Dec. 2, 2016.

Atmos. Chem. Phys.

RE: Manuscript Number: acp-2016-732

Dear Editors:

Thank you very much for your kind decision letter on our paper entitled "**Improving $PM_{2.5}$ forecast over China by the joint adjustment of initial conditions and source emissions with an ensemble Kalman**" (acp-2016-732). We are grateful for the helpful comments from you and the reviewers. We have changed the manuscript according to the reviewer's suggestions. The main changes include: 1) We have used the independent observations to evaluate both the analyses and the forecasts; 2) An experiment of pure assimilation chemical ICs and the corresponding 48-h forecasts experiment were also performed for comparisons; 3) Some discussion about the analyzed emissions. All the scientific questions have been resolved in the revised version (Please see details in it). And our responses to reviewers' comments are detailed in the following document.

The simultaneous adjustment of chemical initial conditions and emission input should be the way forward to improve $PM_{2.5}$ forecast, as large uncertainties in numerical prediction of atmospheric aerosols can usually be attributed to inaccurate aerosol emissions and the initial conditions in addition to the deficiency in the model system. Our research results showed that the forecasts with the optimized initial conditions and emissions were better in the whole, comparing to the forecasts with only the optimized ICs or the control experiment without data assimilation.

In addition, though the analyzed emissions are only the results of a mathematical optimum by utilizing observations, the "top-down" methods were widely used to assess the emissions (e.g. van Loon et al., 2000; Heemink and Segers, 2002; Hakami

et al., 2005; Zhang et al., 2005; Elbern et al., 2007; Peters et al., 2007; Henze et al., 2007, 2009; Yumimoto et al., 2007, 2008; Dubovik et al., 2008; Barbu et al., 2009; Sekiyama et al., 2010; Huneeus et al., 2012; Schutgens et al., 2012; Huneeus et al., 2012, 2013; Miyazaki et al., 2014; Wang et al., 2012; Guerrette and Henze, 2015; etc.). However, the lack of direct or exact emission information to evaluate the analyzed emissions was really a challenging to many emission inversion research teams. We are no exception. Nevertheless, the improvement of our emissions can be verified in terms of two aspects, the diurnal variation and the location of increased emissions (see details in Lines 592 to 617, Page 22).

In this work, $PM_{2.5}$ forecast is the subject of active research. And the simultaneous adjustment of chemical initial conditions and emission input is a promising method to improve $PM_{2.5}$ forecast and to assess the emissions. Plus, our conclusions are more convincing in the revised version. So we hope this manuscript will be published in ACP. We are looking forward to hearing from you soon.

Sincerely Yours,

Zhen Peng

Response to Reviewer #1's comments:

We thank Referee # 1 for their thoughtful comments and suggestions that have helped to improve this manuscript. Our responses to comments (in bold style) and the corresponding changes to the manuscript are detailed below.

General Comments:

1: The authors suggest that the joint adjustment (initial conditions and emissions) provides substantial improvements in from 34- to 48-h forecasts. Do you perform an assimilation and forecasting experiment in which only ICs are adjusted. Comparing between results from the joint adjustment and the IC only adjustment will reinforce your suggestion.

We have performed other two experiments, the assimilation of pure chemical ICs and the corresponding 48-h forecasts experiment. The details are in the revised manuscript (Lines 432 to 434, Page 16; Lines 448 to 452, Page 17; Lines 513 to 533, Page 19; Lines 620 to 622, Page 23; Lines 665 to 704, Page 25).

2: Both analyzed and forecasting results are validated by only observations that used in the assimilation. You should include the independent data, which is not used in the observational constraint, in the validation.

We have used the independent observations to evaluate both the analyses and the forecasts. Please see the details in the revised manuscript (Lines 354 to 355, Page 13; Lines 502 to 515, Page 19; Lines 632, Page 23 to Lines 691, Page25).

Specific comments:

3: Line 40, There are more recent research papers of ensemble-based assimilations with observations derived from in-situ measurements and geostationary satellite.

Dai, T., et al. (2014) Improvement of aerosol optical properties modeling over Eastern Asia with MODIS AOD assim- ilation in a global non-hydrostatic icosahedral aerosol transport model, Environ. Pollut., 195, 319–329.

Ying, X.M., et al. (2016) Estimation of aerosol properties over the Chinese desert

region with MODIS AOD assimilation in a global model, Adv. Clim. Change Res., 7, 90–98.

Yumimoto, K., et al. (2016), Aerosol data assimilation using data from Himawari-8, a next-generation geostationary meteorological satellite, Geophys. Res. Lett., 43, 5886–5894.

We have added those references in Lines 47 to 48, Page 2.

4: Line 90, Does the observation operator (H) include function (conversion) for the emission scaling factor (lambda) or, in other words, does the lambda directly affect the model results in the observation state (Hx) through the observation operator? If no, how does the observations adjust the emission scaling factors in the assimilation process?

In this manuscript, the emission scaling factor $\boldsymbol{\lambda}^{\mathrm{f}}$ is calculated by the persistence forecasting operator $\mathbf{M}_{\mathrm{SF}}$. Then, the emissions are calculated using equation (6) (original Eq. 11). After that, the chemical fields $\mathbf{C}^{\mathrm{f}}$ are forecasted though WRF-Chem. Finally, the model-simulated $PM_{2.5}$ concentration at the observation space is calculated via equation (13) (original Eq. 12) (See details in Section 2.3.1). Therefore, $\boldsymbol{\lambda}^{\mathrm{f}}$ directly affect the model results.

In fact, for the adjustment of the emission scaling factors, $\mathbf{M}_{\mathrm{SF}}$ serves as the forecast model and the observation operator reflects the combined information of emissions (in the format of $\boldsymbol{\lambda}$ in equation (6)), the physics and chemistry processes in WRF-Chem simulations and the transformation of $PM_{2.5}$ from model space to observation space (equation (13)). We have addressed these in Lines 275 to 279, Page 11.

5: Line 139, The ensemble concentration ratio (Kappa) is defined by concentrations of the ensemble forecasting. Can you confirm that the ensemble concentration ratio is random and the ensemble mean of Kappa becomes 1?

The ensemble mean of the concentration ratio is $\overline{\boldsymbol{\kappa}_{i,t}} = \overline{\mathbf{C}_{i,t}^{\mathrm{f}}/\overline{\mathbf{C}_{t}^{\mathrm{f}}}} = \overline{\mathbf{C}_{i,t}^{\mathrm{f}}}/\overline{\mathbf{C}_{t}^{\mathrm{f}}} =$

$\overline{\mathbf{C}_t^{\mathrm{f}}}/\overline{\mathbf{C}_t^{\mathrm{f}}}$=1. We have moved away random variables and revised this sentence in Line 142, Page 5.

6: Line 152 or Equation (10), The denominator in the right hand should be 1/M+1?

No. In Equation (5) (original Eq. 10), j starts from t-M+1. Thus, M scale factors (the prior and M-1 analysis scale factors) are used to calculate $\boldsymbol{\lambda}_{i,t}^{\mathrm{f}}$. For example, in our manuscript, $\mathrm{M} = 4$. Thus, $\boldsymbol{\lambda}_{i,t}^{\mathrm{p}}$, $\boldsymbol{\lambda}_{i,t-1}^{\mathrm{a}}$, $\boldsymbol{\lambda}_{i,t-2}^{\mathrm{a}}$, and $\boldsymbol{\lambda}_{i,t-3}^{\mathrm{a}}$ are used. Therefore, the denominator in the right hand of Equation (5) is 1/M.

7: Line 183, As shown in Equation (12), dust and sea salt aerosols can contribute PM2.5 concentrations. Do you include emissions of dust and sea salt in the assimilation process?

We did not include emissions of dust and sea salt in the assimilation process as our focus is on the major anthropogenic emissions in mega-cities in China.

Emissions of dust and sea salt were parameterized within the GOCART model (Chin et al., 2002). Unlike the approach for anthropogenic emissions, the approach would be different to assimilate dust and sea salt. In addition, only the $PM_{2.5}$ measurements were used in this DA experiment, with such limited observations adding more control variables would cause much more uncertainties in the system which might lead to unreasonable analysis. This is our first attempt to improve $PM_{2.5}$ forecast by the joint adjustment of ICs and source emissions, so we primarily focus on the major anthropogenic sources in heavy polluted regions ($\mathbf{E}_{\mathrm{PM2.5i}}$, $\mathbf{E}_{\mathrm{PM2.5j}}$, $\mathbf{E}_{\mathrm{SO4i}}$, $\mathbf{E}_{\mathrm{SO4j}}$, $\mathbf{E}_{\mathrm{NO3i}}$, $\mathbf{E}_{\mathrm{NO3j}}$). Those emissions have large impacts on the distribution of $PM_{2.5}$, thus are updated in our analysis. In future work, more species of emissions might be included.

We have added some explanations in Lines 300 to 308, Page 12.

8: Line 190, A period may drop in the end of state.

I have revised the text Line 283, Page 11.

9: Figure 1, Could you check figure 1 again? Some characters and numbers of equation are different from those in the manuscript.

I have revised the figure.

10: Line 202, Does this means that you need to perform the 50-member ensemble forecast twice in your assimilation system?

No, we perform the forecast only once. The steps in this workflow are: (1) $\boldsymbol{\lambda}^{\mathrm{f}}_{\mathrm{PM2.5}}$, $\boldsymbol{\lambda}^{\mathrm{f}}_{\mathrm{SO2}}$, $\boldsymbol{\lambda}^{\mathrm{f}}_{\mathrm{NO}}$ and $\boldsymbol{\lambda}^{\mathrm{f}}_{\mathrm{NH3}}$ are calculated using the forecast chemical concentration fields of the previous assimilation cycle; (2) The ensemble members of the emissions are generated;(3) WRF-Chem forecasts the chemical fields; (4) EnSRF assimilates, at this step, the scaling factors and the chemical fields are assimilated; (6) the emissions are updated. So, WRF-Chem runs to forecast only once during a DA cycle.

I have mentioned this in Line 200, Page 8.

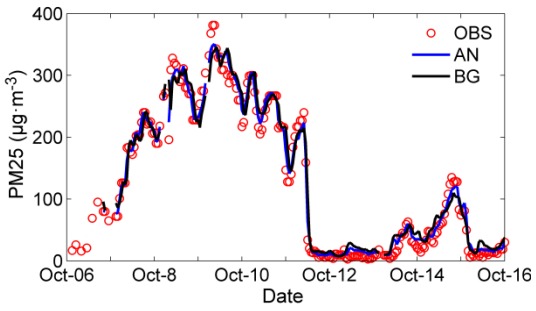
11: Line 254, How often did this exclusion occur? Figure 3a and 8a imply that quite a few large departures occurs in the JJJ region during 9–10 October.

The numbers of the observations were about 17700. Among them 8 observations were discarded because they were larger than 800 $\mu g\ m^{-3}$ and 243 (around 1.5%) were discarded due to the ensemble mean of the first guess departure exceeding 100 $\mu g\ m^{-3}$. In those 243 discarded observations, only 93 were in JJJ.

Figure 3a implied that some ensembles of the $PM_{2.5}$ background may deviate much from the observations during 9–10 October. However, the ensemble mean of the background $PM_{2.5}$ and the ensemble mean of the analysis $PM_{2.5}$ in the assimilation experiments were comparatively near to the observations (see ReFig1.), though the forecast of the $PM_{2.5}$ deviated much from the observations in the CT run and the forecast run. So only a few data were discarded due to the first guess departure

exceeding 100 μg m^{-3}.

We have added this statistics in Lines 373 to 375, Page 14 and in Lines 628 to 629, Page 23.



ReFig1. Time series of the hourly $PM_{2.5}$ obtained from observations (red circle), the ensemble mean of the analysis (blue line) and the ensemble mean of the background (the ensemble mean of the background, black line) in Beijing.

12: Line 281, How do you decide the ensemble member of 50?

We use the same EnSRF following Schwartz et al. (2012), in which the methodology/framework is similar to Whitaker and Hamill (2002). Whitaker and Hamill (2002) indicated the ensemble-mean RMS error is a function of ensemble size. When the ensemble size is larger than 50, the ensemble mean error is close to 0.19. So in this work, 50-member ensemble was chosen, following Schwartz et al. (2012) and Whitaker and Hamill (2002).

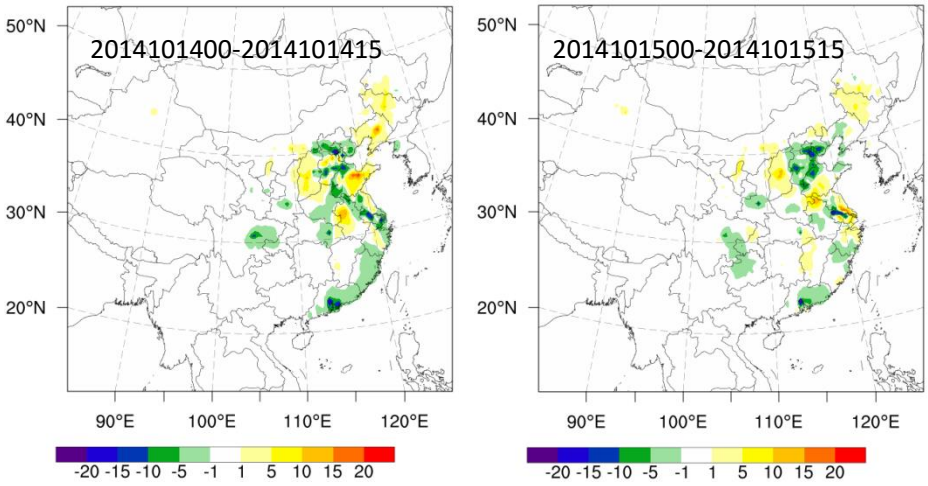
We have added some explanations in in Lines 247 to 248, Page 7.

13: Line 349, Could you add mean distribution of PM2.5 concentration from the control and assimilation simulations in Figure 4? These will make the reader to understand a priori distribution and the adjustment of PM2.5 concentrations easily. Plotting mean observed PM2.5 concentrations on these maps will be even better.

We added the spatial distribution of the $PM_{2.5}$ mass of the observations, the simulation of the control run, the analysis of expJ and expC, and also increments of

192 expJ and expC. The figure of the $PM_{2.5}$ mass differences was removed to save space.
193 It is very clear that the analysis of expJ and that of expC are much different from the
194 simulation of the control run.
195 Then we rewrote paragraph 2 in Section 5.2 in Lines 517, Page 19 to Lines 533,
196 Page 20.
197
198 **14: Line 349, We can find adjustments over the SE Asia and India where you**
199 **have no PM2.5 observation.**
200 The analysis increments (i.e. $\bar{\mathbf{x}}^{a} - \bar{\mathbf{x}}^{b}$) indicate the direct impact of assimilating
201 $PM_{2.5}$ observation. They are determined by both the observation increments and the
202 relative magnitudes of the forecast error and the observation error. From Figure 4 (e)
203 and (f), we can see the increments of both assimilation experiments are distributed
204 around the locations of observations as expected. However, the impact of assimilating
205 $PM_{2.5}$ observations is not limited to the areas where observations were located,
206 observations information is also transported to other areas through the WRF-Chem
207 forecast. Besides, the ensemble forecasts also partly contributed to the $PM_{2.5}$ mass
208 differences (assimilation minus control). Therefore, the spatial distributions of the
209 $PM_{2.5}$ mass in both assimilation experiments were significantly different from the
210 control run. Thus we can find adjustments over the SE Asia and India where no
211 PM2.5 observation is available.
212 We have added the above explanations in Lines 517 to 528, Page 19.
213
214 **15: Figure 5, Overlaying of a priori emissions (it will be flat lines) in Figure 5**
215 **may emphasize that the assimilation can generate the temporal variations in the**
216 **emissions.**
217 I have overlaid a priori emissions (the dash dot line) in Figure 5.
218
219 **16: Line 375, Is the burning of crop residues limited in the JJJ region? Li et al.**
220 **shows that the northern part of YRD also has large emissions from the burning.**
221 We are not sure. In expJ, some larger values for the optimized $E^{a}_{PM2.5}$ were also

obtained in the northern part of YRD region from 0000 UTC to 0015 UTC of 14 October and 15 October (see ReFig2). However, they were much smaller than that in JJJ. In addition, according to the Weekly Crop Residue Burning Monitoring Report traced by Environmental Satellite (data from the satellite Environment Center, Ministry of Environmental Protection), there were only 9 *crop residue burning spots in Anhui province* from 5 to 18 October 2014 and no *crop residue burning spots were reported in YRD*. Thus, we did not mention the burning of crop residues in YRD.



ReFig2 Spatial distribution of the mean differences between the ensemble mean analysis and the prior emissions of the unspeciated primary sources of $PM_{2.5}$ at the lowest model level

Response to Reviewer #2's comments:

We thank Referee # 2 for their thoughtful comments and suggestions that have helped to improve this manuscript. Our responses to comments (in bold style) and the corresponding changes to the manuscript are detailed below. In particular, according to the reviewer's suggestions, we have added two more simulations; we also substantially rewrote the texts in both the major context and summary section to emphasize reviewer's questions.

General Comments:

The authors introduce a DA system based on an ensemble square root filter combined with WRF-Chem that assimilates surface observations of PM2.5 across China. The novelty is that they use both aerosol concentrations and emissions in their DA state vector (although it should be noted they did something very similar for CO2 in Peng et al ACP 2015).

The method in this work is very similar to that used by Peters et al. (2007) and Peng et al. (2015) for CO_2 emission inversion, but it is still of novelty for applications in aerosol anthropogenic emissions. In Peters et al. (2007), $\boldsymbol{\lambda}_{i,t}^{\mathrm{p}}$ were all 1. And only natural CO_2 emissions (i.e., biospheric and oceanic emissions) were assimilated at the ecological scale due to the 'signal-to-noise' problem. Thus, the uncertainty of anthropogenic and other CO_2 emissions were ignored. Besides, the framework is more advanced compared to our previous work. In Peng et al. (2015), in order to generate $\boldsymbol{\lambda}_{i,t}^{\mathrm{p}}$, a set of ensemble forecasts were performed from time t to t+1 to produce the CO_2 concentration fields, forced by the prescribed net CO_2 surface fluxes using the previous assimilated concentration fields as initial conditions. That means that the ensemble forecast were performed twice in that DA system and it was time consuming. However, in order to save computing time, we used the chemical fields $\mathbf{C}_{i,t}^{\mathrm{f}}$ available in the previous assimilation cycle to calculate $\boldsymbol{\lambda}_{i,t}^{\mathrm{p}}$ in this work. Thus, WRF-Chem runs to forecast only once during a DA cycle.

We have added the above paragraph in Lines 187, Page 7 to Lines 200, Page 8.

While the main idea is interesting and the topic is certainly relevant to ACP, I recommend against publication for the following reasons: 1) no independent observations are used to evaluate results. While this is ok for the evaluation of forecasts, this is not good practice for the evaluation of analyses; 2) no proof is offered for the central contention that analyzing emissions *together with* concentrations improves results; 3) no proof is offered for the second central contention that this system improves emissions; 4) many assumptions are merely stated without due reference, deliberation or any kind of sensitivity study; 5) several conclusions are drawn based on irrelevant data (see my comments).

It should be noted that reviewer 1 mentions the first two points as well but is apparently more lenient.

Point 3 I find particularly important as this is a contention made by other authors as well (Tang et al, Miyazaki et al) with little in the form of proof. Models have errors, and analyzing emissions may simply balance out some of these errors without improving the emissions. Note that we do not have observations to evaluate those emissions but this cannot be used as an argument to forego proper scientific reasoning.

In addition I find the structure of the paper illogical, and missed important information on details of their DA system and several references to previous attempts at emission estimation. I hope the authors will continue this work but put more effort in stating their case convincingly, for this research topic is certainly worthwhile. Maybe my comments can be of some help towards improving this manuscript.

Thanks for those comments which did help improving this manuscript. Please see the point-to-point answers as below.

1) We have used the independent observations to evaluate both the analyses and the forecasts. Please see the details in the revised manuscript (Lines 354 to 355, Page 13; Lines 502 to 515, Page 19; Lines 632, Page 23 to Lines 691, Page25).

2) An experiment of pure assimilation chemical ICs and the corresponding 48-h forecasts experiment were also performed for comparisons in the revised manuscript. Please see the details in the revised manuscript (Lines 432 to 434, Page 16; Lines 448 to 452, Page 17; Lines 513 to 533, Page 19; Lines 620 to 622, Page 23; Lines 665 to 704, Page 25).

3) The analyzed emissions are only the results of a mathematical optimum by utilizing observations. They are influenced greatly by the model errors and the observation errors. In addition, only surface $PM_{2.5}$ observations were applied in this work, which may lack abundant constraint on the sources of the secondary aerosol precursors. Moreover, we do not have direct or exact emission information to evaluate the analyzed emissions, which was a challenging to many emission inversion research teams (e.g. Tang et al, 2011; Miyazaki et al., 2012; Ding et al., 2015; Mclinden et al., 2016; etc.). Different from the situations that standard national emission inventories are reported by government as in USA, European or other countries, the rapid economic development and complexity of emission sources in China lead to large uncertainties in the current public available emission inventories. Thus it's impossible for us to conduct the direct evaluation on emissions. For this reason, we weaken our judgment in the text.

Nevertheless, our system considering the emission assimilation provided better simulation results and the improvement of emissions can be verified in terms of two aspects, the diurnal variation and the location of increased emissions. The diurnal variation in the assimilated emissions can be used to verify our judgment to some extent. Especially in the PRD and YRD, $\mathbf{E}^{a}_{PM2.5}$ in the daytime were always larger than those in the night, which agreed well with Olivier et al. (2003), the WRAP (2006) and Wang et al. (2010). In addition, the locations of the larger values for the optimized $E^{a}_{PM2.5}$ in the JJJ region were in good agreement with the places of the crop residues burning traced by the environmental satellite of China. There were 10, 231, 37 and 3 crop residue burning spots in Hebei, Henan, Shandong and Shanxi province respectively from 5 to 11 October 2014 and the numbers are 7, 20, 5 and 21 respectively from 12 to 18 October 2014 (Weekly Crop Residue Burning Monitoring

Report traced by Environmental Satellite, 2015a, 2015b).

We have added the above paragraph in Lines 588, Page 21 to Line 613, Page 22.

4) and 5), we have revised the manuscript according to the reviewer's suggestions.

Abstract

1. P 1, L 13: "The forecast model of emission scaling factors was developed by associating the time smoothing operator with WRF-Chem forecast chemical concentrations". Please rephrase, this sentence is hard to understand without reading the paper first.

This sentence has been rephrased as: "The forecast model of emission scaling factors was developed by using the ensemble concentration ratios of the WRF-Chem forecast chemical concentrations and also the time smoothing operator".

We have rephrased these references in Lines 14 to 16, Page 1.

Introduction

2. P 2, L 40: The authors seem unaware of a lot of previous work on ensemblebased DA: Sekiyama et al ACP 2010, Schutgens et al. ACP 2010a, Schutgens et al ACP 2010b. , Dai et al, *Env. Pol*. 2014, Rubin et al. ACP 2016, , Yumimoto et al GRL 2016. Please include those references.

We have added these references in Lines 46 to 48, Page 2.

3. P 2, L 50: Again, several references seem to be missing i.c. emission estimation. For aerosol: Zhang et al JGR 2005, Sekiyama et al. ACP 2010, Huneeus et al ACP 2012, Schutgens et al. Rem Sens 2012, Huneeus et al ACP 2013

We have added these references in Lines 56 to 58, Page 3.

Methodology

4. **P 3, L 78: Please introduce the ENSRF in context of some other EnKF (EAKF, LEKF, LETKF). What is the reason for this choice of EnKF, what is it main strength/weakness?**

There are different versions of EnKF. The traditional EnKF with perturbed observations (Evensen 1994) introduces sampling errors by perturbing the observations. In contrast to the traditional EnKF, the EnSRF (Whitaker and Hamill, 2002) and the Ensemble Adjustment Kalman Filter (EAKF, developed by Anderson, 2001) obviate the need to perturb the observations. The local ensemble Kalman filtering (LEKF), a kind of EnSRF, was presented by Ott et al. (2002, 2004). It was computationally more efficient compared to the traditional EnKF, since it simultaneously assimilates the observations within a spatially local volume independently. The local Ensemble Transform Kalman Filter (LETKF, Hunt, 2007) integrates the advantages of the Ensemble Transform Kalman Filter (ETKF, developed by Bishop et al., 2001) and the LEKF. The computational cost of LETKF is much lower than that of the original LEKF because the former does not require an orthogonal basis. Though LETKF has more advantages, we still chose the same EnSRF as Schwartz et al. (2014) because we did not need to extend it to analyzing aerosol ICs, very similar to Schwartz et al. (2014).

We have added the above paragraph in Lines 205 to 219, Page 8.

5. P 54 L 94: Change "can be approximated" to "will be approximated". It is by no means certain that this is a good approximation. Part of the evaluation & tuning of an EnKF involves exactly the sampling errors introduced by Eq 5 & 6

We have changed this sentence in Line 235, Page 8.

6. P 3: Since the DA depends on the forecast model's details, I suggest to first discuss the forecast model (and introduce C and λ, and only then the ENSRF)

We have changed the orders of Section 2.1 and 2.2.

7. P 4, L 105: Please provide a bit more information on the base setup of the model: domain size, grid resolution, major aerosol species

We have added more information of the base set up of the model in Lines 101, Page 4 to Lines 114, Page 5.

8. P 4, L 106: "to forecast the emission scaling factors and the aerosol control variables". What are the control variables? I guess the authors mean aerosol concentrations, please change this. Note that both C and λ form the state vector.

We have revised this sentence in Line 88, Page 4.

9. P 5, L 123: "for the lowest eight vertical levels": so the emission inventory included heights at which the emissions were injected? These heights are all within the boundary layer? Why are only the lowest 8 layers considered?

In this work, the lowest 12 vertical levels were at ~ 12 m, 48 m, 98 m, 156 m, 232 m, 300 m, 400 m, 500 m, 600 m, 700 m, 850 m, and 1000 m respectively. So the lowest 12 layers were all within the boundary layer. And the lowest 8 layers were under 500 m.

The emission inventory did not include emission heights at which the emissions were injected, which may cause large uncertainties for model forecast. We prepared the prescribed emissions just following others research (Woo et al., 2003; de meij et al., 2006; Wang et al., 2010): the power generator emissions were interpolated for the lowest eight vertical levels. And other anthropogenic emissions were assigned totally to the 1^{st} level.

Emissions are very small above 500 m for all pollutants. So only the lowest 8 layers are considered.

We have added more discussions about the prescribed emissions in Lines 112 to 114, Page 5; in Lines 117 to 120, Page 5.

10. P 6, L 139: "$\boldsymbol{\kappa}_{i,t}$ are random". I wouldn't call them random. I realize they are distributed around the mean $\overline{\boldsymbol{\kappa}_t}$, , but they were calculated through a short-term forecast of WRF-Chem.

Yes. The ensemble concentration ratio ($\boldsymbol{\kappa}_{i,t}$) are distributed around the ensemble mean ($\overline{\boldsymbol{\kappa}_t}$). And $\overline{\boldsymbol{\kappa}_t} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\kappa}_{i,t} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{C}_{i,t}^{\mathrm{f}}/\overline{\mathbf{C}_t^{\mathrm{f}}} = \frac{1}{N*\overline{\mathbf{C}_t^{\mathrm{f}}}}\sum_{i=1}^{N}\mathbf{C}_{i,t}^{\mathrm{f}} = \overline{\mathbf{C}_t^{\mathrm{f}}}/\overline{\mathbf{C}_t^{\mathrm{f}}} = 1$. So they are actually distributed around 1.

We have removed random variables and changed this sentence as: 'so $\boldsymbol{\kappa}_{i,t}$ are numbers distributed 1 and with ensemble mean values of 1" in Line142, Page 6.

11. P 6, L 144: "$\beta = 1.5$ was chosen in this study": This sounds like an arbitrary choice? Normally β results from tuning a DA but no such exercise was done?

Peters et al (2007) first used the time smooth operator to evaluate the CO_2 fluxed scaling factors: $\boldsymbol{\lambda}_{i,t}^{\mathrm{f}} = (\boldsymbol{\lambda}_{i,t-2}^{\mathrm{a}} + \boldsymbol{\lambda}_{i,t-1}^{\mathrm{a}} + \boldsymbol{\lambda}_{i,t}^{\mathrm{p}})/3$ (P. 8, the last paragraph in Peters et al. 2007. Here, we use the same notation in our manuscript). In that work, $\boldsymbol{\lambda}_{i,t}^{\mathrm{p}}$ were all 1 (P. 11, below S3.3). The time smooth operator was very useful because $\boldsymbol{\lambda}_{i,t}^{\mathrm{f}}$ could gain useful information achieved by previous DA cycle through the using of $\boldsymbol{\lambda}_{i,t-2}^{\mathrm{a}}$ and $\boldsymbol{\lambda}_{i,t-1}^{\mathrm{a}}$. However, they had to assimilate natural CO2 emissions (i.e., biospheric and oceanic) at the ecological scale due to the 'signal-to-noise' problem. Thus, the uncertainty of anthropogenic and other CO2 emissions were ignored.

We used the time smooth operator following Peters et al. (2007). In order to optimize all CO_2 fluxes as a whole at grid scale, we first used the ensemble concentration ratio ($\boldsymbol{\kappa}_{i,t}$) to calculate the ensemble prior emission scaling factors $\boldsymbol{\lambda}_{i,t}^{\mathrm{p}}$ in Peng et al. (2015). $\boldsymbol{\lambda}_{i,t}^{\mathrm{p}}$ were artificial data to generate the ensemble emissions. It was difficult to give the ensemble members of $\boldsymbol{\lambda}_{i,t}^{\mathrm{p}}$ for the ensemble-based emission inversion system. Perhaps it was the simplest way to generate this data at every assimilation cycle by directly using the standard normal distribution function. But the

inversion system failed to optimize the prior fluxes at grid scale due to the 'signal-to-noise' problem (We have done the experiment for CO_2 inversion). From the other aspect, if following Peters et al. (2007) completely, the time smooth operator was applied and $\boldsymbol{\lambda}_{i,t}^{\mathrm{p}} = 1$ was chosen. However, the scaling factors should be perturbed at the first assimilation cycle to generate the ensemble factors. Consequently, this inversion system failed to optimize the prior fluxes at grid scale due to the same 'signal-to-noise' problem (Peng et al., 2015). So other ways should be found to generate $\boldsymbol{\lambda}_{i,t}^{\mathrm{p}}$. In Peng et al., $\boldsymbol{\kappa}_{i,t}$ was used to calculate $\boldsymbol{\lambda}_{i,t}^{\mathrm{p}}$, and it seemed effective.

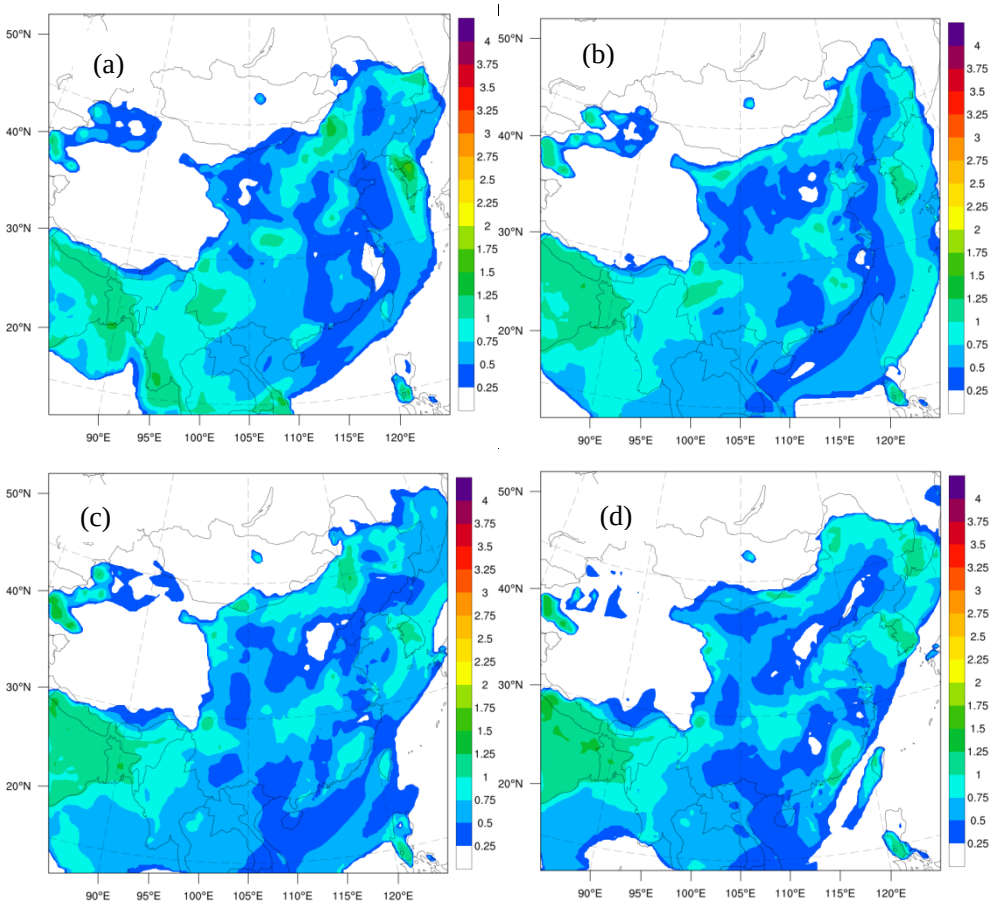
In Peng et al. (2015), the ensemble spread of $\boldsymbol{\kappa}_{i,t}$ was very small (ranging from 0 to 0.08 in most area at model-level 1), though the values of the ensemble spread of $\mathbf{C}_{i,t}^{\mathrm{f}}$ after inflation could reach 1 to 14 ppmv in most area at model-level 1. Therefore, covariance inflation was used to keep it at a certain level. After covariance inflation, the ensemble spread of $\lambda_{\mathrm{i,t}}^{\mathrm{a}}$ ranged from 0.1 to 0.8 in most model area for $\beta = 70$. Besides, several sensitive experiments were performed to investigate β (10, 50, 60, 70, 75, 80, 100). The ensemble spread of $\lambda_{\mathrm{i,t}}^{\mathrm{a}}$ ranged from 0.05 to 1.2 for β =60, 70, 75, 80. And the CO_2 DA system worked comparatively well for β =60, 70, 75, 80. Though CO_2 fluxes inversion was another topic, we mentioned it here because this experience was very helpful for us to develop the joint DA system for aerosol.

As for the $PM_{2.5}$ assimilation, we have done several sensitive experiments to determine the value of β (1.2, 1.5, 1.8, 2, 2.5) by using $PM_{2.5}$ measurements at the five U.S. Embassies stations in China (We did not gain the $PM_{2.5}$ observations from the Ministry of Environmental Protection of China at that time, in August 2015). It showed that the DA system worked comparatively well for $\beta = 1.2$, 1.5 and 1.8. For these cases, the ensemble spread of $\boldsymbol{\lambda}_{\mathrm{PM2.5}}^{\mathrm{f}}$ ranged from 0.1 to 1.25 in most model area. Thus, $\beta = 1.5$ was chosen for latter experiments.

The magnitudes of the ensemble spread of the emission scaling factors were very stable with time. For the joint DA experiment in this manuscript, the ensemble spread

of $\boldsymbol{\lambda}^{\mathrm{f}}_{\mathrm{PM2.5}}$ ranged from 0.25 to 1 in most model area except India where we were not interested in and no observations were available (see details in ReFig. 1). In the manuscript, hourly area-averaged time series of the ensemble spread for $\boldsymbol{\lambda}^{\mathrm{f}}_{\mathrm{PM2.5}}$ over JJJ, YRD, PRD were added in Figure 3d.

It is noted that there were very few negative values for $(\boldsymbol{\kappa}_{i,t})_{\mathrm{inf}}$ after inflation in some cases. A quality control procedure should be performed for $(\boldsymbol{\kappa}_{i,t})_{\mathrm{inf}}$ before further appliance: All these negative data were set as 0.001. Then $(\boldsymbol{\kappa}_{i,t})_{\mathrm{inf}}$ were re-centered to ensure the ensemble mean value of $(\boldsymbol{\kappa}_{i,t})_{\mathrm{inf}}$ were 1. We added this explanation in Lines 146 to 151, Page 6; Lines 489 to 496, Page 18.



ReFig. 1. Spatial distribution of the ensemble spread for $\boldsymbol{\lambda}^{\mathrm{f}}_{\mathrm{PM2.5}}$ at the lowest model level at (a) 0000 UTC 6 October 2014; (b) 0000 UTC 7 October 2014; (c) 0000 UTC 8 October 2014; (d) 0000 UTC 9 October 2014 for $\beta = 1.5$;

12. P 6, L 145: "As the concentrations were closely related to the emissions": if I assume this refers to emissions and concentrations in the same grid-box (given the mathematics of their DA system), this is a bold statement and needs some strong arguments. I can see that during the dust season, Beijing area will be heavily impacted by dust from Eastern China, invalidating your assumptions. Even for pollution emissions, transport may actually be very important.

It is true that transport is very important for aerosol or other air pollution. We corrected the text as "**As the concentrations were closely related to the emissions both locally and in the upwind regions.**"

As stated in Q11 in detail, the prior emission scaling factors $\boldsymbol{\lambda}_{i,t}^{\mathrm{p}}$ were artificial data to generate the ensemble emissions. We chose $\boldsymbol{\lambda}_{i,t}^{\mathrm{p}} = (\boldsymbol{\kappa}_{i,t})_{\mathrm{inf}}$ (4) (original Eq. 9) only as a last resort. Though the concentrations are related to the emissions according to the mass conservation equation, Eq. (4) is not strongly supported. However, same as $(\boldsymbol{\kappa}_{i,t})_{\mathrm{inf}}$, $\boldsymbol{\lambda}_{i,t}^{\mathrm{p}}$ are numbers distributed around 1. From the perspective of generating the ensemble emissions, $\boldsymbol{\lambda}_{i,t}^{\mathrm{p}}$ can play the same role as other data, such as the random numbers created by using the standard normal distribution function. However, there are correlations among the grid-points of $(\boldsymbol{\kappa}_{i,t})_{\mathrm{inf}}$ because $(\boldsymbol{\kappa}_{i,t})_{\mathrm{inf}}$ are calculated through a short-term forecast of WRF-Chem. Thus, $\boldsymbol{\lambda}_{i,t}^{\mathrm{p}}$ have the same correlations as $(\boldsymbol{\kappa}_{i,t})_{\mathrm{inf}}$. While the random numbers are totally different. There are no correlations unless they are generated under certain correlations.

It is noted that the correlations among the grid-points of the prior emissions depend on $\boldsymbol{\lambda}_{i,t}^{\mathrm{p}}$. Maybe these correlations deviate far from the truth. However, the correlations among the grid-points of the forecast emissions maybe come close to the truth due to the appliance of the smooth operator after multiple iterations.

We have revised the sentence "**As the concentrations were closely related to the emissions both locally and in the upwind regions**" in Lines 152 to 153, Page 6 and added the content of the above paragraph in Lines 157, Page 6 to Lines 165, Page 7.

13. P 6, L 147: "concentration ratios $(\boldsymbol{\kappa}_{i,t})_{\mathbf{inf}}$ served as the prior emission scaling factors $\boldsymbol{\lambda}_{i,t}^{\mathrm{p}}$" So the concentrations themselves were not inflated, as is usually done in EnKF? What is the justification for this? Shouldn't the scaling factors be perturbed according to the uncertainty in emission inventories and parametrizations?

Posterior multiplicative inflation was applied for only the concentration analysis aiming to maintain ensemble spread.

As for the emission scaling factors, posterior multiplicative inflation was not used. Besides, they are not perturbed according to the uncertainty in emission inventories and parametrizations. Since $\boldsymbol{\lambda}_{i,t}^{\mathrm{p}}$ are calculated through a short-term forecast of WRF-Chem, $\boldsymbol{\lambda}_{i,t}^{\mathrm{f}}$ have deterministic values from the time smooth operator.

We have addressed the posterior multiplicative inflation, plus the covariance localization, in Lines 247 to 256, Page 10.

14. P 6, L 152: I suspect that Eq 10 is missing a factor 0.5. The prior and analysis scale factors are previous times are averaged.

It is right that the prior and analysis scale factors of previous times are averaged, but a factor 0.5 is not missed. In Equation (5) (original Eq. 10), j starts from t-M+1. Thus, M times of scale factors (the prior and M-1 analysis scale factors) are used to calculate $\boldsymbol{\lambda}_{i,t}^{\mathrm{f}}$. For example, in our manuscript, $\mathrm{M} = 4$. Thus, $\boldsymbol{\lambda}_{i,t}^{\mathrm{p}}$, $\boldsymbol{\lambda}_{i,t-1}^{\mathrm{a}}$, $\boldsymbol{\lambda}_{i,t-2}^{\mathrm{a}}$, and $\boldsymbol{\lambda}_{i,t-3}^{\mathrm{a}}$ are used. Therefore, the denominator in the right hand of Equation (5) is 1/4.

15. P 6, L 153: Again, a rather arbitrary choice (M=4)? How does this relate to the DA cycle?

According to the smooth operator, the ensemble mean values of $\boldsymbol{\lambda}_{i,t}^{\mathrm{f}}$ depend on the ensemble mean of $\boldsymbol{\lambda}_{i,t-M+1}^{\mathrm{a}}$, ⋯, $\boldsymbol{\lambda}_{i,t-2}^{\mathrm{a}}$, $\boldsymbol{\lambda}_{i,t-1}^{\mathrm{a}}$, $\boldsymbol{\lambda}_{i,t}^{\mathrm{p}}$, where the ensemble means of $\boldsymbol{\lambda}_{i,t}^{\mathrm{p}}$ are all 1. After multiple iterations, the smooth operator can give comparatively

good estimation for $\boldsymbol{\lambda}_{i,t}^{\mathrm{f}}$ since anthropogenic emissions are stable at a certain time scale (Mijling et al., 2012).

Peters et al (2007) chose M=3 ($\boldsymbol{\lambda}_{i,t-2}^{\mathrm{a}}, \boldsymbol{\lambda}_{i,t-1}^{\mathrm{a}}$ and $\boldsymbol{\lambda}_{i,t}^{\mathrm{p}}$ were used to calculate $\boldsymbol{\lambda}_{i,t}^{\mathrm{f}}$) for CO_2 fluxes inversion. They indicated that it was a compromise between prescribing prior CO_2 fluxes at each step and letting the system propagate all information from one step to the next without any guidance (in L 3, P 11). They also pointed out that the latter will work fine for the North American fluxes which were strongly constrained by observations. Similar to Peters et al. (2007), fewer states are used to calculate $\boldsymbol{\lambda}_{i,t}^{\mathrm{f}}$ for the joint DA system for aerosol in this manuscript.

In the revised manuscript, we have added some explanation in Lines 171 to 177, Page 7 and some results in Lines 539 to 541, Page 20.

16. P 6, L 159: "emission inventories". Except in the case of dust, sea-salt etc. Or are these not perturbed? If not, why are they not perturbed (surely they are uncertain as well)? Actually, the authors are rather sparse in their information. Is each species perturbed independently from the others? What is the level of perturbation? Are neighbouring grid-points perturbed independently or do you assume correlations?

In the assimilation part, we had applied 4 independent scaling factors: $\boldsymbol{\lambda}_{\mathrm{PM2.5}}$, $\boldsymbol{\lambda}_{\mathrm{SO2}}$, $\boldsymbol{\lambda}_{\mathrm{NO}}$ and $\boldsymbol{\lambda}_{\mathrm{NH3}}$. Both the forecast emissions (perturbed emissions) and the assimilated emissions were calculated according to EQ (6) : $\mathbf{E}_{i,t} = \boldsymbol{\lambda}_{i,t}\mathbf{E}_{t}^{\mathrm{p}}$ (original Eq. 11). $\boldsymbol{\lambda}_{\mathrm{PM2.5}}$ were used to calculated $\mathbf{E}_{\mathrm{PM2.5i}}$, $\mathbf{E}_{\mathrm{PM2.5j}}$, $\mathbf{E}_{\mathrm{SO4i}}$, $\mathbf{E}_{\mathrm{SO4j}}$, $\mathbf{E}_{\mathrm{NO3i}}$, and $\mathbf{E}_{\mathrm{NO3j}}$ (see details in 2.3.1). $\boldsymbol{\lambda}_{\mathrm{SO2}}$, $\boldsymbol{\lambda}_{\mathrm{NO}}$ and $\boldsymbol{\lambda}_{\mathrm{NH3}}$ were used to calculate $\mathbf{E}_{\mathrm{SO2}}$, $\mathbf{E}_{\mathrm{NO}}$ and $\mathbf{E}_{\mathrm{NH3}}$. In this study, only the species of the emission inventories mentioned above were perturbed (or updated according to the assimilated scaling factors).

Other inorganic species of the anthropogenic emission, such as $\mathbf{E}_{\mathrm{EC}}$ and $\mathbf{E}_{\mathrm{ORG}}$, are not perturbed for WRF-Chem, which is a limitation of this manuscript. However, other anthropogenic emissions, such as $\mathbf{E}_{\mathrm{PM2.5}}$, $\mathbf{E}_{\mathrm{SO4}}$ and $\mathbf{E}_{\mathrm{NO3}}$ are much larger

than $\mathbf{E}_{\mathrm{EC}}$ and $\mathbf{E}_{\mathrm{ORG}}$ in most area of China, and the ensemble spreads of the aerosol concentrations largely depend on the uncertainties of those anthropogenic emissions. Besides, model errors arisen from the meteorology, the emissions and the chemical model itself are compensated to some extent through the use of multiplicative inflation. In other words, the ensemble spread of the concentrations can be kept at a certain level though $\mathbf{E}_{\mathrm{EC}}$ and $\mathbf{E}_{\mathrm{ORG}}$, are not perturbed.

Natural emissions, such as dust and sea salt were not perturbed explicitly when the forecast emissions were generated. However, emissions of dust and sea salt were parameterized within the GOCART model (Chin et al., 2002). Within the DA system, varying meteorology across the members implicitly perturbed dust and sea salt emissions.

We have added the above two paragraphs in Lines 320, Page 12 to Lines 334, Page 13.

No other perturbations are added to the scaling factors. And no other correlations are assumed for the scaling factors. As stated above, both the forecast emissions (perturbed emissions) and the assimilated emissions were calculated according to EQ (6) : $\mathbf{E}_{i,t} = \boldsymbol{\lambda}_{i,t}\mathbf{E}_t^{\mathrm{p}}$ (original Eq. 11). The correlations among the grid-points of the forecast emissions depend on the correlations among the grid-points of $\boldsymbol{\lambda}_{i,t}^{\mathrm{f}}$. See some detail in Q.12 and in Line 182 to 186, Page 7.

17. P 7, L 175: "the state variables of the analysis of the ICs were the 15 WRFChem/ GOCART aerosol variables." This should have been mentioned earlier, maybe line 101.

We have moved this to lines 242 to 244, page 9.

18. P 7, L 184: "(λPM2.5, λSO2, λNO and λNH3)" This line and the following paragraph suggest that the authors keep the EEC and EORG constant? They do not matter? I rather think they do. By the way, this paragraph might be

rewritten to improve readability.

Yes, we keep the $\mathbf{E}_{\mathrm{EC}}$ and $\mathbf{E}_{\mathrm{ORG}}$ constant during the joint DA experiment, which is a limitation in this manuscript. It is true that these emissions are also important for the atmosphere aerosol. The reason we did not assimilate $\mathbf{E}_{\mathrm{EC}}$, $\mathbf{E}_{\mathrm{ORG}}$ is that only the $PM_{2.5}$ measurements are used in this DA experiment. However, the sources of the aerosols (especially organic aerosols) are so complex that our knowledge of their formation mechanisms is far from clear. Though it is technically possible to have all emissions assimilated, with such limited observations adding more control variables would cause much more uncertainties in the system which might lead to unreasonable analysis. This is our first attempt to simultaneously optimize the chemical ICs and emission input. In future work, when gas-phase observations of SO_2, NO_2 and O_3 are used and more aerosol species observations are available, perhaps more emissions are assimilated, similar to Tang et al. (2011).

We have added the above paragraph in Lines 300 to 308, Page 12.

We have also rewritten this paragraph in Lines 268 to 276, Page 10.

19. P 8, L 208: The authors never explain how the system is started up. Some initial perturbation in concentrations and/or emissions must be assumed.

We have rewritten some part of in Sec. 4.2 in Lines 424 to 431, Page 16.

20. P 9, L 247: "$\varepsilon_r = r\varepsilon_0\sqrt{\Delta x/\mathrm{L}}$," Can the authors provide a reference for this form of the representation error? Why do they choose L=3 km? How can it be that the representation error is a function of the measurement error? These are two independent error sources.

We calculated the representation errors completely following Schwartz et al. (2012), who followed Elbern et al. (2007) and Pagowski et al. (2010). Elbern et al. (2007) developed this scheme firstly based on the research of the European organizations. In Elbern et al. (2007), L= 20, 10, 4, 2, 1 and 3 km for Remote, Rural, Suburban, Urban, Traffic and Unknown station type (P 3758) respectively. We had

added some information of the scheme in Lines 366 to 367, Page 13.

21. P 9, L 252-255: Some statistics on how often this happened would be appreciated.

The numbers of the observations were about 17700. Among them 8 observations were discarded because they were larger than 800 $\mu g\ m^{-3}$ and 243 (around 1.5%) were discarded due to the ensemble mean of the first guess departure exceeding 100 $\mu g\ m^{-3}$.

We added this statistics in Lines 373 to 375, Page 13.

22. P 10, L 261: "The horizontal grid spacing was 40.5 km and there were 262 57 vertical levels with the model top at 10 hPa." This sort of information should be in Sect 2.2.1

We have moved this sentence in Sect. 2.1.1.

23. P 1, L 265: "initialization and spin-up procedures" Please briefly state the spinup procedure. For how long was the ensemble run before the first DA happened?

We have done initialization experiments from 0000 UTC 1 October to 2300 UTC 4 October 2014. And we have rewritten the last paragraph in Sect. 4.1 in Lines 413 to 416, Page 15.

24. P11, L 279: "clean oceanic conditions." Does this mean that over land you assumed seasalt aerosol as LBC?

Actually the LBCs for chemistry/aerosol fields were idealized profiles embedded within the WRF/Chem model. It's not only for the clean oceanic conditions. We have corrected the text. The differences between the idealized profile and real boundary conditions may bring some errors for the boundary region but since our focus is centered in the JJJ, YRD and PRD regions that far from the boundary region. The impacts would be negligible.

We have corrected the text in Lines 398 to 399, Page 14.

25. P 11, L 280: "standard Gaussian random noise". Please briefly state what standard deviations you assumed, and how you dealt with negative emissions.

We perturbed the anthropogenic emissions following Schwartz et al. (2012).

For possible negative perturbed emissions, they were set as $E_{ip}^{*}(\eta, t) = 0.001 * E_{p}(\eta, t)$. This will increase the prescribed emissions more or less. However, only very few data were negative. So, this influence can be negligible.

It should be noted that the perturbed emissions were only used in the spin-up procedure and expC.

We have rewritten this part in Lines 403 to 412, Page 16.

26. P 13, L 336: "These statistics were calculated against observations over all the analyses" If I understand the authors, the same observations that were assimilated are here used to evaluate the results. This likely explains the high correlations. The authors should make it clear this is not an independent evaluation but merely a sanity check.

We have added the independent observations to evaluate the analysis in Lines 501 to 515, Page 19.

27. P 13, L 356: "These results indicate that DA greatly improved the ICs." This is rather bold as you have not used independent observations to evaluate the ICs. Obviously, if you nudge the model towards observations, the model will do better. Please remove this sentence.

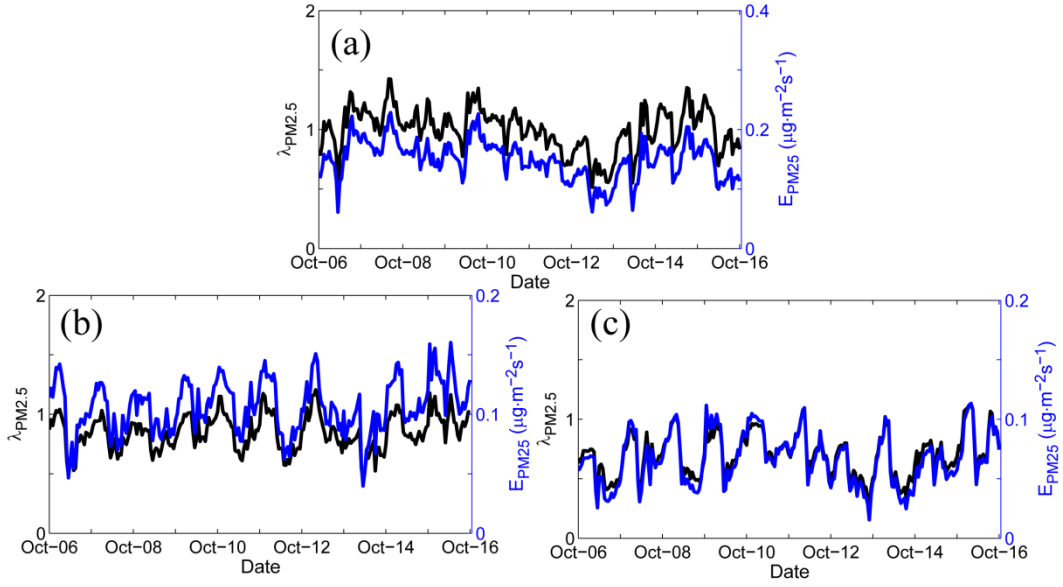
We have used the independent observations to evaluate the analysis. We also removed this sentence.

28. P 13, L 363: "the optimized PM2.5 scaling factor, λPM2.5a, showed an

obvious variation with time, as did the optimized unspeciated primary sources of PM2.5, EPM2.5a" From the authors explanation of how their system works, I do not understand why λPM2.5 and EPM2.5 would have a different (if only slightly) time evolution. Is this because they are regional averages?

Thanks for pointing out this error! The $\boldsymbol{\lambda}^{f}_{PM2.5}$, $\boldsymbol{\lambda}^{f}_{SO2}$, $\boldsymbol{\lambda}^{f}_{NO}$ and $\boldsymbol{\lambda}^{f}_{NH3}$ were 1 hour earlier than the $\mathbf{E}^{f}_{PM2.5}$, $\mathbf{E}^{f}_{SO2}$, $\mathbf{E}^{f}_{NO}$ and $\mathbf{E}^{f}_{NH3}$ in the original plot as I made a mistake when extracting those values.

ReFig. 2 (also updated in the manuscript) shows the right results. It shows that the $\mathbf{E}^{a}_{PM2.5}$ change along with $\boldsymbol{\lambda}^{a}_{PM2.5}$.



ReFig. 2. 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.

29. P 13, L 379: "as the system is optimized based on ambient concentrations in which the transport and transformation processes are not directly taken into account" But surely transport is important? Maybe a Kalman smoother would have been a better system to solve this problem.

We think transport is as important as transformation. In our DA experiments, the

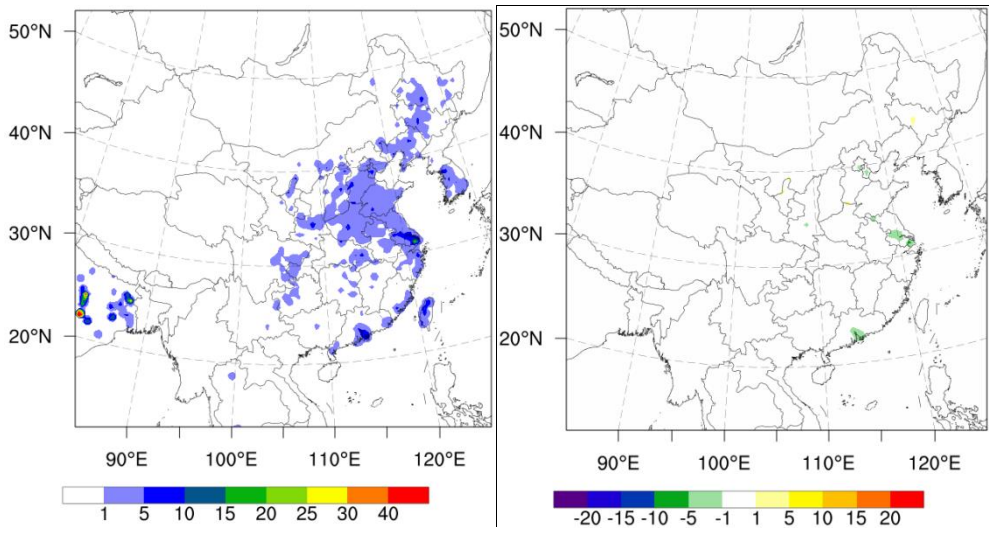
$PM_{2.5}$ measurements network was still spatially sparse and heterogeneous. Almost all measurements were located in the city and no data available in the rural region. However, the crop residues burning always occur in rural region. So the $PM_{2.5}$ measurements network can only capture the burning information a few hours later. It is right that a Kalman smoother would have been a better system to solve this problem.

We have added some explanation in Lines 557 to 565, Page 20.

30. P 14, L 388: "at the lowest model level" Why do you only discuss emissions at lowest level? Are they much larger than those at higher levels? Surely it is the vertically integrated emissions that is important for the amount of particulate matter entering the atmosphere?

Yes, the emissions at lowest level were much larger than those at higher levels. So the time-averaged differences between the ensemble mean analysis and the prior values of the unspeciated primary sources of $PM_{2.5}$ at higher levels were negligible (See ReFig. 3). Thus we only discussed emissions at lowest level.

We have added some explanation in Lines 572 to 574, Page 21.



ReFig. 3 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 analysis and the prior values ($\mu g \cdot m^{-2}\ s^{-1}$) of the vertically integrated emissions from level 2 to level 8 averaged

over all hours from 6 to 16 October 2014.

31. P 15, L 406: "Our assimilated PM2.5 and NOx emissions were in good agreement with this trend". The DA experiments reported here cover a period of a few weeks, so how can you compare that to a trend over 15 years?

This conclusion was really arbitrary. We have removed related sentences.

32. P 17, L 470: "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)." This conclusion is baseless as Jiang et al use a different DA system (3D-VAR) with different observations (PM10) and Schwartz et al use a different domain (USA).

In the revised manuscript, the experiment of pure assimilation chemical ICs and the corresponding 48-h forecasts experiment were also performed for comparison. It seemed that the forecasts with the joint adjustment were always much better than the forecasts with only the optimized ICs for almost all the forecasts in the PRD and YRD. Please see the details in the manuscript (Lines 432 to 434, Page 16; Lines 448 to 452, Page 17; Lines 513 to 533, Page 19; Lines 620 to 622, Page 23; Lines 665 to 704, Page 25).

33. Figure 1: What is F? How is it related to Eq 1?

It was **E**. We have corrected it in Figure 1.

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 using the ensemble concentration ratios of the WRF-Chem forecast chemical concentrations and also the time smoothing operator. Hourly surface fine particulate matter ($PM_{2.5}$) observations were assimilated in this system 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. Besides, we also performed an experiment of pure assimilation chemical ICs and the corresponding 48-h forecasts experiment for comparison. 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

778 delta regions, large reduction of the Root Mean Square Errors (RMSEs) was obtained
779 for almost the entire 48-h forecast range attributed to assimilation. Especially, the
780 relative reduction in RMSE due to assimilation was about 37.5% at nighttime when
781 WRF-Chem performed comparatively worse. In the Beijing–Tianjin–Hebei region,
782 relatively smaller improvements were achieved in the first 24-h forecast. Comparing
783 to the forecasts with only the optimized ICs, the forecasts with the joint adjustment
784 were always much better for almost all the forecasts in the PRD and YRD, although
785 they were very similar in the JJJ region.
786
787 **1. Introduction**
788 Aerosol prediction by regional air quality model in heavy polluted regions is
789 challenging due to many factors. In addition to the deficiency of chemistries, the
790 uncertainties of primary and precursor emissions and the initial conditions (ICs) also
791 limit the forecast accuracy. Data assimilation (DA), which is used to improve the ICs
792 of aerosols and to optimize data on aerosol emissions, has been shown to be one of
793 the most effective ways to improve the forecasting of aerosol pollution.
794 From the perspective of reducing the uncertainties in the ICs for aerosols, recent
795 efforts have focused on assimilating aerosol observations using optimal interpolation
796 (Collins et al., 2001; Yu et al., 2003; Adhikary et al., 2008; Tombette et al., 2009; Lee
797 et al., 2013) or variational (Kahnert, 2008; Zhang et al., 2008; Benedetti et al., 2009;
798 Pagowski et al., 2010; Liu et al., 2011; Schwartz et al., 2012; Li et al., 2013; Jiang et
799 al., 2013; Saide et al., 2013) DA algorithms. Ensemble-based DA algorithms, such as
800 the ensemble Kalman filter (EnKF) (Sekiyama et al., 2010; Schutgens et al., 2010a,
801 2010b; Pagowski and Grell, 2012; Dai et al., 2014; Rubin et al., 2016; Ying, X.M., et
802 al., 2016; Yumimoto et al., 2016) and the hybrid variational-ensemble DA approach
803 (Schwartz et al., 2014) have also been applied to aerosol predictions. All these studies
804 have shown that DA is one of the most effective ways of improving aerosol

805 forecasting through assimilating aerosol observations from multiple sources (e.g.
806 ground-based observations and satellite measurements) to update the chemical ICs.
807 Numerous studies have used DA approaches to estimate or improve source
808 emissions. The EnKF is one of the most popular DA algorithms used to improve
809 estimates of aerosols and gas-phase emissions, such as NO_x, volatile organic
810 compounds, and SO_2 (van Loon et al., 2000; Heemink and Segers, 2002; Zhang et al.,
811 2005; Barbu et al., 2009; Sekiyama et al., 2010; Huneeus et al., 2012; Schutgens et al.,
812 2012; Huneeus et al., 2012, 2013; Miyazaki et al., 2014). Variational DA algorithms
813 have also been applied to constrain emissions of air pollution, such as black carbon,
814 organic carbon, dust, NH_3, SO_x and NO_x (Hakami et al., 2005; Elbern et al., 2007;
815 Henze et al., 2007, 2009; Yumimoto et al., 2007, 2008; Dubovik et al., 2008; Wang et
816 al., 2012; Guerrette and Henze, 2015). These studies have indicated that DA can
817 efficiently reduce the uncertainty in the emission inventories and lead to
818 improvements in the forecasting of air quality (Mijling and van der A, 2012).
819 The optimization of chemical ICs and pollution emissions can improve aerosol
820 forecasts and therefore further improvements are likely to be achieved by
821 simultaneously optimizing the chemical ICs and emissions. Tang et al. (2011)
822 reported that the simultaneous adjustment of the ICs of O_3, NO_x and volatile organic
823 compounds and the emissions of NO_x and volatile organic compounds produced
824 overall better performance in both the 1-h and 24-h ozone forecasts than the
825 adjustment of pure ICs or emissions. Miyazaki et al. (2012) reported that the
826 simultaneous adjustment of emissions and concentrations is a powerful approach to
827 correcting the tropospheric ozone budget and profile analyses.
828 We developed a system to adjust the chemical ICs and source emissions jointly
829 within an EnKF system coupled to the Weather Research and Forecasting with
830 Chemistry (WRF-Chem) model (Grell et al., 2005). We then applied this system to
831 assimilate hourly surface $PM_{2.5}$ measurements over China in early October 2014.
832 The remainder of the paper is organized as follows. Section 2 describes this DA
833 system in detail. Then the experimental designs are introduced in Section 3. Finally,
834 the surface $PM_{2.5}$ observations assimilation results are presented in section 4 before

concluding in section 5.

2. Methodology

2.1 Forecast model

For a chemical model like WRF-Chem, the emissions are the model forcing (or boundary condition), rather than model states. Therefore, a forecasting model, **M**, was developed to forecast the emission scaling factors (representing emissions) as well as the aerosol concentrations. This model combines the WRF-Chem model and the forecast model of emission scaling factors.

2.1.1 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. For the GOCART aerosol scheme, the aerosol species include 14 defined aerosol species and a 15th variable representing unspectiated aerosol contributions (P_{25}). The 14 defined aerosol species are sulfate, 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).

Figure 1 illustrates the model computational domain. It has 120*120 horizontal

grid scales at a 40.5 km spacing by the lambert conform map projection centered at (35 °N, 105 °E). There are 57 vertical levels with the model top at 10 hPa, about 12 layers within the planetary boundary layer (among them the lowest 8 layers were under 500 m), and the first layer centered at ~12 m.

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. The power generator emissions were interpolated for the lowest eight vertical levels (Woo et al., 2003; de meij et al., 2006; Wang et al., 2010). Other anthropogenic emissions were assigned totally to the 1st level. Emissions are very small above 500 m for all pollutants. 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.

2.1.2 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 2a shows the flowchart for the persistence forecasting operator $\mathbf{M}_{\mathrm{SF}}$.

If the ensemble members of the updated chemical fields $\mathbf{C}_{i,t-1}^{\mathrm{a}}$ and the forecast emissions $\mathbf{E}_{i,t-2}^{\mathrm{f}}$ in the previous assimilation cycle are known, then the chemical fields $\mathbf{C}_{i,t}^{\mathrm{f}}$ at time t can be generated via WRF-Chem (Figure 2b). In the actual process, $\mathbf{C}_{i,t}^{\mathrm{f}}$ were available in the previous assimilation cycle, so we did not need to perform the ensemble forecasts again. A dotted box was used in Figure 2a to indicate that the ensemble forecasts were not performed in real process. The ensemble concentration ratios $\boldsymbol{\kappa}_{i,t}, (i = 1, \dots, N)$ are then calculated using

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

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. The ensemble mean of

894 $\boldsymbol{\kappa}_{i,t}$ is,

895 $$\overline{\boldsymbol{\kappa}_t} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\kappa}_{i,t} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{C}_{i,t}^{\mathrm{f}} / \overline{\mathbf{C}_t^{\mathrm{f}}} = 1, \quad (2)$$

896 so $\boldsymbol{\kappa}_{i,t}$ are numbers distributed around 1 and with ensemble mean values of 1.

897 The ensemble spreads of $\boldsymbol{\kappa}_{i,t}, (i = 1, \dots, N)$ may be small and therefore

898 covariance inflation is used to maintain them at a certain level:

899 $$(\boldsymbol{\kappa}_{i,t})_{\mathrm{inf}} = \beta\left(\boldsymbol{\kappa}_{i,t} - \overline{\boldsymbol{\kappa}_t}\right) + \overline{\boldsymbol{\kappa}_t}, (i = 1, \dots, N)\ , \quad (3)$$

900 $\beta = 1.5$ was chosen to make ensure the ensemble spread of $(\boldsymbol{\kappa}_{i,t})_{\mathrm{inf}}$ ranged from

901 0.1 to 1.25 in this study. Same as $\boldsymbol{\kappa}_{i,t}$, the ensemble mean values of $(\boldsymbol{\kappa}_{i,t})_{\mathrm{inf}}$ are 1. It

902 is noted that perhaps there are very few negative values for $(\boldsymbol{\kappa}_{i,t})_{\mathrm{inf}}$ after inflation. A

903 quality control procedure is performed for $(\boldsymbol{\kappa}_{i,t})_{\mathrm{inf}}$ before further appliance. All

904 these negative data were set as 0.001 in this work. Then $(\boldsymbol{\kappa}_{i,t})_{\mathrm{inf}}$ were re-centered to

905 ensure the ensemble mean values of $(\boldsymbol{\kappa}_{i,t})_{\mathrm{inf}}$ were all 1.

906 As the concentrations were closely related to the emissions both locally and in

907 the upwind regions and there is no suitable dynamic model available to forecast the

908 emission scaling factors, the inflated concentration ratios $(\boldsymbol{\kappa}_{i,t})_{\mathrm{inf}}$ serve as the prior

909 emission scaling factors $\boldsymbol{\lambda}_{i,t}^{\mathrm{p}}$:

910 $$\boldsymbol{\lambda}_{i,t}^{\mathrm{p}} = (\boldsymbol{\kappa}_{i,t})_{\mathrm{inf}}, (i = 1, \dots, N), \quad (4)$$

911 The above equation is not supported according to the mass conservation equation but just for

912 the purpose to generate the ensemble emissions. Same as $(\boldsymbol{\kappa}_{i,t})_{\mathrm{inf}}$, $\boldsymbol{\lambda}_{i,t}^{\mathrm{p}}$ are numbers distributed

913 around 1. From the perspective of generating the ensemble emissions, they can play the same role

914 as other data, such as the random numbers created by using the standard normal distribution

915 function. However, there are correlations among the grid-points of $(\boldsymbol{\kappa}_{i,t})_{\mathrm{inf}}$ because $(\boldsymbol{\kappa}_{i,t})_{\mathrm{inf}}$ are

916 calculated through a short-term forecast of WRF-Chem. Thus, $\boldsymbol{\lambda}_{i,t}^{\mathrm{p}}$ have the same correlations as

917 $(\boldsymbol{\kappa}_{i,t})_{\mathrm{inf}}$. While, the random numbers are totally different. There are no correlations unless they are

generated under certain correlations.

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

$$\boldsymbol{\lambda}^{\mathrm{f}}_{i,t} = \frac{1}{M}\left(\sum_{j=t-M+1}^{t-1}\boldsymbol{\lambda}^{\mathrm{a}}_{i,j} + \boldsymbol{\lambda}^{\mathrm{p}}_{i,t}\right), (\, i = 1, \ldots, N, j = t-M+1, \ldots, t-1), \quad (5)$$

Here, M is the time window of the smooth operator. In this study, a value of $M = 4$ (hours) was chosen. According to the smooth operator, the ensemble mean values of $\boldsymbol{\lambda}^{\mathrm{f}}_{i,t}$ depend on the ensemble mean of $\boldsymbol{\lambda}^{\mathrm{a}}_{i,t-M+1}$, ⋯, $\boldsymbol{\lambda}^{\mathrm{a}}_{i,t-2}$, $\boldsymbol{\lambda}^{\mathrm{a}}_{i,t-1}$, $\boldsymbol{\lambda}^{\mathrm{p}}_{i,t}$, where the ensemble means of $\boldsymbol{\lambda}^{\mathrm{p}}_{i,t}$ are all 1. After multiple iterations, the smooth operator can give comparatively good estimation for $\boldsymbol{\lambda}^{\mathrm{f}}_{i,t}$ since anthropogenic emissions are stable at a certain time scale (Mijling et al., 2012). It is a compromise between prescribed prior emissions and letting the system propagate all observation information from one step to the next without any guidance (Peters et al., 2007), for the case $M = 4$.

The ensemble members of the emissions were calculated according to

$$\mathbf{E}_{i,t} = \boldsymbol{\lambda}_{i,t}\mathbf{E}^{\mathrm{p}}_{t}, (i = 1, \ldots, N), \quad (6)$$

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 $\mathbf{E}^{\mathrm{p}}_{t}$ is the prescribed emission, which can be obtained from the emission inventories. It is noted that the correlations among the grid-points of the prior emissions depend on $\boldsymbol{\lambda}^{\mathrm{p}}_{i,t}$. These correlations may deviate far from the truth but we have no other suitable substitute. However, the correlations among the grid-points of the forecast emissions should be more or less close to the truth due to the appliance of the smooth operator after multiple iterations.

It is noted although the method is very similar to that used by Peters et al. (2007) and Peng et al. (2015) for CO_2 emission inversion, it is still of novelty for applications in aerosol anthropogenic emissions. In Peters et al. (2007), $\boldsymbol{\lambda}^{\mathrm{p}}_{i,t}$ were all 1. And only natural CO_2 emissions (i.e., biospheric and oceanic emissions) were assimilated at the ecological scale due to the 'signal-to-noise' problem. Thus, the uncertainty of anthropogenic and other CO_2 emissions were ignored. Besides, the framework is more

advanced compared to our previous work. In Peng et al. (2015), in order to generate $\boldsymbol{\lambda}_{i,t}^{\mathrm{p}}$, a set of ensemble forecasts were performed from time *t* to *t*+1 to produce the CO_2 concentration fields, forced by the prescribed net CO_2 surface fluxes with the previous assimilated concentration fields as initial conditions. That means that the ensemble forecast were performed twice in that DA system and it was time consuming. However, in order to save computing time, we used the chemical fields $\mathbf{C}_{i,t}^{\mathrm{f}}$ available in the previous assimilation cycle to calculate $\boldsymbol{\lambda}_{i,t}^{\mathrm{p}}$ in this work. Thus, WRF-Chem runs to forecast only once during a DA cycle.

2.2 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). The traditional EnKF with perturbed observations (Evensen 1994) introduces sampling errors by perturbing the observations. In contrast to the traditional EnKF, the EnSRF (Whitaker and Hamill, 2002) and the Ensemble Adjustment Kalman Filter (EAKF, developed by Anderson, 2001) obviate the need to perturb the observations. The local ensemble Kalman filtering (LEKF), a kind of EnSRF, was presented by Ott et al. (2002, 2004). It was computationally more efficient compared to the traditional EnKF, since it simultaneously assimilates the observations within a spatially local volume independently. The local Ensemble Transform Kalman Filter (LETKF, Hunt, 2007) integrates the advantages of the Ensemble Transform Kalman Filter (ETKF, developed by Bishop et al., 2001) and the LEKF. The computational cost of LETKF is much lower than that of the original LEKF because the former does not require an orthogonal basis. Though LETKF has more advantages, we still chose the same EnSRF as Schwartz et al. (2014) because we did not need to extend it to analyzing aerosol ICs, very similar to Schwartz et al. (2014).

Following the notation of Ide et al. (1997), given an *m*-dimensional background model forecast vector $\mathbf{x}^{\mathrm{b}}$, a *p*-dimensional observation vector $\mathbf{y}^{\mathrm{o}}$ and an operator $\mathbf{H}$

that converts the model state to the observation states, we expressed the 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

$$\bar{\mathbf{x}}^{\mathrm{a}} = \bar{\mathbf{x}}^{\mathrm{b}} + \mathbf{K}(\mathbf{y}^{\mathrm{o}} - \mathbf{H}\bar{\mathbf{x}}^{\mathrm{b}}), \quad (7)$$

$$\mathbf{x}'^{\mathrm{a}} = \mathbf{x}'^{\mathrm{b}} + \widetilde{\mathbf{K}}(\mathbf{y}'^{\mathrm{o}} - \mathbf{H}\mathbf{x}'^{\mathrm{b}}), \quad (8)$$

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

$$\mathbf{K} = \mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}, \quad (9)$$

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

$$= \left(\mathbf{1} + \sqrt{\mathbf{R}/(\mathbf{H}\mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}} + \mathbf{R})}\right)^{-1}\mathbf{K}, \quad (10)$$

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 $\mathbf{H}\mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}}$ will be approximated using the background ensemble; namely,

$$\mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}} = \frac{1}{N-1}\sum_{i=1}^{N}\mathbf{x}'^{\mathrm{b}}(\mathbf{H}\mathbf{x}'^{\mathrm{b}})^{T} \quad (11)$$

$$\mathbf{H}\mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}} = \frac{1}{N-1}\sum_{i=1}^{N}\mathbf{H}\mathbf{x}'^{\mathrm{b}}(\mathbf{H}\mathbf{x}'^{\mathrm{b}})^{T}. \quad (12)$$

In equations (11) and (12), 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. $\mathbf{x} = [\mathbf{C}, \boldsymbol{\lambda}]^{\mathrm{T}}$. In this study, the state variables of the analysis of the ICs were the 15 WRF-Chem/GOCART aerosol variables, same as that reported by Schwartz et al. (2012). The state variables of the emission scaling factors include $\boldsymbol{\lambda}_{\mathrm{PM2.5}}$, $\boldsymbol{\lambda}_{\mathrm{SO2}}$, $\boldsymbol{\lambda}_{\mathrm{NO}}$ and $\boldsymbol{\lambda}_{\mathrm{NH3}}$ and are described in section 2.3.1. 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.

In this work, a 50-member ensemble was chosen, following Schwartz et al. (2012) and Whitaker and Hamill (2002). Covariance localization forced EnSRF

analysis increments to zero 1280 km from an observation in the horizontal and one scale height to reduce spurious correlations due to sampling error for all control variables, similar to Pagowski et al., (2012) and Schwartz et al., (2012, 2014). In addition, posterior (after assimilation) multiplicative inflation following Whitaker and Hamill (2012) was applied aiming to maintain ensemble spread for only the concentration analysis. The inflation factor $\alpha = 1.2$ was chosen as Pagowski et al., (2012) and Schwartz et al., (2012, 2014). Additive or prior inflation was not employed. As for the emission scaling factor $\boldsymbol{\lambda}$, the inflation was not used at this step.

2.3 Data assimilation system

2.3.1 State variables

As stated in section. 2.2, the state variables of the analysis of the ICs were the 15 WRF-Chem/GOCART aerosol variables. The $PM_{2.5}$ observation operator was the same as that described by Schwartz et al. (2012) and expressed as

$$\mathbf{y}^{\mathrm{f}} = \boldsymbol{\rho}_{\mathrm{d}}[\mathbf{P_{25}} + 1.375\mathbf{S} + 1.8(\mathbf{OC_1} + \mathbf{OC_2}) + \mathbf{BC_1} + \mathbf{BC_2} + \mathbf{D_1} + 0.286\mathbf{D_2} + \mathbf{S_1} + 0.942\mathbf{S_2}], \quad (13)$$

where $\boldsymbol{\rho}_{\mathrm{d}}$ represents the dry air density, which is multiplied by the mixing ratios of aerosol species (in $\mu g\ kg^{-1}$) to convert the units to $\mu g\ m^{-3}$ for consistency with the observations.

From the perspective of the optimization of emissions, four species of emission scaling factors ($\boldsymbol{\lambda}_{PM2.5}$, $\boldsymbol{\lambda}_{SO2}$, $\boldsymbol{\lambda}_{NO}$ and $\boldsymbol{\lambda}_{NH3}$) were also considered as the state variables of the DA system. Atmospheric inorganic aerosols are not only from the primary emissions, but also from secondary processes- 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 $PM_{2.5}$. It is noted that for the optimization of the emission scaling factors, $\mathbf{M}_{SF}$ serves as the forecast model and the observation operator reflects the combined information of emissions (in the format of $\boldsymbol{\lambda}$ in

equation (6)), the physics and chemistry processes in WRF-Chem simulations and the transformation $PM_{2.5}$ from model space to observation space (equation (13)).

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 sulfate and nitrate emissions in the model are as below:

$$\mathbf{E}_{PM2.5i}:\mathbf{E}_{PM2.5j} = 1:4, \quad (14)$$

$$\mathbf{E}_{SO4i}:\mathbf{E}_{SO4j} = 1:4, \quad (15)$$

$$\mathbf{E}_{NO3i}:\mathbf{E}_{NO3j} = 1:4, \quad (16)$$

$$\mathbf{E}_{SO4i} + \mathbf{E}_{SO4j} = a * (\mathbf{E}_{PM2.5i} + \mathbf{E}_{PM2.5j} - \mathbf{E}_{EC} - \mathbf{E}_{ORG}), \quad (17)$$

$$\mathbf{E}_{NO3i} + \mathbf{E}_{NO3j} = b * \left(\mathbf{E}_{PM2.5i} + \mathbf{E}_{PM2.5j} - \mathbf{E}_{EC} - \mathbf{E}_{ORG}\right), \quad (18)$$

where $\mathbf{E}_{EC}$ represents elemental carbon and $\mathbf{E}_{ORG}$ 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 ($\mathbf{E}_{PM2.5i}$, $\mathbf{E}_{PM2.5j}$, $\mathbf{E}_{SO4i}$, $\mathbf{E}_{SO4j}$, $\mathbf{E}_{NO3i}$, and $\mathbf{E}_{NO3j}$), which may have larger uncertainties in heavy polluted events, were updated according to the variation of $\boldsymbol{\lambda}_{PM2.5}$. $\mathbf{E}_{PM2.5i}$ and $\mathbf{E}_{PM2.5j}$ were directly updated according to the variation in $\boldsymbol{\lambda}_{PM2.5}$. The emissions ($\mathbf{E}_{SO4i}$, $\mathbf{E}_{SO4j}$, $\mathbf{E}_{NO3i}$ and $\mathbf{E}_{NO3j}$) were also updated according to the variations in $\mathbf{E}_{PM2.5i}$ and $\mathbf{E}_{PM2.5j}$.

$\mathbf{E}_{EC}$ and $\mathbf{E}_{ORG}$ of the anthropogenic emissions were not assimilated, which is a limitation in this work. Besides, emissions of dust and sea salt were not assimilated. It is true that these emissions are also important for the atmosphere aerosol. The reason we did not assimilate $\mathbf{E}_{EC}$ and $\mathbf{E}_{ORG}$ is that only the $PM_{2.5}$ measurements are used in this DA experiment. However, the sources of the aerosols (especially organic aerosols) are so complex that our knowledge of their formation mechanisms is far from clear. Though it is technically possible to have all emissions

assimilated, with such limited observations adding more control variables would cause much more uncertainties in the system which might lead to unreasonable analysis.

2.3.2 Procedure for the DA system

Figure 2 (b) shows the workflow of the DA system. The steps in this workflow are as follows.

(1) The persistence forecasting operator $\mathbf{M}_{SF}$ is applied to forecast the background fields of the emission scaling factors $\boldsymbol{\lambda}^{f}_{PM2.5}$, $\boldsymbol{\lambda}^{f}_{SO2}$, $\boldsymbol{\lambda}^{f}_{NO}$ and $\boldsymbol{\lambda}^{f}_{NH3}$. 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 $\boldsymbol{\lambda}^{p}_{PM2.5}$, $\boldsymbol{\lambda}^{p}_{SO2}$, $\boldsymbol{\lambda}^{p}_{NO}$ and $\boldsymbol{\lambda}^{p}_{NH3}$. The background scaling factors are then generated using equation (5).

(2) The ensemble members of the emissions, $\mathbf{E}^{f}_{PM2.5i}$, $\mathbf{E}^{f}_{PM2.5j}$, $\mathbf{E}^{f}_{SO2}$, $\mathbf{E}^{f}_{NO}$ and $\mathbf{E}^{f}_{NH3}$, are prepared according to equation (6). The corresponding emissions of $\mathbf{E}^{f}_{SO4i}$, $\mathbf{E}^{f}_{SO4j}$, $\mathbf{E}^{f}_{NO3i}$ and $\mathbf{E}^{f}_{NO3j}$ are obtained based on equations (15–18). Other inorganic species of the anthropogenic emission, such as $\mathbf{E}_{EC}$ and $\mathbf{E}_{ORG}$, are not perturbed for WRF-Chem, which is a limitation of this manuscript. However, other anthropogenic emissions, such as $\mathbf{E}_{PM2.5}$, $\mathbf{E}_{SO4}$ and $\mathbf{E}_{NO3}$, are much larger than $\mathbf{E}_{EC}$ and $\mathbf{E}_{ORG}$ in most area of China, and the ensemble spreads of the aerosol concentrate largely dependent on the uncertainties of those anthropogenic emissions. Besides, model errors raised from the meteorology, the emission and the chemical model itself are compensated to some extent through the use of multiplicative inflation. In other words, the ensemble spread of the concentrations can be kept at a certain level though $\mathbf{E}_{EC}$ and $\mathbf{E}_{ORG}$, are not perturbed.

Natural emissions, such as dust and sea salt emissions were not perturbed explicitly when the forecast emissions were generated. However, emissions of dust and sea salt were parameterized within the GOCART model (Chin et al., 2002). Within the DA system, varying meteorology across the members implicitly perturbed dust and sea salt emissions.

(3) Forced by the changed emissions ($\mathbf{E}_{PM2.5i}$, $\mathbf{E}_{PM2.5j}$, $\mathbf{E}_{SO2}$, $\mathbf{E}_{NO}$, $\mathbf{E}_{NH3}$, $\mathbf{E}_{SO4i}$, $\mathbf{E}_{SO4j}$, $\mathbf{E}_{NO3i}$ 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 (13). At this time, the state vector $\mathbf{x}^{f} = [\mathbf{C}^{f}, \boldsymbol{\lambda}^{f}]^{T}$ was prepared.

(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 (6, 15–18) 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 1 shows the locations of 77 measurement sites used for the $PM_{2.5}$ assimilation experiment and 77 independent sites used for 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 (9) includes contributions from measurement and representation errors. Similar to the work of Schwartz et al. (2012), 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. Following Elbern et al. (2007) and Pagowski et al. (2010), Schwartz et al. (2012), 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, $r = 0.5$), 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 following Elbern et al. (2007), since we do not know the station type that used in this work). 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 $\mathbf{R}$ 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 the ensemble mean of the first guess departure exceeding 100 μg m^{-3} are also omitted. The numbers of the observations were about 17700. Among them 8 observations were discarded because they were larger than 800 μg m^{-3} and 243 (around 1.5%) were discarded due to the latter reasons.

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.

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 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 were idealized profiles embedded within the WRF/Chem model.

Fifty-member emissions were created by adding random noise to the anthropogenic emissions, same as reported by Schwartz et al. (2014),

$$\mathbf{E}_{ip}^{*}(\eta,t) = \mathbf{E}_{p}(\eta,t) + \boldsymbol{W}_{ip}\boldsymbol{\sigma}_{p}^{\mathrm{E}}(\eta,t)$$

where $\mathbf{E}_{ip}^{*}(\eta,t)$ is the ith ensemble member for the pth emissions variable at the ηth grid point and the tth hour, $\mathbf{E}_{p}$ is the unperturbed emissions. The term $\boldsymbol{\sigma}_{p}^{\mathrm{E}}$ is the standard deviation of all $\mathbf{E}_{p}$ values and in the horizontally adjacent points of grid box η at and within 2 h of t. $\boldsymbol{W}$ is a weight that was randomly drawn from a standard Gaussian distribution and varied for each ensemble member and variable but was spatially and temporally constant. No correlations between emissions variables were considered, which is a limitation of this approach. For possible negative perturbed emissions, they were set as $\mathbf{E}_{ip}^{*}(\eta,t) = 0.001 * \mathbf{E}_{p}(\eta,t)$. This will increase the prescribed emissions more or less. However, only very few data were negative. So, this influence can be negligible.

Before the first DA cycle, a 50-member ensemble of four-day WRF-Chem forecasts was performed from 0000 UTC 1 October to 2300 UTC 4 October 2014 using the perturbed ICs at 0000 UTC 1 October 2014, the corresponding perturbed LBCs and the emissions. Then a 50-member ensemble aerosol forecasts at 0000 UTC 5 October 2014 were produced.

4.2 Assimilation experiments

Two DA experiments were performed. One was the pure assimilation of chemical ICs (hereafter expC), the others was the joint adjustment of chemical ICs and source emissions (hereafter expJ). Both DA experiments had same settings except for the emissions. They were conducted from 0000 UTC 5 October 2014 to 0000 UTC 16 October 2014. The assimilation cycle interval was 1 h.

In the first DA cycle in expJ, the first 50 ensemble chemical fields were drawn from the WRF-Chem ensemble forecasts valid at 0000 UTC 5 October 2014, as described in section 4.1. Using the ensemble aerosol forecasts, the prior emission scaling factors $\boldsymbol{\lambda}_{i,t}^{\mathrm{p}}$ at 2300 UTC 4 October 2014 were calculated. $\boldsymbol{\lambda}_{i,t}^{\mathrm{p}}$ were used directly as $\boldsymbol{\lambda}_{i,t}^{\mathrm{f}}$ for the first 5 assimilation cycles (after 5 assimilation cycles, the system has been initialized, all future scaling factors could be created using the persistence forecasting operator $\mathbf{M}_{\mathrm{SF}}$). Then, the state vector $\mathbf{x}^{\mathrm{f}} = [\mathbf{C}^{\mathrm{f}}, \boldsymbol{\lambda}^{\mathrm{f}}]^{\mathrm{T}}$ was prepared. And after that, the DA cycle started.

In expC, the first chemical fields were also drawn from the WRF-Chem ensemble forecasts valid at 0000 UTC 5 October 2014. Then, the state vector $\mathbf{x}^{\mathrm{f}} = [\mathbf{C}^{\mathrm{f}}]^{\mathrm{T}}$ was prepared and the DA cycle started.

At the WRF-Chem forecast step of the subsequent assimilation cycles for both experiments, 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 for all experiments were idealized profiles embedded within the WRF/Chem model. 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.

As stated in the first paragraph in this section, the settings of expC were the same as those in expJ except for the emissions. In expJ, the ensemble anthropogenic emissions were generated by using emission scaling factors. While in expC, the ensemble anthropogenic emissions were prepared by adding random noise, as stated in 4.1.

4.3 Control experiment

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

Statistics for both expJ and expC were computed using the ensemble mean prior (background) and posterior (analysis) fields (average of the 50-member ensemble). The ensemble performances were first examined. Output from the first day of the cycling DA configurations was excluded from all verification statistics to allow the ensemble fields to ‘‘spin up’’ from the initial ensemble.

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 1): 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 for expJ. 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. For expC, the characteristics of the prior ensemble mean RMSE and the total spread for $PM_{2.5}$ were very similar to that for the joint DA experiment.

The magnitudes of the ensemble spread of the emission scaling factors of the joint DA experiment were important for emission inversion. They were very stable throughout the ~10 day experiment period, which indicates that $\mathbf{M}_{SF}$ can generate stable artificial data to generate the ensemble emissions. For $\boldsymbol{\lambda}^{f}_{PM2.5}$, they ranged from 0.25 to 1 in most model area. Figure 3d shows the area-averaged time series extracted from the ensemble spread of $\boldsymbol{\lambda}^{f}_{PM2.5}$. It shows that the ensemble spread was stably distributed around 0.5, which indicates that the uncertainty of the ensemble emissions was about 50%.

5.2 Impact on aerosol ICs

To evaluate quantitatively the impact of the ensemble assimilation system on the ICs, 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 independent observations over all the analyses from 6 to 16 October 2014. Table 1 shows that the bias magnitudes of the control run were 15.9 and 20.6 $\mu g\ m^{-3}$ 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 −18.0 μg m^{-3}. The RMSEs of the control run were 81.6, 30.6 and 31.8 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 for both DA experiment. For expJ, both the maximum bias and the RMSE were obtained in the JJJ region, and were -10.3 and 66.9 μg m^{-3}, respectively. The CORR increased from 0.79, 0.60, and 0.62 to 0.83, 0.85, and 0.80 for the JJJ, YRD and PRD, respectively. The statistics of expC were very similar to those of expJ. The bias and the RMSE in the JJJ region were -12.2 and 64.0 μg m^{-3}, respectively. And the CORR were 0.85, 0.80, and 0.80 for the JJJ, YRD and PRD, respectively. These results indicate that the initial $PM_{2.5}$ fields can be adjusted efficiently by the EnSRF.

Then the analysis increments (i.e. $\bar{\mathbf{x}}^{a} - \bar{\mathbf{x}}^{b}$) were investigated to show the direct impact of $PM_{2.5}$ DA. They are determined by both the observation increments and the relative magnitudes of the forecast error and the observation error, based on Equation (7). From Figure 4(a), (e) and (f), the increments of both assimilation experiments were distributed around the observations as expected. However, the impact of assimilating $PM_{2.5}$ observations was not limited to the areas where observations were located, observations information was also transported to other areas through the WRF-Chem forecast. Besides, the ensemble forecasts also partly contributed to the spatial distribution of the $PM_{2.5}$ mass. Therefore, the spatial distributions of the $PM_{2.5}$ mass in both assimilation experiments were significantly different from the control run (see Figure 4(b), (c) and(d)), which suggest that assimilation $PM_{2.5}$ observations impacts greatly on the aerosol ICs. The $PM_{2.5}$ mass magnitude of both assimilation experiments were smaller than that of the control run at the lowest model level in the YRD, the PRD and in central China. Conversely, positive differences (analysis minus control) were gained in the JJJ region and in northeast China. These indicated the reduction of the overestimation or underestimation of the WRF-Chem simulation over these regions with data assimilation.

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 forecast emission scaling factors, the optimized emission scaling factors, the prior emissions and the optimized emissions. Figure 5 shows that $\boldsymbol{\lambda}^{f}_{PM2.5}$ were changed along with $\boldsymbol{\lambda}^{a}_{PM2.5}$. This indicates that observation information ingested from the previous observations was incorporated through the usage of the time smooth operator.

Figure 5 also shows that although the prior emissions $\mathbf{E}^{p}_{PM2.5}$ 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 $\boldsymbol{\lambda}^{a}_{PM2.5}$ 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 $\mu g\ m^{-2}\ s^{-1}$ over all hours. After assimilation, the time-averaged optimized $\mathbf{E}^{a}_{PM2.5}$ decreased to 0.107 $\mu g\ m^{-2} s^{-1}$, about 15.6% lower than the prior value. In the PRD, the prior $\mathbf{E}^{p}_{PM2.5}$ was about 0.10 $\mu g\ m^{-2}\ s^{-1}$. The time-averaged optimized $\mathbf{E}^{a}_{PM2.5}$ decreased to 0.066 $\mu g \cdot m^{-2}\ s^{-1}$, 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. However, the $PM_{2.5}$ measurements network was still spatially sparse and heterogeneous in this work. Almost all measurements were located in the city and no data available in the rural. Meanwhile, the crop residues burning always occur in the rural region. Therefore, the $PM_{2.5}$ measurements network can only capture the burning information a few hours later. Hence, 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). Maybe a Kalman smoother would have been a better system to solve this problem.

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

1318 the NH_3 emissions increased by 16.7, 7.8 and 7.5% in the JJJ, YRD and PRD regions,
1319 respectively.

1320 Figure 6 shows the spatial distribution of the time-averaged scaling factors
1321 $\boldsymbol{\lambda}^{a}_{PM2.5}$ at the lowest model level over all hours from 6 to 16 October 2014, since the
1322 emissions at higher levels were so small that the impact of assimilating $PM_{2.5}$
1323 observations was negligible. Figure 7 shows the distribution of $\mathbf{E}^{p}_{PM2.5}$ and the
1324 time-averaged differences between the ensemble mean of the assimilation and the
1325 prior values.

1326 These patterns are consistent with those in Figure 5. Negative differences were
1327 obtained in most areas of the YRD and PRD, indicating that the $PM_{2.5}$ DA primarily
1328 decreased the $PM_{2.5}$ emissions. Conversely, positive differences were obtained in
1329 South Hebei, North Henan and Southeast Shanxi provinces, indicating that DA
1330 increased the $PM_{2.5}$ emissions.

1331 As the economy in China has developed, the spatiotemporal distribution of emissions has
1332 changed as a result of changes in energy consumption, the structure of the energy market and
1333 advances in technology. Therefore although this inventory of emissions may have correctly
1334 described anthropogenic emissions in 2006 when it was constructed, it is not representative of the
1335 anthropogenic emissions in 2014. Theoretically, the assimilated emissions should reduce the
1336 uncertainty in the prior emissions as a result of the application of observations. Different from the
1337 situations that standard national emission inventories were reported by government in USA,
1338 European or other countries, the rapid economic development and complexity of emission sources
1339 in China lead to large uncertainties in the current emission inventories even for the latest version.
1340 Thus it's impossible for us to conduct the direct evaluation on emissions.

1341 Although we had no direct emission observations to evaluate the analyzing emissions, which
1342 was a challenging to many emission inversion research teams (e.g. Tang et al, 2011; Miyazaki et
1343 al., 2012; Ding et al., 2015; Mclinden et al., 2016; etc.), the improvement of emissions can be
1344 verified in terms of two aspect, the diurnal variation and the location of increased emissions. The
1345 diurnal variation in the assimilated emissions verified this statement to some extent. Especially in
1346 the PRD and YRD, $\mathbf{E}^{a}_{PM2.5}$ in the daytime were always larger than those in the night, which

agreed well with Olivier et al. (2003), the WRAP (2006) and Wang et al. (2010). In addition, the locations of the larger values for the optimized $E^{a}_{PM2.5}$ in the JJJ region was in good agreement with the place of the crop residues burning *traced by the* environmental satellite of China. There were 10, 231, 37 and 3 *crop residue burning spots in Hebei, Henan, Shandong and Shanxi province respectively* from 5 to 11 October 2014 and 7, 20, 5 and 21 *respectively* from 12 to 18 October 2014 (Weekly Crop Residue Burning Monitoring Report traced by Environmental Satellite, *2015a*, *2015b*).

However, the analyzing emissions are only a mathematical optimum. They are influenced greatly by the model errors and the observation errors. In addition, only surface $PM_{2.5}$ observations were applied in this work, which may lack abundant constraint on the sources of the secondary aerosol precursors. More observations are needed to obtain reliable emissions for the sources of the gas-phase precursors.

5.4 Verification of aerosol forecasting

For the assimilation experiment, 48-h forecasts were performed at each 0000 UTC from 6 to 16 October 2014 with the hourly forecast output for both assimilation experiments. For the verification forecasting experiment for expJ (hereafter fcJ), the ensemble mean of the analyzed ICs and emissions of expJ were used in this longer-range model forecast. For the verification forecasting experiment for expC (hereafter fcC), the ensemble mean of the analyzed ICs of expC and the prescribed anthropogenic emissions were used.

In order to get a more visualized picture of the impact of DA for both assimilation experiments, time series of the hourly $PM_{2.5}$ extracted from the analysis (AN), the control run (CT) and the hourly output of 48-h forecast (fc24 for the first day forecast and fc48 for the second day forecast) were compared with the observations (OBS) for three megacities Beijing, Shanghai and Guangzhou, respectively (Figure 8). As expected, the time series of the analysis (also the background) 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 bias and the RMSE of the surface $PM_{2.5}$ forecasts as a function of forecast range was then calculated against the independent observations for the three sub-regions (Figure 9). Both the bias and 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 29 $\mu g \cdot m^{-3}$ for bias and 37 $\mu g \cdot m^{-3}$ for RMSEs) and at 2300 UTC in the PRD (about 36 $\mu g \cdot m^{-3}$ for bias and 41 $\mu g \cdot m^{-3}$ for RMSEs), likely indicating significant systematic forecast errors at these times. From 0300 to 0900 UTC, the bias (about 1 $\mu g \cdot m^{-3}$ in the YRD and -5 $\mu g \cdot m^{-3}$ in the PRD) and the RMSE values (about 14 $\mu g \cdot m^{-3}$ in the YRD and 16 $\mu g \cdot m^{-3}$ in the PRD) 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 bias (about -20 $\mu g \cdot m^{-3}$)and the RMSEs (about 50 $\mu g \cdot m^{-3}$) were always large as a result of a heavy pollution event. After assimilation, both the magnitude of the bias and the RMSEs decreased sharply. Especially in in YRD and PRD, most bias ranged from -5 to 5 $\mu g \cdot m^{-3}$ and most RMSEs ranged from 11 to 14 $\mu g \cdot m^{-3}$, 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 for expJ. Large reduction of the magnitude of the bias and 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 37.5%. 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, 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.

As for expC, it seemed that large improvements in the surface $PM_{2.5}$ forecasts were gained through the adjustment of the ICs in PRD from 10- to 23-h and from 34- to 47-h. Large reduction of the magnitude of the bias and the RMSEs due to assimilation can be seen during this period. The relative reduction in RMSE ranged from 25% to 37.5%. However, the forecasts deviated much from the observations for 3- to 9-h and 27- to 33-h forecast ranges. One of the reason may be that the adjustment of the ICs decreased the analysis field too much on the whole since the WRF-Chem forecast aerosol mass was systematically overestimated in PRD (see Figure 4, Figure 8f and Figure 9e). While this aerosol mass overestimation might be also due to the possibly overestimated emissions in some time periods (not all-day long) which are not corrected in the simulation. So the over-adjusted ICs compensated the unadjusted emissions in some period but also lead to the negative biases for the periods when emission is not overestimated or underestimated. The other factor was the diurnal variation. It is very clear that $PM_{2.5}$ mass gradually decreased with time from 0000 UTC to 0008 UTC and then obtained the smallest value. After that it increased with time from 0009 UTC to 0023 UTC obtained the largest value at about 0000 UTC. Both reasons led to the systematically underestimation of $PM_{2.5}$ mass of fcC from 3- to 9-h and from 27- to 33-h, though maybe the aerosol ICs were very close to the observations. Therefore, both the magnitude of the bias and the RMSEs of the fcC were larger than those of the control run. In addition, $PM_{2.5}$ forecasts of the fcC were benefit much from the diurnal variation and the adjustment of the ICs from 10- to 23-h and from 34- to 47-h. As a consequence, the magnitude of the corresponding bias and the RMSEs of the fcC were smaller than those of the control run. Similar statics characteristics were also gained in YRD. But the improvements were comparatively small from 10- to 23-h and from 34- to 47-h. However, the performance of fcJ was always much better than that of the fcC for almost the entire 48-h forecast range in the PRD and YRD.

Both DA systems 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 1) and the measurement coverage was much sparser than in the YRD or PRD.

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. Besides, the experiment of pure assimilation chemical ICs and the corresponding 48-h forecasts experiment were also performed for comparison. 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 may be attributed to the sparse measurement coverage and the deficiencies in the model system for forecasting heavy pollution. Comparing to the forecasts with only the optimized ICs, the forecasts with the joint adjustment were always much better for almost all the forecasts in the PRD and YRD. However, In the JJJ region, they were

very similar.

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.

1473 References

1474 Anderson, J.L.: An Ensemble Adjustment Kalman Filter for Data Assimilation, Mon.Weather
1475 Rev., 129, 2884–2903, 2001.

1476 Adhikary, B., Kulkarni, S., Dallura, A., Tang, Y., Chai, T., Leung, L. R., Qian, Y., Chung, C. E.,
1477 Ramanathan, V., and Carmichael, G. R.: A regional scale chemical transport modeling of
1478 Asian aerosols with data assimilation of AOD observations using optimal interpolation
1479 technique, Atmos. Environ., 42, 8600–8615, doi:10.1016/j.atmosenv.2008.08.031, 2008.

1480 Barbu, A. L., Segers, A. J., Schaap, M., Heemink, A.W., and Builtjes, P. J. H.: A multi-component
1481 data assimilation experiment directed to sulphur dioxide and sulphate over Europe, Atmos.
1482 Environ., 43, 1622–1631, 2009.

1483 Benedetti, A., Morcrette, J., Boucher, O., Dethof, A., Engelen, R., Fisher, M., Flentje, H.,
1484 Huneeus, N., Jones, L., and Kaiser, J.: Aerosol analysis and forecast in the European Centre
1485 for Medium-Range Weather Forecasts Integrated Forecast System: 2. Data assimilation, J.
1486 Geophys. Res, 114, D13205, doi:10.1029/2008JD011115, 2009.

1487 Bishop, C. H., Etherton, B. J., and Majumdar, S. J.: Adaptive sampling with the ensemble
1488 transform Kalman filter. Part I: Theoretical aspects, Mon. Weather Rev., 129, 420–436, 2001.

1489 Chen, D., Liu, Z., Fast, J., and Ban, J.: Simulations of Sulfate-Nitrate-Ammonium (SNA) aerosols
1490 during the extreme haze events over Northern China in October 2014, Atmos. Chem. Phys.
1491 Discuss., doi:10.5194/acp-2016-222, in review, 2016.

1492 Chin, M., Rood, R. B., Lin, S. J., Muller, J. F., and Thompson, A. M.: Atmospheric sulfur cycle
1493 simulated in the global model GOCART: Model description and global properties, J.
1494 Geophys. Res.-Atmos., 105, 24671–24687, 2000.

1495 Chin, M., Ginoux, P., Kinne, S., Torres, O., Holben, B.N., Duncan, B. N., Martin, R.V., Logan,
1496 J.A., Higurashi, A., and Nakajima, J.: Tropospheric aerosol optical thickness from the
1497 GOCART model and comparisons with satellite and Sun photometer measurements, J.
1498 Atmos. Sci., 59(3), 461–483, 2002.

1499 Collins, W. D., Rasch, P. J., Eaton, B. E., Khattatov, B. V., and J.-F. Lamarque, J.-F.: Simulating
1500 aerosols using a chemical transport model with assimilation of satellite aerosol retrievals:
1501 Methodology for INDOEX, J. Geophys. Res., 106, 7313–7336, 2001.

1502 de Meij, A., Krol, M., Dentener, F., Vignati, E., Cuvelier, C., and Thunis, P.: The sensitivity of

aerosol in Europe to two different emission inventories and temporal distribution of emissions, Atmos. Chem. Phys., 6, 4287-4309, doi:10.5194/acp-6-4287-2006, 2006.

Dai, T., Schutgens, N.A.J., Goto, D. Shi, G.Y., Nakajima, T.: Improvement of aerosol optical properties modeling over Eastern Asia with MODIS AOD assim- ilation in a global non-hydrostatic icosahedral aerosol transport model, Environ. Pollut., 195, 319–329, 2014.

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.

Ding, J., van der A, R. J., Mijling, B., Levelt, P. F., and Hao, N.: NO_x emission estimates during the 2014 Youth Olympic Games in Nanjing, Atmos. Chem. Phys., 15, 9399-9412, doi:10.5194/acp-15-9399-2015, 2015.

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. Phys., 8, 209–250, doi:10.5194/acp-8-209-2008, 2008

Elbern, H., Strunk, A., Schmidt, H., and Talagrand, O.: Emission rate and chemical state estimation by 4-dimensional variational inversion, Atmos. Chem. Phys., 7, 3749–3769, doi:10.5194/acp- 7-3749-2007, 2007.

Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistcs, J. Geophys. Res., 99(C5), 10143–10162, 1994.

Freitas, S. R.; Longo, K. M.; Alonso, M. F.; Pirre, M.; Marecal, V.; Grell, G.; Stockler, R.; Mello, R. F.; Sánchez Gácita, M.. PREP-CHEM-SRC 1.0: a preprocessor of trace gas and aerosol emission fields for regional and global atmospheric chemistry models. Geoscientific Model Development, v. 4, p. 419-433, 2011.

Ginoux, P., Chin, M. Tegen, I., Prospero, J. M., Holben, B., Dubovik, O., and Lin, S.-J.: Sources and distributions of dust aerosols simulated with the GOCART model, J. Geophys. Res., 106, 20,255–20,273, doi:10.1029/2000JD000053, 2001.

Grell, G., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock, W. C., and Eder, B.: Fully coupled “online” chemistry within the WRF model, Atmos. Environ., 39, 6957–6975, doi:10.1016/j.atmosenv.2005.04.027, 2005.

Guenther, A., Hewitt, C. N., Erickson, D., Fall, R., Geron, C., Graedel, T., Harley, P., Klinger, L.,

1533 Lerdau, M., McKay, W., Pierce, T., Scholes, B., Steinbrecher, R., Tallamraju, R., Taylor, J.,
1534 and Zimmerman, P.: A global model of natural volatile organic compound emissions, J.
1535 Geophys. Res., 100, 8873–8892, doi:10.1029/94JD02950, 1995.
1536 Guerrette, J. J. and Henze, D. K.: Development and application of the WRFPLUS-Chem online
1537 chemistry adjoint and WRFDA-Chem assimilation system, Geosci. Model Dev., 8,
1538 1857-1876, doi:10.5194/gmd-8-1857-2015, 2015.
1539 Hakami, A., Henze, D. K., Seinfeld, J. H., Chai, T., Tang, Y., Carmichael, G. R., and Sandu, A.:
1540 Adjoint inverse modeling of black carbon during the Asian Pacific Regional Aerosol
1541 Characterization Experiment, J. Geophys. Res.-Atmos., 110, D14301,
1542 doi:10.1029/2004JD005671, 2005.
1543 Heemink, A.W., and Segers, A.J.: Modeling and prediction of environmental data in space and
1544 time using Kalman filtering, Stoch. Environ. Res. Risk Assess. 16 (3), 225–240, 2002.
1545 Henze, D. K., Hakami, A., and Seinfeld, J. H.: Development of the adjoint of GEOS-Chem,
1546 Atmos. Chem. Phys., 7, 2413–2433, doi:10.5194/acp-7-2413-2007, 2007.
1547 Henze, D. K., Seinfeld, J. H., and Shindell, D. T.: Inverse modeling and mapping US air quality
1548 influences of inorganic PM2.5 precursor emissions using the adjoint of GEOS-Chem, Atmos.
1549 Chem. Phys., 9, 5877–5903, doi:10.5194/acp-9-5877-2009, 2009.
1550 Houtekamer, P. L., Mitchell, H. L., Pellerin, G., Buehner, M., Charron, M., Spacek, L., and
1551 Hansen, B.: Atmospheric data assimilation with an ensemble Kalman filter: Results with real
1552 observations, Mon. Weather Rev., 133, 604–620, 2005.
1553 Ide, K., Courtier, P., Ghil, M., and Lorenc, A. C.: Unified notation for data assimilation:
1554 operational, sequential and variational, J. Meteorol. Soc. Japan, 75, 181–189, 1997.
1555 Jiang, Z., Liu, Z., Wang, T., Schwartz, C. S., Lin, H.-C., and Jiang, F.: Probing into the impact of
1556 3DVAR assimilation of surface PM10 observations over China using process analysis, J.
1557 Geophys. Res. Atmos., 118, 6738–6749, doi:10.1002/jgrd.50495, 2013.
1558 Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., Miller, J.
1559 B., Bruhwiler, L. M. P., Petron, G., Hirsch, A. I., Worthy, D. E. J., van der Werf, G.
1560 R.,Randerson, J. T., Wennberg, P. O., Krol, M. C., Tans, P. P.: An atmospheric perspective
1561 on North American carbon dioxide exchange: CarbonTracker, P. Natl. Acad. Sci. USA, 104,
1562 18925–18930, 2007.

1563 Kahnert, M.: Variational data analysis of aerosol species in a regional CTM: Background error
1564 covariance constraint and aerosol optical observation operators, Tellus, Ser. B, 60, 753–770,
1565 doi:10.1111/j.1600-0889.2008.00377, 2008.
1566 Kleist, D. T., Parrish, D. F., Derber, J. C., Treadon, R., Wu, W.-S., and Lord, S.: Introduction of
1567 the GSI into the NCEP global data assimilation system, Weather Forecast., 24, 1691–1705,
1568 2009.
1569 Huneeus, N., Chevallier, F., and Boucher, O.: Estimating aerosol emissions by assimilating
1570 observed aerosol optical depth in a global aerosol model, Atmos. Chem. Phys., 12,
1571 4585-4606, doi:10.5194/acp-12-4585-2012, 2012.
1572 Huneeus, N., Boucher, O., and Chevallier, F.: Atmospheric inversion of SO2 and primary aerosol
1573 emissions for the year 2010, Atmos. Chem. Phys., 13, 6555-6573,
1574 doi:10.5194/acp-13-6555-2013, 2013.
1575 Hunt, B., Kostelich, E., and Szunyogh, I.: Efficient data assimilation for spatiotemporal chaos: a Local Ensemble
1576 Transfom Kalman Filter, Physica D, 230, 112–126, 2007.
1577 Lee, E.-H., Ha, J.-C., Lee, S.-S., and Chun, Y.: PM10 data assimilation over South Korea to Asian
1578 dust forecasting model with the optimal interpolation method, Asia-Pacific J. Atmos. Sci.,
1579 49(1), 73–85, doi:10.1007/s13143-013-0009-y, 2013.
1580 Li, Z., Zang, Z., Li, Q. B., Chao, Y., Chen, D., Ye, Z., Liu, Y., and Liou, K. N.: A
1581 three-dimensional variational data assimilation system for multiple aerosol species with
1582 WRF/Chem and an application to $PM_{2.5}$ prediction, Atmos. Chem. Phys., 13, 4265-4278,
1583 doi:10.5194/acp-13-4265-2013, 2013.
1584 Li, J., Li, Y., Bo, Y., and Xie, S.: High-resolution historical emission inventories of crop residue
1585 burning in fields in China for the period 1990–2013, Atmos. Environ., 138, 152–161, 2016.
1586 Liu, Z., Liu, Q., Lin, H. C., Schwartz, C. S., Lee, Y. H., and Wang, T.: Three-dimensional
1587 variational assimilation of MODIS aerosol optical depth: implementation and application to a
1588 dust storm over East Asia, J. Geophys. Res., 116, D23206, doi:10.1029/2011JD016159,
1589 2011.
1590 Liu, F., Zhang, Q., Tong, D., Zheng, B., Li, M., Huo, H., and He, K. B.: High-resolution inventory
1591 of technologies, activities, and emissions of coal-fired power plants in China from 1990 to
1592 2010, Atmos. Chem. Phys., 15, 13299-13317, doi:10.5194/acp-15-13299-2015, 2015.

McLinden, C.A., Fioletov, V., Shephard, M.W., Krotkov, N., Li, C., Martin, R.V., Moran, M.D., and J. Joiner,: Space-based detection of missing sulfur dioxide sources of global air pollution, Nat. Geosci., 9, 496–500, doi:10.1038/ngeo2724, 2016.

Mijling, B. and van der A, R. J.: Using daily satellite observations to estimate emissions of short-lived air pollutants on a mesoscopic scale, J. Geophys. Res., 117, D17302, doi:10.1029/2012JD017817, 2012.

Miyazaki, K., Eskes, H. J., Sudo, K., Takigawa, M., van Weele, M., and Boersma, K. F.: Simultaneous assimilation of satellite NO_2, O_3, CO, and HNO_3 data for the analysis of tropospheric chemical composition and emissions, Atmos. Chem. Phys., 12, 9545– 9579, doi:10.5194/acp-12-9545-2012, 2012.

Miyazaki, K., Eskes, H. J., Sudo, K., and Zhang, C.: Global lightning NOx production estimated by an assimilation of multiple satellite data sets, Atmos. Chem. Phys., 14, 3277–3305, doi:10.5194/acp-14-3277-2014, 2014.

Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E. J., et al.: Exploiting local low dimensionality of the atmospheric dynamics for efficient Kalman filtering, arXiv:physics/0203058, 24 pp., available at: http://arxiv.org/abs/physics/0203058v3/, 2002.

Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E. J., et al.: A local ensemble Kalman filter for atmospheric data assimilation, Tellus A, 56, 415–428, 2004.

Pagowski, M., Grell, G. A., McKeen, S. A., Peckham, S. E., and Devenyi, D.: Three-dimensional variational data assimilation of ozone and fine particulate matter observations: some results using the Weather Research and Forecasting – Chemistry model and Grid-point Statistical Interpolation, Q. J. Roy. Meteor. Soc., 136, 2013–2024, doi:10.1002/qj.700, 2010.

Pagowski, M., and Grell, G. A.: Experiments with the assimilation of fine aerosols using an ensemble Kalman filter, J. Geophys. Res.-Atmos., 117, D21302, doi:10.1029/2012jd018333, 2012.

Peng, Z., Zhang, M., Kou, X., Tian, X., and Ma, X.: A regional carbon data assimilation system and its preliminary evaluation in East Asia, Atmos. Chem. Phys., 15, 1087-1104, doi:10.5194/acp-15-1087-2015, 2015.

Pope, C. A.: Review: Epidemiological basis for particulate air pollution health standards, Aerosol Sci. Tech., 32, 4–14, 2000.

Pope, C. A., Burnett, R. T., Thun, M. J., Calle, E. E., Krewski, D., Ito, K., and Thurston, G. D.: Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution, J. Am. Med. Assoc., 287, 1132–1141, 2002.

Rubin, J. I., Reid, J. S., Hansen, J. A., Anderson, J. L., Collins, N., Hoar, T. J., Hogan, T., Lynch, P., McLay, J., Reynolds, C. A., Sessions, W. R., Westphal, D. L., and Zhang, J.: Development of the Ensemble Navy Aerosol Analysis Prediction System (ENAAPS) and its application of the Data Assimilation Research Testbed (DART) in support of aerosol forecasting, Atmos. Chem. Phys., 16, 3927-3951, doi:10.5194/acp-16-3927-2016, 2016.

Saide, P. E., Carmichael, G. R., Liu, Z., Schwartz, C. S., Lin, H. C., da Silva, A. M., and Hyer, E.: Aerosol optical depth assimilation for a size-resolved sectional model: impacts of observationally constrained, multi-wavelength and fine mode retrievals on regional scale analyses and forecasts, Atmos. Chem. Phys., 13, 10425-10444, doi:10.5194/acp-13-10425-2013, 2013.

Schwartz, C. S., Liu, Z., Lin, H. C., and McKeen, S. A.: Simultaneous three-dimensional variational assimilation of surface fine particulate matter and MODIS aerosol optical depth, J. Geophys. Res., 117, D13202, doi:10.1029/2011JD017383, 2012.

Schwartz, C. S., Liu, Z., Lin, H.-C., and Cetola, J. D.: Assimilating aerosol observations with a "hybrid" variational-ensemble data assimilation system, J. Geophys. Res. Atmos., 119, 4043–4069, doi:10.1002/2013JD020937, 2014.

Sekiyama, T. T., Tanaka, T. Y., Shimizu, A., and Miyoshi, T.: Data assimilation of CALIPSO aerosol observations, Atmos. Chem. Phys., 10, 39-49, doi:10.5194/acp-10-39-2010, 2010.

Schutgens, N. A. J., Miyoshi, T., Takemura, T., and Nakajima, T.: Sensitivity tests for an ensemble Kalman filter for aerosol assimilation, Atmos. Chem. Phys., 10, 6583-6600, doi:10.5194/acp-10-6583-2010, 2010.

Schutgens, N. A. J., Miyoshi, T., Takemura, T., and Nakajima, T.: Applying an ensemble Kalman filter to the assimilation of AERONET observations in a global aerosol transport model, Atmos. Chem. Phys., 10, 2561-2576, doi:10.5194/acp-10-2561-2010, 2010.

Schutgens, N., Nakata, M., and Nakajima, T.: Estimating Aerosol Emissions by Assimilating Remote Sensing Observations into a Global Transport Model, Remote Sensing,4, 3528-3543, 2012.

Tang, X., Zhu, J., Wang, Z. F., and Gbaguidi, A.: Improvement of ozone forecast over Beijing based on ensemble Kalman filter with simultaneous adjustment of initial conditions and emissions, Atmos. Chem. Phys., 11, 12901–12916, doi:10.5194/acp-11-12901-2011, 2011.

Tombette, M., Mallet, V., and Sportisse, B.: PM10 data assimilation over Europe with the optimal interpolation method, Atmos. Chem. Phys., 9, 57-70, doi:10.5194/acp-9-57-2009, 2009.

Torn, R. D., Hakim, G. J., and Snyder, C.: Boundary conditions for limited-area ensemble Kalman filters, Mon. Weather Rev., 134, 2490–2502, 2006.

van Loon, M., Builtjes, P. J. H., and Segers, A. J.: Data assimilation of ozone in the atmospheric transport chemistry model LOTOS, Environ. Model. Softw., 15, 603–609, 2000.

Wang, J., Xu, X., Henze, D. K., Zeng, J., Ji, Q., Tsay, S.-C., and Huang, J.: Top-down estimate of dust emissions through integration of MODIS and MISR aerosol retrievals with the GEOS-Chem adjoint model, Geophys. Res. Lett., 39, L08802, doi:10.1029/2012GL051136, 2012.

Wang, Y. X., Zhang, Q. Q., Jiang, J. K., Zhou, W., Wang, B. Y., He, K. B., Duan, F. K., Zhang, Q., Philip, S., and Xie, Y. Y.: Enhanced sulfate formation during China's severe winter haze episode in January 2013 missing from current models, J.Geophys.Res.-Atmos., 119, 10.1002/2013JD021426, 2014

Wang, X.Y., Liang,. X.Z., Jiang, W.M., Tao, Z.N., Wang, J.X.L., Liu, H.N., Han Z.W., Liu, S.Y., Zhang, Y.Y., Grell, G.A., Peckham, S.E.: WRF-Chem simulation of East Asian air quality: Sensitivity to temporal and vertical emissions distributions, Atmospheric Environment,44(2010) 660-669

Whitaker, J. S., and Hamill, T. M.: Ensemble data assimilation without perturbed observations, Mon. Weather Rev., 130, 1913–1924, 2002.

Woo, J.H., Baek, J.M., Kim, J.W., Carmichaael, G.R., Thongboonchoo, N., Kim, S.T., An, J.H.: Development of a Multi-Resolution Emission Inventory and Its Impact on Sulfur Distribution for Northeast Asia, Water, Air, and Soil Pollution 148: 259–278, 2003.

Weekly Crop Residue Burning Monitoring Report , http://hjj.mep.gov.cn/jgjs/201510/P020151012746205487305.pdf, 2015a (in Chinese).

Weekly Crop Residue Burning Monitoring Report,

http://hjj.mep.gov.cn/jgjs/201510/P020151019568921489639.pdf, 2015b(in Chinese).

Xia Y., Zhao, Y., Nielsen, C.P., Benefits of China's efforts in gaseous pollutant control indicated by the bottom-up emissions and satellite observations 2000-2014, Atmospheric Environment, 136, 43-53, 2016

Yu, H., Dickinson, R. E., Chin, M., Kaufman, Y. J., Geogdzhayev, B., and Mishchenko, M. I.: Annual cycle of global distributions of aerosol optical depth from integration of MODIS retrievals and GOCART model simulations, J. Geophys. Res., 108(D3), 4128, doi:10.1029/2002JD002717, 2003.

Yumimoto, K., Uno, I., Sugimoto, N., Shimizu, A., and Satake, S.: Adjoint inverse modeling of dust emission and transport over East Asia, Geophys. Res. Lett., 34, L00806, doi:10.029/2006GL028551, 2007.

Yumimoto, K., Uno, I., Sugimoto, N., Shimizu, A., Liu, Z., and Winker, D. M.: Adjoint inversion modeling of Asian dust emission using lidar observations, Atmos. Chem. Phys., 8, 2869-2884, doi:10.5194/acp-8-2869-2008, 2008.

Yumimoto, K., Nagao, T.M., Kikuchi, M., Sekiyama, T.T, Murakami, H.,Tanaka, T.Y., Ogi, A., Irie, H., Khatri, P., Okumura, H., Arai, K., Morino, I., Uchino, O., Maki, T.: Aerosol data assimilation using data from Himawari-8, a next-generation geostationary meteorological satellite, Geophys. Res. Lett., 43, 5886–5894, 2016.

Yin, X.M., Dai, T., Xin, J.Y., Gong, D.Y., Yang, J., Teruyuki, N., Shi, G.Y.: Estimation of aerosol properties over the Chinese desert region with MODIS AOD assimilation in a global model, Adv. Clim. Change Res., 7, 90–98, 2016.

Zhang, J., Reid, J. S., Westphal, D., Baker, N., and Hyer, E.: A System for Operational Aerosol Optical Depth Data Assimilation over Global Oceans, J. Geophys. Res., 113, D10208, doi:10.1029/2007JD009065, 2008.

Zhang, Q., Streets, D. G., Carmichael, G. R., He, K. B., Huo, H., Kannari, A., Klimont, Z., Park, I. S., Reddy, S., Fu, J. S., Chen, D., Duan, L., Lei, Y., Wang, L. T., and Yao, Z. L.: Asian emissions in 2006 for the NASA INTEX-B mission, Atmos. Chem. Phys., 9, 5131-5153, doi:10.5194/acp-9-5131-2009, 2009.

Zhang, L., Liu, L. C., Zhao, Y. H., Gong, S. L., Zhang, X. Y., Henze, D. K., Capps, S. L., Fu, T. M., Zhang, Q., and Wang, Y. X.: Source attribution of particulate matter pollution over

North China with the adjoint method, Environ.Res.Lett., 10, Artn 08401110.1088/1748-9326/10/8/084011, 2015.

Zheng, B., Zhang, Q., Zhang, Y., He, K. B., Wang, K., Zheng, G. J., Duan, F. K., Ma, Y. L., and Kimoto, T.: Heterogeneous chemistry: a mechanism missing in current models to explain secondary inorganic aerosol formation during the January 2013 haze episode in North China, Atmos.Chem.Phys., 15, 2031-2049, 10.5194/acp-15-2031-2015, 2015.

List of Figures and Table

Figure 1. Locations of 77 PM2.5 assimilation observation stations (black dot) and the 77 independent observation stations (red triangle) in the model domain. The three colored boxes mark sub-regions with relatively dense coverage for the Beijing–Tianjin–Hebei region (JJJ, 12 assimilation stations and 12 independent stations, red box), the Yangtze River delta (YRD, 24 assimilation stations and 24 independent stations, blue box) and the Pearl River delta (PRD, 9 stations and 9 independent stations, green box).

Figure 2. (a) Framework of M_{SF} and (b) flow chart of the data assimilation system that simultaneously optimizes the chemical initial conditions and emissions.

Figure 3. Time series of prior ensemble mean RMSE and total spread for PM2.5 concentrations aggregated over all observations over the three sub-regions: (a) Beijing–Tianjin–Hebei region; (b) Yangtze River delta; (c) Pearl River delta; and (d) time series of the area mean ensemble spread for $\lambda_{PM2.5}$ over the three sub-regions.

Table 1. Comparison of the surface PM2.5 mass concentrations from the control and assimilation experiments to observations over all analysis times from 6 to 16 October 2014.

Figure 4. Spatial distribution of the PM2.5 mass (μg m−3) of the (a) observations; (b) simulation of the control run; (c) analysis of expJ; (d) analysis of expC; (e) increments of expJ; (f) increments of expC; at the lowest model level averaged over all hours from 6 to 16 October 2014.

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

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

Figure 7. Spatial distribution of (a) the prior unspeciated primary sources of PM2.5 (μg m−2 s−1) and (b) the time-averaged differences between the ensemble mean analysis and the prior values (μg · m−2 s−1) at the lowest model level averaged over all hours from 6 to 16 October 2014.

Figure 8. Time series of the hourly PM2.5 obtained from observations (circle), analysis (blue line), control run (black line) and hourly output of 48-h forecast in three megacities: (a) Beijing; (c) Shanghai; and (e) Guangzhou in expJ and (b) Beijing; (d) Shanghai; and (f) Guangzhou in expC. See text in section 5.4.

Figure 9. Bias of surface PM2.5 as a function of forecast range calculated against independent observations over the three sub-regions: (a) Beijing–Tianjin–Hebei region; (c) Yangtze River delta; (e) Pearl River delta and RMSE over (b) Beijing–Tianjin–Hebei region; (d) Yangtze River delta; (f) Pearl River delta; (g) Normalized RMSE (assimilation divided by control) for expJ and (h) (g) Normalized RMSE for expC.

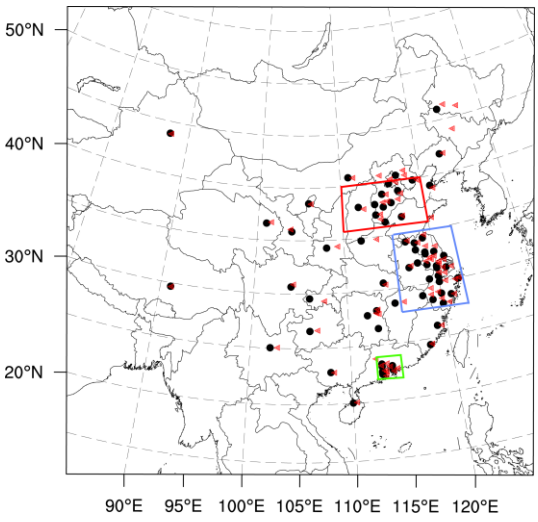


Figure 1. Locations of 77 $PM_{2.5}$ assimilation observation stations (black dot) and the 77 independent observation stations (red triangle) in the model domain. The three colored boxes mark sub-regions with relatively dense coverage for the Beijing–Tianjin–Hebei region (JJJ, 12 assimilation stations and 12 independent stations, red box), the Yangtze River delta (YRD, 24 assimilation stations and 24 independent stations, blue box) and the Pearl River delta (PRD, 9 stations and 9 independent stations, green box).

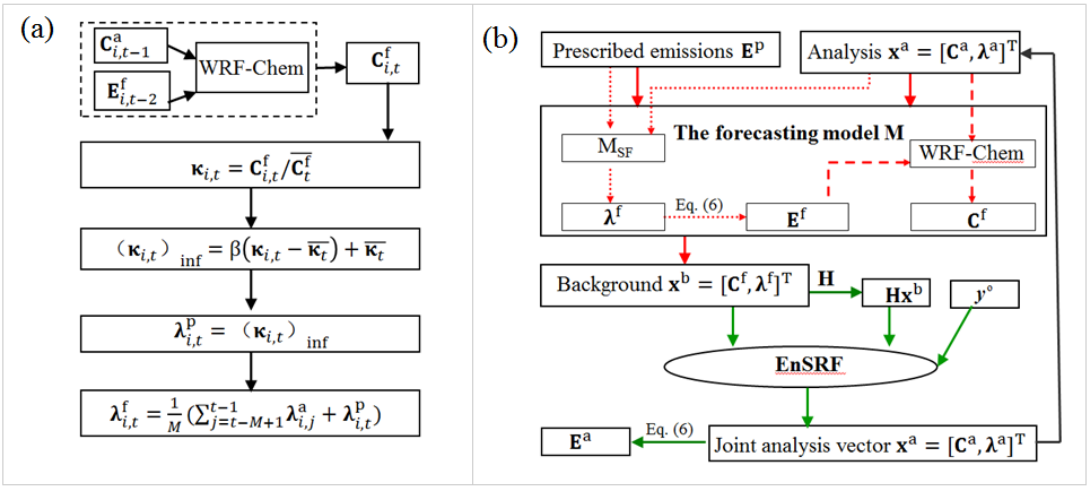


Figure 2. (a) Framework of $\mathbf{M}_{SF}$ and (b) flow chart of the data assimilation system that simultaneously optimizes the chemical initial conditions and emissions.

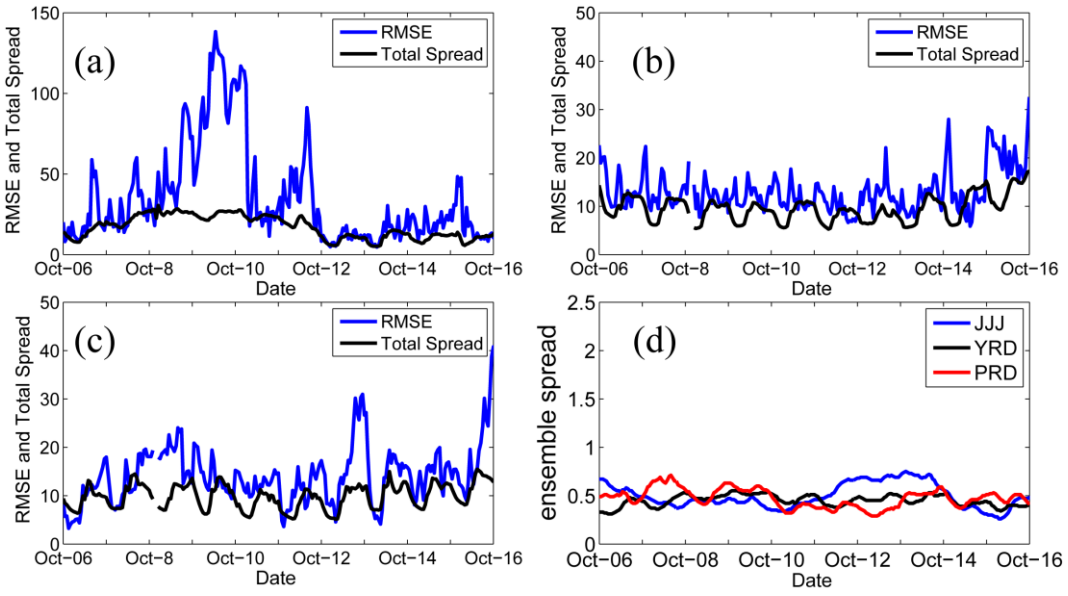


Figure 3. Time series of prior ensemble mean RMSE and total spread for $PM_{2.5}$ concentrations aggregated over all observations over the three sub-regions: (a) Beijing–Tianjin–Hebei region; (b) Yangtze River delta; (c) Pearl River delta; and (d) time series of the area mean ensemble spread for $\boldsymbol{\lambda}_{PM2.5}$ over the three sub-regions.

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

Region	Experiment	Mean observed value	Mean simulated value	BIAS	RMSE	CORR
Beijing–Tianjin–Hebei	Control	116.3	98.3	−18.0	81.6	0.790
	expJ		106.0	−10.3	66.9	0.827
	expC		104.1	-12.2	64.0	0.845
Yangtze River delta	Control	48.5	64.4	15.9	30.6	0.593
	expJ		46.9	-1.6	15.3	0.846
	expC		46.1	-2.4	17.3	0.803
Pearl River delta	Control	61.8	82.4	20.6	31.8	0.624
	expJ		66.5	4.7	16.1	0.800
	expC		64.1	-2.3	15.6	0.797

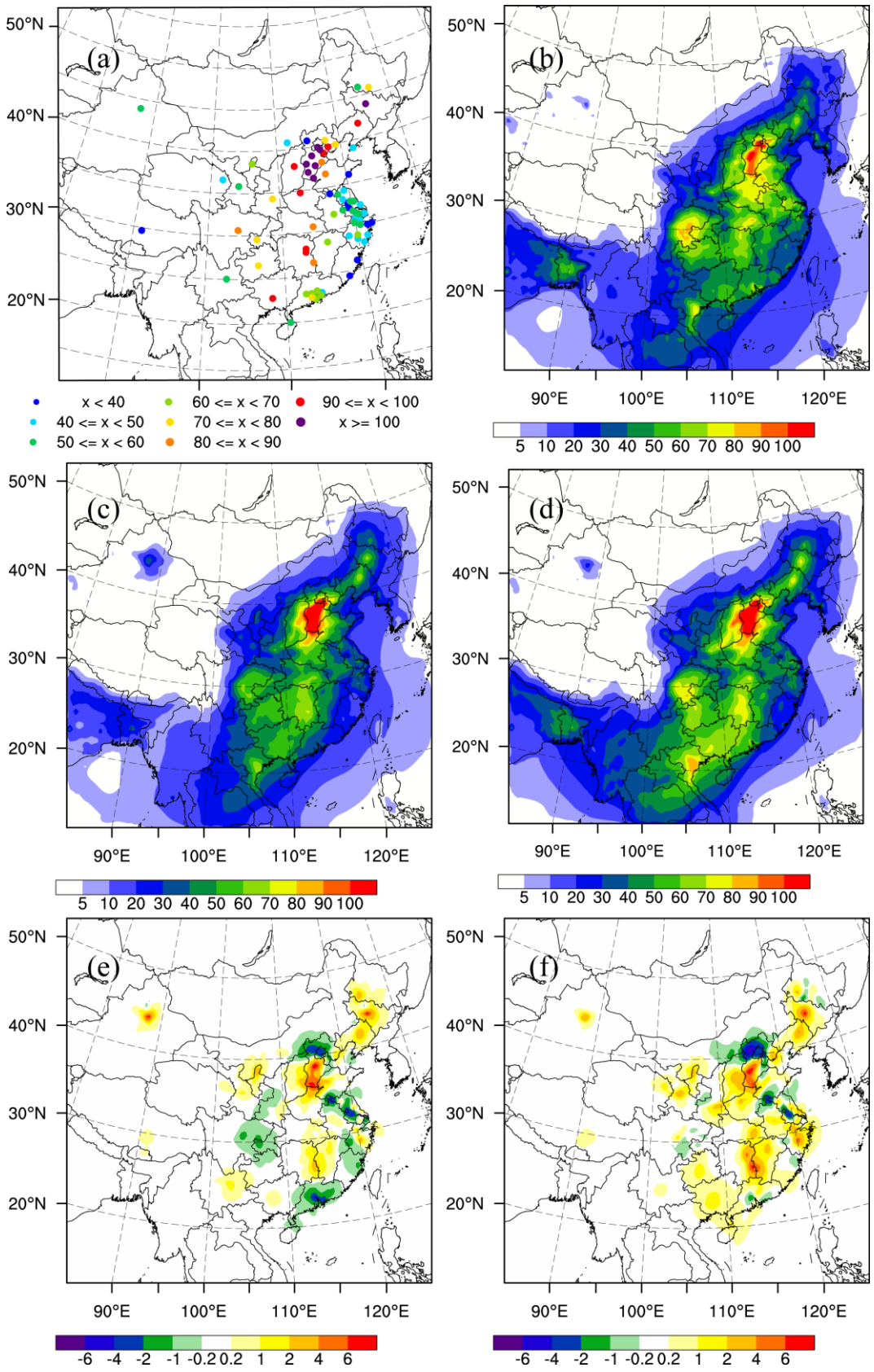


Figure 4. Spatial distribution of the $PM_{2.5}$ mass (μg m^{-3}) of the (a) observations; (b) simulation of the control run; (c) analysis of expJ; (d) analysis of expC; (e) increments of expJ; (f) increments of expC; at the lowest model level averaged over all hours

from 6 to 16 October 2014.

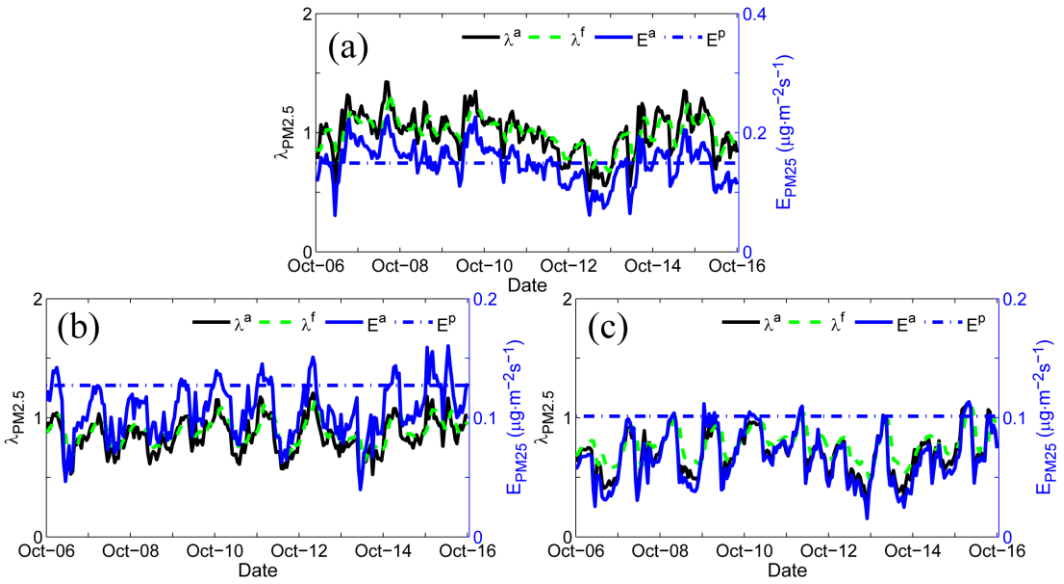


Figure 5. Hourly area-averaged time series of emission scaling factors (black) extracted from the ensemble mean of the analyzed $\boldsymbol{\lambda}^{a}_{PM2.5}$ and the corresponding analyzed unspeciated primary $PM_{2.5}$ emissions $\mathbf{E}^{a}_{PM2.5}$ (blue) over the three sub-regions: (a) Beijing–Tianjin–Hebei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

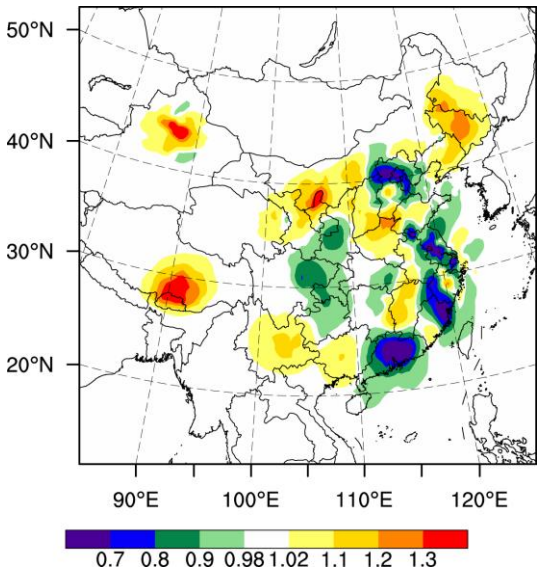


Figure 6. Spatial distribution of $\boldsymbol{\lambda}_{\mathrm{PM2.5}}$ at the lowest model level averaged over all hours from 6 to 16 October 2014.

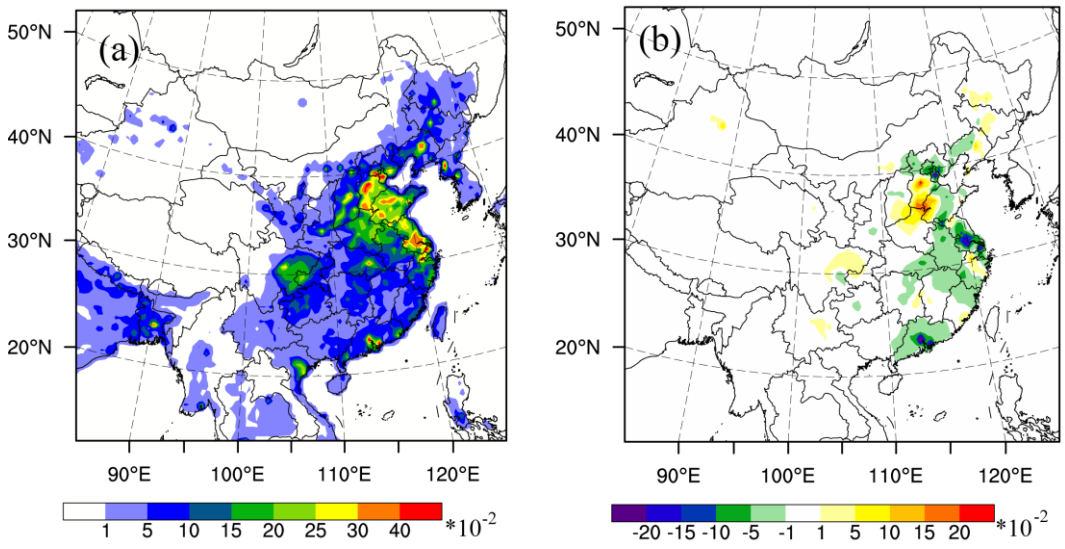


Figure 7. Spatial distribution of (a) the prior unspeciated primary sources of $PM_{2.5}$ ($\mu g\ m^{-2}\ s^{-1}$) and (b) the time-averaged differences between the ensemble mean analysis and the prior values ($\mu g \cdot m^{-2}\ s^{-1}$) at the lowest model level averaged over all hours from 6 to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

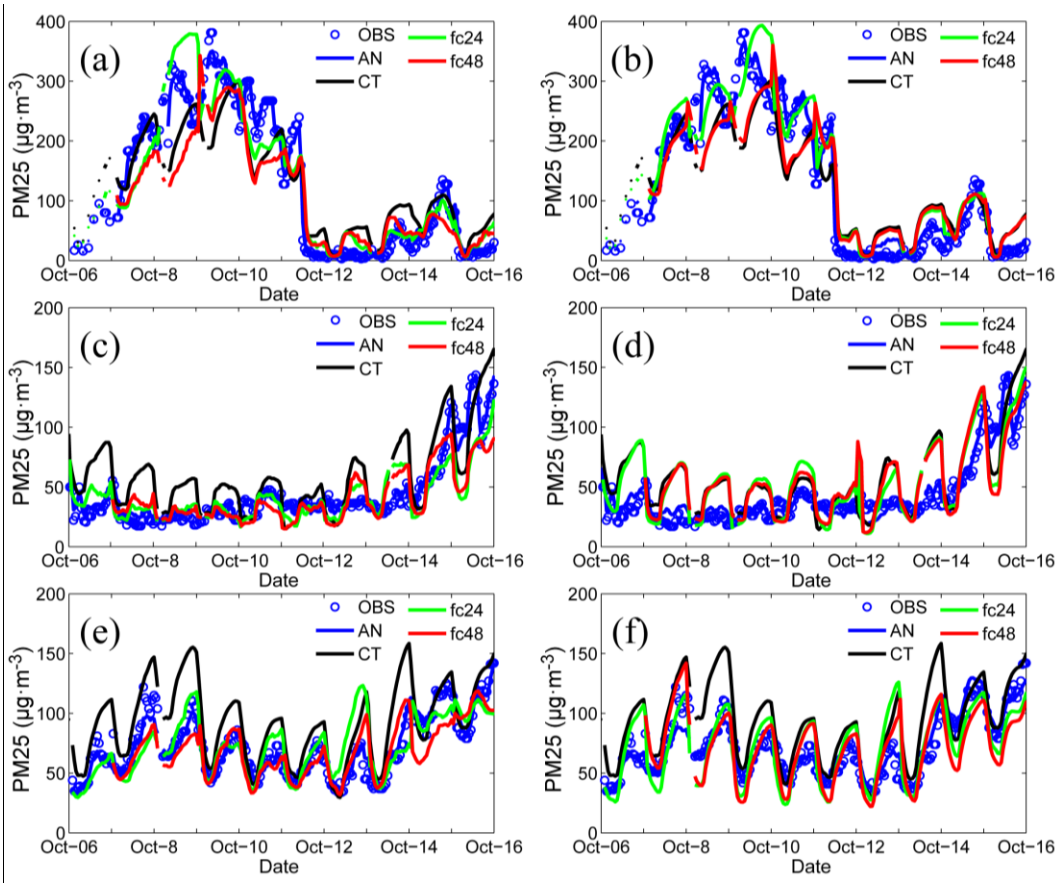


Figure 8. Time series of the hourly $PM_{2.5}$ obtained from observations (circle), analysis (blue line), control run (black line) and hourly output of 48-h forecast in three megacities: (a) Beijing; (c) Shanghai; and (e) Guangzhou in expJ and (b) Beijing; (d) Shanghai; and (f) Guangzhou in expC. See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

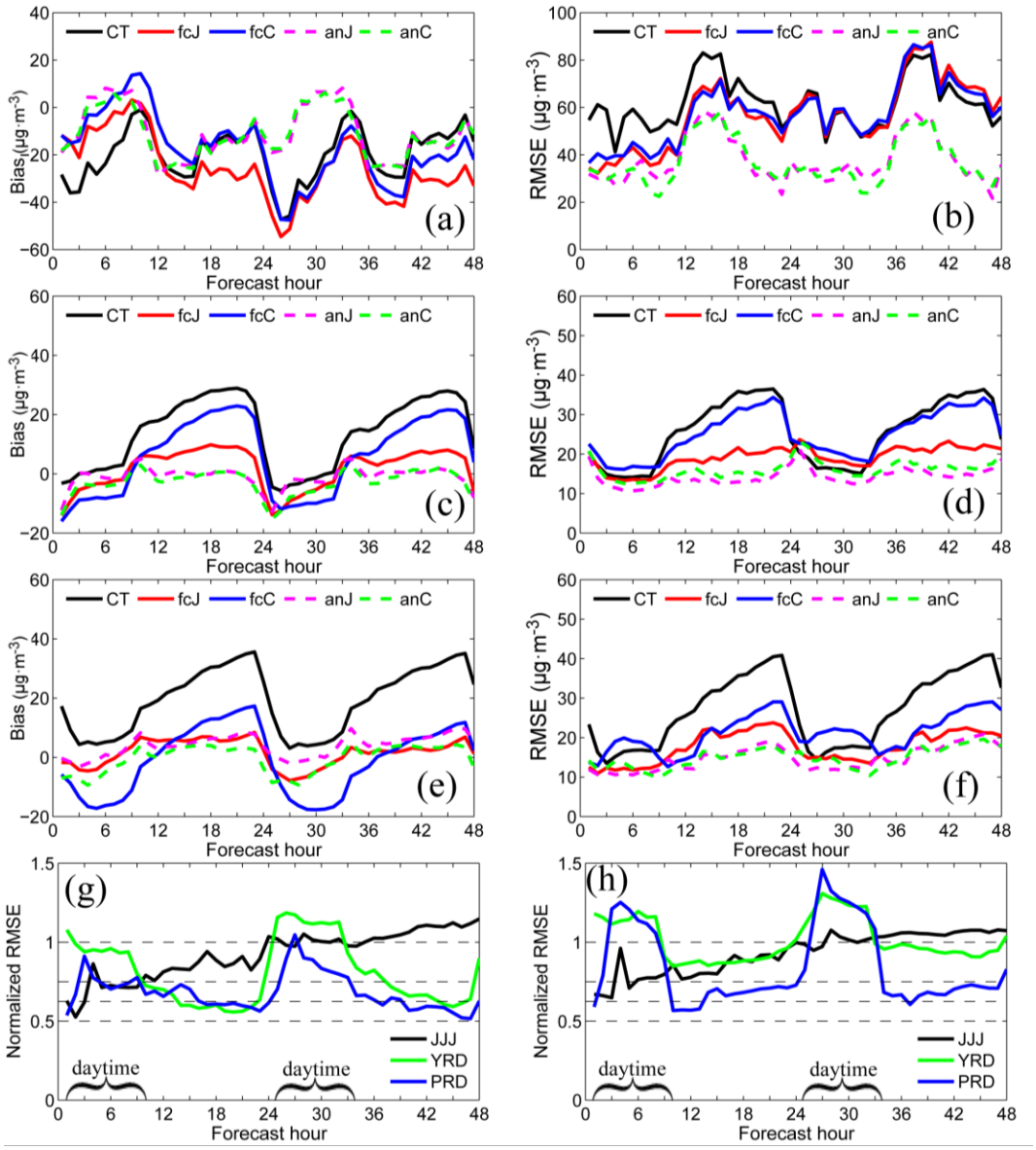


Figure 9. Bias of surface $PM_{2.5}$ as a function of forecast range calculated against independent observations over the three sub-regions: (a) Beijing–Tianjin–Hebei region; (c) Yangtze River delta; (e) Pearl River delta and RMSE over (b) Beijing–Tianjin–Hebei region; (d) Yangtze River delta; (f) Pearl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