

## Data assimilation of CALIPSO aerosol observations

T. T. Sekiyama<sup>1</sup>, T. Y. Tanaka<sup>1</sup>, A. Shimizu<sup>2</sup>, and T. Miyoshi<sup>3,\*</sup>

<sup>1</sup>Meteorological Research Institute, Tsukuba, Japan

<sup>2</sup>National Institute for Environmental Studies, Tsukuba, Japan

<sup>3</sup>Japan Meteorological Agency, Tokyo, Japan

\* now at: Department of Atmospheric and Oceanic Science, University of Maryland, College Park, Maryland, USA

Received: 17 December 2008 – Published in Atmos. Chem. Phys. Discuss.: 4 March 2009

Revised: 1 December 2009 – Accepted: 4 December 2009 – Published: 5 January 2010

**Abstract.** We have developed an advanced data assimilation system for a global aerosol model with a four-dimensional ensemble Kalman filter in which the Level 1B data from the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) were successfully assimilated for the first time, to the best of the authors' knowledge. A one-month data assimilation cycle experiment for dust, sulfate, and sea-salt aerosols was performed in May 2007. The results were validated via two independent observations: 1) the ground-based lidar network in East Asia, managed by the National Institute for Environmental Studies of Japan, and 2) weather reports of aeolian dust events in Japan. Detailed four-dimensional structures of aerosol outflows from source regions over oceans and continents for various particle types and sizes were well reproduced. The intensity of dust emission at each grid point was also corrected by this data assimilation system. These results are valuable for the comprehensive analysis of aerosol behavior as well as aerosol forecasting.

### 1 Introduction

Natural and anthropogenic aerosols have a considerable impact on air quality and climate. Large uncertainty, however, still exists in our estimates of aerosol emission and distribution. To improve our understanding of aerosol behavior, further observations and better numerical simulations are essential. Moreover, data assimilation must make optimal use of both observations and numerical simulations to obtain the best possible estimate of aerosol behavior. Data assimilation has played an essential role in numerical weather predictions (NWP), generating accurate initial conditions for better

forecasts (cf. Kalnay, 2003). Additionally, some data assimilation analyses have been applied to atmospheric chemical species (e.g., stratospheric ozone by Geer et al., 2006; tropospheric carbon monoxide by Arellano et al., 2007, and references therein). Compared with the long history and solid examination of data assimilation in operational NWP, data assimilation methods have only recently been applied to aerosol studies. The European Centre for Medium-Range Weather Forecasts (ECMWF) has launched the Global and regional Earth-system Monitoring using Satellite and in-situ data (GEMS) project, in which a four-dimensional variational (4D-Var) data assimilation system is adapted to monitor and forecast, among others, sea-salt, dust, organic, and black-carbon aerosols (Hollingsworth et al., 2008). The observations used in the GEMS project are retrieved aerosol optical depth data from the Moderate Resolution Imaging Spectro-radiometer (MODIS) onboard the Terra and Aqua satellites. Zhang et al. (2008) reported their 3D-Var data assimilation system for aerosol, using MODIS optical depth observations, to improve Naval Research Laboratory (NRL) aerosol analysis and prediction. A major limitation of the previous studies is that, because the MODIS aerosol optical depth is a column-integrated amount, the observational information does not include aerosol vertical profiles. In addition, MODIS observations cannot discriminate the type, size, and shape of the aerosols.

Dust aerosol in East Asia, where severe dust storms are frequent, was explored early using data assimilation methods. Yumimoto et al. (2007, 2008) developed a 4D-Var data assimilation system for Asian dust (yellow sand, or *Kosa*) with a regional chemistry-transport model and extinction coefficient data retrieved from ground-based lidar observations of several observatories in Japan. Niu et al. (2008) reported their dust storm forecast over China using a 3D-Var data assimilation system. They compiled two sets of observations for the data assimilation; one is a satellite-retrieved index of column amounts of dust aerosol; the other is the



Correspondence to: T. T. Sekiyama  
(tsekiyam@mri-jma.go.jp)

surface visibility observed by the meteorological stations of the Chinese Meteorological Administration (CMA). Lin et al. (2008a, b) developed an Ensemble Kalman Filter (EnKF) data assimilation system for Asian dust storms and have demonstrated the usefulness of EnKF in aerosol studies. These pioneering works (Yumimoto et al., 2007, 2008; Niu et al., 2008; Lin et al., 2008a, b), however, cover only the East Asian region and are not applicable to aerosols other than dust. Furthermore, the available observational data are very sparse and probably contain retrieval errors.

Recently, the first satellite lidar observations of aerosols have been made available by the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) mission (e.g., Winker et al., 2007). The Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) carried by CALIPSO provides continuous global measurements using a two-wavelength and polarization-sensitive backscattering lidar with very high vertical and horizontal resolution. The CALIPSO polar orbit passes through a longitudinal interval of approximately 1000 km per day at mid-latitudes. These measurements are sufficient to discriminate the type, size, and shape of the aerosols. The CALIPSO information released to the public as Level 1B data (<http://www-calipso.larc.nasa.gov/>) contains the total attenuated backscattering coefficients at 532 and 1064 nm and the volume depolarization ratio at 532 nm. These values are not contaminated by retrieval errors because they were directly measured and have not been processed by low-accuracy retrieval algorithms. On the basis of the availability of CALIPSO data, we developed an advanced data assimilation system for aerosol, which consists of a global chemistry-transport model (cf. Tanaka et al., 2003), a four-dimensionally expanded EnKF (cf. Miyoshi and Yamane, 2007, and references therein), and an observational operator for the CALIPSO Level 1B data.

In contrast with other data assimilation schemes, the EnKF explicitly and continuously provides a flow- and time-dependent approximation of the background error statistics to integrate the observations into model simulations. Other schemes, such as optimal interpolation and 3D-Var, practically assume that the background error statistics are spatially homogeneous, horizontally isotropic, and temporally stationary (cf. Kalnay, 2003). This assumption conflicts with the actual errors that are significantly flow- and time-dependent. Although the 4D-Var implicitly evolves the background error statistics, the evolution is effective only during each assimilation time window. Additionally, the 4D-Var involves the complexity of constructing linearized operators or adjoint matrices, which are unnecessary for the EnKF. The EnKF requires only forward model integrations. These are clear advantages of EnKF when it is applied to atmospheric chemistry analyses, such as aerosol studies. Furthermore, the four-dimensionally expanded EnKF used in this study works as a pseudo Ensemble Kalman Smoother (EnKS), which will provide more reliable reanalyses and emission estimates of aerosol via application of both past and future observations.

In this paper, we present preliminary results for the global EnKF data assimilation experiments for dust, sulfate, and sea-salt aerosols. To the authors' knowledge, this is the first attempt to perform state-of-the-science data assimilation with CALIPSO aerosol observations. The assimilation results were validated by two independent observations: data obtained from the ground-based lidar network (Shimizu et al., 2004) in East Asia, which is managed by the National Institute for Environmental Studies of Japan (NIES), and weather reports regarding aeolian dust events. The next section details our data assimilation system and experiments. The results of the data assimilation are discussed in Sect. 3, taking May 2007 as an example. The summary and future perspectives are presented in Sect. 4.

## 2 Description of the data assimilation system

### 2.1 Observational data

CALIPSO was launched on 28 April 2006 as part of the NASA A-train (cf. Winker et al., 2007). All satellites of the A-train are in a 705-km sun-synchronous polar orbit between 82° N and 82° S with a 16-day repeat cycle, which is an approximately 1000-km longitudinal interval per day at mid-latitudes. The primary instrument CALIOP carried by CALIPSO was the first satellite lidar to be optimized for aerosol and cloud measurements. During both day and night, the CALIOP continuously provides vertical profiles of the total attenuated backscattering coefficient  $\beta'_{\lambda}(\zeta)$  at 532 and 1064 nm and the volume depolarization ratio ( $\delta = \beta'_{\lambda}(\zeta)_{\text{perpendicular}}/\beta'_{\lambda}(\zeta)_{\text{parallel}}$ ) at 532 nm, with a horizontal resolution between 333 m and 1 km and a vertical resolution of 30–60 m in the troposphere. The attenuated backscattering coefficient  $\beta'_{\lambda}(\zeta)$  at wavelength  $\lambda$  is expressed as follows:

$$\beta'_{\lambda}(\zeta) = \zeta^2 P_{\lambda}(\zeta) / C_{\lambda} = (\beta_{\lambda,m}(\zeta) + \beta_{\lambda,p}(\zeta)) T_{\lambda,m}^2(\zeta) T_{\lambda,p}^2(\zeta), \quad (1)$$

where  $P_{\lambda}(\zeta)$  is the raw signal intensity from altitude  $\zeta$ , and  $C_{\lambda}$  is the instrument constant. Backscattering coefficients are represented by  $\beta(\zeta)$ , and the two-way transmittance due to scattering (or absorbing) species is given by  $T^2(\zeta)$ ; the subscripts m and p specify the molecular and particulate (either aerosol or cloud), respectively, contributions to the signal. In the present study, we assimilated the value  $\beta'_{\lambda}(\zeta)$  and its depolarization ratio  $\delta(\zeta)$  to a global chemistry-transport model. This direct assimilation prevents unnecessary and unnatural errors due to data retrieval processes. In contrast, extinction coefficients and optical thickness, which have been generally used for aerosol assimilation studies, require erroneous assumptions or prior estimations to be retrieved from “raw” measurement data, and thus these secondary quantities are contaminated by retrieval errors.

The values of  $\beta'_{\lambda}(\zeta)$  and  $\delta(\zeta)$  are contained in CALIPSO Level 1B data with their time-references and geo-locations. In this study, we used version 2.01 of CALIPSO Level 1B data and assumed the measurement uncertainties to be 20%,

in reference to Winker et al. (2007). The data products generated from the CALIPSO/CALIOP measurements are not only Level 1B data but also Level 2 data, which include geophysical variables derived from Level 1B data by the CALIPSO science team. Version 2.01 of the Level 2 data (<http://www-calipso.larc.nasa.gov/>) includes Cloud-Aerosol Discrimination (CAD) scores (Liu et al., 2004) with 5-km horizontal resolution. The CAD score is an indicator that enables discrimination of target contents between aerosols and clouds as an integer value ranging from  $-100$  (most likely to be aerosols) to  $+100$  (most likely to be clouds). To identify the aerosol signals and screen out cloud signals, we used CAD scores with  $\beta'_{\lambda}(\zeta)$  and  $\delta(\zeta)$  measurements. These measurements were selected only when the CAD score was less than or equal to  $-33$ , and then the selected measurements were horizontally and vertically averaged along each satellite orbit to approximately model resolution prior to data assimilation. This data selection markedly decreased the number of measurements used for data assimilation. After selection and averaging, the total number of  $\beta'_{\lambda=532}$ ,  $\beta'_{\lambda=1064}$ , and  $\delta_{\lambda=532}$  measurements to be assimilated was 15 000–25 000 points per day in the global troposphere. The geographical coordinate (i.e., longitude, latitude and altitude) of each data point is the barycenter of averaged measurements.

## 2.2 EnKF method

The basic idea of the EnKF is that the Monte Carlo ensemble of state vectors represents the probability distribution function of the system's state. The EnKF is mathematically equivalent to the original Kalman filter (KF) when the simulation model is linear and the EnKF employs an infinite ensemble size, because the background error covariance matrix can be absolutely calculated with infinite samples under those perfect conditions. Meanwhile, the KF is mathematically equivalent to the 4D-Var when the model is perfect and the error covariance matrices are the same given (e.g., Bouttier and Courtier, 1999). Altogether, the EnKF is mathematically equivalent to the 4D-Var under the ideal conditions. There is no fundamental discrepancy of the two data assimilation methods, and their differences arise from non-Gaussianity and nonlinearity of the real world (= non-ideal conditions). Furthermore, as in the 4D-Var or 3D-Var, the EnKF can treat nonlinear observational operators, which enables direct assimilation of measured physical quantities (e.g., satellite-measured radiances) that are generally nonlinear with respect to the model variables (e.g., aerosol concentrations or extinction coefficients).

Research examining the EnKF technique was initiated by Evensen (1994), and the first application to an atmospheric system was conducted by Houtekamer and Mitchell (1998). Their method is classified as a perturbed-observation EnKF, and the perturbed observations are a source of sampling errors. Whitaker and Hamill (2002) then proposed a Square Root Filtering (SRF) method of EnKF to avoid perturbing

the observations. Tippett et al. (2003) summarized several methods of ensemble SRF (EnSRF), all of which are efficient only when the observations are serially assimilated. Alternatively, Ott et al. (2002, 2004) proposed a Local Ensemble Kalman Filter (LEKF), a kind of EnSRF, which simultaneously assimilates the observations within a spatially local volume. Since analysis at each grid point is conducted independently of that of the other grid points, the LEKF promotes computational efficiency with parallel implementation. Furthermore, Hunt et al. (2007) applied the Ensemble Transform Kalman Filter (ETKF, developed by Bishop et al., 2001) approach to LEKF; this method is known as a Local Ensemble Transform Kalman Filter (LETKF; cf. Harlim, 2006). The computational cost of LETKF is much lower than that of the original LEKF because the former does not require an orthogonal basis. In addition, Hunt et al. (2004) expanded the EnKF four-dimensionally to assimilate observations asynchronously. This expansion allows the EnKF to assimilate observations at the appropriate time and, when available, to use future observations as with an Ensemble Kalman Smoother (EnKS) or 4D-Var.

As these EnKF techniques improved and became a viable choice in the field of operational NWP, the Japan Meteorological Agency (JMA) developed a four-dimensionally expanded LETKF (4D-LETKF) and applied it experimentally to NWP models (Miyoshi and Aranami, 2006; Miyoshi and Sato, 2007; Miyoshi and Yamane, 2007; Miyoshi et al., 2007a, b). As well as LEKF and LETKF, the 4D-LETKF of JMA has an advantage over most other EnKF implementations in its computational efficiency with simultaneous assimilation of increasing observations. Furthermore, one of major differences between LETKF (or the 4D-LETKF) and LEKF is that LETKF does not require local patches, and thus it allows a flexible choice of observations to be assimilated at each grid point (Hunt et al., 2007). The original implementation of LEKF separates the global model grid into local patches uniformly in the model grid space. The physical length between two successive grid points in the longitudinal direction is proportional to cosine of latitude. Therefore, the physical size of the local patch in the Polar Regions is much smaller than that in lower latitudes. This causes discontinuity in the analysis, which is not desirable. Moreover, Miyoshi and Yamane (2007) indicate that computational time is increased quadratic with the local patch size. LETKF removes local patches and implements natural localization weighting determined only by the physical distance (Miyoshi et al., 2007b). Miyoshi et al. (2007b) indicate that the computation without local patches is accelerated and more robust with choices of localization scales. These advantages of LETKF are suitable for not only higher resolution NWP models but also chemistry-transport models which includes much more prognostic variables than NWP models. Additionally, four-dimensional expansion of LETKF enables continuous observations, such as those of polar-orbit satellites (e.g., CALIPSO), to be assimilated effectively.

In principle, EnKF core modules are applicable to any numerical models, including not only NWP models but also aerosol models. In this study, we applied the 4D-LETKF of JMA to an aerosol chemistry-transport model, the Model of Aerosol Species in the Global Atmosphere (MASINGAR), which was developed by the Meteorological Research Institute (MRI) of Japan (cf. Tanaka et al., 2003). In order to assimilate attenuated backscattering and its depolarization measured by CALIPSO/CALIOP, we used an observational operator that emulates atmospheric optics induced by molecules (Rayleigh scattering), particles such as sulfate and sea-salt aerosols (Mie scattering), and dust particles. Specifically,  $\beta_{\lambda,m}(\zeta)$ ,  $\beta_{\lambda,p}(\zeta)$ ,  $T_{\lambda,m}(\zeta)$ , and  $T_{\lambda,p}(\zeta)$  in Eq. (1) are calculated from model variables (i.e., pressure, temperature, and aerosol concentrations) using the formulas of Rayleigh scattering and Mie scattering at each wavelength  $\lambda$ .  $\beta_{\lambda,m}(\zeta)$  is the backscattering coefficient of atmospheric molecules of which concentrations can be estimated from model pressure and temperature using the equation of the gas law. The transmittance  $T_{\lambda,m}(\zeta)$  can be estimated by accumulation of the extinction coefficients of atmospheric molecules between the lidar instrument and altitude  $\zeta$ . The backscattering coefficient of aerosol particles  $\beta_{\lambda,p}(\zeta)$  is the sum of the backscattering coefficients of sulfate, sea-salt, and dust aerosols of which concentrations are model prognostic variables. These backscattering coefficients are calculated dependently on the aerosol type and size using the equations of the Mie scattering theory. The transmittance of aerosol particles  $T_{\lambda,p}(\zeta)$  can be estimated by accumulation of all the extinction coefficients of aerosol particles as in the estimation of  $T_{\lambda,m}(\zeta)$ . It is noted that the dust extinction coefficient is empirically approximated from the Mie scattering theory, and the dust backscattering coefficient is estimated with the extinction coefficient divided by an empirical value of 50 sr. It is assumed that depolarization of the 532-nm backscattering is induced only by dust aerosol in this study, and that the depolarization ratio  $\delta$  is equal to 0.35 (Shimizu et al., 2004). Equation (1) with these  $\beta_{\lambda,m}(\zeta)$ ,  $\beta_{\lambda,p}(\zeta)$ ,  $T_{\lambda,m}(\zeta)$ , and  $T_{\lambda,p}(\zeta)$  derives the attenuated backscattering  $\beta'_{\lambda}(\zeta)$  which should be observed by CALIPSO/CALIOP under the clear sky condition if the model simulation were real. This observational operator needs only forward calculation. When the observational operator is nonlinear as this study, EnKF has a distinct advantage over 3D-Var or 4D-Var in that EnKF does not require a linearized observational operator or its adjoint.

### 2.3 Experimental design

The global chemistry-transport model MASINGAR has been successfully used for aerosol studies (Tanaka et al., 2005; Tanaka and Chiba, 2005, 2006; Uno et al., 2006; Tanaka et al., 2007). MASINGAR includes the emission, advection, diffusion, gravitational settling, wet/dry deposition, and chemical processes of SO<sub>2</sub>, dimethyl sulfide (DMS), sulfate aerosol, sea-salt aerosol (partitioned into 10 size bins), and

dust aerosol (partitioned into 10 size bins) in this study. The meteorological field in MASINGAR is nudged to a 6-h interval reanalysis of JMA using a Newtonian relaxation scheme in which dynamic tendencies are added at each time step to reproduce realistic meteorological conditions of the global atmosphere. MASINGAR has 30 vertical layers in a hybrid sigma-pressure coordinate from the earth's surface to the stratopause (approximately 7 layers below 800 hPa and 15 layers above 150 hPa). We incorporated MASINGAR with the 4D-LETKF data assimilation system having approximately 2.8° × 2.8° horizontal resolution (T42 spectrum truncation). Details of this model are described by Tanaka et al. (2003).

We applied this 4D-LETKF data assimilation system to the global aerosol analysis of May 2007. Data assimilation was initiated at 00:00 UTC on 1 May and terminated at 00:00 UTC on 1 June. A total of 918 CALIPSO orbit paths from pole to pole were obtained and used for this one-month analysis period; each path contains approximately 50 minutes long data. The initial conditions at 00:00 UTC on 1 May were prepared by a 2-year simulation that was carried out by MASINGAR without any aerosol assimilation as a spin-up. The initial ensemble spreads were generated by adding random Gaussian noise to this initial field. The time window of the 4D-LETKF is 48 h long, and consists of 49 time points with a 1-h interval. The analysis target time is chosen to be at the center of the 48-h window. Namely, an analysis is performed within the past 24-h and future 24-h measurements. The localization scale was set to horizontally 1000 km and vertically 15 grids with Gaussian localizations. With the 48-h time window and the 1000 km horizontal localization, aerosol clouds are probably detected and assimilated without omission globally, as long as the sky is clear. The vertical 15-grid length was based on the number of layers in the modeled troposphere. The ensemble size was set to 20 members following Miyoshi and Yamane (2007). The multiplicative spread inflation parameter was fixed at 10%. The assimilated model variables/parameters in this system represent the concentrations of aerosols (sulfate aerosol, 10-partitioned sea-salt aerosol, and 10-partitioned dust aerosol) and dust emission factors. These control variables and parameters are updated respectively by the 4D-LETKF at each step of the analysis. The dust emission flux  $\Phi_{i,j,n}$  of the  $n$ -th size bin at location  $(i, j)$  is corrected by

$$\Phi_{i,j,n} = \alpha_{i,j,n} \times F_{i,j,n}, \quad (2)$$

where  $F_{i,j,n}$  is the original dust emission flux of MASINGAR at each time step without assimilation, and  $\alpha_{i,j,n}$  is the dust emission factor given by the 4D-LETKF. The dust emission factor  $\alpha_{i,j,n}$  is estimated at each land location and for each size bin.

A model run was performed as a “reference” and included the same conditions and period as for the 4D-LETKF experiment, except that the data assimilation process was excluded. The data assimilation results were then compared

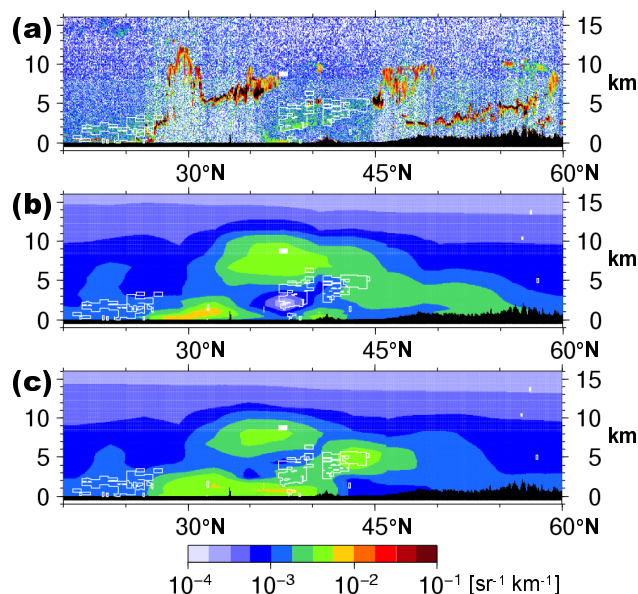
and verified with independent observations obtained from the ground-based NIES lidar network in East Asia (cf. Shimizu et al. 2004; <http://www-lidar.nies.go.jp/>). NIES has been continuously operating or co-operating this lidar observation network at more than 15 stations in Japan, South Korea, China, and Mongolia. It provides high-resolution vertical and temporal information for both spherical (sea-salt, sulfate, or pollutant) and non-spherical (dust) aerosols. One of the lidar stations, in western Japan near the Korean Peninsula, is used for this comparison because dust storms transported from the Gobi or Taklimakan Desert were clearly detected at this station in May 2007.

### 3 Assimilation results and discussion

#### 3.1 Comparison with CALIPSO data

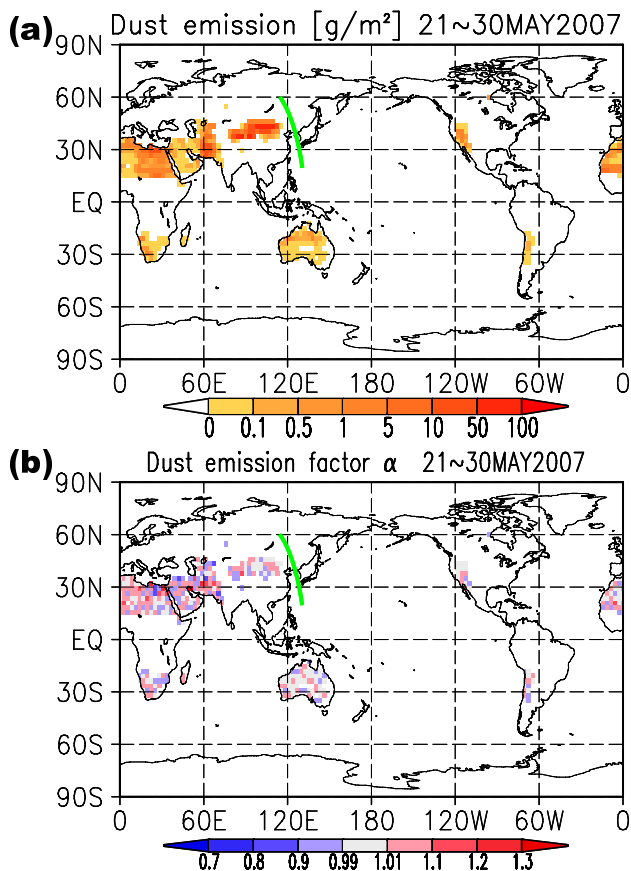
A sample comparison of model results obtained with CALIPSO/CALIOP attenuated backscattering measurements is shown in Fig. 1 as an overview of the correction performance of the 4D-LETKF data assimilation system. It is noted that the intensity of attenuated backscattering at each grid point depends not only on the aerosol concentration at the grid point but also on the aerosol and molecule concentrations in the light path between the grid point and the lidar instrument. Therefore, the intensity distribution of attenuated backscattering does not directly reflect the distribution of aerosols. Additionally, aerosols cannot be observed in or across clouds by the lidar. Figure 1a, for example, indicates that much of the area scanned by CALIPSO/CALIOP is masked or whitewashed by thick clouds (red or dark red area). This figure shows the total attenuated backscattering coefficients at 532 nm, cross-sectioned by a portion of CALIPSO's orbit path on 27 May 2007 (around 05:200 UTC or 14:00 LT in the daytime), which extends from 20° N to 60° N over East Asia (Japan, Korea, and Manchuria) indicated with a green line in Fig. 2. The white contours in Fig. 1 indicate regions with Cloud-Aerosol Discrimination (CAD) scores of less than or equal to  $-33$ , signifying a high probability of aerosol existence. This plot provides a good example for comparison of measurements with the model results because heavy dust storms continuously occurred in the Gobi Desert and Mongolia during late May 2007, and vast areas of East Asia were swept by the dust aerosol. Although this plot is slightly noisy because of the daylight and clouds, it is clear that some particles are detected in the low-CAD score regions near 40° N at an altitude of 2–5 km. According to Hara et al. (2008), these signals represent dust aerosol that was emitted in the Gobi Desert several days earlier and then transported eastward by a low-pressure system.

In comparison with this measurement, the dust signals near 40° N at an altitude of 2–5 km are not reproduced in the “reference” model run without data assimilation (Fig. 1b). In contrast, the 4D-LETKF assimilation reproduces dust signals



**Fig. 1.** Comparison of the total attenuated backscattering coefficients [ $\text{sr}^{-1} \text{km}^{-1}$ ] at 532 nm cross-sectioned by one of the CALIPSO orbit paths over East Asia (Japan, Korea, and Manchuria) on 27 May 2007. (a) CALIPSO/CALIOP measurements, (b) “reference” model run without assimilation, and (c) 4D-LETKF assimilation results. Red or yellow shades indicate relatively high values, and blue or blue-gray shades indicate relatively low values of the total attenuated backscattering coefficients. The X-axis shows the latitude ( $^{\circ}\text{N}$ ), and the Y-axis shows the altitude (km). White contours indicate regions in which the Cloud-Aerosol Discrimination (CAD) scores were less than or equal to  $-33$  (likely to be aerosol).

as part of the uplifting structure from the planetary boundary layer (PBL) near 30–40° N to the free troposphere at higher latitudes (Fig. 1c). The total attenuated backscattering coefficients shown in Fig. 1b and c are calculated only from model variables, such as aerosol concentrations, humidity, temperature, and pressure, using the same scheme as that employed in the observational operator mentioned above, excluding clouds. By comparison of Fig. 1b and c, it is clear that the correction process utilized by 4D-LETKF works well in low-CAD score regions near 40° N at an altitude of 2–5 km. In other areas, however, it is difficult to evaluate the data assimilation performance because thick clouds prevent the detection of aerosols, as shown in CALIPSO/CALIOP measurements (Fig. 1a). For example, the aerosol layer in the PBL near 25–40° N reproduced by the 4D-LETKF, shown in Fig. 1c, cannot be ascertained because of thick clouds above the aerosol layer in Fig. 1a. In this case, part of the aerosol layer in the PBL might be obscured by another dust layer near 35–40° N at an altitude of 2–5 km. The thin aerosol layer observed near 25° N at an altitude of 1–3 km is not reproduced both in Fig. 1b and c. This indicates the limitations of model simulation and data assimilation.



**Fig. 2.** (a) Simulated dust emission intensity in MASINGAR without assimilation from 21 to 30 May 2007; all 10 size bins are accumulated. Red indicates relatively high values, and yellow indicates relatively low values of the intensity. (b) Dust emission factor  $\alpha$  for each grid point estimated by 4D-LETKF assimilation; averaged from 21 to 30 May 2007 across all 10 size bins. Red indicates relatively high values, and blue indicates relatively low values of the dust emission factor. A green line over East Asia indicates the CALIPSO orbit path shown in Fig. 1.

In this study, the attenuated backscattering coefficients of CALIPSO/CALIOP were used for the 4D-LETKF data assimilation only when the CAD score indicated the presence of aerosol (e.g., white-contoured regions in Fig. 1a). On the other hand, continuous operation of the 4D-LETKF allows accumulation of information obtained during the past analysis in every assimilation step, and the four-dimensional assimilation simultaneously utilizes the 24-h past and 24-h future measurements together. Consequently, aerosol particles that are not detected by one measurement are detected by another measurement and thus are certainly used in the assimilation. For example, the aerosol layer in/above the PBL near 25–40° N in Fig. 1c is likely corrected by the past and/or future measurements. It is no surprise that Fig. 1b of the “reference” model result is very different from Fig. 1c of the data assimilation result, even outside the white-contoured (prob-

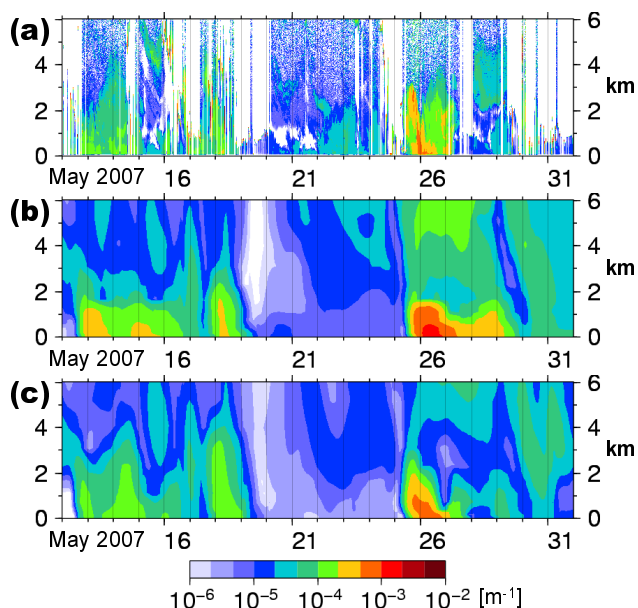
ably aerosol) regions. Furthermore, this correction can be traced back to the most upstream factors, i.e., the intensity of dust emission, even though there is no observation just above the dust-emitting areas. Figure 2b shows the dust emission factor  $\alpha$  in Eq. (2) for each grid point estimated by the 4D-LETKF assimilation, with the original dust emission intensity of the “reference” model run shown in Fig. 2a. The dust emission quantity and dust emission factors were accumulated and averaged, respectively, from 21 to 30 May 2007 across all 10 size bins. Parameter  $\alpha$  can be statistically used for correction of model biases (e.g., Lin et al., 2008b).

### 3.2 Comparison with independent observations

In the previous section, we compared the assimilation results with the CALIPSO/CALIOP measurements. These measurements were used directly in the assimilation process, and thus it is naturally expected that the assimilation results will agree well with measurements performed at the moment of analysis. Furthermore, it is uncertain whether the aerosol type and size are properly discriminated by the 4D-LETKF from the viewpoint of an attenuated backscattering coefficient only. Thus, other observational indices of aerosol, which are independent of the assimilation process, are presented here to compare with the model results.

First, the time-altitude cross-section of the extinction coefficients at 532 nm, which were observed by a ground-based lidar from 12 to 31 May 2007, is shown in Figs. 3a (for dust aerosol) and 4a (for spherical particles). This lidar observatory is located at Matsue (133° E, 35° N) in western Japan near the Korean Peninsula and is continuously operated by the NIES (cf. <http://www-lidar.nies.go.jp/>). The extinction coefficients were retrieved from the “raw” measurement data (cf. Shimizu et al., 2004), cloud signals were removed, and the contribution of dust aerosol to the total extinction was determined by the depolarization ratio, assuming all non-spherical particles to be dust aerosol. According to the weather report of the Japan Meteorological Agency, aeolian dust events (yellow sand, or *Kosa*) were observed at/near Matsue on 14, 26, and 27 May 2007. It is clear that these dust events were detected by the NIES lidar, as shown in Fig. 3a. Based on this plot, it can be seen that the dust extinction coefficients derived from the “reference” model result without data assimilation (Fig. 3b) partly disagree with that of the real case. For example, the dust extinction coefficients from 13 to 18 May are much larger in the PBL and smaller in the free troposphere than the NIES lidar measurements; although the large dust event on 26 and 27 May is reproduced in the PBL, it does not end on 28 May but continues until 30 May, as shown in Fig. 3b.

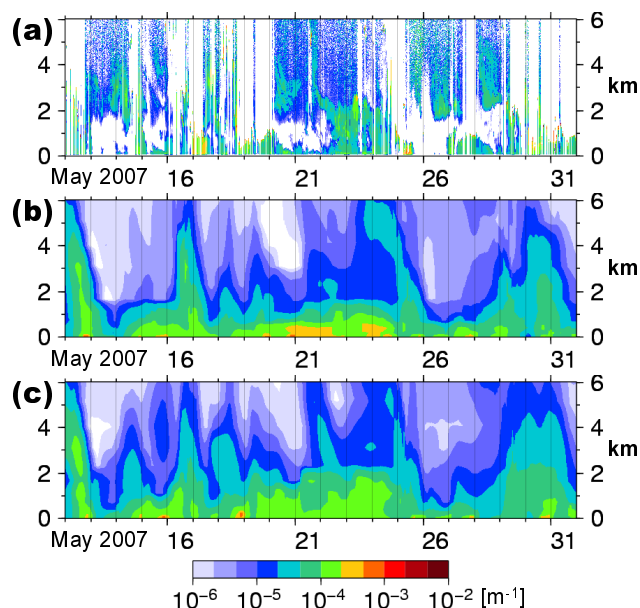
In contrast to the “reference” model result, the assimilation result (Fig. 3c) demonstrates much closer agreement with real events. For example, the dust extinction coefficients in the PBL from 13 to 18 May are closer to the NIES lidar measurements; the peak of the dust event in the PBL on



**Fig. 3.** Comparison of NIES-lidar observed and simulated extinction coefficients for non-spherical particles (dust aerosol) at 532 nm [ $\text{m}^{-1}$ ] at the Matsue observatory in western Japan, near the Korean Peninsula ( $133^\circ \text{E}$ ,  $35^\circ \text{N}$ ), from 12 to 31 May 2007. The X-axis shows the date, in which each tick grid line indicates 00:00 UTC. The Y-axis shows the altitude (km). (a) NIES-lidar measurements, (b) “reference” model run without assimilation, and (c) 4D-LETKF assimilation results. Red or yellow shades indicate relatively high values, and blue or blue-gray shades indicate relatively low values of the extinction coefficients.

14 May is reproduced; the dust clouds rising to 4 km above sea level (a.s.l.) on 14 and 18 May are clearly reproduced; the dust cloud on 26 May rises higher to 4 km a.s.l.; the dust event on 26 May ends on 27 May; the dust cloud in the free troposphere above 4 km from 26 to 28 May is much thinner than the “reference” model result. It is not obvious that the assimilation result agrees with the NIES lidar measurements because the NIES lidar network data were not used for the assimilation process and the CALIPSO orbit rarely passes through or near the NIES lidar observatory. This result shows, however, that the 4D-LETKF assimilation clearly improves the spatial distribution of dust aerosol, even in areas lacking observation.

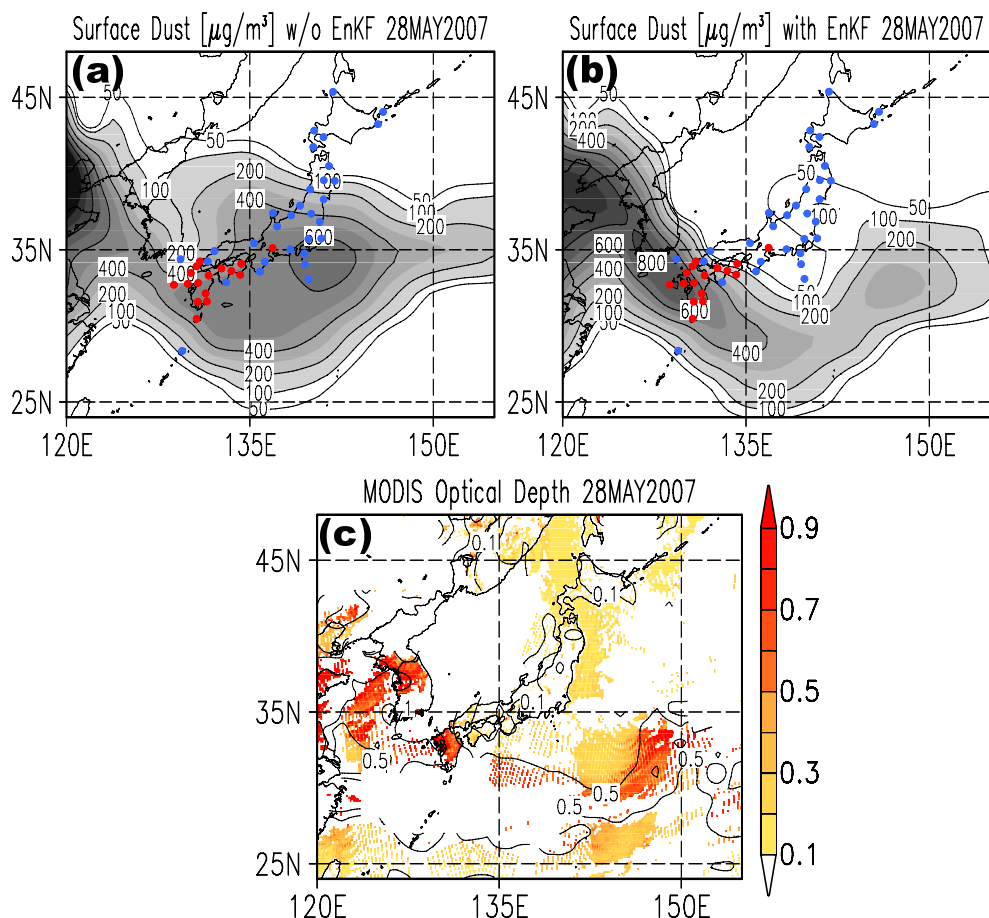
On the other hand, there is some disagreement between the lidar measurement (Fig. 3a) and the assimilation result (Fig. 3c): a thin dust cloud in/above the PBL from 21 to 24 May, evident in Fig. 3a, is not reproduced in Fig. 3c. By comparison with Fig. 4a, one can immediately observe a spherical-particle cloud at the same location and in the same shape as the dust cloud in Fig. 3a; the spherical-particle cloud is thicker than the dust cloud. This event in Fig. 4a is well reproduced by the 4D-LETKF assimilation shown in Fig. 4c, which shows the extinction coefficients for sulfate and sea-salt aerosols in the model, more precisely than the



**Fig. 4.** Comparison of NIES-lidar observed and simulated extinction coefficients for spherical particles (sulfate and sea-salt aerosols in MASINGAR) at 532 nm [ $\text{m}^{-1}$ ] at the Matsue observatory in western Japan, near the Korean Peninsula ( $133^\circ \text{E}$ ,  $35^\circ \text{N}$ ), from 12 to 31 May 2007. The X-axis shows the date, in which each tick grid line indicates 00:00 UTC. The Y-axis shows the altitude (km). (a) NIES-lidar measurements, (b) “reference” model run without assimilation, and (c) 4D-LETKF assimilation results. Red or yellow shades indicate relatively high values, and blue or blue-gray shades indicate relatively low values of the extinction coefficients.

“reference” model run shown in Fig. 4b. This discrepancy may be attributed to: 1) error in the depolarization ratio measured by the lidar, 2) the assumption that all non-spherical particles are dust aerosol, 3) the lack of elemental carbon aerosol in the model, 4) subgrid scale events, or 5) the failure of 4D-LETKF to discriminate the type of aerosol during the data assimilation process. Further investigation is needed to ascertain this discrepancy.

Second, a comparison of the modeled surface dust concentrations with weather reports of aeolian dust is shown in Fig. 5, demonstrating that the horizontal distribution of dust aerosol is improved by the 4D-LETKF assimilation. Daily mean concentrations of dust aerosol, including all 10 size bins, in the lowermost layer (from the surface to 100 m) on 28 May 2007 are plotted in Fig. 5a and b for the “reference” model result and the assimilation result, respectively. Roughly speaking, a concentration of  $100 \mu\text{g}/\text{m}^3$  is recognized as the environmental quality standard of aeolian dust (cf. the Japanese Ministry of the Environment, <http://www.env.go.jp/en/>). No aeolian dust events were observed in the eastern part of Japan on this day, but most of Japan was swept by historically heavy dust on the previous days, 26 and 27 May 2007, according to the Japan Meteorological Agency (cf. <http://www.jma.go.jp/jma/indexe.html>).



**Fig. 5.** Comparison of simulated surface dust concentrations on 28 May 2007 (contours and gray shades; daily mean) and the weather stations (red circles) that observed aeolian dust events, so-called *Kosa*, on the same day (plotted only in Japan). Blue circles indicate the weather stations that did not observe any aeolian dust events on the day. (a) Reference model run without assimilation and (b) 4D-LETKF assimilation results. (c) MODIS optical depth measured on 28 May 2007. Red shades indicate relatively high values, and yellow shades indicate relatively low values of the optical depth.

The red circles in Fig. 5a and b indicate the JMA weather stations that observed an aeolian dust event on 28 May 2007 with less than 10-km visibility. All of these stations are located in the western part of Japan. In contrast, the weather stations that reported that they did not observe any aeolian dust event on the day (plotted as blue circles in Fig. 5a and b) are mainly located in the eastern part of Japan. It is noted that this weather report is only within Japan; no dust events are plotted for Korea or China.

While the “reference” model result without data assimilation suggests that most of Japan, even its eastern region, is covered with a high-concentration dust plume (Fig. 5a), the dust plume of the assimilation result is mostly limited to the area of the weather stations that observed the aeolian dust event (Fig. 5b). Table 1 shows the number of weather stations in Fig. 5a and b where aeolian dust more than  $100 \mu\text{g}/\text{m}^3$  was analyzed or not, and observed or not. From this table, the threat score and the hit score (= percentage correct) were

calculated for both the “reference” model result and the assimilation result (Table 2). These scores range between 0 and 1, and the nearer to 1 the score is, the better the performance is. They are often used to qualify NWP’s performance. The assimilation makes the scores remarkably improved on 28 May 2007. In contrast, the dust plume covers all of Japan the day before, i.e., 27 May 2007, in both results (data not shown), equivalent to the weather report. It is evident that the assimilation result is superior to the “reference” model result. This feature is also shown by the MODIS optical depth (Fig. 5c). Although the optical depth depends on not only the surface dust concentration but also the column amount of all particles, the thick optical depth areas (more than 0.5) well fit into the thick dust distribution in Fig. 5b (the assimilation result). As with the NIES lidar extinction coefficients, these weather reports and MODIS measurements are not used for the assimilation process. This means that the 4D-LETKF assimilation makes it possible to monitor the



**Table 1.** The number of weather stations in Fig. 5a and b.

Dust event	Reference result (Fig. 5a)	Assimilation result (Fig. 5b)
Analyzed and Observed (AO)	16	14
Not Analyzed and Not observed (NANO)	9	28
Analyzed but Not observed (ANO)	22	3
Not analyzed but Observed (NAO)	0	2
Total	47	47

**Table 2.** Analysis scores of reference and assimilation results.

Score	Reference result (Fig. 5a)	Assimilation result (Fig. 5b)
Threat score = $AO/(AO+ANO+NAO)$	0.42	0.74
Hit score = $(AO+NANO)/Total$	0.53	0.89

horizontal distribution of aerosol continuously and to supply the initial conditions for aerosol forecasting with high accuracy, even when it is difficult to acquire observational data at remote areas or across clouds.

#### 4 Summary and future work

In the present study, we applied the 4D-LETKF data assimilation system to CALIPSO aerosol observations and successfully performed a one-month experiment in May 2007. The analysis estimates the emission intensity of dust aerosol and the spatial distribution of dust, sulfate, and sea-salt aerosols. The 4D-LETKF deals directly with the attenuated backscattering coefficients and depolarization ratios contained in the CALIPSO Level 1B dataset using an observational operator that emulates the atmospheric optics. Consequently, the type and size of aerosols have been discriminated and assimilated by the 4D-LETKF without retrieval errors. The assimilation results have been validated by two independent observations: the extinction coefficient profiles of dust and spherical aerosols, which are measured at a ground-based lidar observatory in East Asia, and the weather reports of aeolian dust events in Japan. The validation results indicated that this assimilation system can potentially provide reanalyses for the detailed three-dimensional and time-variable structure of aerosol outflows from source regions over oceans and continents for various types and sizes of aerosol particles and correct the intensity of dust emission at each grid point. These reanalyses and corrected parameters will surely be useful for comprehensive analysis of aerosol behavior. Furthermore,

this system makes it possible to supply the initial conditions for aerosol forecasting with high accuracy, even in remote areas and across clouds lacking observation.

In the next step of the 4D-LETKF experiments, we are going to install organic and elemental carbon aerosols into MASINGAR and simultaneously assimilate them with the other aerosols. In a future study, we will try to correct sulfate aerosol emission as well as dust emission. Furthermore, not only emission corrections, but also deposition corrections, will be performed in the data assimilation system. Reanalysis precision and aerosol predictability should be assessed globally, not only within East Asia. We have not yet sufficiently inspected the sensitivity of the model resolution and the influence of the ensemble size because of their high computational load. Miyoshi and Yamane (2007) suggested that LETKF is significantly stabilized with more than 20 ensemble members in their realistic weather forecasting system. However, it is not obvious that the properties are applicable to our aerosol LETKF assimilation system. Experiments assessing higher resolutions and larger ensemble sizes should also be carried out as soon as possible. In addition, we are going to evaluate the sensitivity of the emission factor analysis to the atmospheric concentration analysis.

*Acknowledgements.* Special thanks are extended to all individuals and organizations that provided the data used in this study. The CALIPSO data were obtained via the NASA Langley Research Center website (<http://www-calipso.larc.nasa.gov/>). The MODIS data were obtained via the NASA Goddard Space Flight Center website (<http://modis.gsfc.nasa.gov/>). The NIES lidar data were obtained from the NIES website (<http://www-lidar.nies.go.jp/>). The

weather reports of aeolian dust were provided by the Atmospheric Environment Division of JMA. This assimilation study was supported by a Grant-in-Aid for Young Scientists (20740275) and a Grant-in-Aid for Scientific Research (20244078) from the Japanese Ministry of Education, Culture, Sports, Science, and Technology, and partly supported by the Global Environment Research Fund (B-093) and (C-091) by the Japanese Ministry of the Environment.

Edited by: W. Lahoz

## References

- Arellano Jr., A. F., Raeder, K., Anderson, J. L., Hess, P. G., Emons, L. K., Edwards, D. P., Pfister, G. G., Campos, T. L., and Sachse, G. W.: Evaluating model performance of an ensemble-based chemical data assimilation system during INTEX-B field mission, *Atmos. Chem. Phys.*, 7, 5695–5710, 2007, <http://www.atmos-chem-phys.net/7/5695/2007/>.
- Bishop, C. H., Etherton, B. J., and Majumdar, S. J.: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects, *Mon. Weather Rev.*, 129, 420–436, 2001.
- Bouttier, F. and Courtier, P.: Data assimilation concepts and methods, *Meteorological Training Course Lecture Series*, ECMWF, 75 pp., 1999.
- Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, 99(C5), 10143–10162, 1994.
- Geer, A. J., Lahoz, W. A., Bekki, S., Bormann, N., Errera, Q., Eskes, H. J., Fonteyn, D., Jackson, D. R., Juckes, M. N., Massart, S., Peuch, V.-H., Rharmili, S., and Segers, A.: The ASSET inter-comparison of ozone analyses: method and first results, *Atmos. Chem. Phys.*, 6, 5445–5474, 2006, <http://www.atmos-chem-phys.net/6/5445/2006/>.
- Hara, Y., Yumimoto, K., Uno, I., Shimizu, A., Sugimoto, N., Liu, Z., and Winker, D. M.: Asian dust outflow in the PBL and free atmosphere retrieved by NASA CALIPSO and an assimilated dust transport model, *Atmos. Chem. Phys.*, 9, 1227–1239, 2009, <http://www.atmos-chem-phys.net/9/1227/2009/>.
- Harlim, J.: Errors in the initial conditions for numerical weather prediction: A study of error growth patterns and error reduction with ensemble filtering, Ph.D. dissertation, University of Maryland, 76 pp., 2006.
- Hollingsworth, A., Engelen, R. J., Textor, C., Benedetti, A., Boucher, O., Chevallier, F., Dethof, A., Elbern, H., Eskes, H., Flemming, J., Granier, C., Kaiser, J. W., Morcrette, J. J., Rayner, P., Peuch, V. H., Rouil, L., Schultz, M. G., Simmons, A. J., and The GEMS Consortium: Toward a Monitoring and Forecasting System For Atmospheric Composition: The GEMS Project, *B. Am. Meteorol. Soc.*, 89, 1147–1164, 2008.
- Houtekamer, P. L. and Mitchell, H. L.: Data assimilation using an ensemble Kalman filter technique, *Mon. Weather Rev.*, 126, 796–811, 1998.
- Hunt, B. R., Kalnay, E., Kostelich, E. J., Ott, E., Patil, D. J., and coauthors: Four-dimensional ensemble Kalman filtering, *Tellus A*, 56, 273–277, 2004.
- Hunt, B. R., Kostelich, E. J., and Szunyogh, I.: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter, *Physica D*, 230, 112–126, 2007.
- Kalnay, E.: *Atmospheric Modeling, Data Assimilation and Predictability*, Cambridge University Press, Cambridge, UK, 341 pp., 2003.
- Liu, Z., Vaughan, M. A., Winker, D. M., Hostetler, C. A., Poole, L. R., Hlavka, D., Hart, W., and McGill, M.: Use of probability distribution functions for discriminating between cloud and aerosol in lidar backscatter data, *J. Geophys. Res.*, 109, D15202, doi:10.1029/2004JD004732, 2004.
- Lin, C., Wang, Z., and Zhu, J.: An Ensemble Kalman Filter for severe dust storm data assimilation over China, *Atmos. Chem. Phys.*, 8, 2975–2983, 2008a, <http://www.atmos-chem-phys.net/8/2975/2008/>.
- Lin, C., Zhu, J., and Wang, Z.: Model bias correction for dust storm forecast using ensemble Kalman filter, *J. Geophys. Res.*, 113, D14306, doi:10.1029/2007JD009498, 2008b.
- Miyoshi, T. and Aranami, K.: Applying a Four-dimensional Local Ensemble Transform Kalman Filter (4D-LETKF) to the JMA Nonhydrostatic Model (NHM), *SOLA*, 2, 128–131, doi:10.2151/sola.2006-033, 2006.
- Miyoshi, T. and Sato, Y.: Assimilating Satellite radiances with a Local Ensemble Transform Kalman Filter (LETKF) applied to the JMA Global Model (GSM), *SOLA*, 3, 037–040, doi:10.2151/sola.2007-010, 2007.
- Miyoshi, T. and Yamane, S.: Local Ensemble Transform Kalman Filtering with an AGCM at a T159/L48 resolution, *Mon. Weather Rev.*, 135, 3841–3861, 2007.
- Miyoshi, T., Yamane, S., and Enomoto, T.: The AFES-LETKF experimental ensemble reanalysis: ALERA, *SOLA*, 3, 045–048, doi:10.2151/sola.2007-012, 2007a.
- Miyoshi, T., Yamane, S., and Enomoto, T.: Localization the error covariance by physical distances within a Local Ensemble Transform Kalman Filter (LETKF), *SOLA*, 3, 089–092, doi:10.2151/sola.2007-023, 2007b.
- Niu, T., Gong, S. L., Zhu, G. F., Liu, H. L., Hu, X. Q., Zhou, C. H., and Wang, Y. Q.: Data assimilation of dust aerosol observations for the CUACE/dust forecasting system, *Atmos. Chem. Phys.*, 8, 3473–3482, 2008, <http://www.atmos-chem-phys.net/8/3473/2008/>.
- Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E. J., et al.: Exploiting local low dimensionality of the atmospheric dynamics for efficient Kalman filtering, arXiv:physics/0203058, 24 pp., available at: <http://arxiv.org/abs/physics/0203058v3/>, 2002.
- Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E. J., et al.: A local ensemble Kalman filter for atmospheric data assimilation, *Tellus A*, 56, 415–428, 2004.
- Shimizu, A., Sugimoto, N., Matsui, I., Arao, K., Uno, I., Murayama, T., Kagawa, N., Aoki, K., Uchiyama, A., and Yamazaki, A.: Continuous observations of Asian dust and other aerosols by polarization lidars in China and Japan during ACE-Asia, *J. Geophys. Res.*, 109, D19S17, doi:10.1029/2002JD003253, 2004.
- Tanaka, T. Y. and Chiba, M.: Global simulation of dust aerosol with a chemical transport model, *MASINGAR*, *J. Meteorol. Soc. Jpn.*, 83A, 255–278, 2005.
- Tanaka, T. Y. and Chiba, M.: A numerical study of the contributions of dust source regions to the global dust budget, *Global Planet. Change*, 52, 88–104, doi:10.1016/j.gloplacha.2006.02.002, 2006.
- Tanaka, T. Y., Orito, K., Sekiyama, T. T., Shibata, K., Chiba, M.,

- and Tanaka, H.: MASINGAR, a global tropospheric aerosol chemical transport model coupled with MRI/JMA98 GCM, *Pap. Meteorol. Geophys.*, 53, 119–138, 2003.
- Tanaka, T. Y., Kurosaki, Y., Chiba, M., Matsumura, T., Nagai, T., Yamazaki, A., Uchiyama, A., Tsunematsu, N., and Kai, K.: Trans-continental dust transport from North Africa and the Middle East to East Asia, *Atmos. Environ.*, 39(21), 3901–3909, 2005.
- Tanaka, T. Y., Aoki, T., Takahashi, H., Shibata, K., Uchiyama, A., and Mikami, M.: Study of the sensitivity of optical properties of mineral dust to the direct aerosol radiative perturbation using a global aerosol transport model, *SOLA*, 3, 33–36, doi:10.2151/sola.2007-009, 2007.
- Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M., and Whitaker, J. S.: Ensemble square root filters, *Mon. Weather Rev.*, 131, 1485–1490, 2003.
- Uno, I., Wang, Z., Chiba, M., et al.: Dust model intercomparison (DMIP) study over Asia: Overview, *J. Geophys. Res.*, 111, D12213, doi:10.1029/2005JD006575, 2006.
- Whitaker, J. S. and Hamill, T. M.: Ensemble data assimilation without perturbed observations, *Mon. Weather Rev.*, 130, 1913–1924, 2002.
- Winker, D. M., Hunt, H. H., and McGill, M. J.: Initial performance assessment of CALIOP, *Geophys. Res. Lett.*, 34, L19803, doi:10.1029/2007GL030135, 2007.
- Yumimoto, K., Uno, I., Sugimoto, N., Shimizu, A., and Satake, S.: Adjoint inverse modeling of dust emission and transport over East Asia, *Geophys. Res. Lett.*, 34, L00806, doi:10.1029/2006GL028551, 2007.
- Yumimoto, K., Uno, I., Sugimoto, N., Shimizu, A., Liu, Z., and Winker, D. M.: Adjoint inversion modeling of Asian dust emission using lidar observations, *Atmos. Chem. Phys.*, 8, 2869–2884, 2008, <http://www.atmos-chem-phys.net/8/2869/2008/>.
- Zhang, J., Reid, J. S., Westphal, D. L., Baker, N. L., and Hyer, E. J.: A system for operational aerosol optical depth data assimilation over global oceans, *J. Geophys. Res.*, 113, D10208, doi:10.1029/2007JD009065, 2008.