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 Jin[1*], Ting Yang[3], Xi Chen[1], Arjo Segers[4], Batjargal Buyantogtokh[5], Yixuan Gu[1], Jiandong Li[1], Hai Xiang Lin[2,6], Hong Liao[1], and Wei Han[7*]

[1]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
[2]Delft Institute of Applied Mathematics, Delft University of Technology, Delft, the Netherlands
[3]State Key Laboratory of Atmospheric Boundary Layer Physics and Atmospheric Chemistry (LAPC), Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China
[4]TNO, Department of Climate, Air and Sustainability, The Netherlands
[5]Information and Research Institute of Meteorology, Hydrology and Environment, Ulaanbaatar, Mongolia
[6]Institute of Environment Sciences, Leiden University, the Netherlands
[7]CMA Earth System Modeling and Prediction Centre, Chinese Meteorological Administration, Beijing, China

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

Abstract. Modeling 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 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 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 (PM_{10}) 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 versus 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 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 owning to 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 three-dimensional aerosol measurements on a large scale remains impractical and costly. Measurements are limited to surface level concentrations of total particle matters, column-integrated aerosol optical depths (AOD), 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 model 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 also have demonstrated superior performance in sequential forecasting without the adjoint or tangent linear model that 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 quantifies the difference between observation field and simulation field (Ma et al., 2020). These two usually have different degrees of freedom. With that, the optimized state then could be calculated considering the integration of background and observational penalties. In perspective of aerosol data assimilation, the main object is to reproduce the optimal aerosol states (Liu et al., 2011), concerning their spatial, vertical, aerosol species and size distribution features. While the available observations commonly measure their mixed state. For example, AOD that is column-integrated optical extinction of all aerosols and high uncertainties exist in the

optical properties of aerosol (Tsikerdekis et al., 2021). Ground PM_{10} concentration measurement that is additive sum of all particles with diameter less than $10\,\mu m$. They are not directly comparable to the aerosol state. For ground measurements, the monitoring stations are sparsely distributed. High data vacancy is inevitable in AOD product as a result of retrieval failure.

55 The order of the observations is therefore commonly much lower than the model state or the degree of freedom is limited compared to that of model state. Mathematically, the assimilation result highly depends on the background error covariance for transferring the increments from the low-degree observational space into 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

60 from prior model gives rise to worsen posterior (Jin et al., 2023b). In addition, the position error in the dust plume simulation can cause failure in assimilating valid observations (Jin et al., 2021). As to the vertical structure, most studies have focused on proposing relevant assimilation methods to assimilate observations like Cloud-Aerosol Lidar with Orthogonal Polarization (CALIPSO) (Sekiyama et al., 2010; Cheng et al., 2019; Ye et al., 2021), Light Detection and Ranging (LiDAR) (Wang et al., 2022). The sensitivities of assumptions in the prior simulations need to be explored.

65 Take dust aerosol data assimilation for instance, the most representative observation sources are direct observation like PM_{10}, $PM_{2.5}$ concentrations and indirect observation like dust optical depth (DOD), extinction coefficient (Jin et al., 2022). Since they all are a combination of different particles, it is necessary to perform the 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, the ground stations produce scattered concentrations on ground, satellite and LiDAR receive column-

70 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 a continuous vertical information about the aerosol with large spatial coverage. When there is incorrect information about the dust aerosol structure in model prior, due to the considerably greater degree of freedom of model state, data assimilation algorithms may fail to correct it or even worse, further degrades the vertical dust aerosol loading.

75 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 on 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 lead assimilation methods to maintain or exacerbate inaccuracies.

80 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 PM_{10} 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 PM_{10} data from ground stations and DOD retrievals from the Himawari-8 satellite.

85 To validate, we employ aerosol profile observations from CALIPSO and LiDAR systems. By presenting both favorable 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 described the model used for 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 validate the results with the CALIPSO and LiDAR observations. Finally, Section 5 summaries this study.

2 Dust aerosol simulation and observations

This section introduces the model used for simulating the dust aerosol. Then describe the observation methods applied for monitoring the dust including ground monitor station, Himawari-8, CALIPSO and LiDAR. Among these, ground PM_{10} from ground monitor stations and DOD from Himawari-8 are used for assimilation. Extinction profile from CALIPSO and LiDAR are used to validate the posterior profile.

2.1 Dust simulation

2.2 LOTOS-EUROS

The LOTOS-EUROS v2.2 is used to simulate the dust aerosol. The LOTOS-EUROS model is a 3D chemistry transport model aimed for air quality forecasting (Manders et al., 2017). It has also been applied in source apportionment and emission inversion worldwide. In this study, The modeling domain spans from $15°$ N to $50°$ N and from $70°$ E to $140°$ E with spatial resolution of $0.25° \times 0.25°$. Vertically, it comprises 21 layers with a top level at 10 km, which is adequate for recognizing the vertical structure. 3 hourly ECMWF operational forecast 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 one day prior to the initial assimilation time point, during which no dust emission occurs. The dust emission process is modeled using the *Zender03* emission parameterization scheme (Zender et al., 2003). In general, we assign the dust simulation uncertainty to the dust emission. Ensemble emission field $[\boldsymbol{f}_1, ..., \boldsymbol{f}_N]$ are generated randomly following the emission uncertainty choice $\boldsymbol{f}_{\mathrm{priori}}$ 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 $[\boldsymbol{x}_1, ..., \boldsymbol{x}_N]$ as:

$$[\boldsymbol{x}_1, \cdots, \boldsymbol{x}_\mathbf{N}] = [\mathcal{M}(\boldsymbol{f}_1), \cdots, \mathcal{M}(\boldsymbol{f}_\mathbf{N})] \tag{1}$$

Here, N refers to the total ensemble number.

2.3 Dust observations for assimilation

2.3.1 Ground PM$_{10}$

115 There are over 1600 ground air quality monitor stations standing all over China. Their spatial distribution can be seen in fig. 1 (blue scatters). They provide hourly air pollutant concentrations including PM$_{2.5}$, PM$_{10}$, O$_3$, NO$_2$, SO$_2$ and CO. PM$_{10}$ can appropriately represent the dust load for it is under the size definition of fine and coarse dust (Adebiyi et al., 2023). Its uncertainty is also relatively lower than some remote sensing instruments. However, PM$_{10}$ is a combination of various species. Assimilating such observations with inaccurate or incorrect representativeness can diverge the model. It is then necessary to

120 perform bias-correction (BC) for removing the non-dust part before they are assimilated. Details concerning the method can be found in Jin et al. (2022). Afterwards, it 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 Meteorology Agency (JMA), provides aerosol optical information covering East Asia and Western Pacific Region at a time-resolution of 10-min (Yumimoto et al., 2016). This allows for

125 monitoring the source and movement of dust at high spatial-temporal resolution. Its products, Ångström exponent ($\mathring{A}$) and Aerosol Optical Thickness (AOT) at 550 nm, are selected to measure the dust optical depth (DOD). These data are merged to model grid ($0.25° \times 0.25°$) and hourly resolution.

To remove the fine-mode non-dust AOD in total AOD, an empirical function concerning Ångström exponent ($\mathring{A}$) is used to calculate the sub-micron fraction (SMF) Anderson et al. (2005); Di Tomaso et al. (2022). Then the dust optical depth (DOD)

130 can be obtained by the SMF.

$$SMF = -0.0512 \times \mathring{A}^2 + 0.5089 \times \mathring{A} + 0.02 \tag{2}$$

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

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

135 ### 2.4 Dust observations for validation

Aerosol extinction coefficient from the 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), onboard the CALIPSO satellite, provides 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

140 of 30 m. A normal extinction profile sample can be seen above the black dash line on map in fig. 1 (a). It has been wildly used

in investigating the aerosol distribution and evolution (Xu et al., 2020; Zhang et al., 2020; Han et al., 2022; Chen et al., 2023), 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

145 Polarization LiDAR deployed on the ground can provide continuous measurements about the vertical details of the atmospheric environment with high resolution (Hofer et al., 2017). Recently, Wang et al. (2021) collected the dust extinction coefficient data through a dual-wavelength aerosol LiDAR built in the tower of the Institute of Atmospheric Physics (39°58′35″N, 116°22′41″E, red star in fig. 1 (a)). The device provided the 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 occur in spring, 2021 was

150 captured thoroughly by the instrument, thus make it a crucial measurement source to learn about the vertical structure and its development during the dust storm. In this study, these data are employed to compare the vertical structure from model simulation and assimilation analysis.
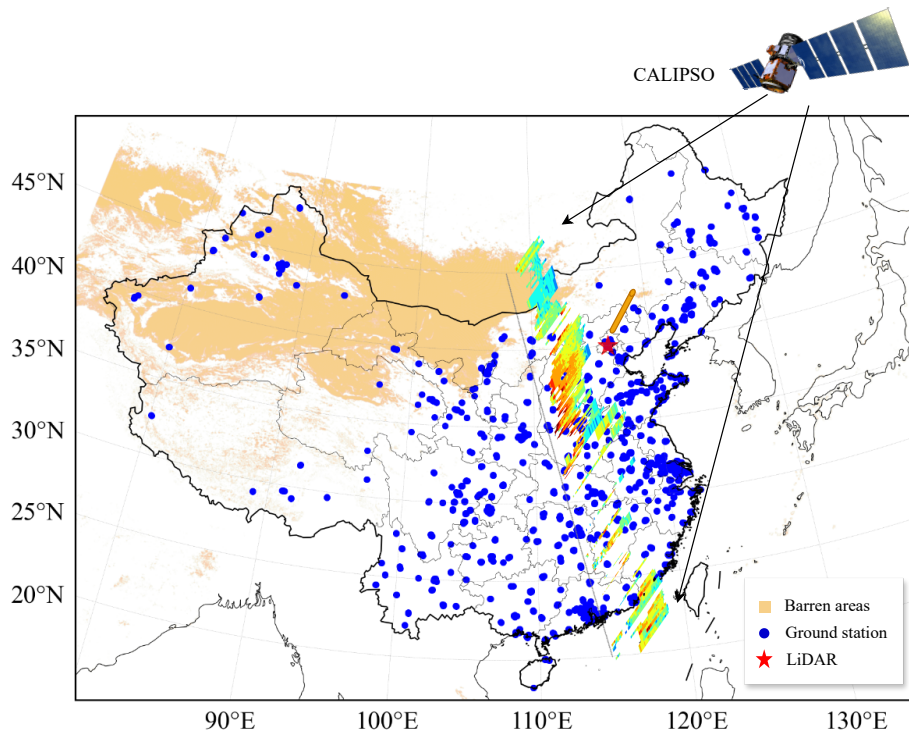
Figure 1. Spatial distribution of ground monitor stations (blue scatters), location of the LiDAR (red star) with upward scanning ray (orange cylinder) and aerosol extinction profile sample scanned by the CALIPSO satellite (trajectory is plotted as the gray line on ground).
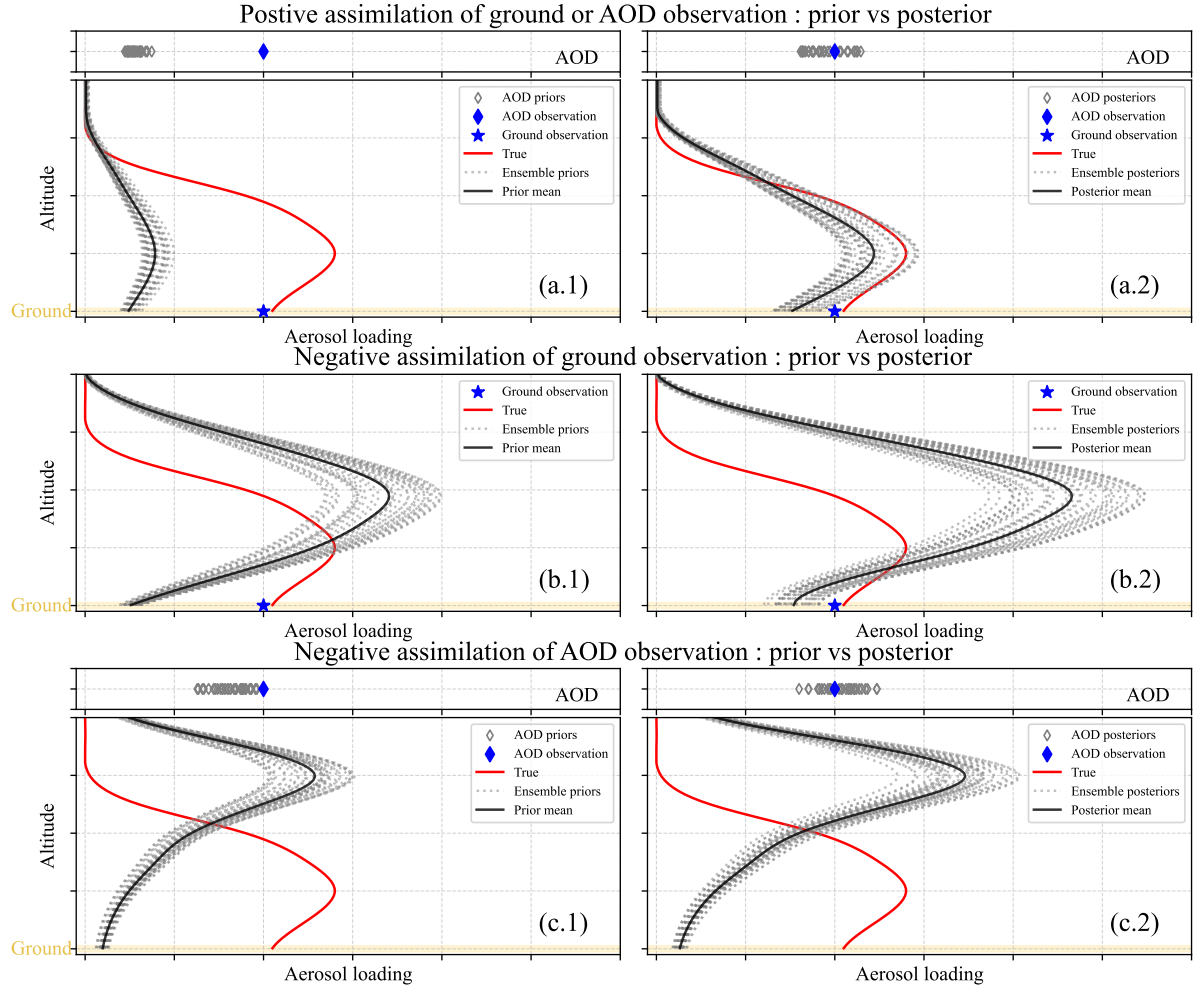
Figure 2. Schematic diagram of the sensitivity of data assimilation to aerosol vertical structure. Assimilation of ground and AOD observation under positive condition with AOD priors (gray hollow diamond), AOD observation (blue solid diamond) in upper box and true profile (red line), ensemble priors (gray dash line), prior mean (black line), ground observation (blue star) in lower box (**a.1**). Figure on the right are for the posteriors (**a.2**). Figures in the second row are prior (**b.1**) and posterior (**b.2**) of the ground assimilation under negative condition. Figures in the third row are prior (**c.1**) and posterior (**c.2**) of the AOD assimilation under negative condition.

3 Assimilation methodology and experiments

This Section first illustrates the methodology of the classic EnKF that were used in section 3.1, then sensitivity of the aerosol posterior to the vertical structure were emphasised when assimilating the ground PM_{10} concentration in section 3.2 and when assimilating the DOD in section 3.3. Cases experiments for evaluating the sensitiveness are described in section 3.4.

3.1 EnKF

The data assimilation algorithms used in this study is the stochastic EnKF formulated by Burgers et al. (1998). It is implemented in our self-developed PyFilter toolbox (Pang, last access: May. 2024). Its capability of improving dust forecasting had been proven in our recent research (Pang et al., 2023). We didn't adopt any other variant of EnKF for the reason that the problem we are going to explain 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 $x_{t-1}^{a,i}$ at previous time $t-1$ is integrated forward according to the model dynamics $\mathcal{M}$ to generate a prior forecast $x_t^{f,i}$ at the next moment t. i refers to the ensemble member here.

$$x_t^{f,i} = \mathcal{M}(x_{t-1}^{a,i}) \tag{4}$$

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

$$\mathbf{P}^f = \frac{1}{N-1}(x_i^f - \overline{x}^f)(x_i^f - \overline{x}^f)^{\mathrm{T}} \tag{5}$$

N is the number of ensembles. x_i^f is the ith individual of the ensemble prior states and $\overline{x}^f$ means the average of ensemble members.

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

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

The a posteriori ensemble individual x_i^a is calculated by :

$$x_i^a = x_i^f + \mathbf{K}(y + \epsilon_i - \mathcal{H}x_i^f) \tag{7}$$

ϵ_i is the extra perturbation and it's variance subjecting to the diagonal of observation error covariance $\mathbf{R}$. It serves to maintain the ensemble spread van Leeuwen (2020).

Meanwhile, the a posteriori x^a can be updated via

$$x^a = \overline{x}^f + \mathbf{K}(y - \mathcal{H}\overline{x}^f) \tag{8}$$

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

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

3.2 Ground PM$_{10}$ assimilation

In this section, impact of ground measurements data assimilation on the aerosol vertical structure is explained. In practice, each of the PM$_{10}$ observations could be assimilated to optimize the 3D states that are correlated. The increments is transferred from the the observation space in the model space via the Kalman Gain $\mathbf{K}$ shown above. To easily evaluate the posterior, we solely perform the assimilation analysis at the pixels where PM$_{10}$ measurements are available. So here we confine the states only in vertical direction in one pixel to better identify the sensitivity, and assuming the prior states $\boldsymbol{x}^f$ with total k levels:

$$\boldsymbol{x}^f = [x_1^f, \cdots\cdots, x_k^f]^{\mathrm{T}} \tag{9}$$

In which 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 ground level:

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

$$\mathcal{H} = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix} \tag{11}$$

Calculation of $\mathbf{K}$ follows the eq. (6) and substitute eq. (10) and eq. (11) into it:

$$\mathbf{K} = \mathbf{P}^f \mathcal{H}^{\mathrm{T}} (\mathcal{H} \mathbf{P}^f \mathcal{H}^{\mathrm{T}} + \mathbf{R})^{-1} = [C_{11}, \cdots\cdots, C_{k1}]^{\mathrm{T}} (C_{11} + \sigma_y^2)^{-1} \tag{12}$$

σ_y^2 is the error of the ground observation.

Substitute $\mathbf{K}$ into the update function:

$$[x_1^a, \cdots\cdots, x_k^a]^{\mathrm{T}} = [x_1^f, \cdots\cdots, x_k^f]^{\mathrm{T}} + [C_{11}, \cdots\cdots, C_{k1}]^{\mathrm{T}} (C_{11} + \sigma_y^2)^{-1} (y - x_1^f) \tag{13}$$

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

On the ground level, the posterior state is obtained as:

$$\boldsymbol{x}_1^a = \frac{\sigma_y^2}{C_{11} + \sigma_y^2} \boldsymbol{x}_1^f + \frac{C_{11}}{C_{11} + \sigma_y^2} y \tag{14}$$

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

$$x_0^a \approx y \tag{15}$$

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

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

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

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

$$v_k = \frac{x_i}{x_1} \tag{17}$$

which is the ratio of state on ith level to ground level.

Divide x_i^f on both sides of eq. (16):

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

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 $v_i^f \ll v_i^{true}$ or $v_i^f \gg v_i^{true}$. Then, the vertical structure of posterior is high 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 enhance the incorrect prior structure v_i^f.

Figure 2 (b) illustrates this kind of pattern. In panel (b.1), the priors have greatly underestimated the aerosol concentration on the ground level. Meanwhile, over the upper space, it has overestimated 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 (gray dash line) is tuned to a large extent and much closer to the observation on ground layer. While for the vertical distribution, a greater overestimation than the prior can be noticed. The original incorrect profile is significantly amplified.

3.3 AOD assimilation

Assimilation of AOD-related observations is carried out by connecting the 3-D aerosol priors $\boldsymbol{x}^f$ and AOD priors $\boldsymbol{\tau}^f$. It is assumed that the vertical structure of aerosol field remains the same for the AOD lacks the vertical information. Then, a 3-D mass concentration field can be calculated by finding the optimal AOD field.

To begin with, value of AOD is related to the aerosol mass concentration, aerosol type, humidity, et al. Here, it's assumed to be linearly related to the aerosol mass concentration field $\boldsymbol{x}^f$ solely. The column-integrated AOD is calculated by the summary

of aerosol mass concentration x_i^f on each level:

$$\tau^f = \sum_{i=1}^{k} \mathcal{M} x_i^f \tag{19}$$

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

Calculation of posterior AOD τ^a follows the eq. (8):

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

Then, substitute eq. (19) into eq. (20). We can bridge the mass concentration field and AOD field:

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

Same for AOD, let's confine the states in vertical direction in one pixel. We have one prior AOD τ^f and aerosol mass prior x_i^f on total k levels. Define a vertical factor u_i which is the ratio of mass of ith level to the mass of all levels first:

$$u_i = \frac{x_i}{\displaystyle\sum_{i=1}^{k} x_i} \tag{22}$$

Substitute it into eq. (21):

$$\mathcal{M} \frac{x_i^a}{u_i} = \mathcal{M} \frac{x_i^f}{u_i} + \mathbf{K}(y - \mathcal{H}\tau^f) \tag{23}$$

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

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

Substitute them into eq. (23) and we have the mass posterior x_i^a on each layer:

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

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

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 can notice that $u_i \gg u_i^{true}$ or $u_i \ll u_i^{true}$. The analysis increment can be reallocated on the wrong level. Thus amplify the incorrect structure.

Figure 2 (c) demonstrates the negative pattern of assimilating AOD. In the upper row of panel (c.1), we have the ensemble AOD priors (gray hollow diamond) and AOD observation (blue solid diamond) that are combined to optimize the 3-D dust

mass field (gray dash 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 overall amplified since the prior DOD simulation implies the model underestimates the total dust column compared to the observation. Another pattern which the prior mass is incorrectly situated on ground level is also illustrated in Supplementary fig. S1. It is notable that the sensitivity of AOD assimilation to the vertical structure is not as significant as the ground assimilation since the AOD contains information of total columns.

3.4 Case settings

To demonstrate the sensitivity of aerosol data assimilation on vertical structure, five assimilation analysis cases were selected and name after *P-Gd-CAL*, *N-Gd-CAL*, *N-Gd-Li*, *N-DOD-CAL* and *NP-DOD-CAL*, respectively. These cases were selected from the ensemble model run starting from 0:00 14th to 0:00 17th March and 0:00 14th to 0:00 17th April 2021. Dust classification is carried out by the CALIPO depolarization ratio data to ensure that dust is the dominant particle during these time range. The procedures and results can be found in Supplementary. The background run consists of 32 ensembles and each of them are driven by perturbated emission inventories. Details concerning the generation of the ensembles can be found in Jin et al. (2022).

Table 1 lists information about the experiment cases. *P-Gd-CAL* is the positive case which shows how the data assimilation correct the surface concentrations and maintain the correct aerosol structure. *N-Gd-CAL* and *N-Gd-Li* are negative cases which shows how the assimilation deteriorates the profile when assimilating ground observations solely. The time of the former two cases is fixed, and their profiles follow the trajectory of CALIPSO. Hence, by comparison we can tell the impact of assimilation on vertical structure. The last case focuses on a fixed site. A continuous time-line 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 the CALIPSO.

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

Table 1. Experiment configurations

Name	Time (CST)	Type	Assimilated observations	Validation data
P-Gd-CAL	3:00 16th April 2021	positive	BC-PM$_{10}$	CALIPSO
N-Gd-CAL	3:00 16th March 2021	negative	BC-PM$_{10}$	CALIPSO
N-Gd-Li	15th March 2021	negative	BC-PM$_{10}$	LiDAR
N-DOD-CAL	14:00 16th March 2021	negative	Himawari-8 DOD	CALIPSO
NP-DOD-CAL	14:00 29th March 2021	negative & positive	Himawari-8 DOD	CALIPSO

4 Results and discussions

This section delves into the sensitivities of data assimilation from ground and DOD observation, with a particular emphasis on the impact of prior vertical profile. Regarding 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 from 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 a pair of instances, highlighting favorable and unfavorable 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 aimed to show how the assimilation works properly. Figure 3 (a.1) is time series of ground PM_{10} and BC-PM_{10} from several ground stations that is close to CALIPSO trajectory (< 0.25°). The assimilation time point, 3:00 16th April 2021, is selected in this case (highlighted in green). Although it is on the later stage of the dust storm, the concentrations were as high as around 400 μgm^{-3}. Through comparison between raw and bias-corrected PM_{10} (solid and dash line), it is clear that the scanned region was still dominated by coarse dust aerosol. Figure 3 (c.1 and d.1) are snapshots of average of ensemble priors and posteriors. We can see that the priors had underestimated the dust load to the south of the domain and overestimated it to the east. After assimilation analysis, the dust load had been adjusted that better fits the observations. The ground observations had been consistent with the posterior. RMSE has been reduced from 254.83 μgm^{-3} to 118.03 μgm^{-3}.

Take a further lookup to the vertical structure. Figure 3 (b) illustrates the extinction coefficient profile from CALIPSO. From the profile we can tell that the dust was concentrated on the ground over latitude of 34° to 38° (circled by dash line). Vertically, it extended upwards accompanied with terrain from 1 km at 34.5° to 2.5 km at 37.5°. Figure 3 (c.1) shows the prior dust profile following the same trajectory of the CALIPSO. The color bar here is rescaled to show more detail. At around 41° to 42°, a heavy dust load was simulated (600-800 μgm^{-3}) while CALIPSO didn't capture the data there. Hence, it can't be validated, and we will focus on the dust at 34° to 38°. At 34° to 38°, the prior model had successfully reproduced the dust structure indicated by CALIPSO. The concentrations were about 100-300 μgm^{-3}, which is underestimated as shown in fig. 3 (d.2). After assimilation, the ground concentrations are increased to 400-500 μgm^{-3} as shown in (c.1) and (d.1). The vertical structure is also reserved, which is identical to the structure shown by CALIPSO.

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 4 (a.1) is time series of ground PM_{10} and BC-PM_{10} from several ground stations that is close to CALIPSO trajectory. As shown in
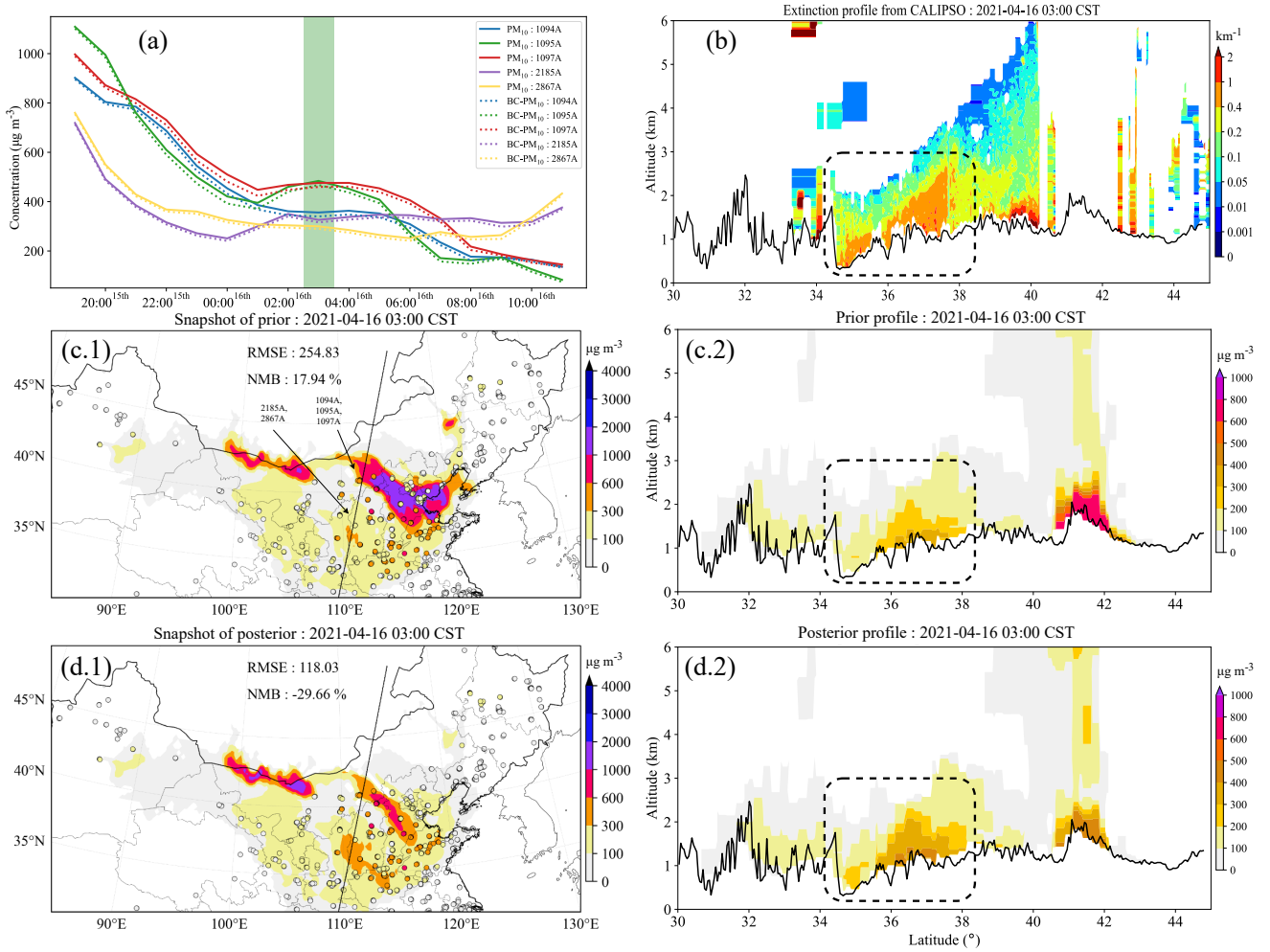
Figure 3. Time series of PM_{10} and BC-PM_{10} concentrations during the case from several ground stations that are close to the CALIPSO trajectory (**a**). Spatial distribution of ground dust concentrations from the average of the ensemble priors and the posteriors with scatters of ground BC-PM_{10} observations (**c.1** and posterior **d.1**). The black line inside is where the CALIPSO scanned through. Figures on the right column is the extinction profile from CALIPSO (**b**), dust concentration profile following the CALIPSO scanning trajectory from the prior (**c.2**) and the posterior (**d.2**). The color bars in **c.2** and **d.3** are rescaled to show more detail. Black line at the bottom is the terrain altitude. The case time is at 3:00 on 16th April 2021 (CST).

panel (a), ground dust aerosol loading from ground stations in *N-Gd-CAL* was under rapid increase since two hours before assimilation instance. Meaning that the region in the below rectangle in (c.1) was dust concentrated region.

Evidence from CALIPSO also indicated that there were heavy dust loading. As the fig. 4 (b) shows, extinction coefficient profile from 37° to 42° (circled by black dash rectangle) exhibits extremely high values (maximum exceeded 2 km^{-1}). It spread upwards for 1 km at an altitude of 1 km. This region is highlighted by light blue dash rectangle in (c.1) and (d.1). Inferred from the farther ground stations (concentrations over 1000 µgm^{-3}), this region was dominated by dust aerosols and is situated on ground level.

The dust profile in (c.2) tells that the ensemble priors agreed that the aerosols were mainly concentrated upwards. The plume was floating up to 4 km with the latitude. Only a few dust aerosols over 36° to 37° on ground (300-600 µgm^{-3}) show consistence with the observations (shown in (c.1) and (c.2)). Great discrepancy exits between the prior and independent observations at 37° to 42°. After assimilating BC-PM$_{10}$ concentrations, the overall loading are increased. Great underestimation in east was improved. However, The dust concentrated region circled in (d.1) are only improved to a small extent. This is due to lack of observations nearby and the much lower uncertainties exhibited by ensemble priors. In terms of the vertical structure, a disturbing profile is obtained. As fig. 4 (d.2) shows, ground concentrations at 36° to 37° are aligned to over 600 µgm^{-3}. This is consistent with surrounding stations. While for the reason mentioned earlier, dust loads over 37° to 42° are barely corrected. Yet, these trivial increments significantly enhance the erroneous vertical structure. On ground level, the dust concentrations rise from less than 100 µgm^{-3} to nearly 300 µgm^{-3} at 37° to 39°. The vertical loading are almost 3 times of loading from prior profile. What's more, the extremely low region at 39° to 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 soaring of the vertical loading. In this case, this phenomenon can be vividly noticed through the comparison between prior, posterior and measurements for validation.

4.1.3 Negative case validated by LiDAR

N-Gd-Li is another negative case. In this case, we don't focus on single time point. A series of assimilation analyses are conducted to investigate the effect. As fig. 5 (a) shows, a high-resolution profile of dust extinction coefficient from LiDAR is given to demonstrate the changing structure of dust loading on a fix location. Entering at about 5:00 and then landing at 7:00, this dust storm had lasted for about 20 hours in where the LiDAR located. Time range from 8:00 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 km to 1.4 km. Afterwards, the intensity descended gradually.

Figure 5 (b) illustrates the hourly prior and posterior dust profile during the selected period. The profiles of extinction coefficient are also plotted in blue dash dot line. The x-axis which represents concentrations is logarithmic rescaled here. In terms of prior profiles, dust loading have extended up to 5 km at 8:00 and declined to 3 km at 16:00. This structure is inconsistent with the LiDAR profile, in which time only little dust was observed. The dust storm has been overestimated to a large extent in height. The ground aerosol concentrations are lower than $200\,\mu g\,m^{-3}$ throughout eight time points, which are much lower than the observations (over $1000\,\mu g\,m^{-3}$). After assimilation, this underestimation has been mitigated. The ground
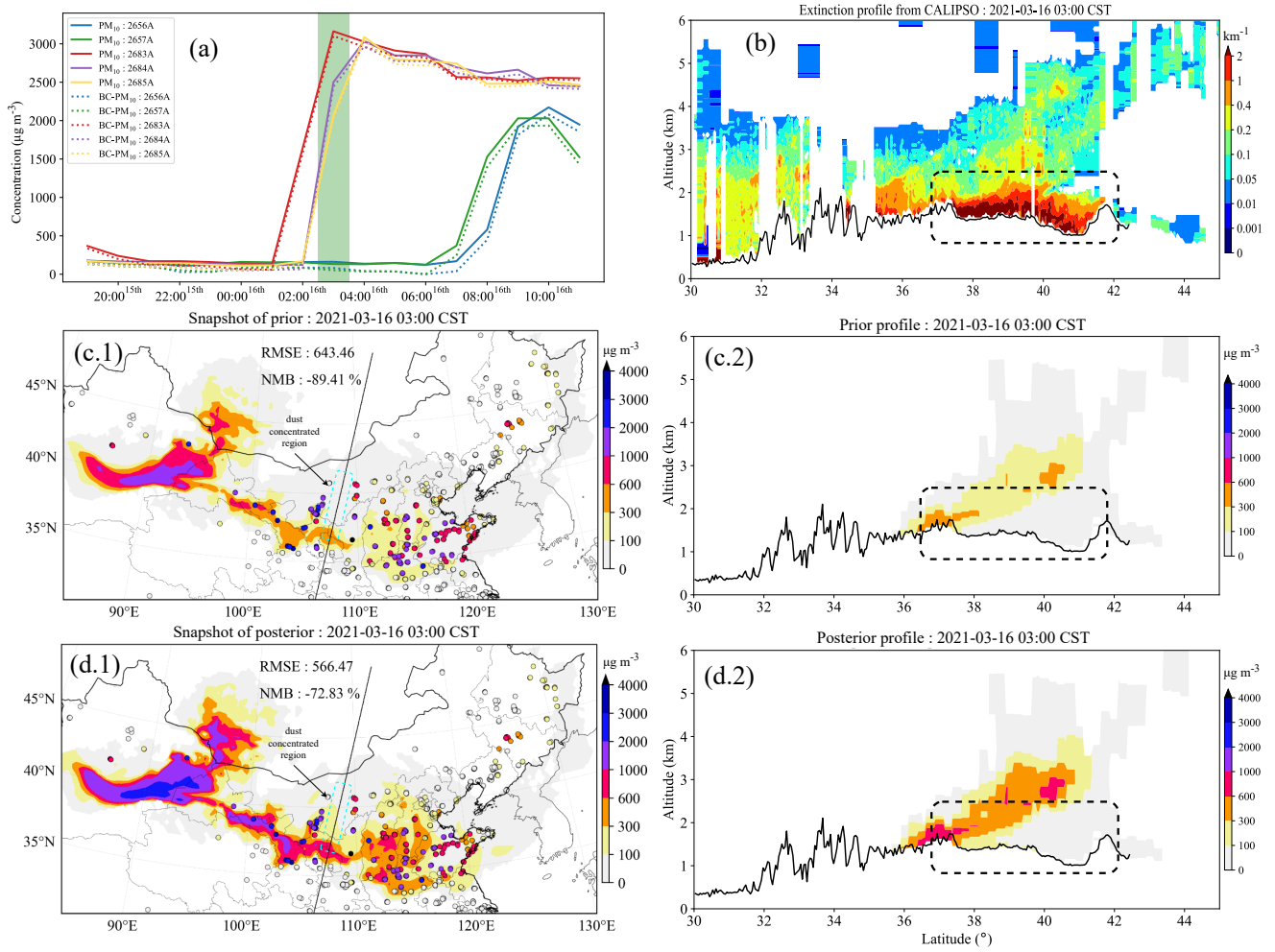
Figure 4. Time series of PM_{10} and $BC-PM_{10}$ concentrations during the case from several ground stations that are close to the CALIPSO trajectory (**a**). Spatial distribution of ground dust concentrations from the average of the ensemble priors and the posteriors with scatters of ground $BC-PM_{10}$ observations (**c.1** and posterior **d.1**). The black line inside is where the CALIPSO scanned through. Figures on the right column is the extinction profile from CALIPSO (**b**), dust concentration profile following the CALIPSO scanning trajectory from the prior (**c.2**) and the posterior (**d.2**). Black line at the bottom is the terrain altitude. The case time is at 3:00 on 16th March 2021 (CST).

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 seen in the comparison between concentration and extinction coefficient. Especially in 9:00, the dust loading above 2 km have been amplified to over $1000\,\mu\mathrm{g\,m^{-3}}$, which will give out utterly incorrect information about the dust storm structure and impact the further forecasting.
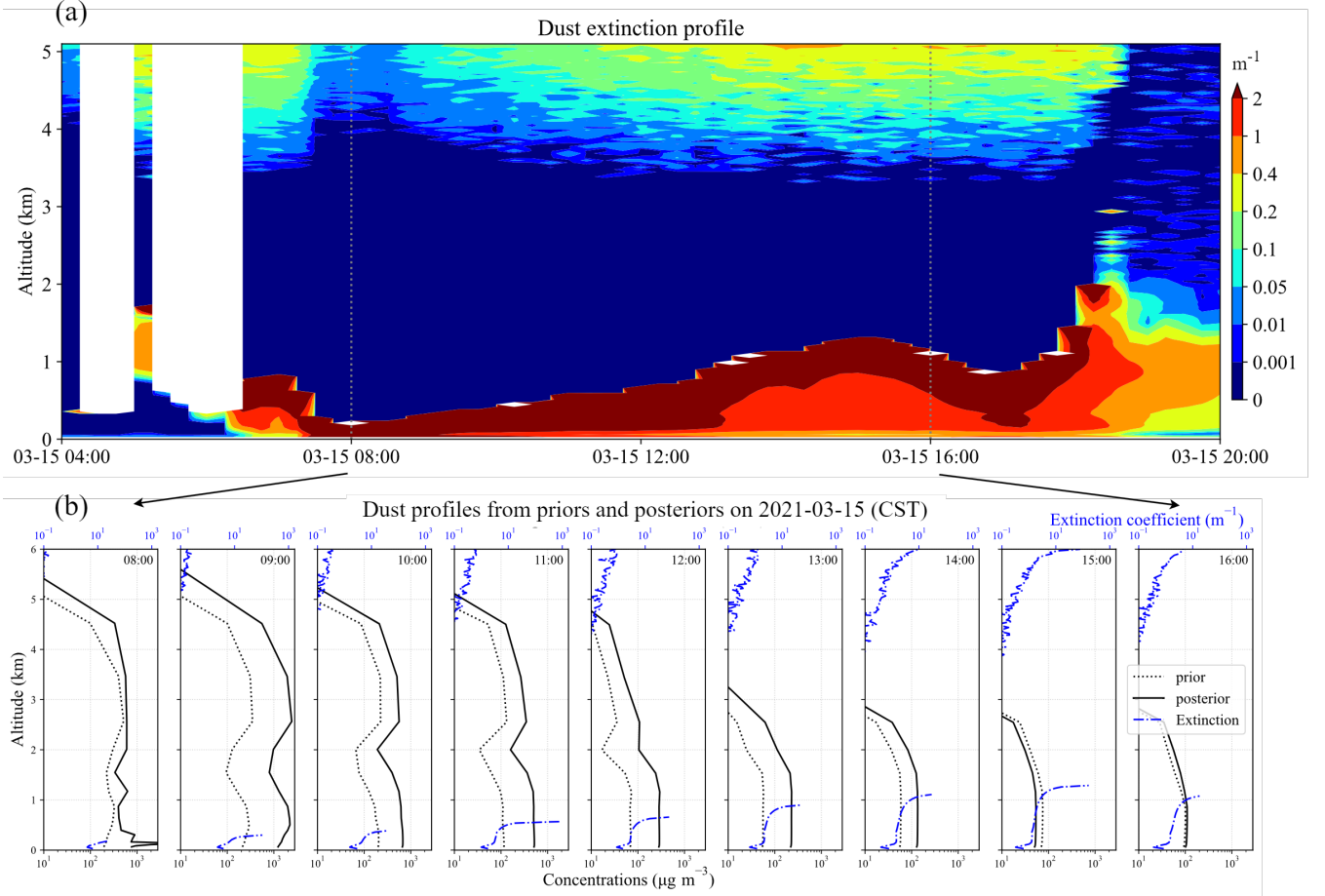
Figure 5. Time series of dust extinction coefficient profile obtained from LiDAR (**a**). The assimilation analysis is performed hourly from 8:00 to 16:00. Figure below is hourly dust profile line from prior (dash line) and posterior (solid line) (**b**). The extinction coefficient is also plotted (blue dash dot line). The profile data is extracted from the closest grid point to the LiDAR location. The x-axis is logarithmic rescaled. Note that instead of using the posterior in previous time to propagate the model, analysis here is separately conducted on the static background.

4.2 Cases on DOD assimilation

In this section, the sensitivity of assimilating DOD on the aerosol vertical profile is presented. Starts with the *NP-DOD-CAL* case. Figure 6 (a) show the spatial distribution of DOD observed by Himawari-8 at 14:00 29th March 2021. The dust plume then

355 were approaching the east of China and reaching out to the East China Sea. The prior model has reproduced the distribution while the overall concentration is underestimated (less than $300\,\mu\mathrm{g\,m}^{-3}$) as shown in panel (c.1). By assimilating the Himawari-8 DOD, the overall dust concentration field is increased to over $1000\,\mu\mathrm{g\,m}^{-3}$. The posterior dust plume is consistent with the DOD observations.

As the extinction profile in panel (b) shows, the dust was concentrated around $32°$ to $38°$ and extended upwards to altitude

360 of 3 km. In terms of the prior, this structure is partly correct. The dust in red box is located on lower layer and aligns with the extinction profile. While the dust in blue box reports 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 red box is maintained and the intensity is enhanced. However, the incorrect dust structure in blue box is also amplified, causing an obvious inaccurate dust loading which should

365 located on lower level.

N-DOD-CAL is another case that illustrates the negative aspects of assimilating DDO. Figure 7 comprises the same assimilation experiment results as in *NP-DOD-CAL* except at 14:00 16th March 2021. The DOD observations report high values densely distributed in east and sparsely distributed in north. The prior model simulates the dust plume that concentrated in north-west. As to the east, only rare dust is noticed. It is caused by instability of model to reproduce the long-term transport of

370 dust. If we focus on where the CALIPSO scanned through, an increase of dust can be seen.

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

375 ## 5   Conclusions

Data assimilation has been wildly used in construction of reanalysis dataset and improving the predictability of atmospheric model. The performance of assimilation algorithms relies not only on the methodology themselves but also on the observations. Nowadays, the complexity and dimensions of atmospheric model have been growing fast. For atmospheric aerosol model, high spatial and temporal resolution are becoming feasible. While in terms of observations, they are either outnumbered in space

380 or discontinuous in time compared to the model. Particularly in vertical direction, the available aerosol measurements are still limited both in space and time. Through assimilating ground or satellite measurements, the aerosol field can be improved while in 3D space, it can also amplify the erroneous estimation about the aerosol structure.

In this paper, we explore the sensitivity of aerosol data assimilation to the vertical profile via carrying out five assimilation experiment cases. Dust aerosol, which is simulated by LOTOS-EUROS, is selected as the optimization target. The data assim-

385 ilation methodology, EnKF, is adopted. Bias-corrected ground PM_{10} concentrations obtained from ground monitoring stations and DOD from the Himawari-8 are assimilated. Extinction coefficient profile from CALIPSO and Polarization LiDAR are used to validate the aerosol vertical structure. *P-Gd-CAL* is aimed to show how ground data assimilation tune the vertical structure
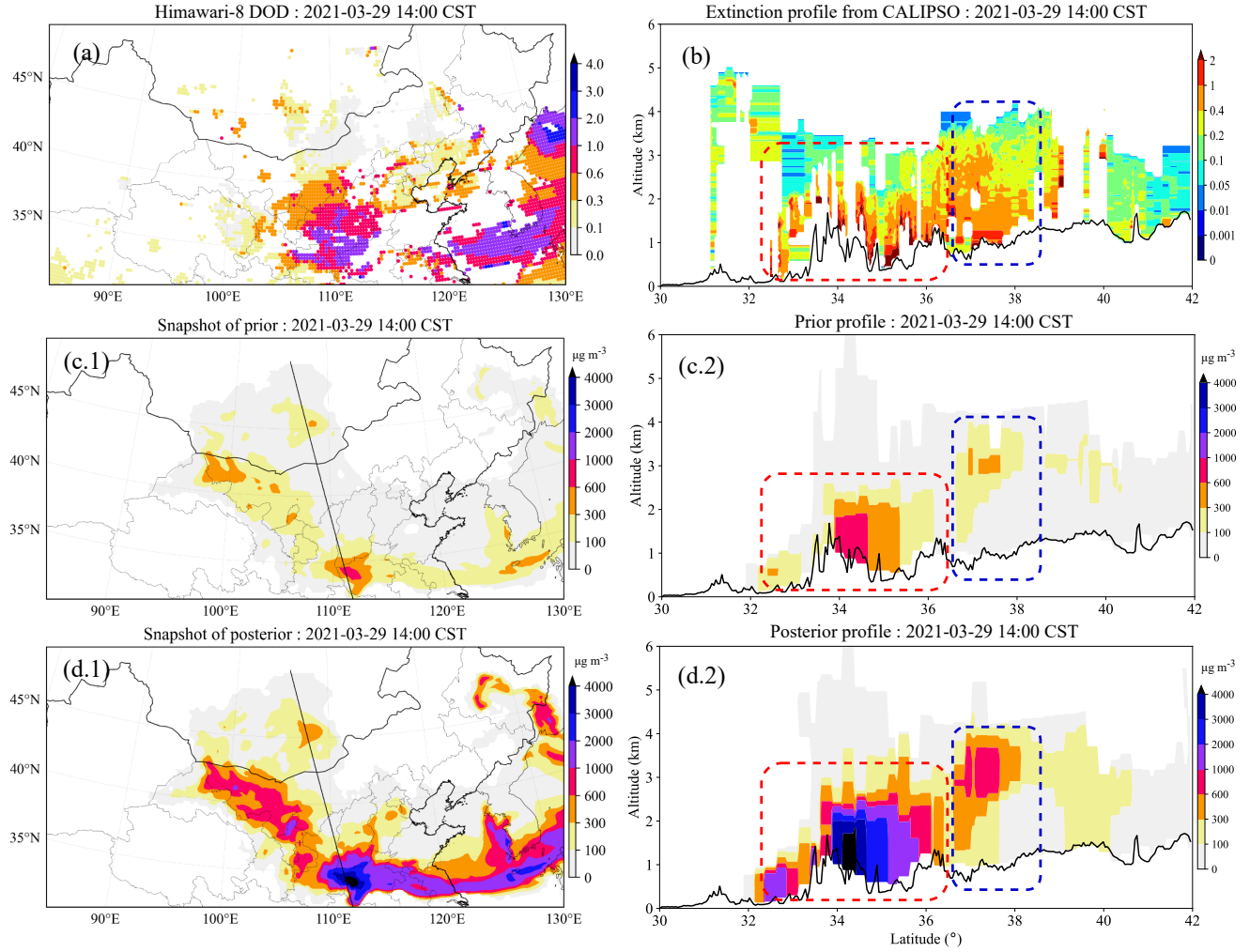
Figure 6. Time series of PM_{10} and $BC\text{-}PM_{10}$ concentrations during the case from several ground stations that are close to the CALIPSO trajectory (**a**). Spatial distribution of ground dust concentrations from the average of the ensemble priors and the posteriors with scatters of ground $BC\text{-}PM_{10}$ observations (**c.1** and posterior **d.1**). The black line inside is where the CALIPSO scanned through. Figures on the right column is the extinction profile from CALIPSO (**b**), dust concentration profile following the CALIPSO scanning trajectory from the prior (**c.2**) and the posterior (**d.2**). Black line at the bottom is the terrain altitude. The case time is at 14:00 on 29th March 2021 (CST).
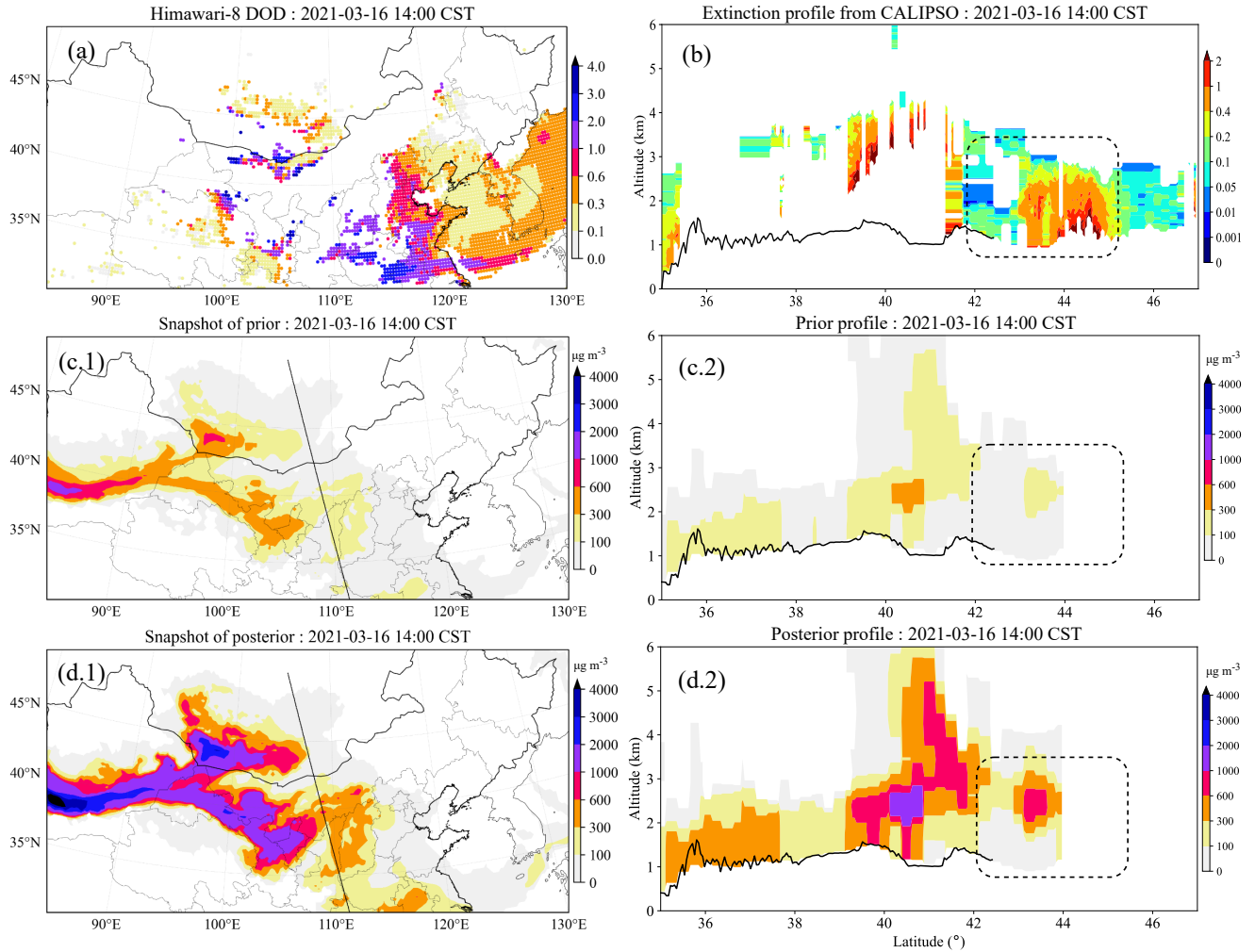
Figure 7. Snapshot of Himawari-8 DOD at the assimilation instance (**a**). Spatial distribution of ground dust concentrations from the average of the ensemble priors (**c.1** and posterior **d.1**). The black line inside is where the CALIPSO scanned through. Figures on the right column is the extinction profile from CALIPSO (**b**), dust concentration profile following the CALIPSO scanning trajectory from the prior (**c.2**) and the posterior (**d.2**). Black line at the bottom is the terrain altitude. The case time is at 14:00 on 16th March 2021 (CST).

positively. After assimilation, the ground dust field has been 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 drastic negative effect. In this case, even small increments on ground can greatly enhance the incorrect aerosol structure. *N-Gd-Li* is another negative case which displays hourly prior and posterior profile. 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 degrades the vertical dust loading. *N-DOD-CAL* and *NP-DOD-CAL* are aimed to show how DOD assimilation deteriorates the vertical structure. Different from the ground assimilation, the posterior vertical structure here is 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 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 amplified 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:

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

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:

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

Here, λ_0 represents the reference wavelength where the AOD is known, and λ is the target wavelength. The wavelengths $550\,\mathrm{nm}$ and $532\,\mathrm{nm}$ are very close in the visible spectrum, with a difference of only $18\,\mathrm{nm}$. 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). Is defined as:

$$\tau = \sum_{k=1}^{n} \epsilon_d^k z^k \tag{B1}$$

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

$$\epsilon_d^k = Q_{ext} S C \tag{B2}$$

where Q_{ext} is the sum of scattering and absorption efficiency. It's decided by the ratio of aerosol radius, incident wavelength and chemical composition (H. C., 1958). S depends on the particle size and aerosol mass density. The dust bins and diameter ranges are shown in table A1. Detailed descriptions concerning the calculation of Q_{ext} and S can be found in Section 2, Jin et al. (2023b).

Table A1. 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), also open at GitHub (https://github.com/xxcvvv/open-PyFilter). The ensemble initial fields, prior and posterior fields used in this paper are archived in https://doi.org/10.5281/zenodo.14846965. 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://www-calipso.larc.nasa.gov/about/ (NASA, last access: August 2023). The ground PM$_{10}$ observations can be obtained at https://quotsoft.net/air/ (Wang, last access: May. 2024). The Himawari-8 aerosol product is available at https://www.eorc.jaxa.jp/ptree.

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

Acknowledgements. This study was supported by the National Key Research and Development Program of China [grant number 2024YFE0113700] and the National Natural Science Foundation of China [grant number 42475150].

Competing interests

The authors declare that they have no conflict of interest.

References

Adebiyi, 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, 100 849, 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, https://doi.org/10.1029/2005JD005978, 2005.

Bannister, R. N.: A Review of Operational Methods of Variational and Ensemble-Variational Data Assimilation, Q. J. R. Meteorolog. 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, 10 615–10 643, 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, 13 243–13 268, 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, 1998.

Buseck, P. R. and Pósfai, M.: Airborne Minerals and Related Aerosol Particles: Effects on Climate and the Environment, PNAS, 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, e2022JD038 066, 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, 13 445–13 467, 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., Castrillo, 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, 10 143, 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, 117 216, 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, Solar 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.

H. C., v. d. H.: 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, Quarterly Journal of the Royal Meteorological Society, 84, 198–199, https://doi.org/10.1002/qj.49708436025, 1958.

Hamill, T. M.: Ensemble-Based Atmospheric Data Assimilation, in: Predictability of Weather and Climate, edited by Palmer, T. and Hagedorn, R., pp. 124–156, Cambridge University Press, 1 edn., https://doi.org/10.1017/CBO9780511617652.007, 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, 112 906, 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, 14 559–14 577, 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.

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, 10 009–10 026, 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.: 4DEnVar-based Inversion System for Ammonia Emission Estimation in China through Assimilating IASI Ammonia Retrievals, Environ. Res. Lett., 18, 034 005, 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.

520 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, 12 253–12 273, 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, https://doi.org/10.1029/2011JD016159, 2011.

525 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, e2019JD031 465, 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., 530 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– 535 2019), Remote Sens. Environ., 259, 112 410, https://doi.org/10.1016/j.rse.2021.112410, 2021.

NASA: Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations, last access: August 2023.

Pang, M.: Xxcvvv/Open-PyFilter: Pyfilter_v1.1, Zenodo, https://doi.org/10.5281/zenodo.14036308, 2024.

Pang, M.: Source Code of PyFilter, Zenodo, https://doi.org/10.5281/zenodo.7611976, last access: May. 2024.

Pang, M., Jin, J., Segers, A., Jiang, H., Fang, L., Lin, H. X., and Liao, H.: Dust Storm Forecasting through Coupling LOTOS-EUROS with 540 Localized Ensemble Kalman Filter, Atmos. Environ., 306, 119 831, 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, 545 104 039, 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, 118 684, 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, 2002.

550 Segers, A.: LOTOS-EUROS-v2.2 for GMD-2024-113, Zenodo, 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., Huneeus, 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 Leeuwen, P. J.: A Consistent Interpretation of the Stochastic Version of the Ensemble Kalman Filter, Q. J. R. Meteorolog. 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., 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: May. 2024.

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

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., Trepte, 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, 117 670, 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, 105 877, https://doi.org/10.1016/j.atmosres.2021.105877, 2021.

590 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, https://doi.org/10.1029/2002JD002775, 2003.

595 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., 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
600 Human: A Review, Environ. Res., 198, 111 281, https://doi.org/10.1016/j.envres.2021.111281, 2021.