Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-732, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 26 August 2016

Atmospheric Chemistry and Physics Discussions

Improving $PM_{2.5}$ forecast over China by the joint adjustment of initial conditions and source emissions with an ensemble Kalman filter

Zhen Peng[1,2], Zhiquan Liu[2], Dan Chen[2], Junmei Ban[2]

1 School of Atmospheric Sciences, Nanjing University, Nanjing, China

2 National Center for Atmospheric Research, Boulder, Colorado, USA

Abstract. In an attempt to improve the forecasting of atmospheric aerosols, the ensemble square root filter algorithm was extended to simultaneously optimize the chemical initial conditions and emission input. The forecast model, which was expanded by combining the Weather Research and Forecasting with Chemistry (WRF-Chem) model and a forecast model of emission scaling factors, generated both chemical concentration fields and emission scaling factors. The forecast model of emission scaling factors was developed by associating the time smoothing operator with WRF-Chem forecast chemical concentrations. Hourly surface fine particulate matter ($PM_{2.5}$) observations were assimilated over China from 5 to 16 October 2014. A series of 48-h forecasts were then carried out with the optimized initial conditions and emissions on each day at 0000 UTC and a control experiment was performed without data assimilation. The results showed that the forecasts with the optimized initial conditions and emissions typically outperformed those from the control experiment. In the Yangtze River delta and the Pearl River delta regions, large reduction of the Root Mean Square Errors (RMSEs) was obtained for almost the entire 48-h forecast range attributed to assimilation. Especially, the relative reduction

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-732, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 26 August 2016

Atmospheric Chemistry and Physics Discussions

in RMSE due to assimilation was about 40% at nighttime when WRF-Chem performed comparatively worse. In the Beijing–Tianjin–Hebei region, relatively smaller improvements were achieved in the first 24-h forecast.

1. Introduction

Aerosol prediction by regional air quality model in heavy polluted regions is challenging due to many factors. In addition to the deficiency of chemistries, the uncertainties of primary and precursor emissions and the initial conditions (ICs) also limit the forecast accuracy. Data assimilation (DA), which is used to improve the ICs of aerosols and to optimize data on aerosol emissions, has been shown to be one of the most effective ways to improve the forecasting of aerosol pollution.

From the perspective of reducing the uncertainties in the ICs for aerosols, recent efforts have focused on assimilating aerosol observations using optimal interpolation (Collins et al., 2001; Yu et al., 2003; Adhikary et al., 2008; Tombette et al., 2009; Lee et al., 2013) or variational (Kahnert, 2008; Zhang et al., 2008; Benedetti et al., 2009; Pagowski et al., 2010; Liu et al., 2011; Schwartz et al., 2012; Li et al., 2013; Jiang et al., 2013; Saide et al., 2013) DA algorithms. Ensemble-based DA algorithms, such as the ensemble Kalman filter (EnKF) (Pagowski and Grell, 2012) and the hybrid variational-ensemble DA approach (Schwartz et al., 2014) have also been applied to aerosol predictions. All these studies have shown that DA is one of the most effective ways of improving aerosol forecasting through assimilating aerosol observations from multiple sources (e.g. ground-based observations and satellite measurements) to update the chemical ICs.

Numerous studies have used DA approaches to estimate or improve source emissions. The EnKF is one of the most popular DA algorithms used to improve estimates of aerosols and gas-phase emissions, such as NO_x, volatile organic compounds, and SO_2 (van Loon et al., 2000; Heemink and Segers, 2002; Barbu et al., 2009; Miyazaki et al., 2014). Variational DA algorithms have also been applied to constrain emissions of air pollution, such as black carbon, organic carbon, dust, NH_3,

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-732, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 26 August 2016

Atmospheric Chemistry and Physics Discussions

SO_x and NO_x (Hakami et al., 2005; Elbern et al., 2007; Henze et al., 2007, 2009; Yumimoto et al., 2007, 2008; Dubovik et al., 2008; Wang et al., 2012; Guerrette and Henze, 2015). These studies have indicated that DA can efficiently reduce the uncertainty in the emission inventories and lead to improvements in the forecasting of air quality (Mijling and van der A, 2012).

The optimization of chemical ICs and pollution emissions can improve aerosol forecasts and therefore further improvements are likely to be achieved by simultaneously optimizing the chemical ICs and emissions. Tang et al. (2011) reported that the simultaneous adjustment of the ICs of O_3, NO_x and volatile organic compounds and the emissions of NO_x and volatile organic compounds produced overall better performance in both the 1-h and 24-h ozone forecasts than the adjustment of pure ICs or emissions. Miyazaki et al. (2012) reported that the simultaneous adjustment of emissions and concentrations is a powerful approach to correcting the tropospheric ozone budget and profile analyses.

We developed a system to adjust the chemical ICs and source emissions jointly within an EnKF system coupled to the Weather Research and Forecasting with Chemistry (WRF-Chem) model (Grell et al., 2005). We then applied this system to assimilate hourly surface $PM_{2.5}$ measurements over China in early October 2014.

The remainder of the paper is organized as follows. Section 2 describes this DA system in detail. Then the experimental designs are introduced in Section 3. Finally, the surface $PM_{2.5}$ observations assimilation results are presented in section 4 before concluding in section 5.

2. Methodology

2.1 Ensemble square root filter

The ensemble square root filter (EnSRF) algorithm was introduced by Whitaker and Hamill (2002) and its expansion to analyzing aerosol ICs was described by Schwartz et al. (2014). Following the notation of Ide et al. (1997), given an *m*-dimensional background model forecast vector $\mathbf{x}^b$, a *p*-dimensional observation vector $\mathbf{y}^o$ and an operator $\mathbf{H}$ that converts the model state to the observation states, we expressed the

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-732, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 26 August 2016

Atmospheric Chemistry and Physics Discussions

variables as an ensemble mean (denoted by an over-bar) and a deviation from the mean (denoted by a prime). Thus, the ensemble mean $\bar{\mathbf{x}}^{\mathrm{a}}$ of the analyzed state $\mathbf{x}^{\mathrm{a}}$ and the deviations $\mathbf{x}'^{\mathrm{a}}$ from the ensemble mean are updated separately by

$$\bar{\mathbf{x}}^{\mathrm{a}} = \bar{\mathbf{x}}^{\mathrm{b}} + \mathbf{K}(\mathbf{y}^{\mathrm{o}} - \mathbf{H}\bar{\mathbf{x}}^{\mathrm{b}}), \quad (1)$$

$$\mathbf{x}'^{\mathrm{a}} = \mathbf{x}'^{\mathrm{b}} + \widetilde{\mathbf{K}}(\mathbf{y}'^{\mathrm{o}} - \mathbf{H}\mathbf{x}'^{\mathrm{b}}), \quad (2)$$

where $\mathbf{K}$ is the traditional Kalman gain matrix and $\widetilde{\mathbf{K}}$ is the gain used to update the deviations from the ensemble mean. These are given by

$$\mathbf{K} = \mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1} \quad (3)$$

$$\widetilde{\mathbf{K}} = \mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}}\left[\left(\sqrt{\mathbf{H}\mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}} + \mathbf{R}}\right)^{-1}\right]^{\mathrm{T}}\left(\sqrt{\mathbf{H}\mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}} + \mathbf{R}} + \sqrt{\mathbf{R}}\right)^{-1}$$

$$= \left(\mathbf{1} + \sqrt{\mathbf{R}/(\mathbf{H}\mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}} + \mathbf{R})}\right)^{-1}\mathbf{K}, \quad (4)$$

where $\mathbf{P}^{\mathrm{b}}$ is the $m * m$-dimensional background error covariance matrix and $\mathbf{R}$ is the $p * p$-dimensional diagonal observation error covariance matrix. In real applications, $\mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}}$ and $\mathbf{H}\mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}}$ can be approximated using the background ensemble; namely,

$$\mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}} = \frac{1}{N-1}\sum_{i=1}^{N}\mathbf{x}'^{\mathrm{b}}(\mathbf{H}\mathbf{x}'^{\mathrm{b}})^{T} \quad (5)$$

$$\mathbf{H}\mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}} = \frac{1}{N-1}\sum_{i=1}^{N}\mathbf{H}\mathbf{x}'^{\mathrm{b}}(\mathbf{H}\mathbf{x}'^{\mathrm{b}})^{T}. \quad (6)$$

In equations (5) and (6), N is the ensemble size.

Note that for the joint analysis of ICs and emissions, the state vector $\mathbf{x}$ is the joint vector of the mass concentration $\mathbf{C}$ and the emission scaling factor $\boldsymbol{\lambda}$, i.e. $\mathbf{x} = [\mathbf{C}, \boldsymbol{\lambda}]^{\mathrm{T}}$. After each ensemble analysis, the ensemble forecasts were performed with the corresponding models to advance $\mathbf{C}$ and $\boldsymbol{\lambda}$ to the next analysis time. The forecast models are described in section 2.2.

2.2 Forecast model

A forecasting model, $\mathbf{M}$, was developed to forecast the emission scaling factors and the aerosol control variables. This model combines the WRF-Chem model and the forecast model of emission scaling factors.

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-732, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 26 August 2016

Atmospheric Chemistry and Physics Discussions

2.2.1 WRF-Chem model

Version 3.6.1 of the WRF-Chem model (Grell et al., 2005) was used to forecast the aerosol and chemical species. WRF-Chem is an online model with the fully coupled chemical and meteorological components.

Most of the WRF-Chem settings were the same as those reported in Liu et al. (2011): the Goddard Chemistry Aerosol Radiation and Transport (GOCART) aerosol scheme coupled with the Regional Atmospheric Chemistry Mechanism for gaseous chemical mechanisms; the WRF single-moment five-class microphysics scheme; the Rapid Radiative Transfer Model longwave and Goddard shortwave radiation schemes; the Yonsei University (YSU) boundary layer scheme; the Noah land surface model; and the Grell-3D cumulus parameterization.

With respect to the emissions, the hourly prior anthropogenic emissions were based on the monthly regional emission inventory in Asia (Zhang et al., 2009) for the year 2006 interpolated to the model grid (40.5 km) for the lowest eight vertical levels. In order to keep objective for the prior anthropogenic emissions, no time variation was added. Thus, the hourly prior anthropogenic emissions were constant. The biogenic (Guenther et al., 1995), dust (Ginoux et al., 2001), dimethylsulfide and sea salt emissions (Chin et al., 2000, 2002) were calculated online.

2.2.2 Forecast model of scaling factors

As no suitable dynamic model was available to forecast the emission scaling factors, a persistence forecasting operator served as the forecast model for the scaling factors, similar to the method used by Peng et al. (2015) for CO_2 emission inversion. Figure 1a shows the flowchart for the persistence forecasting operator $\mathbf{M}_{\mathrm{SF}}$.

If the ensemble members of the updated chemical fields $\mathbf{C}^{\mathrm{a}}_{i,t-1}$ and the forecast emissions $\mathbf{E}^{\mathrm{f}}_{i,t-2}$ in the previous assimilation cycle are known, then the chemical fields $\mathbf{C}^{\mathrm{f}}_{i,t}$ at time t can be generated via WRF-Chem (Figure 1b). The ensemble concentration ratios $\boldsymbol{\kappa}_{i,t}, (i = 1, \dots, N)$ were then calculated using

$$\boldsymbol{\kappa}_{i,t} = \mathbf{C}^{\mathrm{f}}_{i,t} / \overline{\mathbf{C}^{\mathrm{f}}_{t}}, (i = 1, \dots, N) \quad (7)$$

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-732, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 26 August 2016

Atmospheric Chemistry and Physics Discussions

where $\overline{\mathbf{C}_t^{\mathrm{f}}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{C}_{i,t}^{\mathrm{f}}$ is the ensemble mean of the forecast. $\boldsymbol{\kappa}_{i,t}$ are random variables with a mean values of 1.

The ensemble spreads of $\boldsymbol{\kappa}_{i,t}, (i = 1, \dots, N)$ may be small and therefore covariance inflation was used to maintain them at a certain level:

$$(\boldsymbol{\kappa}_{i,t})_{\mathrm{inf}} = \beta\left(\boldsymbol{\kappa}_{i,t} - \overline{\boldsymbol{\kappa}_t}\right) + \overline{\boldsymbol{\kappa}_t}, (i = 1, \dots, N) \quad (8)$$

$\beta = 1.5$ was chosen in this study. $\overline{\boldsymbol{\kappa}_t}$ is the ensemble mean of $\boldsymbol{\kappa}_{i,t}$, and is equal to 1. As the concentrations were closely related to the emissions and there was no suitable dynamic model available to forecast the emission scaling factors, the inflated concentration ratios $(\boldsymbol{\kappa}_{i,t})_{\mathrm{inf}}$ served as the prior emission scaling factors $\boldsymbol{\lambda}_{i,t}^{\mathrm{p}}$:

$$\boldsymbol{\lambda}_{i,t}^{\mathrm{p}} = (\boldsymbol{\kappa}_{i,t})_{\mathrm{inf}}, (i = 1, \dots, N) \quad (9)$$

To incorporate the useful information from the previous times, the previous DA cycles' analysis scaling factors, $\boldsymbol{\lambda}_{i,t-M+1}^{\mathrm{a}}$, ⋯, $\boldsymbol{\lambda}_{i,t-2}^{\mathrm{a}}$, $\boldsymbol{\lambda}_{i,t-1}^{\mathrm{a}}$ and the prior scaling factor $\boldsymbol{\lambda}_{i,t}^{\mathrm{p}}$ were used to estimate $\boldsymbol{\lambda}_{i,t}^{\mathrm{f}}$ by the time smooth operator; namely,

$$\boldsymbol{\lambda}_{i,t}^{\mathrm{f}} = \frac{1}{M}\left(\sum_{j=t-M+1}^{t-1}\boldsymbol{\lambda}_{i,j}^{\mathrm{a}} + \boldsymbol{\lambda}_{i,t}^{\mathrm{p}}\right), (i = 1, \dots, N, j = t - M + 1, \dots, t - 1) \quad (10)$$

Here, M is the time window of the smooth operator. In this study, a value of $M = 4$ (hours) was chosen.

The ensemble members of the emissions were calculated according to

$$\mathbf{E}_{i,t} = \boldsymbol{\lambda}_{i,t}\mathbf{E}_t^{\mathrm{p}}, (i = 1, \dots, N), \quad (11)$$

where $\mathbf{E}_{i,t}$ is the ith ensemble member of the emissions for each grid at time t, $\boldsymbol{\lambda}_{i,t}$ represents the scaling factors and $\mathbf{E}_t^{\mathrm{p}}$ is the prescribed emission, which can be obtained from the emission inventories.

2.3 Data assimilation system

2.3.1 State variables

The EnSRF algorithm was expanded to optimize the emissions of WRF-Chem.

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-732, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 26 August 2016

Atmospheric Chemistry and Physics Discussions

For the GOCART aerosol scheme, the aerosol species include unspecified P25 (P_{25}), sulfate (S), hydrophobic and hydrophilic organic carbon (OC_1 and OC_2, respectively), hydrophobic and hydrophilic black carbon (BC_1 and BC_2, respectively), dust in five particle size bins (effective radii of 0.5, 1.4, 2.4, 4.5 and 8.0 μm; referred to as D_1, D_2, D_3, D_4 and D_5, respectively) and sea salt in four particle size bins (effective radii of 0.3, 1.0, 3.25 and 7.5 μm for dry air; referred to as S_1, S_2, S_3 and S_4, respectively). The $PM_{2.5}$ observation operator was the same as that described by Schwartz et al. (2012) and expressed as

$$\mathbf{y}^{f} = \boldsymbol{\rho}_{d}[\mathbf{P_{25}} + 1.375\mathbf{S} + 1.8(\mathbf{OC_1} + \mathbf{OC_2}) + \mathbf{BC_1} + \mathbf{BC_2} + \mathbf{D_1} + 0.286\mathbf{D_2} + \mathbf{S_1} + 0.942\mathbf{S_2}], \quad (12)$$

where $\boldsymbol{\rho}_d$ represents the dry air density, which is multiplied by the mixing ratios of aerosol species (in μg kg^{-1}) to convert the units to μg m^{-3} for consistency with the observations. As reported by Schwartz et al. (2012), the state variables of the analysis of the ICs were the 15 WRF-Chem/GOCART aerosol variables.

From the perspective of the optimization of emissions, atmospheric inorganic aerosols are not only from the primary emissions, but also secondary process-chemical and thermodynamic transformations from the gas-phase precursors. Therefore, not only the primary sources of $PM_{2.5}$, but also the sources of the gas-phase precursors, need to be optimized. In this study, the sources of SO_2, NO_x and NH_3 ($\mathbf{E}_{SO2}$, $\mathbf{E}_{NO}$ and $\mathbf{E}_{NH3}$), which have a large impact on the distribution of $PM_{2.5}$, were also optimized in addition to the primary sources of $PM_{2.5}$. Therefore, the four species of emission scaling factors ($\boldsymbol{\lambda}_{PM2.5}$, $\boldsymbol{\lambda}_{SO2}$, $\boldsymbol{\lambda}_{NO}$ and $\boldsymbol{\lambda}_{NH3}$) were considered as the state variables of the DA system in addition to the mass concentration of 15 aerosol variables.

The direct sources of $PM_{2.5}$ include the unspeciated primary sources of $PM_{2.5}$ $\mathbf{E}_{PM2.5}$, sulfate $\mathbf{E}_{SO4}$, nitrate $\mathbf{E}_{NO3}$, organic compounds E_{org} and elemental compounds E_{BC}; all of them are given in two modes (the nuclei and accumulation modes, represented as i and j in the subscripts respectively) The ratios between the nuclei and accumulation modes were the same as in the suggested emission process for National Emission Inventory in WRF-Chem (Freitas et al., 2011). The formula of

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-732, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 26 August 2016

Atmospheric Chemistry and Physics Discussions

sulfate and nitrate emissions in the model are as below:

$$\mathbf{E}_{\mathrm{PM2.5i}}:\mathbf{E}_{\mathrm{PM2.5j}} = 1:4, \quad (13)$$

$$\mathbf{E}_{\mathrm{SO4i}}:\mathbf{E}_{\mathrm{SO4j}} = 1:4, \quad (14)$$

$$\mathbf{E}_{\mathrm{NO3i}}:\mathbf{E}_{\mathrm{NO3j}} = 1:4, \quad (15)$$

$$\mathbf{E}_{\mathrm{SO4i}} + \mathbf{E}_{\mathrm{SO4j}} = a * (\mathbf{E}_{\mathrm{PM2.5i}} + \mathbf{E}_{\mathrm{PM2.5j}} - \mathbf{E}_{\mathrm{EC}} - \mathbf{E}_{\mathrm{ORG}}), \quad (16)$$

$$\mathbf{E}_{\mathrm{NO3i}} + \mathbf{E}_{\mathrm{NO3j}} = b * \left(\mathbf{E}_{\mathrm{PM2.5i}} + \mathbf{E}_{\mathrm{PM2.5j}} - \mathbf{E}_{\mathrm{EC}} - \mathbf{E}_{\mathrm{ORG}}\right), \quad (17)$$

where $\mathbf{E}_{\mathrm{EC}}$ represents elemental carbon and $\mathbf{E}_{\mathrm{ORG}}$ organic compounds, and $a = 0.074$ and $b = 0.038$ were chosen based on the internal emissions and observational data. In the DA process, the first 6 species of direct sources of emissions ($\mathbf{E}_{\mathrm{PM2.5i}}$, $\mathbf{E}_{\mathrm{PM2.5j}}$, $\mathbf{E}_{\mathrm{SO4i}}$, $\mathbf{E}_{\mathrm{SO4j}}$, $\mathbf{E}_{\mathrm{NO3i}}$, and $\mathbf{E}_{\mathrm{NO3j}}$), which may have larger uncertainties in heavy polluted events, were updated according to the variation of $\boldsymbol{\lambda}_{\mathrm{PM2.5}}$. $\mathbf{E}_{\mathrm{PM2.5i}}$ and $\mathbf{E}_{\mathrm{PM2.5j}}$ were directly updated according to the variation in $\boldsymbol{\lambda}_{\mathrm{PM2.5}}$. The emissions ($\mathbf{E}_{\mathrm{SO4i}}$, $\mathbf{E}_{\mathrm{SO4j}}$, $\mathbf{E}_{\mathrm{NO3i}}$ and $\mathbf{E}_{\mathrm{NO3j}}$) were also updated according to the variations in $\mathbf{E}_{\mathrm{PM2.5i}}$ and $\mathbf{E}_{\mathrm{PM2.5j}}$.

2.3.2 Procedure for the DA system

Figure 1 (b) shows the workflow of the DA system. The steps in this workflow are as follows.

(1) The persistence forecasting operator $\mathbf{M}_{\mathrm{SF}}$ is applied to forecast the background fields of the emission scaling factors $\boldsymbol{\lambda}^{\mathrm{f}}_{\mathrm{PM2.5}}$, $\boldsymbol{\lambda}^{\mathrm{f}}_{\mathrm{SO2}}$, $\boldsymbol{\lambda}^{\mathrm{f}}_{\mathrm{NO}}$ and $\boldsymbol{\lambda}^{\mathrm{f}}_{\mathrm{NH3}}$. The forecast chemical fields of P_{25}, SO_2, NO and NH_3 of the previous assimilation cycle are used to create the prior emission scaling factors $\boldsymbol{\lambda}^{\mathrm{p}}_{\mathrm{PM2.5}}$, $\boldsymbol{\lambda}^{\mathrm{p}}_{\mathrm{SO2}}$, $\boldsymbol{\lambda}^{\mathrm{p}}_{\mathrm{NO}}$ and $\boldsymbol{\lambda}^{\mathrm{p}}_{\mathrm{NH3}}$. The background scaling factors are then generated using equation (10).

(2) The ensemble members of the emissions, $\mathbf{E}^{\mathrm{f}}_{\mathrm{PM2.5i}}$, $\mathbf{E}^{\mathrm{f}}_{\mathrm{PM2.5j}}$, $\mathbf{E}^{\mathrm{f}}_{\mathrm{SO2}}$, $\mathbf{E}^{\mathrm{f}}_{\mathrm{NO}}$ and $\mathbf{E}^{\mathrm{f}}_{\mathrm{NH3}}$, are prepared according to equation (11). The corresponding emissions of $\mathbf{E}^{\mathrm{f}}_{\mathrm{SO4i}}$, $\mathbf{E}^{\mathrm{f}}_{\mathrm{SO4j}}$, $\mathbf{E}^{\mathrm{f}}_{\mathrm{NO3i}}$ and $\mathbf{E}^{\mathrm{f}}_{\mathrm{NO3j}}$ are obtained based on equations (14–17).

(3) Forced by the changed emissions ($\mathbf{E}_{\mathrm{PM2.5i}}$, $\mathbf{E}_{\mathrm{PM2.5j}}$, $\mathbf{E}_{\mathrm{SO2}}$, $\mathbf{E}_{\mathrm{NO}}$, $\mathbf{E}_{\mathrm{NH3}}$,

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-732, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 26 August 2016

Atmospheric Chemistry and Physics Discussions

$\mathbf{E}_{SO4i}$, $\mathbf{E}_{SO4j}$, $\mathbf{E}_{NO3i}$ and $\mathbf{E}_{NO3j}$ were substituted by $\mathbf{E}^{f}_{PM2.5i}$, $\mathbf{E}^{f}_{PM2.5j}$, $\mathbf{E}^{f}_{SO2}$, $\mathbf{E}^{f}_{NO}$, $\mathbf{E}^{f}_{NH3}$, $\mathbf{E}^{f}_{SO4i}$, $\mathbf{E}^{f}_{SO4j}$, $\mathbf{E}^{f}_{NO3i}$ and $\mathbf{E}^{f}_{NO3j}$; the other emissions such as $\mathbf{E}_{EC}$ and $\mathbf{E}_{ORG}$ remained unchanged), WRF-Chem is run again to forecast the chemical fields $\boldsymbol{\rho}^{f}$ with the updated chemical fields of the previous assimilation cycle as the ICs. The state variables, i.e., 15 aerosol species and four scaling factors, are then prepared.

(4) The model-simulated $PM_{2.5}$ concentration at the observation space is then calculated via equation (12).

(5) In the assimilation step, the state variables, the concentrations of 14 defined aerosol species and a 15th unspeciated aerosol, and the four species of emission scaling factors $\boldsymbol{\lambda}^{f}_{PM2.5}$, $\boldsymbol{\lambda}^{f}_{SO2}$, $\boldsymbol{\lambda}^{f}_{NO}$ and $\boldsymbol{\lambda}^{f}_{NH3}$, were optimized through EnSRF.

(6) After the assimilation step, the optimized emissions ($\mathbf{E}^{a}_{PM2.5i}$, $\mathbf{E}^{a}_{PM2.5j}$, $\mathbf{E}^{a}_{SO2}$, $\mathbf{E}^{a}_{NO}$, $\mathbf{E}^{a}_{NH3}$, $\mathbf{E}^{a}_{SO4i}$, $\mathbf{E}^{a}_{SO4j}$, $\mathbf{E}^{a}_{NO3i}$ and $\mathbf{E}^{a}_{NO3j}$) were calculated according to equations (11, 14–17) using the optimized scaling factors ($\boldsymbol{\lambda}^{a}_{PM2.5}$, $\boldsymbol{\lambda}^{a}_{SO2}$, $\boldsymbol{\lambda}^{a}_{NO}$ and $\boldsymbol{\lambda}^{a}_{NH3}$).

3. $PM_{2.5}$ observation data and errors

Hourly averaged surface $PM_{2.5}$ observations from the Ministry of Environmental Protection of China were assimilated. Figure 2 shows the locations of 77 measurement sites used for the $PM_{2.5}$ assimilation experiment and forecast verification. The observation sites spanned most of central and eastern China and were primarily located in urban and suburban areas.

The observation error covariance matrix $\mathbf{R}$ in equation (3) includes contributions from measurement and representation errors. Similar to the work of Schwartz et al. (2012), who followed Elbern et al. (2007) and Pagowski et al. (2010), the measurement error ε_0 is defined as $\varepsilon_0 = 1.5 + 0.0075 * \Pi_0$, where Π_0 denotes the observational values for $PM_{2.5}$ ($\mu g\ m^{-3}$). Thus, higher $PM_{2.5}$ values were associated with larger measurement errors. The representativeness error ε_r depends on the resolution of the model and the characteristics of the observation locations and is calculated as $\varepsilon_r = r\varepsilon_0\sqrt{\Delta x/\mathrm{L}}$, where r is an adjustable parameter (here, $r = 0.5$),

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-732, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 26 August 2016

Atmospheric Chemistry and Physics Discussions

Δx is the grid spacing (here, 40.5 km), and L is the radius of influence of an observation (here, L was set to 3 km). The total $PM_{2.5}$ error (ε_t) is defined as $\varepsilon_t = \sqrt{\varepsilon_0{}^2 + \varepsilon_r{}^2}$. The observation errors are assumed to be uncorrelated so that **R** is a diagonal matrix.

The $PM_{2.5}$ observations were subject to quality control to ensure data reliability before DA. $PM_{2.5}$ values larger than 800 μg m^{-3} are classified as unrealistic and were not assimilated; observations with a first guess departure exceeding 100 μg m^{-3} are also omitted.

4. Experimental design

Two parallel experiments were performed to evaluate the impact of $PM_{2.5}$ DA on the analyses and forecasts of aerosols over China: an assimilation experiment and a control experiment. Both experiments used identical WRF-Chem settings and physical parameterizations. The horizontal grid spacing was 40.5 km and there were 57 vertical levels with the model top at 10 hPa.

4.1 Spin-up ensemble forecast with perturbed Initial and boundary conditions

The initialization and spin-up procedures were identical to those reported by Schwartz et al. (2014). The ICs and lateral boundary conditions (LBCs) for the meteorological fields were provided by the National Centers for Environmental Prediction Global Forecast System (GFS).

The initial meteorological fields were created at 0000 UTC 1 October 2014 by interpolating the GFS analyses onto the model domain. The 50 ensemble members were then generated by adding Gaussian random noise with a zero mean and static background error covariances (Torn et al., 2006) to the temperature, water vapor, velocity, geopotential height and dry surface pressure fields. The ICs of each member were zero in the initial aerosol fields, representing clean conditions as described by Liu et al. (2011).

The LBCs for the meteorological fields were then interpolated from the GFS

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-732, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 26 August 2016

Atmospheric Chemistry and Physics Discussions

analyses from 0000 UTC 1 October 2014 to 0000 UTC 16 October 2014 and perturbed similarly to the initial fields at 0000 UTC 1 October 2014. The aerosol LBCs of each member for all experiments represented clean oceanic conditions.

Fifty-member emissions were created by adding standard Gaussian random noise to the anthropogenic emissions, as reported by Schwartz et al. (2014). A 50-member ensemble of four-day WRF-Chem forecasts was then produced using the perturbed ICs at 0000 UTC 1 October 2014, the corresponding perturbed LBCs and the emissions.

4.2 Assimilation experiment

The assimilation experiment was conducted from 0000 UTC 5 October 2014 to 0000 UTC 16 October 2014. The assimilation cycle interval was 1 h. The first initial chemical fields were drawn from the WRF-Chem ensemble forecasts valid at 0000 UTC 5 October 2014, as described in section 4.1. In the subsequent assimilation cycles, the ICs for the chemical variables of each member were drawn from the updated chemical fields of the previous cycle. The aerosol LBCs of each member represented clean oceanic conditions. As for the meteorological ensemble fields, the LBCs were prepared in advance as depicted in section 4.1; the ICs of each member of the meteorological fields were drawn from the forecast meteorological fields of the previous cycle before re-centering with the GFS analysis because we do not do meteorological analysis:

$$\boldsymbol{\pi}_{i_{\mathrm{new}}} = \boldsymbol{\pi}_i + (\boldsymbol{\pi}_{\mathrm{GFS}} - \overline{\boldsymbol{\pi}}), \quad (18)$$

where $\boldsymbol{\pi}_i$ is the *i*th member of the forecast meteorological fields of the previous cycle, $\overline{\boldsymbol{\pi}}$ is the ensemble mean of the forecast meteorological fields of the previous cycle, $\boldsymbol{\pi}_{\mathrm{GFS}}$ is the meteorological field interpolated from the GFS analyses and $\boldsymbol{\pi}_{i_{\mathrm{new}}}$ is the new meteorological field used as the IC in WRF-Chem in the next cycle.

4.3 Control experiment

The control experiment was conducted for the same period as the assimilation

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-732, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 26 August 2016

Atmospheric Chemistry and Physics Discussions

experiment and the simulation cycle period was 1 h, as in the assimilation experiment. The first initial chemical fields were extracted from the ensemble mean valid at 0000 UTC 5 October 2014. In the subsequent simulation process, the ICs for the chemical fields were from the previous cycle's 1-h forecast. The LBCs and ICs for the meteorological fields were updated by interpolating the GFS analyses. The emissions were the prescribed emissions $\mathbf{E}_t^{\mathrm{p}}$ without any perturbation.

5. Results

As the measurement coverage is an important factor that may determine the performance in DA, we primarily focused our attention on the results from three sub-regions with comparatively dense observational coverage (Figure 2): the Beijing–Tianjin–Hebei region (JJJ, 12 stations); the Yangtze River delta (YRD, 24 stations); and the Pearl River delta (PRD, 9 stations).

5.1 Ensemble performance

It is important to assess the ensemble performance for an ensemble-based DA system. In a well-calibrated system, a comparison of the prior ensemble mean root-mean-square error (RMSE) with respect to the observations should equal the prior "total spread" (square root of the sum of ensemble variance and observation error variance) (Houtekamer et al., 2005). Figure 3 shows the time series for the prior ensemble mean RMSE and the total spread for $PM_{2.5}$ aggregated over all observations in the three sub-regions. It indicates that the magnitudes of both the total spread and the RMSE were influenced by the diurnal cycle and heavy air pollution. Almost all the total spreads were smaller than the RMSE, showing an insufficient spread of $PM_{2.5}$ ensemble forecasts, which is especially evident for heavy polluted period with much larger RMSEs.

5.2 Impact on aerosol ICs

To evaluate quantitatively the impact of the ensemble assimilation system on the ICs,

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-732, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 26 August 2016

Atmospheric Chemistry and Physics Discussions

the mean errors (bias), RMSEs and correlation coefficient (CORR) of the assimilation experiment and the control run were first analyzed. These statistics were calculated against observations over all the analyses from 6 to 16 October 2014. Table 1 shows that the bias magnitudes of the control run were 19.1 and 23.5 μg m^{-3} for the YRD and the PRD, respectively, suggesting a significant overestimation of the WRF-Chem aerosol mass in these two sub-regions. However, a significant underestimation of the aerosol mass occurred in the JJJ region, where the model bias was −16.6 μg m^{-3}. The RMSEs of the control run were 76.8, 32.2 and 34.5 μg m^{-3} for the JJJ, YRD and PRD regions, respectively. After assimilation, the statistics showed an apparent improvement and the magnitude of the bias and the RMSE decreased. Both the maximum bias and the RMSE were obtained in the JJJ region, and were -5.2 and 41.6 μg m^{-3}, respectively. The CORR increased from 0.79, 0.60, and 0.62 to 0.93, 0.93, and 0.87 for the JJJ, YRD and PRD, respectively. These results indicate that the initial $PM_{2.5}$ fields can be adjusted efficiently by the EnSRF.

The spatial distribution of the time-averaged differences in $PM_{2.5}$ (assimilation minus control) illustrated the impact on the aerosol ICs. Figure 4 shows that the average difference in $PM_{2.5}$ at the lowest model level was negative in the YRD, the PRD and in central China, indicating the reduction of the overestimation of the WRF-Chem simulation over these regions with data assimilation. The smallest value of −40 μg m^{-3} was seen around Guangzhou. Conversely, positive differences were seen in the JJJ region and in northeast China. The largest value (>40 μg m^{-3}) was seen in South Hebei Province. These results indicate that DA greatly improved the ICs.

5.3 Impact on emissions

To determine the impact of assimilating $PM_{2.5}$ observations on the chemical emissions, we analyzed the area-averaged time series extracted from the optimized emission scaling factors and the optimized emissions. Figure 5 shows that although the prior emissions had no diurnal variation when the experiments were designed, the optimized $PM_{2.5}$ scaling factor, $\boldsymbol{\lambda}^{a}_{PM2.5}$, showed an obvious variation with time, as did the optimized unspeciated primary sources of $PM_{2.5}$, $\mathbf{E}^{a}_{PM2.5}$. Moreover, the values of

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-732, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 26 August 2016

Atmospheric Chemistry and Physics Discussions

$\boldsymbol{\lambda}^{a}_{PM2.5}$ were <1 at almost all times in the YRD and PRD, which resulted that the analyzed emission $\mathbf{E}^{a}_{PM2.5}$ were lower than the prior $PM_{2.5}$ emissions $\mathbf{E}^{p}_{PM2.5}$. In the YRD, the prior $\mathbf{E}^{p}_{PM2.5}$ was about 0.127 μg m^{-2} s^{-1} over all hours. After assimilation, the time-averaged optimized $\mathbf{E}^{a}_{PM2.5}$ decreased to 0.107 μg m^{-2} s^{-1}, about 15.6% lower than the prior value. In the PRD, the prior $\mathbf{E}^{p}_{PM2.5}$ was about 0.10 μg m^{-2} s^{-1}. The time-averaged optimized $\mathbf{E}^{a}_{PM2.5}$ decreased to 0.066 μg·m^{-2} s^{-1}, leading to a decrease of 35.0%. However, larger values for the optimized $\mathbf{E}^{a}_{PM2.5}$ were obtained in the JJJ region in three periods, from 1600 UTC 6 October to 0000 UTC 8 October, from 1600 UTC 9 October to 0000 UTC 10 October, and from 1600 UTC 13 October to 0000 UTC 15 October as a result of the increased optimized scaling factor $\boldsymbol{\lambda}^{a}_{PM2.5}$. This may have been caused by the burning of crop residues during harvesting in this region (Li et al., 2016), which was not taken into account in the prior emissions. Although the system is able to detect the emission changes caused by burning events, the time that the system started to show increased scaling factors might be not accurate enough (may shift a few hours later); as the system is optimized based on ambient concentrations in which the transport and transformation processes are not directly taken into account.

The NO, SO_2 and NH_3 emissions were all adjusted to some extent by our DA approach. The NO emissions increased by 41.3, 43.7 and 20.3% in the JJJ, YRD and PRD regions, respectively. The SO_2 emissions increased by 16.3, 10.0 and 18.3% and the NH_3 emissions increased by 16.7, 7.8 and 7.5% in the JJJ, YRD and PRD regions, respectively.

Figure 6 shows the spatial distribution of the time-averaged scaling factors $\boldsymbol{\lambda}^{a}_{PM2.5}$ at the lowest model level over all hours from 6 to 16 October 2014. Figure 7 shows the distribution of $\mathbf{E}^{p}_{PM2.5}$ and the time-averaged differences between the ensemble mean of the assimilation and the prior values. These patterns are consistent with those in Figure 5. Negative differences were obtained in most areas of the YRD and PRD, indicating that the $PM_{2.5}$ DA primarily decreased the $PM_{2.5}$ emissions.

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-732, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 26 August 2016

Atmospheric Chemistry and Physics Discussions

393 Conversely, positive differences were obtained in South Hebei, North Henan and
394 Southeast Shanxi provinces, indicating that DA increased the $PM_{2.5}$ emissions.
395 As the economy in China has developed, the spatiotemporal distribution of
396 emissions has changed as a result of changes in energy consumption, the structure of
397 the energy market and advances in technology. Therefore although this inventory of
398 emissions may have correctly described anthropogenic emissions in 2006 when it was
399 constructed, it is not representative of the anthropogenic emissions in 2014.
400 Theoretically, the assimilated emissions should reduce the uncertainty in the prior
401 emissions as a result of the application of observations. The diurnal variation in the
402 assimilated emissions verified this statement to some extent. In addition, Liu et al.
403 (2015) reported that $PM_{2.5}$ and SO_2 emissions in China decreased from 2006 to 2010,
404 whereas NO_x emissions increased over the same time. Besides, Xia et al. (2016) also
405 reported that NO_x emissions increased from 2000 to 2011 then decreased slowly from
406 2012 to 2014. Our assimilated $PM_{2.5}$ and NO_x emissions were in good agreement with
407 this trend, but not the SO_2 emissions. One possible reason for this may be that only
408 surface $PM_{2.5}$ observations were applied in this work, which may have less constraint
409 on the sources of the secondary aerosol precursors, such as SO_2 emissions.
410 Nevertheless, although we had no direct emission observations to use as a reference,
411 we concluded that the assimilated $PM_{2.5}$ emissions were a better reflection of the
412 distribution of the spatiotemporal variations in the real discharge than the emissions
413 inventory. More observations are needed to obtain reliable emissions for the sources
414 of the gas-phase precursors.
415
416 5.4 Verification of aerosol forecasting
417 For the assimilation experiment, 48-h forecasts were performed at each 0000
418 UTC from 6 to 16 October 2014 with the hourly forecast output. The ensemble mean
419 of the analyzed ICs and emissions was used in this longer-range model forecast.
420 Time series of the hourly $PM_{2.5}$ extracted from the analysis (AN), the control run
421 (CT) and the hourly output of 48-h forecast (fc24 for the first day forecast and fc48
422 for the second day forecast) were compared with the observations (OBS) for three

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-732, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 26 August 2016

megacities Beijing, Shanghai and Guangzhou, respectively (Figure 8). As expected, the time series of the analysis were consistent with the observations. The control run showed large deviations from the observations, especially in Shanghai and Guangzhou. Benefit from DA on both the first day and the second day forecasts can be clearly seen.

The RMSE of the surface $PM_{2.5}$ forecasts as a function of forecast range was then calculated against the observations for the three sub-regions (Figure 9). The RMSEs of the control run were characterized by the diurnal cycle in the YRD and PRD. The largest errors were seen at 2100 UTC in the YRD (about 43 $\mu g \cdot m^{-3}$) and at 2200 UTC in the PRD (about 46 $\mu g \cdot m^{-3}$), likely indicating significant systematic forecast errors at these times. From 0300 to 0900 UTC, the RMSE values (about 15 $\mu g \cdot m^{-3}$) were much smaller than at other times in both the YRD and PRD, showing that WRF-Chem performed well during this period. However, in the JJJ region, the RMSEs (about 50 $\mu g \cdot m^{-3}$) were always large as a result of a heavy pollution event. After assimilation, the RMSEs decreased sharply. They remained at about 10 $\mu g \cdot m^{-3}$ for all three sub-regions during the whole experimental period, further indicating that DA greatly affected the ICs.

The improvements in the surface $PM_{2.5}$ forecasts by the joint adjustment of the ICs and emissions were dramatic in the YRD and PRD. Large reduction of the RMSEs due to assimilation can be seen for almost the entire 48-h forecast range. From 10- to 23-h and from 34- to 47-h, in particular, the relative reduction in RMSE was about 40%. However, the DA impact was much smaller for 3- to 9-h forecast ranges, which are at daytime of the first day forecast. This may be because WRF-Chem performed sufficiently well during this period and therefore the further improvement was more difficult. From the perspective of the DA impact, the differences between the optimized $PM_{2.5}$ emissions and the prior emissions from 0000 to 0700 UTC each day were always smaller than those for other periods. In addition, the improvements were nearly negligible from 27- to 33-h, the daytime of the second day forecast, suggesting that the benefit gained from adjusting the ICs decreased progressively and eventually disappeared with model integration. Nevertheless,

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-732, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 26 August 2016

attributed greatly to the large adjustment of chemical emissions, substantial improvements were still achieved from 34- to 47- h. These results revealed that joint adjustment of the ICs and emissions can improve surface $PM_{2.5}$ forecasts up to 48 h in the YRD and PRD.

The DA system did not perform as well in the JJJ region as in the YRD and RRD and relatively smaller improvements were achieved in the first 24-h forecast. One possible reason for this result may be systematic errors due to chemistry mechanism in WRF-Chem. The sources of the aerosols are so complex that our knowledge of their formation mechanisms is far from clear and large uncertainties still exist in the model simulations. Chemical transport models have a tendency to underestimate PM concentrations, especially during episodes of heavy pollution (Denby et al., 2007) due to some missing reactions (Wang et al., 2014; Zhang et al., 2015, Zheng et al., 2015; Chen et al., 2016). As a result, a large bias may be obtained in forecasts of heavy pollution given the ICs and emission inventories achieved from the joint assimilation. Another reason may be the sparse coverage of measurements. There were only 12 sites in the JJJ region (Figure 2) and the measurement coverage was much sparser than in the YRD or PRD. However, these results are still better than those obtained with the pure adjustment of ICs that lead to improvements in the first 12-h forecasts (Jiang et al., 2013; Schwartz et al., 2014).

6. Summary

The EnSRF algorithm was extended to adjust the chemical ICs and the primary and precursor emissions to improve forecasts for surface $PM_{2.5}$. This system was applied to assimilate hourly surface $PM_{2.5}$ measurements from 5 to 16 October 2014 over China. To evaluate the effectiveness of DA, 48-h forecasts were performed using the optimized ICs and emissions, together with a control experiment without DA. The results indicated that the forecasts with the optimized ICs and emissions performed much better than the control simulations. Large improvements were achieved for almost all the 48-h forecasts, particularly in the YRD and PRD. However, relatively smaller improvements were achieved in the first 24-h forecast in the JJJ region, which

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-732, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 26 August 2016

Atmospheric Chemistry and Physics Discussions

may be attributed to the sparse measurement coverage and the deficiencies in the model system for forecasting heavy pollution.

This study represents the first step in the simultaneous optimization of chemical ICs and emissions and only surface $PM_{2.5}$ measurements were assimilated. In future work, gas-phase observations of SO_2, NO_2 and CO will be used to further improve the performance of this DA system.

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-732, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 26 August 2016

Atmospheric Chemistry and Physics Discussions

References

Adhikary, B., Kulkarni, S., Dallura, A., Tang, Y., Chai, T., Leung, L. R., Qian, Y., Chung, C. E., Ramanathan, V., and Carmichael, G. R.: A regional scale chemical transport modeling of Asian aerosols with data assimilation of AOD observations using optimal interpolation technique, Atmos. Environ., 42, 8600–8615, doi:10.1016/j.atmosenv.2008.08.031, 2008.

Barbu, A. L., Segers, A. J., Schaap, M., Heemink, A.W., and Builtjes, P. J. H.: A multi-component data assimilation experiment directed to sulphur dioxide and sulphate over Europe, Atmos. Environ., 43, 1622–1631, 2009.

Benedetti, A., Morcrette, J., Boucher, O., Dethof, A., Engelen, R., Fisher, M., Flentje, H., Huneeus, N., Jones, L., and Kaiser, J.: Aerosol analysis and forecast in the European Centre for Medium-Range Weather Forecasts Integrated Forecast System: 2. Data assimilation, J. Geophys. Res, 114, D13205, doi:10.1029/2008JD011115, 2009.

Chen, D., Liu, Z., Fast, J., and Ban, J.: Simulations of Sulfate-Nitrate-Ammonium (SNA) aerosols during the extreme haze events over Northern China in October 2014, Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-222, in review, 2016.

Chin, M., Rood, R. B., Lin, S. J., Muller, J. F., and Thompson, A. M.: Atmospheric sulfur cycle simulated in the global model GOCART: Model description and global properties, J. Geophys. Res.-Atmos., 105, 24671–24687, 2000.

Chin, M., Ginoux, P., Kinne, S., Torres, O., Holben, B.N., Duncan, B. N., Martin, R.V., Logan, J.A., Higurashi, A., and Nakajima, J.: Tropospheric aerosol optical thickness from the GOCART model and comparisons with satellite and Sun photometer measurements, J. Atmos. Sci., 59(3), 461–483, 2002.

Collins, W. D., Rasch, P. J., Eaton, B. E., Khattatov, B. V., and J.-F. Lamarque, J.-F.: Simulating aerosols using a chemical transport model with assimilation of satellite aerosol retrievals: Methodology for INDOEX, J. Geophys. Res., 106, 7313–7336, 2001.

Denby, B., Schaap, M., Segers, A.J., Builtjes, P.J.H., Horalek, J.: Comparison of two data assimilation methods for assessing PM10 exceedances on the European scale, Atmos. Environ., 42 (30), 7122–7134, 2007.

Dubovik, O., Lapyonok, T., Kaufman, Y. J., Chin, M., Ginoux, P., Kahn, R. A., and Sinyuk, A.: Retrieving global aerosol sources from satellites using inverse modeling, Atmos. Chem.

520 Phys., 8, 209–250, doi:10.5194/acp-8-209-2008, 2008
521 Elbern, H., Strunk, A., Schmidt, H., and Talagrand, O.: Emission rate and chemical state
522 estimation by 4-dimensional variational inversion, Atmos. Chem. Phys., 7, 3749–3769,
523 doi:10.5194/acp- 7-3749-2007, 2007.
524 Freitas, S. R.; Longo, K. M.; Alonso, M. F.; Pirre, M.; Marecal, V.; Grell, G.; Stockler, R.; Mello,
525 R. F.; Sánchez Gácita, M.. PREP-CHEM-SRC 1.0: a preprocessor of trace gas and aerosol
526 emission fields for regional and global atmospheric chemistry models. Geoscientific Model
527 Development, v. 4, p. 419-433, 2011.
528 Ginoux, P., Chin, M. Tegen, I., Prospero, J. M., Holben, B., Dubovik, O., and Lin, S.-J.: Sources
529 and distributions of dust aerosols simulated with the GOCART model, J. Geophys. Res., 106,
530 20,255–20,273, doi:10.1029/2000JD000053, 2001.
531 Grell, G., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock, W. C., and Eder, B.:
532 Fully coupled "online" chemistry within the WRF model, Atmos. Environ., 39, 6957–6975,
533 doi:10.1016/j.atmosenv.2005.04.027, 2005.
534 Guenther, A., Hewitt, C. N., Erickson, D., Fall, R., Geron, C., Graedel, T., Harley, P., Klinger, L.,
535 Lerdau, M., McKay, W., Pierce, T., Scholes, B., Steinbrecher, R., Tallamraju, R., Taylor, J.,
536 and Zimmerman, P.: A global model of natural volatile organic compound emissions, J.
537 Geophys. Res., 100, 8873–8892, doi:10.1029/94JD02950, 1995.
538 Guerrette, J. J. and Henze, D. K.: Development and application of the WRFPLUS-Chem online
539 chemistry adjoint and WRFDA-Chem assimilation system, Geosci. Model Dev., 8,
540 1857-1876, doi:10.5194/gmd-8-1857-2015, 2015.
541 Hakami, A., Henze, D. K., Seinfeld, J. H., Chai, T., Tang, Y., Carmichael, G. R., and Sandu, A.:
542 Adjoint inverse modeling of black carbon during the Asian Pacific Regional Aerosol
543 Characterization Experiment, J. Geophys. Res.-Atmos., 110, D14301,
544 doi:10.1029/2004JD005671, 2005.
545 Heemink, A.W., and Segers, A.J.: Modeling and prediction of environmental data in space and
546 time using Kalman filtering, Stoch. Environ. Res. Risk Assess. 16 (3), 225–240, 2002.
547 Henze, D. K., Hakami, A., and Seinfeld, J. H.: Development of the adjoint of GEOS-Chem,
548 Atmos. Chem. Phys., 7, 2413–2433, doi:10.5194/acp-7-2413-2007, 2007.
549 Henze, D. K., Seinfeld, J. H., and Shindell, D. T.: Inverse modeling and mapping US air quality

influences of inorganic PM2.5 precursor emissions using the adjoint of GEOS-Chem, Atmos. Chem. Phys., 9, 5877–5903, doi:10.5194/acp-9-5877-2009, 2009.

Houtekamer, P. L., Mitchell, H. L., Pellerin, G., Buehner, M., Charron, M., Spacek, L., and Hansen, B.: Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations, Mon. Weather Rev., 133, 604–620, 2005.

Ide, K., Courtier, P., Ghil, M., and Lorenc, A. C.: Unified notation for data assimilation: operational, sequential and variational, J. Meteorol. Soc. Japan, 75, 181–189, 1997.

Jiang, Z., Liu, Z., Wang, T., Schwartz, C. S., Lin, H.-C., and Jiang, F.: Probing into the impact of 3DVAR assimilation of surface PM10 observations over China using process analysis, J. Geophys. Res. Atmos., 118, 6738–6749, doi:10.1002/jgrd.50495, 2013.

Kahnert, M.: Variational data analysis of aerosol species in a regional CTM: Background error covariance constraint and aerosol optical observation operators, Tellus, Ser. B, 60, 753–770, doi:10.1111/j.1600-0889.2008.00377, 2008.

Kleist, D. T., Parrish, D. F., Derber, J. C., Treadon, R., Wu, W.-S., and Lord, S.: Introduction of the GSI into the NCEP global data assimilation system, Weather Forecast., 24, 1691–1705, 2009.

Lee, E.-H., Ha, J.-C., Lee, S.-S., and Chun, Y.: PM10 data assimilation over South Korea to Asian dust forecasting model with the optimal interpolation method, Asia-Pacific J. Atmos. Sci., 49(1), 73–85, doi:10.1007/s13143-013-0009-y, 2013.

Li, Z., Zang, Z., Li, Q. B., Chao, Y., Chen, D., Ye, Z., Liu, Y., and Liou, K. N.: A three-dimensional variational data assimilation system for multiple aerosol species with WRF/Chem and an application to $PM_{2.5}$ prediction, Atmos. Chem. Phys., 13, 4265-4278, doi:10.5194/acp-13-4265-2013, 2013.

Li, J., Li, Y., Bo, Y., and Xie, S.: High-resolution historical emission inventories of crop residue burning in fields in China for the period 1990–2013, Atmos. Environ., 138, 152–161, 2016.

Liu, Z., Liu, Q., Lin, H. C., Schwartz, C. S., Lee, Y. H., and Wang, T.: Three-dimensional variational assimilation of MODIS aerosol optical depth: implementation and application to a dust storm over East Asia, J. Geophys. Res., 116, D23206, doi:10.1029/2011JD016159, 2011.

Liu, F., Zhang, Q., Tong, D., Zheng, B., Li, M., Huo, H., and He, K. B.: High-resolution inventory

580 of technologies, activities, and emissions of coal-fired power plants in China from 1990 to
581 2010, Atmos. Chem. Phys., 15, 13299-13317, doi:10.5194/acp-15-13299-2015, 2015.
582 Mijling, B. and van der A, R. J.: Using daily satellite observations to estimate emissions of
583 short-lived air pollutants on a mesoscopic scale, J. Geophys. Res., 117, D17302,
584 doi:10.1029/2012JD017817, 2012.
585 Miyazaki, K., Eskes, H. J., Sudo, K., Takigawa, M., van Weele, M., and Boersma, K. F.:
586 Simultaneous assimilation of satellite NO_2, O_3, CO, and HNO_3 data for the analysis of
587 tropospheric chemical composition and emissions, Atmos. Chem. Phys., 12, 9545– 9579,
588 doi:10.5194/acp-12-9545-2012, 2012.
589 Miyazaki, K., Eskes, H. J., Sudo, K., and Zhang, C.: Global lightning NOx production estimated
590 by an assimilation of multiple satellite data sets, Atmos. Chem. Phys., 14, 3277–3305,
591 doi:10.5194/acp-14-3277-2014, 2014.
592 Pagowski, M., Grell, G. A., McKeen, S. A., Peckham, S. E., and Devenyi, D.: Three-dimensional
593 variational data assimilation of ozone and fine particulate matter observations: some results
594 using the Weather Research and Forecasting – Chemistry model and Grid-point Statistical
595 Interpolation, Q. J. Roy. Meteor. Soc., 136, 2013–2024, doi:10.1002/qj.700, 2010.
596 Pagowski, M., and Grell, G. A.: Experiments with the assimilation of fine aerosols using an
597 ensemble Kalman filter, J. Geophys. Res.-Atmos., 117, D21302, doi:10.1029/2012jd018333,
598 2012.
599 Peng, Z., Zhang, M., Kou, X., Tian, X., and Ma, X.: A regional carbon data assimilation system
600 and its preliminary evaluation in East Asia, Atmos. Chem. Phys., 15, 1087-1104,
601 doi:10.5194/acp-15-1087-2015, 2015.
602 Pope, C. A.: Review: Epidemiological basis for particulate air pollution health standards, Aerosol
603 Sci. Tech., 32, 4–14, 2000.
604 Pope, C. A., Burnett, R. T., Thun, M. J., Calle, E. E., Krewski, D., Ito, K., and Thurston, G. D.:
605 Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air
606 pollution, J. Am. Med. Assoc., 287, 1132–1141, 2002.
607 Saide, P. E., Carmichael, G. R., Liu, Z., Schwartz, C. S., Lin, H. C., da Silva, A. M., and Hyer, E.:
608 Aerosol optical depth assimilation for a size-resolved sectional model: impacts of
609 observationally constrained, multi-wavelength and fine mode retrievals on regional scale

610 analyses and forecasts, Atmos. Chem. Phys., 13, 10425-10444,
611 doi:10.5194/acp-13-10425-2013, 2013.
612 Schwartz, C. S., Liu, Z., Lin, H. C., and McKeen, S. A.: Simultaneous three-dimensional
613 variational assimilation of surface fine particulate matter and MODIS aerosol optical depth, J.
614 Geophys. Res., 117, D13202, doi:10.1029/2011JD017383, 2012.
615 Schwartz, C. S., Liu, Z., Lin, H.-C., and Cetola, J. D.: Assimilating aerosol observations with a
616 "hybrid" variational-ensemble data assimilation system, J. Geophys. Res. Atmos., 119, 4043–
617 4069, doi:10.1002/2013JD020937, 2014.
618 Tang, X., Zhu, J., Wang, Z. F., and Gbaguidi, A.: Improvement of ozone forecast over Beijing
619 based on ensemble Kalman filter with simultaneous adjustment of initial conditions and
620 emissions, Atmos. Chem. Phys., 11, 12901–12916, doi:10.5194/acp-11-12901-2011, 2011.
621 Tombette, M., Mallet, V., and Sportisse, B.: PM10 data assimilation over Europe with the optimal
622 interpolation method, Atmos. Chem. Phys., 9, 57-70, doi:10.5194/acp-9-57-2009, 2009.
623 Torn, R. D., Hakim, G. J., and Snyder, C.: Boundary conditions for limited-area ensemble Kalman
624 filters, Mon. Weather Rev., 134, 2490–2502, 2006.
625 van Loon, M., Builtjes, P. J. H., and Segers, A. J.: Data assimilation of ozone in the atmospheric
626 transport chemistry model LOTOS, Environ. Model. Softw., 15, 603–609, 2000.
627 Wang, J., Xu, X., Henze, D. K., Zeng, J., Ji, Q., Tsay, S.-C., and Huang, J.: Top-down estimate of
628 dust emissions through integration of MODIS and MISR aerosol retrievals with the
629 GEOS-Chem adjoint model, Geophys. Res. Lett., 39, L08802, doi:10.1029/2012GL051136,
630 2012.
631 Wang, Y. X., Zhang, Q. Q., Jiang, J. K., Zhou, W., Wang, B. Y., He, K. B., Duan, F. K., Zhang,
632 Q., Philip, S., and Xie, Y. Y.: Enhanced sulfate formation during China's severe winter haze
633 episode in January 2013 missing from current models, J.Geophys.Res.-Atmos., 119,
634 10.1002/2013JD021426, 2014
635 Whitaker, J. S., and Hamill, T. M.: Ensemble data assimilation without perturbed observations,
636 Mon. Weather Rev., 130, 1913–1924, 2002.
637 Xia Y., Zhao, Y., Nielsen, C.P., Benefits of China's efforts in gaseous pollutant control indicated
638 by the bottom-up emissions and satellite observations 2000-2014, Atmospheric Environment,
639 136, 43-53, 2016

640 Yu, H., Dickinson, R. E., Chin, M., Kaufman, Y. J., Geogdzhayev, B., and Mishchenko, M. I.:
641 Annual cycle of global distributions of aerosol optical depth from integration of MODIS
642 retrievals and GOCART model simulations, J. Geophys. Res., 108(D3), 4128,
643 doi:10.1029/2002JD002717, 2003.
644 Yumimoto, K., Uno, I., Sugimoto, N., Shimizu, A., and Satake, S.: Adjoint inverse modeling of
645 dust emission and transport over East Asia, Geophys. Res. Lett., 34, L00806,
646 doi:10.029/2006GL028551, 2007.
647 Yumimoto, K., Uno, I., Sugimoto, N., Shimizu, A., Liu, Z., and Winker, D. M.: Adjoint inversion
648 modeling of Asian dust emission using lidar observations, Atmos. Chem. Phys., 8,
649 2869-2884, doi:10.5194/acp-8-2869-2008, 2008.
650 Zhang, J., Reid, J. S., Westphal, D., Baker, N., and Hyer, E.: A System for Operational Aerosol
651 Optical Depth Data Assimilation over Global Oceans, J. Geophys. Res., 113, D10208,
652 doi:10.1029/2007JD009065, 2008.
653 Zhang, Q., Streets, D. G., Carmichael, G. R., He, K. B., Huo, H., Kannari, A., Klimont, Z., Park, I.
654 S., Reddy, S., Fu, J. S., Chen, D., Duan, L., Lei, Y., Wang, L. T., and Yao, Z. L.: Asian
655 emissions in 2006 for the NASA INTEX-B mission, Atmos. Chem. Phys., 9, 5131-5153,
656 doi:10.5194/acp-9-5131-2009, 2009.
657 Zhang, L., Liu, L. C., Zhao, Y. H., Gong, S. L., Zhang, X. Y., Henze, D. K., Capps, S. L., Fu, T.
658 M., Zhang, Q., and Wang, Y. X.: Source attribution of particulate matter pollution over
659 North China with the adjoint method, Environ.Res.Lett., 10, Artn
660 08401110.1088/1748-9326/10/8/084011, 2015.
661 Zheng, B., Zhang, Q., Zhang, Y., He, K. B., Wang, K., Zheng, G. J., Duan,
662 F. K., Ma, Y. L., and Kimoto, T.: Heterogeneous chemistry: a mechanism
663 missing in current models to explain secondary inorganic aerosol formation
664 during the January 2013 haze episode in North China, Atmos.Chem.Phys., 15, 2031-2049,
665 10.5194/acp-15-2031-2015, 2015.
666
667

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-732, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 26 August 2016

Atmospheric Chemistry and Physics Discussions

668 **List of Figures and Table**
669
670 Figure 1. (a) Framework of $\mathbf{M}_{SF}$ and (b) flow chart of the data assimilation system
671 that simultaneously optimizes the chemical initial conditions and emissions.
672
673 Figure 2. Locations of 77 $PM_{2.5}$ observation stations in the model domain. The three
674 colored boxes mark sub-regions with relatively dense coverage for the Beijing–
675 Tianjin–Hebei region (JJJ, 12 stations, red), the Yangtze River delta (YRD, 24
676 stations, blue) and the Pearl River delta (PRD, 9 stations, green).
677
678 Figure 3. Time series of prior ensemble mean RMSE and total spread for $PM_{2.5}$
679 concentrations aggregated over all observations over the three sub-regions: (a)
680 Beijing–Tianjin–Hebei region; (b) Yangtze River delta; and (c) Pearl River delta.
681
682 Table 1. Comparison of the surface $PM_{2.5}$ mass concentrations from the control and
683 assimilation experiments to observations over all analysis times from 6 to 16 October
684 2014.
685
686 Figure 4. $PM_{2.5}$ mass differences (assimilation minus control, $\mu g\ m^{-3}$) at the lowest
687 model level averaged over all hours from 6 to 16 October 2014.
688
689 Figure 5. Hourly area-averaged time series of emission scaling factors (black)
690 extracted from the ensemble mean of the analyzed $\boldsymbol{\lambda}^{a}_{PM2.5}$ and the corresponding
691 analyzed unspeciated primary $PM_{2.5}$ emissions $\mathbf{E}^{a}_{PM2.5}$ (blue) over the three
692 sub-regions: (a) Beijing–Tianjin–Hebei region; (b) Yangtze River delta; and (c) Pearl
693 River delta.
694
695 Figure 6. Spatial distribution of $\boldsymbol{\lambda}_{PM2.5}$ at the lowest model level averaged over all
696 hours from 6 to 16 October 2014.
697
698 Figure 7. Spatial distribution of (a) the prior unspeciated primary sources of $PM_{2.5}$
699 ($\mu g\ m^{-2}\ s^{-1}$) and (b) the time-averaged differences between the ensemble mean
700 analyses and the prior values ($\mu g \cdot m^{-2}\ s^{-1}$) at the lowest model level averaged over all
701 hours from 6 to 16 October 2014.
702
703 Figure 8. Time series of the hourly $PM_{2.5}$ obtained from observations (circle), analysis
704 (blue line), control run (black line) and hourly output of 48-h forecast in three
705 megacities: (a) Beijing; (b) Shanghai; and (c) Guangzhou. See text in section 5.4.
706
707 Figure 9. RMSE of surface $PM_{2.5}$ as a function of forecast range calculated against
708 observations over the three sub-regions: (a) Beijing–Tianjin–Hebei region; (b)
709 Yangtze River delta; and (c) Pearl River delta. (d) Normalized RMSE (assimilation
710 divided by control).
711

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-732, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 26 August 2016

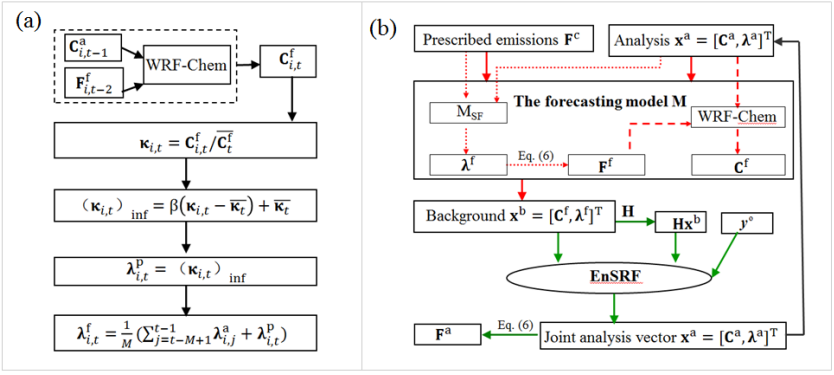


Figure 1. (a) Framework of $\mathbf{M}_{SF}$ and (b) flow chart of the data assimilation system that simultaneously optimizes the chemical initial conditions and emissions.

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-732, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 26 August 2016

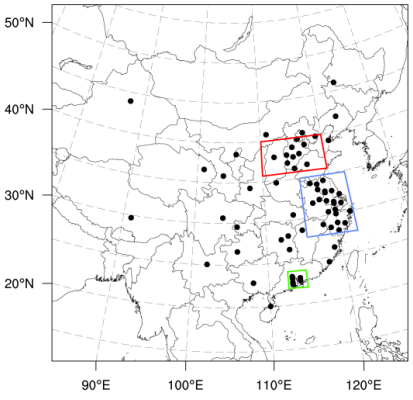
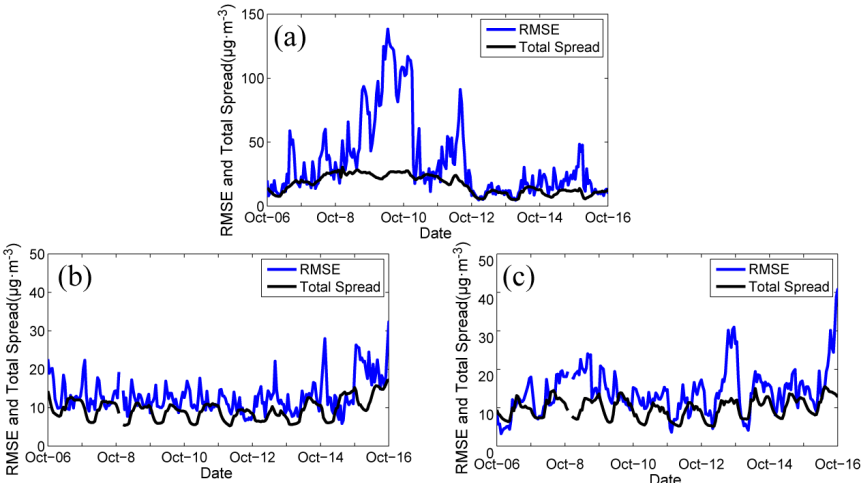


Figure 2. Locations of 77 $PM_{2.5}$ observation stations in the model domain. The three colored boxes mark sub-regions with relatively dense coverage for the Beijing–Tianjin–Hebei region (JJJ, 12 stations, red), the Yangtze River delta (YRD, 24 stations, blue) and the Pearl River delta (PRD, 9 stations, green).

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-732, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 26 August 2016

Atmospheric Chemistry and Physics Discussions

725



726

727 Figure 3. Time series of prior ensemble mean RMSE and total spread for $PM_{2.5}$

728 concentrations aggregated over all observations over the three sub-regions: (a)

729 Beijing–Tianjin–Hebei region; (b) Yangtze River delta; and (c) Pearl River delta.

730

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-732, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 26 August 2016

Atmospheric Chemistry and Physics Discussions

731
732 Table 1. Comparison of the surface $PM_{2.5}$ mass concentrations from the control and
733 assimilation experiments to observations over all analysis times from 6 to 16 October
734 2014.

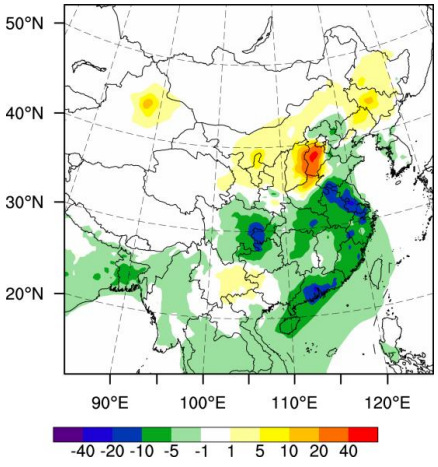
Region	Experiment	Mean observed value	Mean simulated value	BIAS	RMSE	CORR
Beijing–Tianjin–Hebei	Control	117.7	101.1	−16.6	76.8	0.785
	Assimilation		112.5	−5.2	41.6	0.932
Yangtze River delta	Control	47.9	67.0	19.1	32.2	0.603
	Assimilation		49.0	1.1	10.5	0.928
Pearl River delta	Control	63.7	87.2	23.5	34.5	0.618
	Assimilation		66.1	2.4	12.9	0.866

735
736

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-732, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 26 August 2016

Atmospheric Chemistry and Physics Discussions

737



738

739 Figure 4. $PM_{2.5}$ mass differences (assimilation minus control, μg m^{-3}) at the lowest
740 model level averaged over all hours from 6 to 16 October 2014.
741

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-732, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 26 August 2016

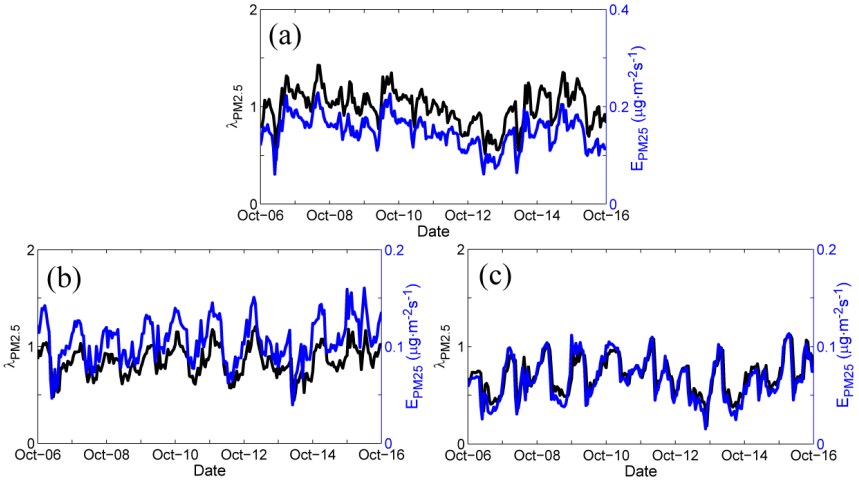


Figure 5. Hourly area-averaged time series of emission scaling factors (black) extracted from the ensemble mean of the analyzed $\boldsymbol{\lambda}^{a}_{PM2.5}$ and the corresponding analyzed unspeciated primary $PM_{2.5}$ emissions $\mathbf{E}^{a}_{PM2.5}$ (blue) over the three sub-regions: (a) Beijing–Tianjin–Hebei region; (b) Yangtze River delta; and (c) Pearl River delta.

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-732, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 26 August 2016

Atmospheric Chemistry and Physics Discussions

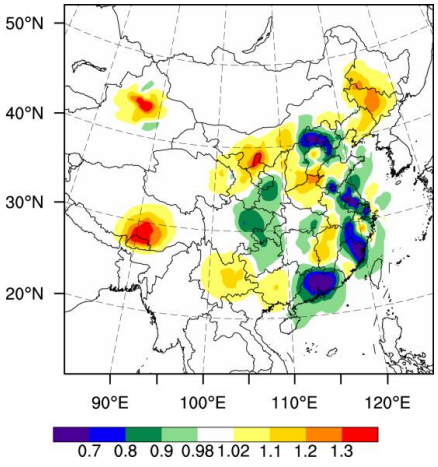


Figure 6. Spatial distribution of $\boldsymbol{\lambda}_{PM2.5}$ at the lowest model level averaged over all hours from 6 to 16 October 2014.

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-732, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 26 August 2016

Atmospheric Chemistry and Physics Discussions

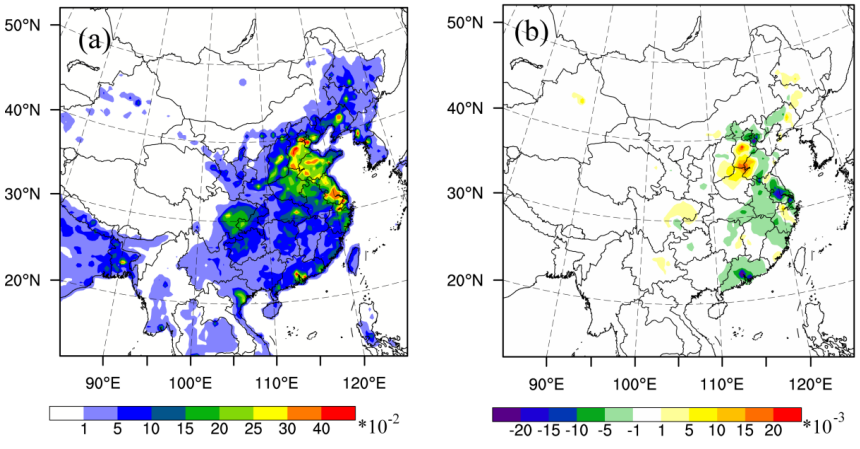
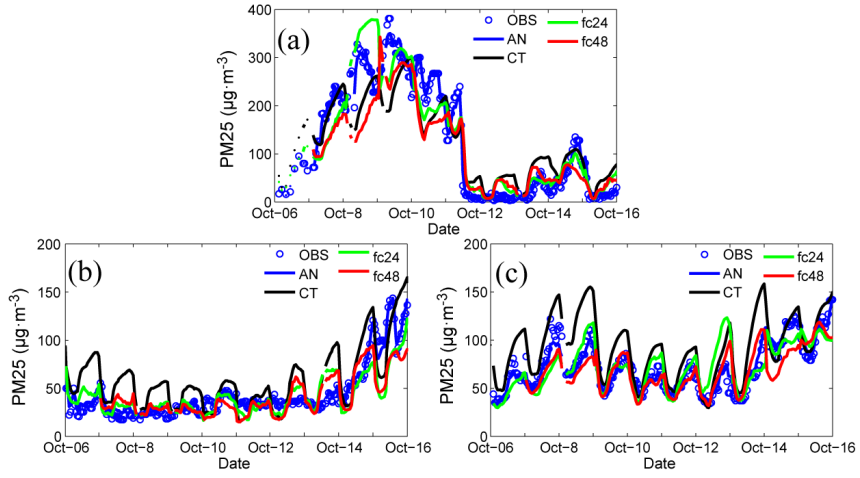


Figure 7. Spatial distribution of (a) the prior unspeciated primary sources of $PM_{2.5}$ ($\mu g\ m^{-2}\ s^{-1}$) and (b) the time-averaged differences between the ensemble mean analyses and the prior values ($\mu g \cdot m^{-2}\ s^{-1}$) at the lowest model level averaged over all hours from 6 to 16 October 2014.

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-732, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 26 August 2016

Atmospheric Chemistry and Physics Discussions

764



765

766 Figure 8. Time series of the hourly $PM_{2.5}$ obtained from observations (circle), analysis
767 (blue line), control run (black line) and hourly output of 48-h forecast in three
768 megacities: (a) Beijing; (b) Shanghai; and (c) Guangzhou. See text in section 5.4.

769

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-732, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 26 August 2016

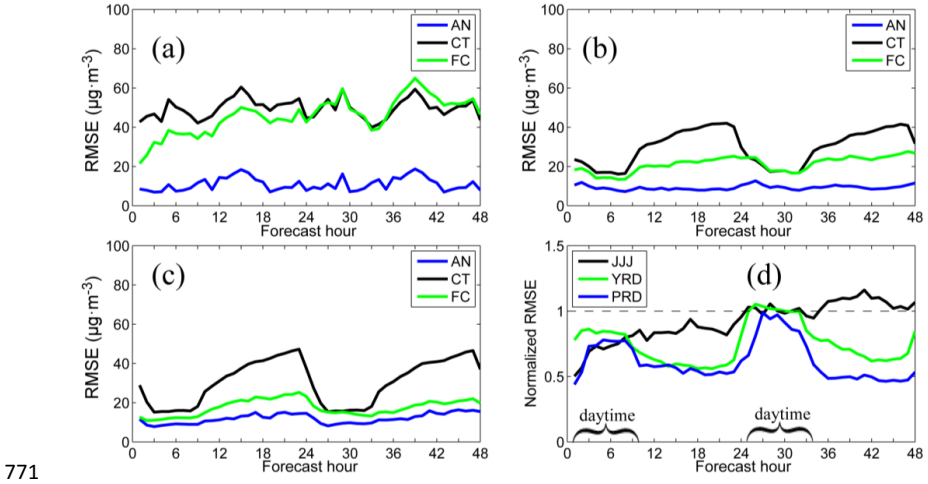


Figure 9. RMSE of surface $PM_{2.5}$ as a function of forecast range calculated against observations over the three sub-regions: (a) Beijing–Tianjin–Hebei region; (b) Yangtze River delta; and (c) Pearl River delta. (d) Normalized RMSE (assimilation divided by control).