ADVANCES IN ATMOSPHERIC SCIENCES, VOL. 26, NO. 6, 2009, 1053–1070

Dust Storm Ensemble Forecast Experiments in East Asia

ZHU Jiang*(朱 江), LIN Caiyan (林彩燕), and WANG Zifa (王自发)

State Key Laboratory of Atmospheric Boundary Layer Physics and Atmospheric Chemistry,

Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029

(Received 26 December 2008; revised 8 March 2009)

ABSTRACT

The ensemble Kalman filter (EnKF), as a unified approach to both data assimilation and ensemble forecasting problems, is used to investigate the performance of dust storm ensemble forecasting targeting a dust episode in the East Asia during 23–30 May 2007. The errors in the input wind field, dust emission intensity, and dry deposition velocity are among important model uncertainties and are considered in the model error perturbations. These model errors are not assumed to have zero-means. The model error means representing the model bias are estimated as part of the data assimilation process. Observations from a LIDAR network are assimilated to generate the initial ensembles and correct the model biases. The ensemble forecast skills are evaluated against the observations and a benchmark/control forecast, which is a simple model run without assimilation of any observations. Another ensemble forecast experiment is also performed without the model bias correction in order to examine the impact of the bias correction. Results show that the ensemble-mean, as deterministic forecasts have substantial improvement over the control forecasts and correctly captures the major dust arrival and cessation timing at each observation site. However, the forecast skill decreases as the forecast lead time increases. Bias correction further improved the forecasts in down wind areas. The forecasts within 24 hours are most improved and better than those without the bias correction. The examination of the ensemble forecast skills using the Brier scores and the relative operating characteristic curves and areas indicates that the ensemble forecasting system has useful forecast skills.

Key words: dust storm, ensemble forecast, data assimilation, bias correction

Citation: Zhu, J., C. Y. Lin, and Z. F. Wang, 2009: Dust storm ensemble forecast experiments in East Asia. *Adv. Atmos. Sci.*, **26**(6), 1053–1070, doi: 10.1007/s00376-009-8218-0.

1. Introduction

Since the late 1990s, dust emission and transport models have been developed and have played an important role in understanding the characteristics of dust phenomena (e.g., Marticorena and Bergametti, 1995; Marticorena et al., 1997; Wang et al., 2000; Nickovic et al., 2001; Uno et al., 2001, 2003; Park and In, 2003; Gong et al., 2003; Liu et al., 2003; Shao et al., 2003; Pérez et al., 2006). Recently, encouraged by reasonable simulation results of several models in the Asia-Pacific region (e.g., Gong et al., 2003; Liu et al., 2003; Chin et al., 2003; Uno et al., 2004), numerical dust storm forecasts have been put into operation. One example is the CUACE/Dust (Chinese Unified Atmospheric Chemistry Environment for Dust) developed by coupling dust aerosol microphysics onto a mesoscale meteorological model at the Chinese Meteorological Administration. CUACE/Dust has been issuing three-day SDS real time deterministic forecasts and early warning information in the springtime from March to May since 2005 (Zhou et al., 2008).

However, deterministic forecasts could suffer from large uncertainties in the dust models. In an intercomparison effort made by developers of eight models from different countries, quite different performances for Asian dust episodes in March and April of 2002 were displayed in terms of flux, transport, and horizontal and vertical distribution outputs (Uno et al., 2006). One of the primary differences is the dust emission intensity. The eight models produced largely different dust emission intensities. For example, the total dust emission ranged from 27 to 336 Tg during the period from 15 to 25 March 2002 (ten days). Though

*Corresponding author: ZHU Jiang, jzhu@mail.iap.ac.cn

different dust emission schemes with various complexities contributed greatly to the differences in the modeled emission intensities, the lack of reliable surface land-use conditions and soil/surface information also brought large biases in the modeled dust emission intensities. Another factor is the uncertainty in the meteorological fields, such as wind speed and precipitation. This kind of uncertainty also contains bias (nonzero errors averaged over a certain period of time and spatial domain) and random errors. The initial conditions for dust storm forecasts can be obtained by assimilating in situ observations, such as surface PM_{10} concentrations (Lin et al., 2008a,b), LIDAR (Light Detection and Ranging) observations (Yumimoto et al., 2007a,b), or satellite-based observations (Niu et al., 2008). However, the initial conditions can only be estimated within a certain accuracy. During a forecast period, the errors in the modeled dust emission intensity, meteorological fields, and the initial condition can accumulate and result in significant forecast errors.

Ensemble forecasting is a useful approach to deal with the limitations imposed by these uncertainties. Numerical weather prediction has over a decade of experience in ensemble forecasting, aiming at predicting quantitatively the probability density of the future atmospheric states. The Monte-Carlo technique that samples the probability density function (p.d.f.) at an initial time and evolves the sampled initial states with the forecast model is the only feasible approach nowadays. A comprehensive and in-depth review of ensemble weather forecasting can be found in Leutbecher and Palmer (2008).

To develop high quality ensemble forecast systems of dust storms, problems such as identifying major sources of uncertainties and examination of their impacts on ensemble forecasting should be investigated. Due to the sparseness of dust observations, Lin et al. (2006) first compared the performance of an ensemble Kalman filter (EnKF), a Monte-Carlo technique with the classic Kalman equation to update the p.d.f. of the initial condition, with optimal interpolation for different observation networks, and showed the good performance of EnKF for sparse observations. Then, Lin et al. (2008a,b) applied EnKF to the assimilation of surface PM_{10} concentration observations in China. They found that there exists large bias in modeled dust emission intensity, and applied the EnKF to correct the bias while assimilating the surface PM_{10} observations to improve the initial conditions. It showed that the model bias correction during the assimilation process played an important role in improving the deterministic forecast using the ensemble mean. Though not examining the forecast results in the context of ensemble forecasting, Lin et al. (2008a,b) in fact performed ensemble forecasts of dust storms, in which the random model errors and model biases were assumed to be in the forecasted wind fields and the emission intensity. In this study, we further perform ensemble forecast experiments for the East Asia region during 23–30 May 2007 when a strong dust episode occurred and was observed by a LIDAR network in China, Korea, and Japan. The three-dimensional structure of the transport for this case has been analyzed by Hara et al. (2008) and showed that the dust moved to Japan late on the 25 to 26 May 2007.

The setup of the ensemble forecast is as follows. The initial p.d.f. is produced by the EnKF that assimilates the LIDAR dust extinction coefficient observations. The random model errors come from the modeled dust emission intensity, dry deposition velocity, and wind fields. These errors are assumed to be Gaussian but not to have zero means. Their means are treated as model biases and are corrected during the data assimilation process performed as parameter estimation. Unlike the previous studies of Lin et al. (2008a,b), we consider the error and bias in the dry deposition process and high level wind speeds (excluding surface winds), which have strong influences on the long-distance transport of dust. In Lin et al. (2008a,b), the studied area is limited to North China, which is closer to the dust source regions in West China and Mongolia. In this study, the area is extended to include Korea and Japan. The aim of this study is to examine the feasibility of this setup for dust storm ensemble forecasting over East Asia and to identify problems for further studies in the field of dust storm ensemble forecasting.

The paper is arranged as follows. Section 2 gives a brief description of the dust transport model and the data used in this study. Section 3 lists uncertainty sources, and their quantifications in this study. Section 4 describes the EnKF data assimilation scheme used to provide the ensemble initial conditions and correct model biases. Experiment setups are given in section 5 and their results are shown in section 6. Section 7 is the conclusions and discussions.

2. Model and data

2.1 *Model*

The regional dust transport model included deflation, transport, diffusion, and removal processes during the life cycle of the yellow sand particles. This model has been successfully used to study atmospheric trace gases and particles, such as SO_x, dust, O_3, and acid rain over East Asia (Wang et al., 2000, 2002; Uematsu et al., 2003). In this study, 9 dust particle size

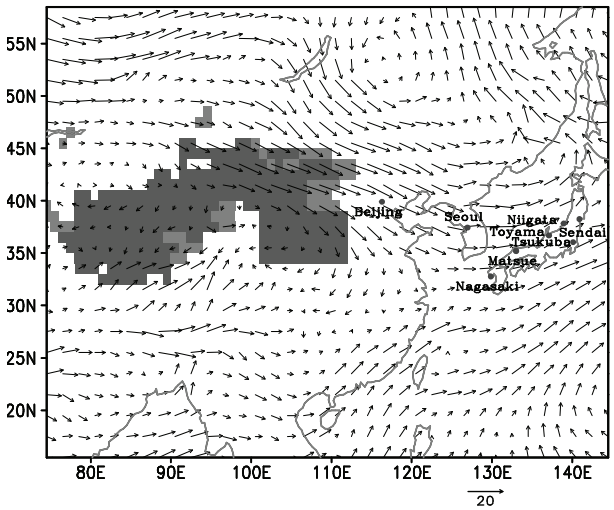

Fig. 1. Modeling domain, the shaded areas are the major dust source regions (i.e., desert and bare ground) defined in the model. The winds and the eight NIES LIDAR sites (black dots). The vectors are the three-day averaged winds at 1000-m height over the period of 24–26 May.

bins (0.5, 0.96, 1.83, 3.5, 6.7, 12.8, 24.5, 46.7, 90.0 μm in radius) are modeled.

The dust emission intensity $Q_{i.j,l}$ in the l-th size bin at location (i, j) depends on the conditions at the lowest model level and is given by

$$Q_{i,j,l} = C_1 C_2 u^{*2}_{i,j,l}(1 - u^*_{0,i,j,l}/u^*_{i,j,l})W_{i,j,l}R_{i,j,l} \quad (1)$$

where $Q_{i.j,l}$ is given as kg m^{-2} s^{-1}, C_1 is the weighting factor for different land types (C_1 for desert and bare ground are shown in Fig. 1), and C_2 is the empirical constant set as 2.9×10^{-2}. $u^*_{i,j,l}$ is the friction velocity, and $u^*_{0,i,j,l}$ is the threshold value of the friction velocity (0.4 m s^{-1}). $R_{i,j,l}$ is the fraction of the l-th bin of the deflating yellow sand, $W_{i,j,l}$ is the humidity factor, which is assumed to be linearly dependent on the relative humidity if the surface relative humidity is larger than its critical value (40%) or to be zero otherwise.

The mechanism of dry removal due to sedimentation and turbulent diffusion to the surface constructed by Zhang et al. (1998) is used to estimate the dry removal in which the dry deposition velocities are calculated by

$$V_{\rm d} = V_{\rm g} + u^{*2}/[k\hat{u}(Sc^{0.6} + 10^{-3/St})] \,, \quad (2)$$

for the surface level and

$$V_{\rm d} = V_{\rm g} \quad (3)$$

for other model levels. u^* is the friction velocity, $\hat{u}$ is the surface wind speed, k is the von Karman constant (equal to 0.4), Sc is the particle Schmidt number, and St is the particle Stokes number. St is the ratio of the particle stopping distance to a turbulent length scale. The gravitational settling velocity $V_{\rm g}$ is determined by the Stokes law for a particle with radius r:

$$V_{\rm g} = \frac{2(\rho_{\rm p} - \rho)r^2 g}{9\gamma} \quad (4)$$

where $\rho_{\rm p}$ is the particle density, ρ is the air density, g is the gravitational constant, and γ is the air viscosity. For desert dust particles, we have the mean value of the clay density $\rho_{\rm p} = 2.5$ g cm^{-3}.

The dust extinction coefficient at a wavelength of 0.55 μm is calculated at each time step in this model with the formula

$$E_{i,j,k} = \sum_{l=1}^{L=9} \frac{Q_{\rm ext}(l)^* C_{\rm dust}(i,j,k,l)}{\rho R_{\rm eff}(l)} \quad (5)$$

where $Q_{\rm ext}(l)$ is the extinction efficiency factor, which is determined by the Mie theory, for dust size bin l. $C_{\rm dust}(i, j, k, l)$ is the dust concentration for dust size bin l at location i, j and level k. $R_{\rm eff}(l)$ is the effective radius of the dust particle given above.

The simulation domain ranges from (16°N, 75°E) to (60°N, 146°E) consisting of 72 by 45 grid cells horizontally (see Fig. 1) and 18 vertical layers in terrain-following coordinates. Details about the model can be found in Wang et al. (2000).

2.2 *Data*

The dust observations used in this paper are the extinction coefficients obtained from the National Institute of Environemental Science (NIES) LIDAR network (Sugimoto et al., 2006), which measures the vertical profiles of dust with high spatial and temporal resolutions. Their vertical resolution is 30 m and the temporal resolution is 15 minutes. The extinction coefficient was derived based on the backward Fernald's method (Fernald, 1984) by setting a boundary condition at 6 km and a LIDAR ratio S1 = 50 sr (Liu et al., 2002). Contribution of soil dust in the total extinction coefficient is discriminated with the method using the particle depolarization ratio (Shimizu et al., 2004).

In this study we select eight LIDAR sites (Beijing, Seoul, Nagasaki, Matsue, Toyama, Niigata, Tsukuba, and Sendai; see Fig. 1) from the network for assimilation. Since the spatial resolution of the model is low, it cannot resolve information with very high spatial resolution. Due to very high spatial and temporal resolution of the observations, the computational cost of assimilating them would be expensive. These observations are thinned to reduce the computational cost. The observations are locally averaged to a lower temporal-spatial resolution with 150 m vertical resolution and a 3-hour temporal resolution. The average

is operated on each 3-hour/120 m temporal/spatial bin. Another motivation of thinning the observations is that a high data density may violate the assumption of independent measurement errors, which is made in most practical assimilation schemes. Liu and Rabier (2002) showed that there is a connection between the observation density and the resolution of the model grid. They showed theoretically that the analysis quality decreases, if the density of the observational data set is too large and the error correlations are neglected. Before thinning, some poor data would be discarded through quality control mainly considering the spatial and temporal continuity.

The meteorological fields (including the winds, temperature, humidity, pressure, etc.) used to drive the dust transport model in this study are the NCEP reanalysis data.

3. Model errors

Preliminary sensitivity analysis of model uncertainty (Lin et al., 2009), including their biases, of the dust transport model (Wang et al., 2000) over North China, suggests that dust emission, surface wind fields, and dry deposition velocity have strong impacts on surface dust prediction and have large biases. Among these, the impact of dry deposition velocity is relatively smaller than the other two kinds of error within short ranges of transport. In this study, we extend our study area to East Asia including Japan and Korea. Therefore, we add the dry deposition velocity as an error source. For errors in wind field, we do not consider the surface wind field. The surface wind has little impact on the long range dust transport but has a strong impact on the dust emission intensity. Since we explicitly consider the error in the net dust emission intensity, the impact of the surface wind field is included indirectly.

Ideally, the wind perturbations should come from a set of ensemble numerical weather forecasts. However in this study, we are not able to obtain large size ensemble wind forecast products but only a single wind analysis (from the NCEP 6-hourly reanalysis product). We have to make a set of ensembles from this single realization. The error in the wind field includes an overall error in the wind direction and spatially varying error of the wind speed. The overall error (the same at all levels and all grid-points) in the wind direction is defined as a Gaussian random number with a zero mean and the standard deviation of 20°. The error in the wind speed, also a Gaussian distribution, is defined with a non-zero mean (proportional to the input wind speed) and the standard deviation of 30% of the input wind speed at each model grid-point. The perturbation magnitudes are based on the reanalysis error statistics. In the NCEP reanalysis page, the stationary background error (6-hour forecast error) variances for the component wind at the middle latitudes range from $2\ \mathrm{m\ s^{-1}}$ to $4\ \mathrm{m\ s^{-1}}$ (http://wwwt.emc.ncep.noaa.gov/gmb/bkistler/oberr/fcsterr.html). The actual analysis error variances will change from time to time, but statistically are slightly less than the background error variances, depending on observation availability and observational errors. In this study most component wind ranges from $-9\ \mathrm{m\ s^{-1}}$ to $9\ \mathrm{m\ s^{-1}}$ with extreme values of about $\pm 20\ \mathrm{m\ s^{-1}}$, the 30% of wind speed perturbation magnitudes and 20° of wind direction perturbation magnitudes roughly agree with the analysis error statistics. The ratio of the non-zero mean and the input wind speed at each grid-point is assumed to be spatially homogenous and denoted by α_{w}. The dimensionless coefficient α_{w} determines the error bias in the wind speed field. This wind perturbation scheme is very simple and is not based on the original wind analysis error statistics. However, this scheme represents two most important uncertainty factors that impact the dust transport: speed and direction. This kind of simple wind perturbation scheme has been used in uncertainty analysis and EnKF of air quality models (e.g., Hanna et al., 2001; Wu et al., 2008).

The error in the net dust emission intensity is defined at every grid-point in the dust source regions in the model. The error is assumed to be Gaussian with a standard deviation of 50% of the calculated net dust emission intensity $Q_{i,j,l}$ by the original model. Similar to the wind speed field, the non-zero mean of the dust emission error at each grid-point in the source regions is assumed to be proportional to $Q_{i,j,l}$. The ratio of the non-zero mean and $Q_{i,j,l}$ at each grid-point is assumed to be spatially homogenous and denoted by α_Q. The dimensionless coefficient α_Q determines the bias in the net dust emission intensity.

The error in the dry deposition velocity is defined in a similar manner as that in the wind speed and dust emission. The standard deviation is 50% of the original model calculated dry deposition velocity V_{g}. The dimensionless overall coefficient that defines the bias in the dry deposition velocity at each grid-point is denoted by α_{d}.

Note that the three parameters α_{w}, α_Q, and α_{d} are not constant in time and will be estimated in the data assimilation process. But, for a given ensemble member, they have the same values over the model domain at a given time. Before convergence to a single value, the values of the parameter are different over ensemble members. More description is given in the next section.

4. EnKF data assimilation with parameter estimation

EnKF is used to provide the initial dust concentration field and to correct the model bias. The basic idea behind the EnKF (Evensen, 1994) is to construct a Monte Carlo ensemble such that the mean of the ensemble is the best estimate, and the ensemble error covariance is a good estimate of the forecast error covariance. In the EnKF, the background error covariance estimated directly from an ensemble of forecasts propagated forward from an ensemble of analyses using the fully nonlinear forecast model. The algorithm used here is basically the one described by Evensen (2003, 2004). The detailed description of EnKF is omitted here.

Several studies (Anderson, 1997; Hamill et al., 2000; Bowler, 2006; Descamps and Talagrand, 2007) showed the advantages of EnKF for providing the initial conditions of ensemble forecasts. The EnKF also can readily include parameter estimation by state space augmentation in the same framework (Annan and Hargreaves, 2004; Annan et al., 2005a,b; Evensen, 2006). The principle is to consider the parameters as part of the model state alongside the conventional variables, and then using the covariance estimated from the ensemble to update parameters directly in the same manner as for the state variables.

In this study, we apply the EnKF to estimate the above defined three dimensionless parameters $\alpha_{\rm w}$, α_Q, and $\alpha_{\rm d}$. We need to give a set of initial ensembles for each of them. That is done by sampling from three Gaussian distributions with the mean value of their first guesses, which are zero in this study (meaning that there are no biases). Ideally, the initial standard deviation of the prior distribution of a parameter should represent the error in the first guess. However, the initial error is usually unknown in reality. In this study, the ensemble spread of each parameter is initialized to be 0.3, 0.5, and 0.5 for $\alpha_{\rm w}$, α_Q, and $\alpha_{\rm d}$, respectively.

The choice of the first guesses of a set of parameters will not affect the finally estimated values if the set of parameters are identifiable. However, the first guess values will affect the convergence rate during the data assimilation process. If the first guess values are selected close to the finally estimated values, the convergence will be fast. As discussed by Tong and Xue (2008a,b), whether the uncertain parameters are identifiable is ultimately determined by whether the inverse problem has a unique solution. The quality of observational data is another factor that decides the identifiability of the parameters. The estimation could be sensitive to small errors in the observations. In this paper, we do not focus on the detailed issues of identifiability of the three parameters $\alpha_{\rm w}$, α_Q, and $\alpha_{\rm d}$. However, their identifiability can be discussed from perspective of the independence of their impacts on observable variables. The model bias in the dust emission intensity has a strong impact on dust concentrations near the source regions, while the model bias in the dry velocity has a relatively small impact on dust concentrations near the source regions and has a large impact on the dust concentrations after long-distance transport. The high level wind speed bias has an impact on the arrival times of dust storm peaks at the observation stations. The observation network in this study has one observation site (Beijing) near the source regions and several sites in Japan where the long-distance transported dust concentrations can be measured.

5. Experiment setups

A series of dust storms occurred in the Gobi Desert and Mongolia and swept across a wide area including China, Korea, and Japan during the end of May 2007, which was an unusual phenomenon in late spring over the past years and was also observed as the latest spring record in Japan within the Japan Meteorological Agency observation history (Hara et al., 2008). A low-pressure system deeply developed on 23 May in northeastern Mongolia with very strong surface winds, which triggered the dust mobilization over the Gobi desert. The observations (dust records) from the Chinese Meteorological Administration and LIDAR dust observations showed that a dust storm moved into eastern China on 24 May and then arrived in Korea on 25 May and late in Japan around 25 to 26 May. The three-day averaged winds at the 1000 m height during 24–26 May 2007 (see Fig. 1) show that the northwest winds prevailed over East Mongolia and swept across North China, Korea, and Japan, which drove the dust transport. In this study, we target this case to investigate the performance of dust ensemble forecasting over eastern Asia.

We designed three experiments (shown in Table 1). A control experiment (Ctrl), a deterministic forecast, is performed without any assimilation of data. In two ensemble forecast experiments (Exp1 and Exp2), the EnKF is used to assimilate LIDAR observations. The difference is that in Exp1, only the initial conditions are analyzed while in Exp2 both the initial conditions and the model biases are analyzed or estimated. The model integration starts from 22 May and ends on 30 May in all three experiments. The assimilation cycles are performed during 23 May to 30 May with a 3-hour interval and the vertical profiles of the dust extinction

Table 1. Experimental configuration.

Experiment	Configuration
Ctrl	Control run, no assimilation
Exp1	Analyze only the initial conditions
Exp2	Analyze both the initial conditions and the model biases

coefficients at the above eight observational sites are assimilated.

We used the formulation of the EnKF with observations perturbed as in Houtekamer and Mitchell (1998). We used the same observation error covariance in a 4DVAR dust storm data assimilation study by Yumimoto et al. (2007a) that assimilated the same set of Lidar observations as this study. The observation errors were assumed to be temporally and spatially independent and set as 0.125 km^{-1} at the Beijing site and 0.0625 km^{-1} at the other sites (about 10%–40% of the maximum of 3-h averaged values during the dust period). Both Exp1 and Exp2 use 50 ensemble members. Many studies have suggested that the so-called localization technique is a feasible solution to reduce the effect of the sampling error for applications of an EnKF (e.g., Hamill et al., 2001), especially when the ensemble size is small. Localization aims to delete those long distance correlations in the gain matrix and thus to limit the influence of a single observation by the Kalman update equation within a fixed region around the observation location. Localization also can increase the rank of the forecast error covariance and improve the performance (Oke et al., 2002; Hunt et al., 2007). The localization can be implemented either by canceling the covariance between the long-distance elements (Hamill et al., 2001; Houtekamer and Mitchell, 2001) or by applying filters locally in the physical space (Ott et al., 2004). We adopt the covariance localization approach specified in Gaspari and Cohn (1999) for horizontal and vertical localization. The horizontal and vertical localizations are used in Exp1 and Exp2 in this study with vanishing radii of 1200 km (horizontally) and 1.5 km (vertically), respectively.

6. Results

6.1 *Ensemble mean as deterministic forecasting*

The ensemble mean, as the best estimate, should have better forecast skill than any of the other ensemble members because it averages over the many possible forecasts and essentially smoothes the chaotic nature of the atmosphere or ocean. First, we made a comparison between the ensemble means from two ensemble forecast experiments and the control experiment to examine the basic performance of Exp1 and Exp2, respectively.

Figure 2 shows the evolution of the means and standard deviations of the model biases in wind speed, dust emission intensity, and the dry deposition velocity during the assimilation process of Exp2. They are normalized by the original model calculated or input values, that the negative (positive) values indicate that the original ones are reduced (increased). The adjustments of the model biases will be justified below by comparing the forecast skills of Exp1 (without any bias adjustment) and Exp2.

Figure 3 shows the time-height cross section of the dust extinction coefficient validated at the same time

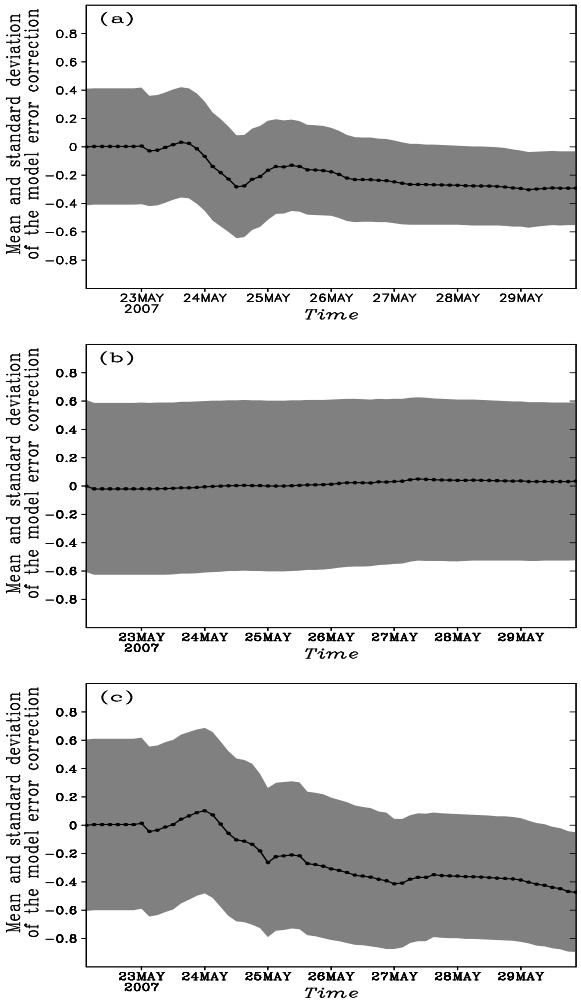

Fig. 2. The evolution of the means and standard deviations of the model errors in (a) wind speed; (b) dust emission intensity; (c) dry deposition velocity. They are normalized by the original model calculated or input values.

at Beijing, Seoul, Nagasaki, and Toyama, respectively: (a) observed by LIDAR; (b) predicted by the experiment Ctrl and (c) 3-hour forecast from Exp2 (i.e., LIDAR observations were assimilated 3 hours before the validation time). The 3-hour forecasts from Exp1 are very similar to that from Exp2 and not shown. Contrasting to the observations, results from the control experiment have a false dust peak around the 1000 m height at Beijing on 27 May [similar result also appeared in Hara et al. (2008)], while the 3-hour forecasts from Exp2 successfully removed the false peak. To explain this improvement, we plot a snapshot of the dust concentration fields on the end of 26 May from the control experiment and Exp2, respectively (Fig. 4). It indicates that the assimilation of LIDAR observations at Beijing reduced the false high dust concentration around Beijing.

Though there are some missing data at Seoul, there are two dust centers around the middle of 25 May and early 26 May as indicated by the observations. However, the control experiment predicted only one big peak on 25 May. The assimilation experiment, Exp2 forecasted correctly, the two-peak pattern but with smaller amplitudes. This underestimate may be due to the underestimate in the upstream area (for example, Beijing) on 24 May between the 1 km and 1.5 km height after assimilation (see Fig. 3a for Beijing). At Nagasaki, the control experiment predicted a peak at the end of 25 May that is too weak and about 6 hours earlier than the observed. The assimilation experiment, Exp2 improved the forecast, by increasing the peak amplitude and delaying the arrival time. Overall, the assimilation of the observations improves the timing of the arrival, the extinction coefficient level, and the dust layer height in all observation sites for 3-hour forecasts.

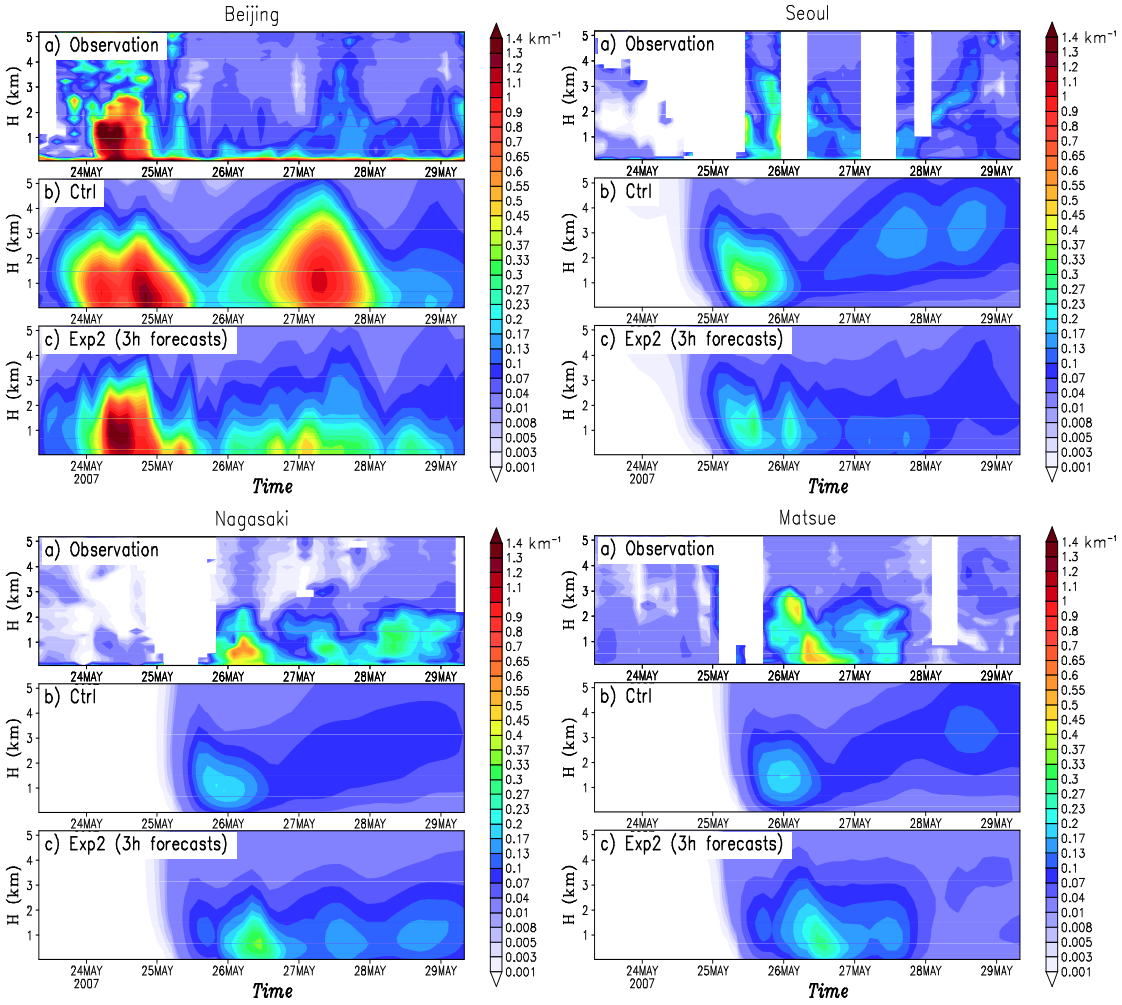

Fig. 3. The time-height cross section of (a) the dust extinction coefficient (km^{-1}) observed by LIDAR, (b) the simulated dust extinction coefficient (km^{-1}) by the control run and of (c) the 3 h ensemble mean forecast after assimilation at eight observational sites.

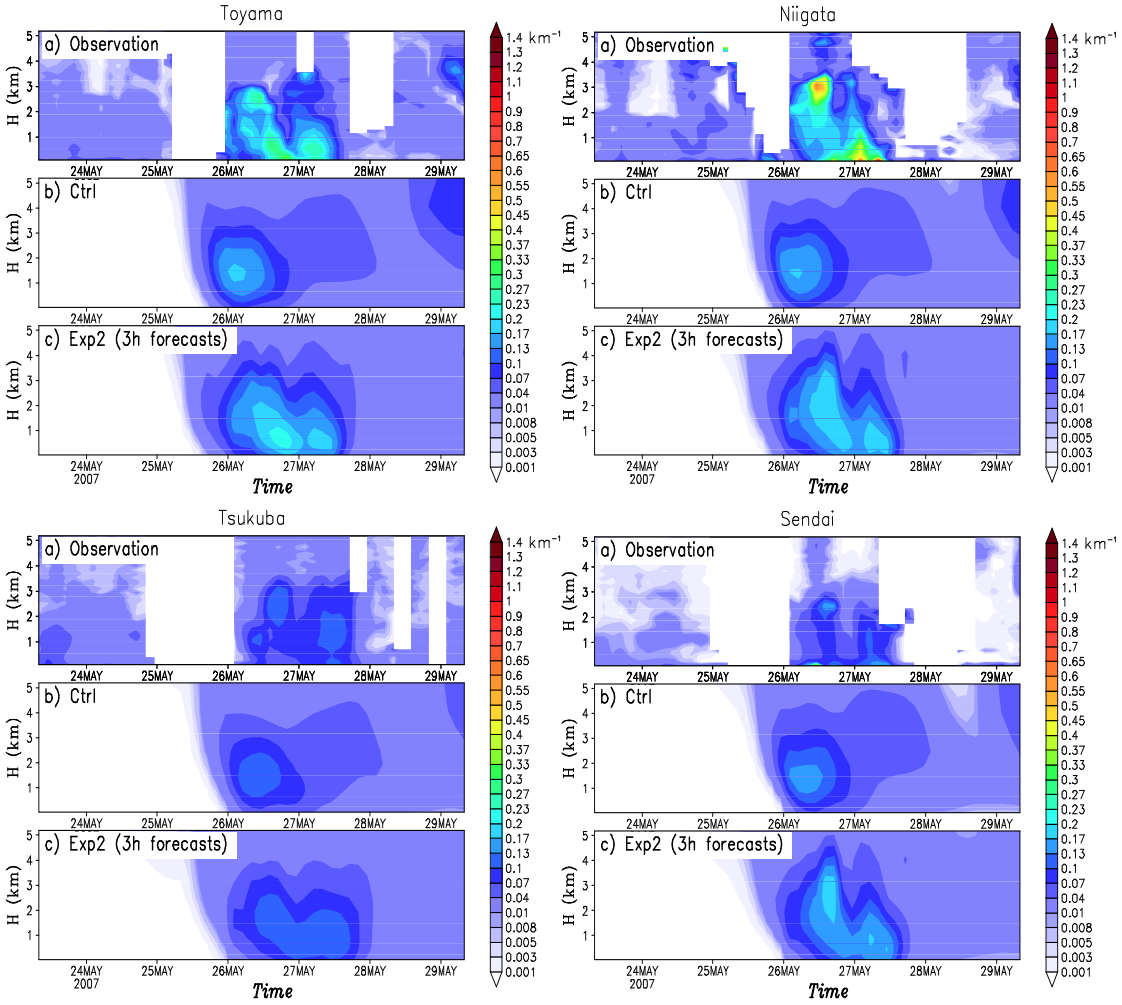

Fig. 3. Continued.

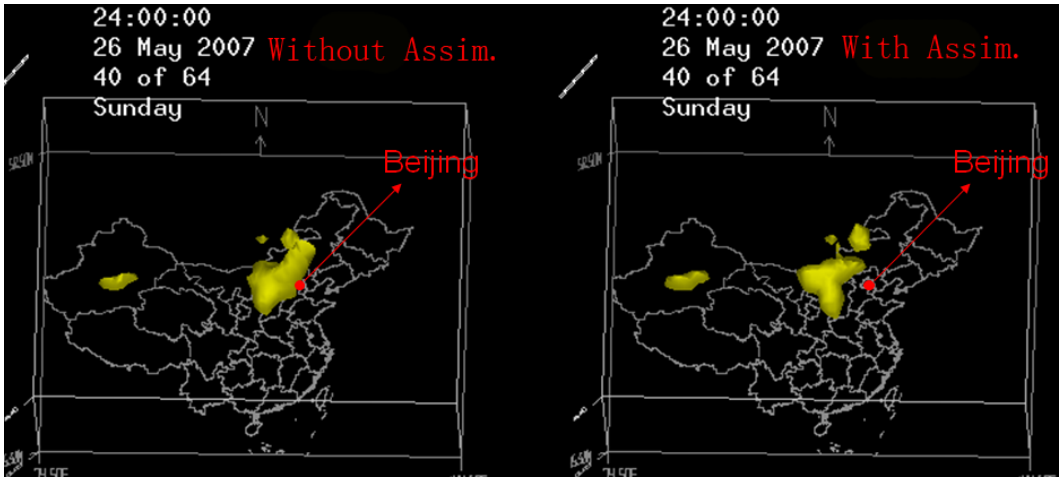

Fig. 4. A snapshot of dust concentration fields at the end of 24 May from the control experiment (the left panel) and Exp2 (the right panel), respectively.

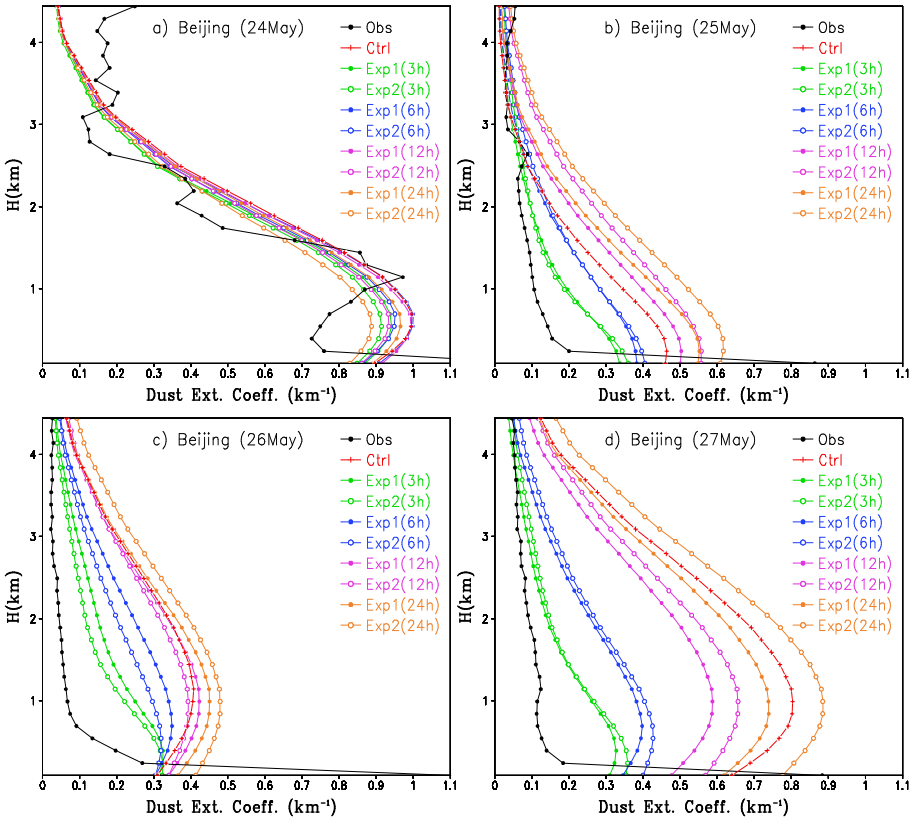

Fig. 5. The daily-averaged vertical profiles of the dust extinction coefficient observed by LIDAR and predicted by the model without and with assimilation in Beijing.

Figure 5 gives the daily-averaged vertical profiles of the dust extinction coefficient observed by LIDAR and predicts by the control experiment and the ensemble means of the ensemble forecast experiments in Beijing. The daily-averaged profiles in the three experiments match the observed well on 24 May though some fine structural differences exist. On 25 May, the forecasts within a 6 h forecast lead time are improved and closer to the observations than the control run, and there is no obvious difference between Exp1 and Exp2 within the 6 h lead time. On 26 and 27 May, the 12-h forecasts by Exp1 and Exp2 are improved over that of the control experiment. In addition, the forecast within a 12 h lead time with the bias correction (Exp2) are better than those without the bias correction (Exp1) on 26 May, while it is opposite on 27 May. The fact that the 24-h forecasts made by Exp1 and Exp2 did not improve that by the control experiment is mainly caused by the lack of observations at the upstream areas of Beijing. As shown by Fig. 4, the dust concentration field is only changed near Beijing by assimilating the observations at Beijing. We also noticed that the 24-h forecasts made by Exp2 are worse than that made by Exp1. The overall correction of biases of the high level wind speed and dry deposition velocity (the same for each model grid-point), which is mainly determined by observations in down-stream areas, would contribute to this.

The forecast profiles in Seoul, Nagasaki, and Toyama on 26 and 27 May are shown in Fig. 6. In down wind areas, the forecasts after assimilation are mostly improved up to a 24-h lead time and the improvement decreases with the forecast lead time. The forecasts after correcting the model biases are closer to the observations than those without the bias correction. This indicates that the biases of the high level wind speed and dry deposition velocity are important for forecasts in down wind areas and the EnKF system can correct them reasonably.

The forecast skill is related to the level of dust concentrations. The lower level of dust concentration in the downstream sites often yields a better forecast skill such as RMSE. The straightforward comparison of skills at downstream sites and near the dust source site (Beijing) may not be fair. In the future, some skills defined at different levels of concentrations should be introduced to evaluate dust storm forecasting. However, the forecast at the downstream sites benefit from the assimilation of the upstream observations at the site near the source, while the forecasts at the site

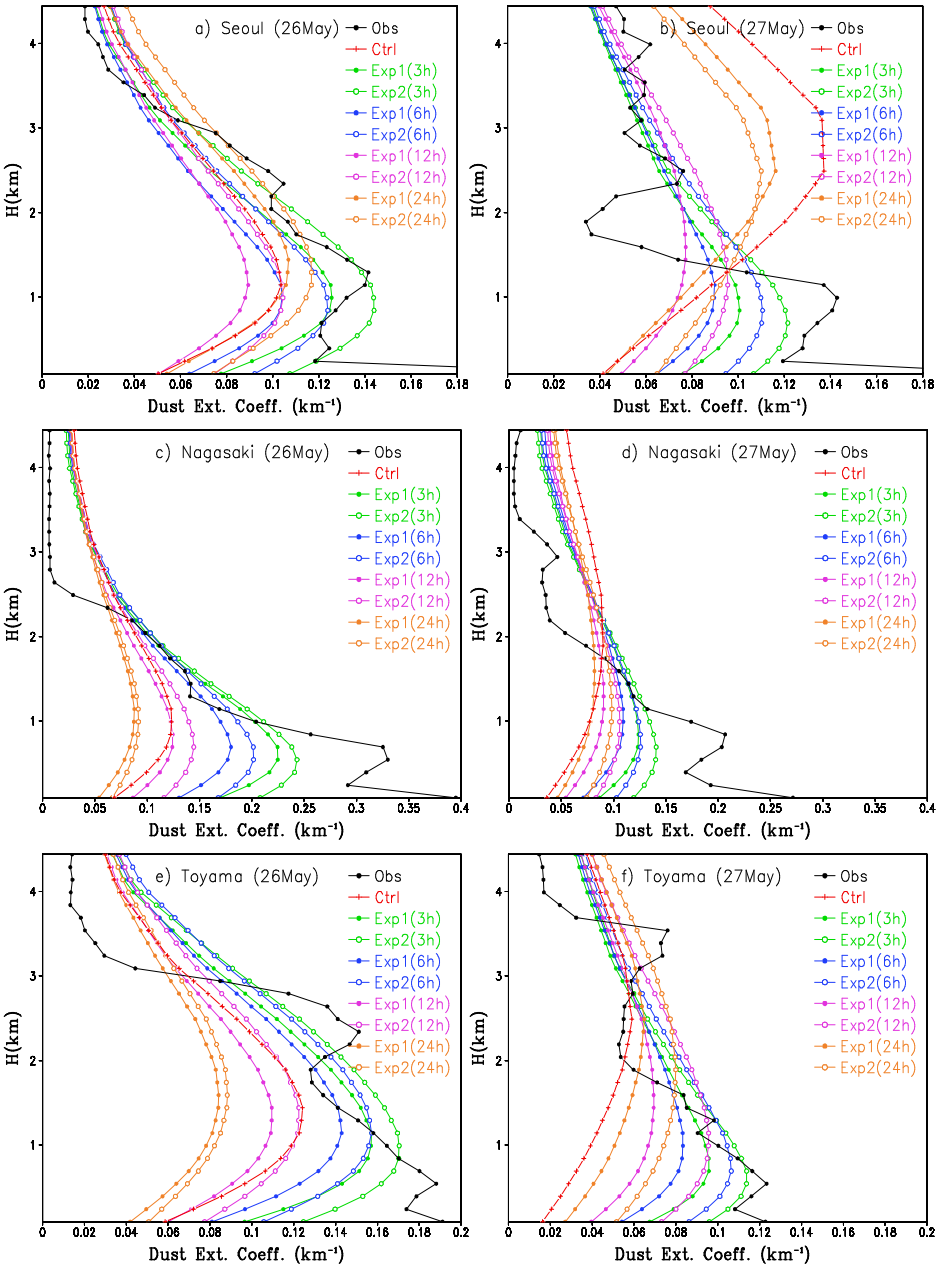

Fig. 6. As in Fig. 5 but for observation sites at Seoul, Nagasaki, and Toyama on 26 and 27 May.

near the source cannot get the same benefit from the downstream sites.

6.2 *Ensemble forecast skills*

The ensemble forecasts have a natural advantage over deterministic forecasts because it can provide many possible forecasts of the future scenario and can be used to estimate the probability distribution of the future scenario. An ensemble forecast system is thereby also a probabilistic forecast system (PFS), which can provide the likelihood of the occurrence of an event from a given set of ensemble forecasts rather than a categorical statement. Therefore, it naturally contains more information. The probability of an event occurrence can be calculated as the frequency of the ensemble members to predict the event over the total ensemble. For signal detection, the event could be the frequency of occurrence in the top or the bottom tercile of the climatological frequency. For a continuous predictant such as dust storm PFS, the event could be dust concentration (or dust extinction coefficient) above a certain threshold. In this study, dust extinction coefficients are used to sort the observations into separate groups.

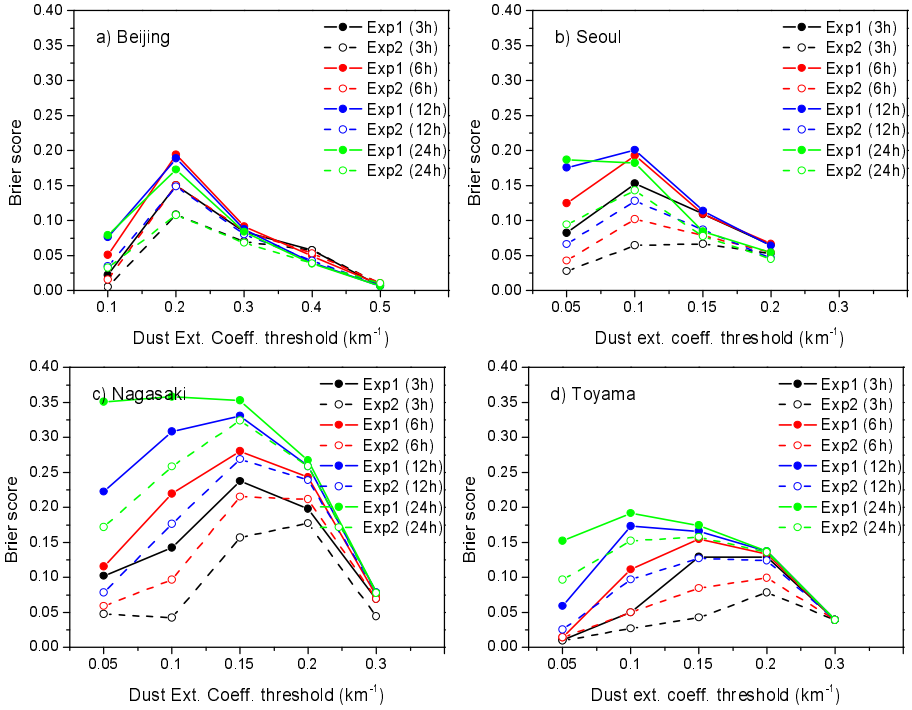

Fig. 7. Brier scores for 5 different dust extinction coefficient thresholds with different forecast lead times (3 h, 6 h, 12 h, and 24 h) in (a) Beijing, (b) Seoul, (c) Nagasaki, and (d) Toyama.

Ensemble forecast skill can be evaluated by determining the predictive accuracy of a forecast distribution. Two important forecast attributes can be computed to evaluate the accuracy and the usefulness of ensemble forecasting: a Brier score (Brier, 1950) and the relative operating characteristic (Swets, 1973; Mason, 1982; Harvery et al., 1992; Mason and Graham, 1999).

6.2.1 *Brier score*

A Brier score proposed by Brier (1950) measures the accuracy of a set of probability assessments for a binary event. It is the average deviation between predicted probabilities for a set of events and their outcomes, so a lower score represents a higher accuracy. Suppose there is a forecasted probability f_i that the event will occur and let $o_i = 1$ if the event did occur and $o_i = 0$ if it didn't, for each (i) of the total number of forecasts. Then, the Brier score (BS) is given by

$$\mathrm{BS} = \frac{1}{N}\sum_{i=1}^{N}(f_i - o_i)^2 \qquad (6)$$

Brier scores range from 0 to 1, taking the value of 0 for perfect forecasting and the value of 1.0 for the worst possible forecasting.

Figure 7 shows Brier scores for five thresholds in Beijing, Seoul, Nagasaki and Toyama for the forecasts during 23–30 May 2007 in two ensemble forecasting experiments with different forecast lead times (3 h, 6 h, 12 h, and 24 h). The Brier scores become worse as the forecast lead time at each station with the best forecast (smallest Brier score) for 3-h forecasts. It is interesting to notice that the Brier score values are not monotone functions of the threshold values. The Brier score in Beijing for the threshold value of 0.2 km^{-1} is the largest, and indicates that the probabilistic forecasts are the worst among the five thresholds. The difference of the Brier scores between the two experiments and different forecast lead times decreases as the threshold values increase, which may imply that the difference of the impact of the corrected initial conditions and bias correction on ensemble forecasts become slight for larger threshold values. In down wind areas (Seoul, Nagasaki, and Toyama), scores in Exp2 with the bias correction are mostly smaller than those in Exp1, which indicates the positive impact of the bias correction on improved forecasts.

6.2.2 *Relative operating characteristic*

The relative operating characteristic (ROC) curve (Swets, 1973; Mason, 1982; Harvery et al., 1992; Mason and Graham, 1999) is a useful method of representing forecast skill based on a contingency table, which can be used to estimate the quality of probabilistic forecast systems that express output in continuous, categorical,

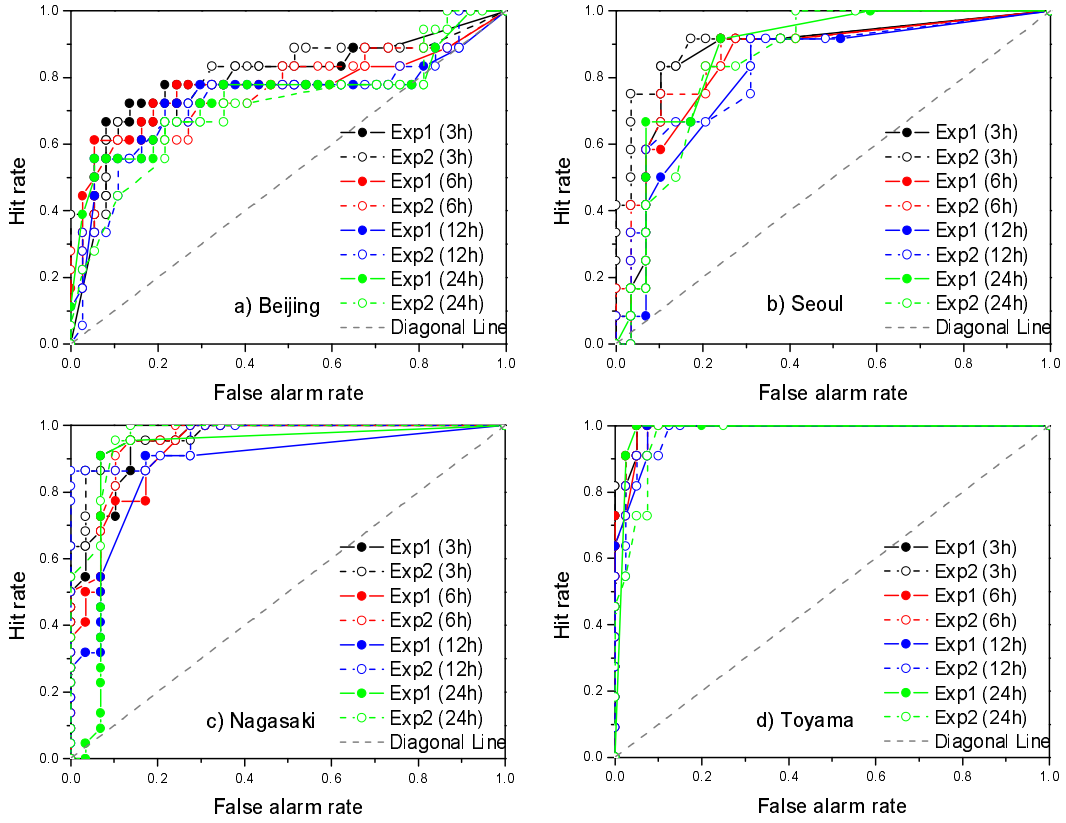

Fig. 8. Hit rates vs false-alarm rates (relative operating characteristic, ROC, curve) for the observed dust extinction coefficients above (a) 0.3 km^{-1} in Beijing, 0.1 km^{-1} in (b) Seoul, (c) Nagasaki, and (d) Toyama for the forecasts during the period of 23–30 May 2007 in two data assimilation experiments with different forecast lead times (3 h, 6 h, 12 h, and 24 h). The better the probabilistic forecast, the closer the ROC curve is to the upper left corner. The dashed line is the ROC curve for a chance forecast. The ROC curves are generated for each frequency threshold which assumes values from 0/50 to 50/50, with increments of 1/50.

or binary mode (events). A two-by-two contingency table for a binary system is illustrated in Table 2. A hit, if an event occurred and a warning was provided (h is the number of hits) with the likelihood greater than or equal to the given probability threshold; a false alarm, if an event did not occur but a warning was provided (f is the number of false alarms) with the likelihood greater than or equal to the given probability threshold; a miss, occurred but not forecast (m is the number of misses); a correct rejection, neither occurred nor forecast (c is the number of correct rejections).

The hit rate and the false rate are computed from the contingency table (Table 2):

$$\text{Hit rate} = h/(h+m) = h/e\,; \tag{7}$$

$$\text{False alarm rate} = f/(f+c) = f/e'\,. \tag{8}$$

Then, the hit rates are plotted on the ordinate against the corresponding false alarm rates on the abscissa to generate the ROC curve (see Fig. 8). The better the probabilistic forecast, the closer the ROC curve is to the upper left corner. The curve approaches the lower right corner when there is negative skill. The diagonal line represents the ROC curve for a chance forecast from climatology, which indicates that the forecast system has no skill.

Figure 8 shows the ROC curves for observed dust extinction coefficients above 0.3 km^{-1} in Beijing, 0.1 km^{-1} in Seoul, Nagasaki, and Toyama for the forecasts during 23–30 May 2007 in two ensemble forecasting experiments with different forecast lead times (3 h, 6 h, 12 h, and 24 h). The ROC curves are generated assum-

Table 2. Two by two contingency table.

Observations	Forecast		
	Warning	No warning	Total
Event	h	m	e
No event	f	c	e'

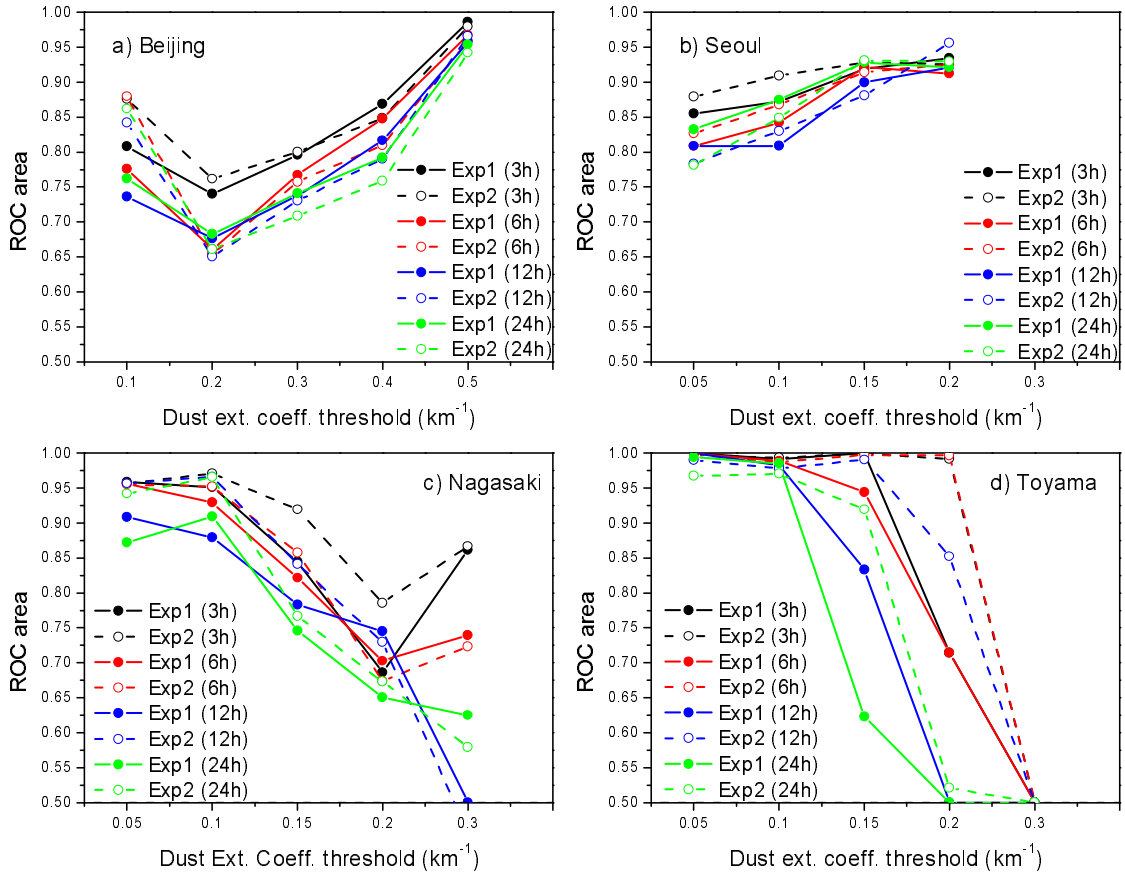

Fig. 9. ROC areas for 5 different dust extinction coefficient thresholds with different forecast lead times (3 h, 6 h, 12 h, and 24 h) in (a) Beijing, (b) Seoul, (c) Nagasaki, and (d) Toyama.

ing probability threshold values from 0/50 to 50/50, with increments of 1/50. Most of the ROC curves are located in the upper left corner, which indicates the ensemble forecasting has some skill. A more clear comparison with different thresholds for two experiments and different forecast lead times can be obtained from Fig. 9.

The ROC area is the area under the ROC curve representing the usefulness of an ensemble to discriminate between events. The area is larger than 0.5 when there is skill and equal to 0.5 with no skill. The closer the area approaches one, the more useful the forecast. The ROC areas under the corresponding ROC curves and their variations with five different dust extinction coefficient thresholds (five events) with different forecast lead times (3 h, 6 h, 12 h, and 24 h) are shown in Fig. 9. No values in Seoul with a threshold of 0.3 km^{-1} are shown because there are no observation events above 0.3 km^{-1} during this period. The ROC areas mostly decrease with the forecast lead time. Most of the 3-h forecasts have larger ROC area values (black lines with symbols) than those for longer forecast lead times. In Beijing, the variation of ROC areas with thresholds is opposite to those of the Brier scores, which indicates that the implication of ROC areas agrees well with that of the Brier scores in Fig. 7. It shows the worst probabilistic forecasts for the threshold value of 0.2 km^{-1} in Beijing as indicated by the Brier scores. The probabilistic forecasts (in Beijing) for larger threshold values ($\geqslant 0.3$ km^{-1}) are better (the larger relative operating characteristic areas) in Exp1 than those in Exp2. In down wind areas (Seoul, Nagasaki, and Toyama), the probabilistic forecasts in Exp2 are mostly better (larger areas) than those in Exp1. The areas generally increase in Seoul and decrease in Nagasaki and Toyama with the threshold values. The above results also indicate that the bias correction of the high level wind and dry deposition velocity have positive impacts on the ensemble forecasts in downwind areas, especially at Japanese stations. In general, the implication of ROC areas agrees with that of the Brier scores.

6.3 *Ensemble width vs. observations*

A "good" ensemble forecasting system should have the truth as a plausible member of the ensembles.

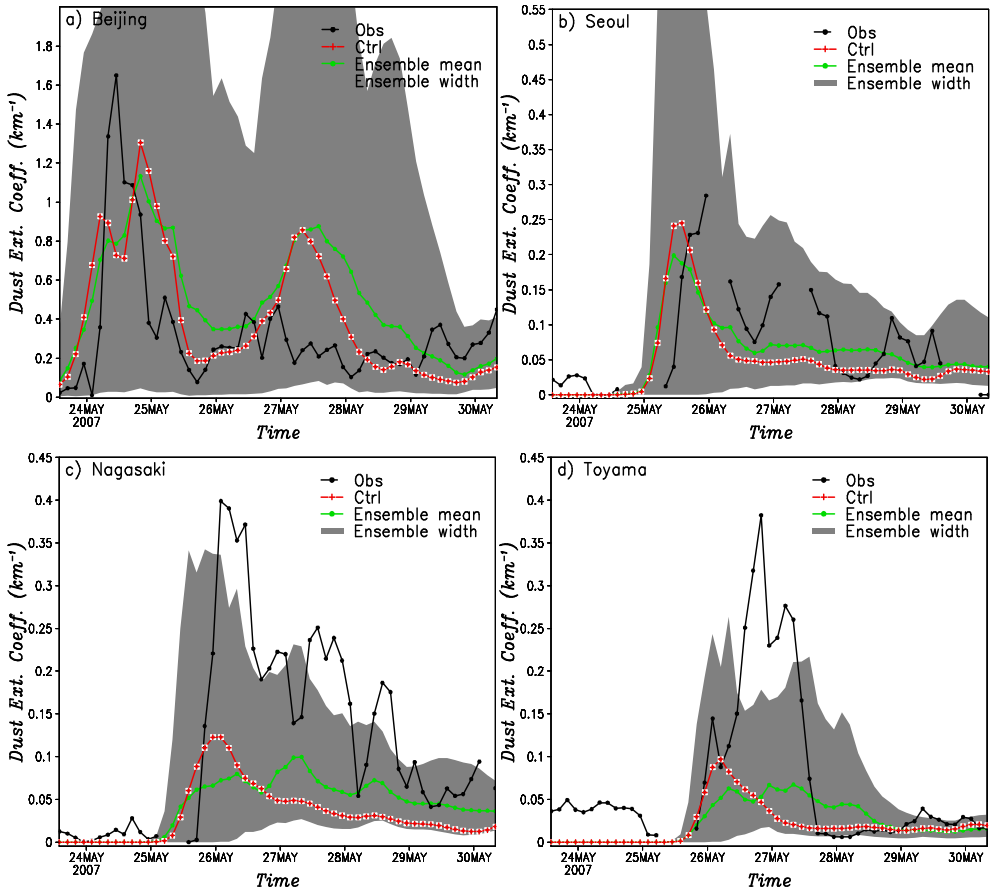

Fig. 10. Time series of dust extinguish coefficients at 180 m height of 24-h ensemble forecasting (ensemble mean and width) in Exp2, along with observations and the forecasts from the control experiment at four sites.

Figure 10 shows the time evolution of the mean and width of the 24-h ensemble forecast in Exp2 in terms of the dust extinction coefficient at the 180 m height, along with the observations and the forecasts from the control experiment at Beijing, Seoul, Nagasaki, and Toyama, respectively. The ensembles are generally wide enough to cover the observations as the approximate truth. However, ensembles at Beijing are over dispersive (confirmed by a U-shaped Talagrand diagram, not shown) while at Nagasaki and Toyama, the bias problems exist. This indicates the insufficient representations of model error and initial condition error in the present study.

7. Conclusions and discussions

An ensemble forecast is a collection of various possible forecasts verifying at the same time. Each of the forecasts is regarded as one of the possible scenarios given the uncertainty associated with the forecasting. The key to good ensemble forecasting is the correct presentation of the errors in the forecast models and the initial conditions. The EnKF method, as an unified approach to both the data assimilation and ensemble forecasting problems (Kalnay et al., 2006), is used in this study that focuses on the performance of dust storm ensemble forecasting targeting a dust episode from China to Korea and Japan during 23–30 May 2007. Based on our previous studies (Lin et al., 2008a,b), the errors in the input wind field, dust emission intensity, and dry deposition velocity are among important model uncertainties. Other studies (e.g., Uno et al., 2006) also indicated that some of these errors could have large biases. Therefore, the model bias correction is also considered in this study as part of the data assimilation process. The NIES LIDAR observations are assimilated to generate the initial ensembles and correct the model biases. The ensemble forecast skills are evaluated against the observations and a benchmark/control forecast that is a simple model run without assimilation of any observations. Results from this study are not verified over a large number of cases and must be considered preliminary.

Results show that the ensemble mean forecasts

have substantial improvement over the control forecasts after correcting the initial conditions by assimilating the LIDAR dust extinction coefficients. The forecasts after assimilation correctly capture the major dust arrival and cessation timing at each observation site and improve the distribution and the coefficient levels especially for short time forecasts (e.g. 3-h forecasts). The improvement decreases with the forecast lead times. The improved forecast lead time is shorter in Beijing, which is closer to the dust source regions, where the forecasts with more than a 6-hour lead time have little improvement. While the improved forecast lead time is longer in down wind areas (e.g. in Korea and Japan), the 12-h forecasts still need much improvement. It indicates that the closer to the dust source regions, the smaller impact of the corrected initial conditions on improved dust forecasts and the shorter lead time of improved forecasts. This may be due to the lack of observations in the dust source regions in this study.

Bias correction of wind speed and dry deposition velocity further improved the forecasts in down wind areas. The forecasts within 24 hours are mostly improved and better than those without the bias correction. This indicates that the EnKF system can correct the bias accurately and the biases of wind speed and dry deposition velocity have strong impacts on forecasts in down wind areas. In addition, due to the opposite modeled results in areas closer to the dust source regions (e.g. Beijing, overestimated) and in down wind areas (underestimated), the overall correction of the biases of the high level wind speed and dry deposition velocity (the same for each model grid-point), which are mainly determined by observations in down wind areas, may contribute oppositely and worsen the results in upstream areas.

The verification of the ensemble forecasts by Brier scores and ROC curves (or ROC areas) indicates that the ensemble forecasting system has useful forecast skills The skill and usefulness weaken with the forecast lead time. It also shows that there is better forecast skill and more useful ensemble forecasts with the bias correction in down wind areas than those without the bias correction. However, there still exists bias in the forecasts after correcting the major model biases, especially for a sudden increase of dust, which may be due to model deficiencies or other model biases not considered in this study.

The study suggests that better initial conditions in upstream areas are important for improving dust ensemble forecasts in downstream areas. One of the important weak points in this study is that the LIDAR network has only one LIDAR (at Beijing) relatively close to the dust source region. Some other LIDARs did exist in China, but were not functional during the period. Assimilation of observations of this LIDAR is limited in providing unbiased initial ensembles in the upwind areas that are crucial for 24 h or longer forecasts of down-wind areas. In future studies, assimilation of observations with comprehensive coverage near dust source regions is a priority. The surface PM_{10} concentrations are operationally monitored around the dust source regions in China. The satellite remote sensing retrievals made by the China Meteorological Administration are also an important source of data (Niu et al., 2008).

The result shows that the benefit from ensembling decreases with the forecast time. Usually, the opposite is true in ensemble weather prediction. This result may not be generally true for other cases. It may also reflect that the model errors in the ensemble forecast still miss some important uncertainties or/and drawbacks in the current model error perturbations. This should be further investigated.

The model errors considered in this study certainly are not sufficient. The way of presenting them is much simplified. The use of ensemble wind forecasts from operational ensemble weather forecasting should be tested in future studies. Considerations of multiple dust emission parameterization schemes and the use of several sources of land use and soil information in the emission will enhance the presentation of the model error in the dust emission. The multi-model approach (i.e., super-ensemble method) could be potential ways to better represent the model errors.

Though the bias correction implemented in this study shows a positive impact, there is much free space to improve the implementation. The biases in the high level wind errors and dry deposition velocity errors have stronger impacts on dust concentrations in down wind areas (e.g., Korea and Japan) than in the area near the dust source regions, while the bias in the dust emission intensity error has a larger impact on areas relatively closer to the dust source regions. This fact may lead to a better implementation of correcting error biases by putting different weights on observations based on their locations.

The fact that the observations contain some error should be considered when comparing them to the ensemble members. A quantitative evaluation of the ensemble spread and bias should be performed, as described by Candille et al. (2007). Unfortunately the sample size (i.e. number of observations and dates) is too small in this study to enable a useful result from such an approach. In future studies, more cases should be used in such an approach.

One deficiency of not performing a full coupled analysis of dust with wind and other meteorological

fields, as in the present study, is that the wind perturbations could be unrealistically larger/smaller than the wind analysis error characteristics, then one may get an undesirably large spread. On the other hand, this larger/smaller spread may be compensating for model errors. To carry out a full coupled analysis of dust with wind and other meteorological fields is one of the directions of future research.

An important issue that is not addressed in this study is the adaptive identifying model errors and quantifying their statistical characteristics. This is a tough problem. However, the innovation information (i.e., forecast minus observation) can provide some hints. For example, if the forecast has a large difference with the observations near the dust source regions, it is quite certain that the dust emission intensity contains large errors or the wind field between the source regions and the observation sites has large errors.

Acknowledgements. The authors would like to express their great appreciation to Dr. Sugimoto and Shimizu for providing the LIDAR data. This research was supported by the Chinese Academy of Sciences (Grant Nos. KZCX2-YW-202, KZCX2-YW-205), the National Basic Research Program of China (2006CB403600), and the National Natural Science Foundation of China (Grant No. 40221503).

REFERENCES

Anderson, J. L., 1997: Impact of dynamical constraints on the selection of initial conditions for ensemble predictions: Low order perfect model results. *Mon. Wea. Rev.*, **125**, 2 969–2 983.

Annan, J. D., and J. C. Hargreaves, 2004: Efficient parameter estimation for a highly chaotic system. *Tellus*, **56A**, 520–526.

Annan, J. D., J. C. Hargreaves, N. R. Edwards, and R. Marsh, 2005a: Parameter estimation in an intermediate complexity earth system model using an ensemble Kalman filter. *Ocean Modelling*, **8**, 135–154.

Annan, J. D., D. J. Lunt, J. C. Hargreaves, and P. J. Valdes, 2005b: Parameter estimation in an atmospheric GCM using the ensemble Kalman filter. *Nonlinear Processes Geophys.*, **12**, 363–371.

Bowler, N., 2006: Comparison of error breeding, singular vectors, random perturbations and ensemble Kalman filter perturbation strategies on a simple model. *Tellus*, **58A**, 538–548.

Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.*, **78**, 1–3.

Candille, G., C. Côté, P. L. Houtekamer, and G. Pellerin, 2007: Verification of an ensemble prediction system against observations. *Mon. Wea. Rev.*, **135**, 2 688–2 699.

Chin, M., and Coauthors, 2003: A global aerosol model forecast for the ACE-Asia field experiment. *J. Geophys. Res.*, **108**(D23), 8654, doi: 10.1029/2003JD003642.

Descamps, L., and O. Talagrand, 2007: On some aspects of the definition of initial conditions for ensemble prediction. *Mon. Wea. Rev.*, **135**, 3 260–3 272.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**(C5), 10 143–10 162.

Evensen, G., 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, **53**, 343–367.

Evensen, G., 2004: Sampling strategies and square root analysis schemes for the EnKF. *Ocean Dynamics*, 54, 539–560.

Evensen, G., 2006: *Data Assimilation: The Ensemble Kalman Filter*. Springer, German, 280pp.

Fernald, F. G., 1984: Analysis of atmospheric LIDAR observations: Some comments. *Appl. Opt.*, **23**, 652–653.

Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723–757.

Gong, S. L., X. Y. Zhang, T. L. Zhao, I. G. McKendry, D. A. Jaffe, and N. M. Lu, 2003: Characterization of soil dust aerosol in China and its transport and distribution during 2001 ACE-Asia: 2. Model simulation and validation. *J. Geophys. Res.*, **108**(D9), 4262, doi: 10.1029/2002JD002633.

Hamill, T. M., C. Snyder, and R. E. Morss, 2000: A comparison of probabilistic forecasts from bred, singular-vector, and perturbed observation ensembles. *Mon. Wea. Rev.*, **128**, 1 835–1 851.

Hamill, T. M., J. S. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background error covariances in an ensemble Kalman Filter. *Mon. Wea. Rev.*, **129**, 2 776–2 790.

Hanna, S. R., and Coauthors, 2001: Uncertainties in predicted ozone concentrations due to input uncertainties for the UAM-V photochemical grid model applied to the July 1995 OTAG domain. *Atmos. Environ.*, **35**, 891–903.

Hara, Y., K. Yumimoto, I. Uno, A. Shimizu, N. Sugimoto, Z. Liu, and D. M. Winker, 2008: Asian dust outflow in the PBL and free atmosphere retrieved by NASA CALIPSO and an assimilated dust transport model. *Atmospheric Chemistry and Physics Discussions*, **8**, 8 715–8 742.

Harvey, L. O., K. R. Hammond, C. M. Lusk, and E. F. Mross, 1992: The application of signal detection theory to weather forecasting behavior. *Mon. Wea. Rev.*, **120**, 863–833.

Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811.

Houtekamer, P. L., and H. L. Mitchell, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **129**, 123–137.

Hunt, B. R., E. J. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica* (*D*), **230**, 112–126.

Kalnay, E., B. Hunt, E. Ott, and I. Szunyogh, 2006: Ensemble forecasting and data assimilation: two problems with the same solution? Chapter 7, *Predictability of Weather and Climate*, T. Palmer and R. Hagedron, Eds., Cambridge University Press, 702pp.

Leutbecher, M., and T. N. Palmer, 2008: Ensemble forecasting. *J. Comput. Phys.*, **227**, 3 515–3 539.

Lin, C. Y., J. Zhu, and C. G, Lu, 2006: Comparison of Ensemble Kalman Filter and optimal interpolation in different observational networks. *Climatic and Environmental Research*, **11**, 553–564. (in Chinese)

Lin, C. Y., Z. F. Wang and J. Zhu, 2008a: An Ensemble Kalman Filter for severe dust storm data assimilation over China. *Atmospheric Chemistry and Physics*, **8**, 2 975–2 983.

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

Lin, C. Y., J. Zhu, and Z. F. Wang, 2009: Uncertainty analysis of a dust transport model. *Chinese J. Atmos. Sci.*, **33**(2). 232–240. (in Chinese)

Liu, M., D. L. Westphal, S. Wang, A. Shimizu, N. Sugimoto, J. Zhou, and Y. Chen, 2003: A high-resolution numerical study of the Asian dust storms of April 2001. *J. Geophys. Res.*, **108**(D23), 8653, doi: 10.1029/2002JD003178.

Liu, Z., N. Sugimoto, and T. Murayama, 2002: Extinction-to-backscatter ratio of Asian dust observed with high-spectral-resolution LIDAR and Raman LIDAR. *Appl. Opt.*, **41**, 2 760–2 767.

Liu, Z.-Q., and F. Rabier, 2002: The interaction between model resolution, observation resolution and observation density in data assimilation: A one-dimensional study. *Quart. J. Roy. Meteor. Soc.*, **128**, 1 367–1 386.

Marticorena, B., and G. Bergametti, 1995: Modeling the atmospheric dust cycle: 1. Design of a soil-derived emission scheme. *J. Geophys. Res.*, **100**(D8), 16 415–16 430.

Marticorena, B., G. Bergametti, B. Aumont, Y. Callot, C. N'Doume, and M. Legrand, 1997: Modeling the atmospheric dust cycle: 2. Simulation of Saharan dust sources. *J. Geophys. Res.*, **102**(D4), 4 387–4 404.

Mason, S. J., 1982: A model for assessment of weather forecasts. *Australian Meteorology Magazine*, **30**, 291–303.

Mason, S. J., and N. E. Graham, 1999: Conditional probabilities, relative operating characteristics, and relative operating levels. *Wea. Forecasting*, **14**, 713–725.

Nickovic, S., G. Kallos, A. Papadopoulos, and O. Kalaliagou, 2001: A model for prediction of desert dust cycle in the atmosphere, *J. Geophys. Res.*, **106**(D16), 18 113–18 129.

Niu, T., S. L. Gong, G. F. Zhu, H. L. Liu, X. Q. Hu, C. H. Zhou, Y. Q. Wang, and X. Y. Zhang, 2008: Data assimilation of dust aerosol observations for CUACE/Dust forecasting system. *Atmospheric Chemistry and Physics*, **8**, 3 473–3 482,

Oke, P. R., J. S. Allen, R. N. Miller, G. D. Egbert, P. M. Kosro, 2002: Assimilation of surface velocity data into a primitive equation coastal ocean model. *J. Geophys. Res.*, **107**, 3122, doi: 10.1029/2000JC000511.

Ott, E., and Coauthors, 2004: A local ensemble Kalman filter for atmospheric data assimilation. *Tellus*, **56A**, 415–428.

Park, S., and H. In, 2003: Parameterization of dust emission for the simulation of the yellow sand (Asian dust) event observed in March 2002 in Korea. *J. Geophys. Res.*, **108**(D19), 4618, doi: 10.1029/2003JD003484.

Pérez, C., S. Nickovic, J. M. Baldasano, M. Sicard, F. Rocadenbosch, and V. E. Cachorro, 2006: A long Saharan dust event over the western Mediterranean: Lidar, Sun photometer observations, and regional dust modeling. *J. Geophys. Res.*, **111**, D15214, doi: 10.1029/2005JD006579.

Shao, Y., and Coauthors, 2003: Northeast Asian dust storms: Real-time numerical prediction and validation. *J. Geophys. Res.*, **108**(D22), 4691, doi: 10.1029/2003JD003667.

Shimizu, A., N. Sugimoto, I. Matsui, K. Arao, I. Uno, T. Murayama, N. Kagawa, and K. Aoki，2004：Continuous observations of Asian dust and other aerosols by dual-polarization lidars in China and Japan during ACE-Asia. *J. Geophys. Res.*, **109**, D19S17, doi: 10.1029/2002JD003253.

Sugimoto, N., and coauthors, 2006: Network observations of Asian dust and air pollution aerosols using two-wavelength polarization lidars. *23rd International Laser Radar Conference*, July 2006, Nara, Japan (23ILRC, ISBN 4-9902916-0-3), 851–854.

Swets, J. A., 1973: The relative operating characteristic in psychology. *Science*, **182**, 990–1 000.

Tong, M, and M. Xue, 2008a: Simultaneous estimation of microphysical parameters and atmospheric state with simulated radar data and ensemble square root Kalman filter. Part I: Sensitivity analysis and parameter identifiability. *Mon. Wea. Rev.*, **136**, 1 630–1 648.

Tong, M, and M. Xue, 2008b: Simultaneous estimation of microphysical parameters and atmospheric state with simulated radar data and ensemble square root Kalman filter. Part II: Parameter estimation experiments. *Mon. Wea. Rev.*, **136**, 1 649–1 668.

Uematsu, M., Z. Wang, and I. Uno, 2003: Atmospheric input of mineral dust to the western North Pacific region based on direct measurements and a regional chemical transport model. *Geophys. Res. Lett.*, **30**, 1342, doi: 10.1029/2002GL016645.

Uno, I., H. Amano, S. Emori, K. Kinoshita, I. Matsui, and N. Sugimoto, 2001: Trans-Pacific yellow sand transport observed in April 1998: A numerical simulation. *J. Geophys. Res.*, **106**(D16), 18 331–18 344.

Uno, I., and Coauthors, 2003: Regional chemical weather forecasting system CFORS: Model descriptions and analysis of surface observations at Japanese island stations during the ACE-Asia experiment. *J. Geophys. Res.*, **108**(D23), 8668, doi: 10.1029/2002JD002845.

Uno, I., and Coauthors, 2004: Numerical study of Asian dust transport during the springtime of 2001 simulated with the Chemical Weather Forecasting System (CFORS) model. *J. Geophys. Res.*, **109**, D19S24, doi: 10.1029/2003JD004222.

Uno, I., and Coauthors, 2006: Dust model intercomparison (DMIP) study over Asia: Overview. *J. Geophys. Res.*, **111**, D12213, doi: 10.1029/2005JD006575.

Wang, Z., H. Ueda, and M. Huang, 2000: A deflation module for use in modeling long-range transport of yellow sand over East Asia. *J. Geophys. Res.*, **105**(D22), 26 947–26 960.

Wang, Z., H. Akimoto, and I. Uno, 2002: Neutralization of soil aerosol and its impact on the distribution of acid rain over East Asia: Observations and model results. *J. Geophys. Res.*, **107**, 4389, doi: 10.1029/2001JD001040.

Wu, L., V. Mallet, M. Bocquet, and B. Sportisse, 2008: A comparison study of data assimilation algorithms for ozone forecasts. *J. Geophys. Res.*, **113**, D20310, doi: 10.1029/2008JD009991.

Yumimoto, K., I. Uno, N. Sugimoto, A. Shimizu, and S. Satake, 2007a: Adjoint inverse modeling of dust emission and transport over East Asia. *Geophys. Res. Lett.*, **34**, L08806, doi: 10.1029/2006GL028551.

Yumimoto, K., I. Uno, N. Sugimoto, A. Shimizu, Z. Liu, and D. M. Winker, 2007b: Numerical modeling of Asian dust emission and transport with adjoint inversion using LIDAR network observations. *Atmospheric Chemistry and Physics Discussions*, **7**, 15 955–15 987.

Zhang, X. Y., R. Akimoto, G. H. Zhu, T. Chen, and G. Y. Zhang, 1998: Concentrations, size-distribution and deposition of mineral aerosol over Chinese desert regions. *Tellus*, **50B**, 317–330.

Zhou, C. H., and Coauthors, 2008: Development and evaluation of an operational SDS forecasting system for East Asia: CUACE/DUST. *Atmospheric Chemistry and Physics*, **8**, 787–798.