Wang, Z., Li, J., Chai, W., Tang, G., Kong, L., and Chen, X.: An aerosol vertical data assimilation system (NAQPMS-PDAF v1.0): development and application, *Geosci. Model Dev.*, 15, 3555–3585, <https://doi.org/10.5194/gmd-15-3555-2022>, 2022.
- Wang, X.: Historical data on air quality in china, <https://quotsoft.net/air/>, last access: 23 June 2025.
- 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](https://doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2), 2002.
- Whitaker, J. S., Compo, G. P., and Thépaut, J.-N.: A comparison of variational and ensemble-based data assimilation systems for reanalysis of sparse observations, *Mon. Weather Rev.*, 137, 1991–1999, <https://doi.org/10.1175/2008MWR2781.1>, 2009.
- Winker, D. M., Pelon, J., Coakley, J. A., Ackerman, S. A., Charlson, R. J., Colarco, P. R., Flamant, P., Fu, Q., Hoff, R. M., Kittaka, C., Kubar, T. L., Treut, H. L., McCormick, M. P., Mégie, G., Poole, L., Powell, K., Treppe, C., Vaughan, M. A., and Wielicki, B. A.: The CALIPSO mission: a global 3D view

- of aerosols and clouds, *B. Am. Meteorol. Soc.*, 91, 1211–1230, <https://doi.org/10.1175/2010BAMS3009.1>, 2010.
- Wu, X., Vu, T. V., Shi, Z., Harrison, R. M., Liu, D., and Cen, K.: Characterization and source apportionment of carbonaceous PM_{2.5} particles in China – a review, *Atmos. Environ.*, 189, 187–212, <https://doi.org/10.1016/j.atmosenv.2018.06.025>, 2018.
- Xu, X., Wu, H., Yang, X., and Xie, L.: Distribution and transport characteristics of dust aerosol over Tibetan Plateau and Taklimakan Desert in China using MERRA-2 and CALIPSO data, *Atmos. Environ.*, 237, 117670, <https://doi.org/10.1016/j.atmosenv.2020.117670>, 2020.
- Ye, H., Pan, X., You, W., Zhu, X., Zang, Z., Wang, D., Zhang, X., Hu, Y., and Jin, S.: Impact of CALIPSO profile data assimilation on 3-D aerosol improvement in a size-resolved aerosol model, *Atmos. Res.*, 264, 105877, <https://doi.org/10.1016/j.atmosres.2021.105877>, 2021.
- Yumimoto, K., Nagao, T., Kikuchi, M., Sekiyama, T., Murakami, H., Tanaka, T., Ogi, A., Irie, H., Khatri, P., Okumura, H., Arai, K., Morino, I., Uchino, O., and Maki, T.: Aerosol data assimilation using data from Himawari-8, a next-generation geostationary meteorological satellite, *Geophys. Res. Lett.*, 43, 5886–5894, <https://doi.org/10.1002/2016GL069298>, 2016.
- Zender, C. S., Bian, H., and Newman, D.: Mineral Dust Entrainment and Deposition (DEAD) model: description and 1990s dust climatology, *J. Geophys. Res.*, 108, 4416, <https://doi.org/10.1029/2002JD002775>, 2003.
- Zhang, M., Su, B., Bilal, M., Atique, L., Usman, M., Qiu, Z., Ali, M. A., and Han, G.: An investigation of vertically distributed aerosol optical properties over Pakistan using CALIPSO satellite data, *Remote Sens.-Basel*, 12, 2183, <https://doi.org/10.3390/rs12142183>, 2020.
- Zhao, C., Yang, Y., Fan, H., Huang, J., Fu, Y., Zhang, X., Kang, S., Cong, Z., Letu, H., and Menenti, M.: Aerosol characteristics and impacts on weather and climate over the Tibetan Plateau, *Natl. Sci. Rev.*, 7, 492–495, <https://doi.org/10.1093/nsr/nwz184>, 2020.
- Zhu, C., Maharajan, K., Liu, K., and Zhang, Y.: Role of atmospheric particulate matter exposure in COVID-19 and other health risks in human: a review, *Environ. Res.*, 198, 111281, <https://doi.org/10.1016/j.envres.2021.111281>, 2021.