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 $\overline{\kappa_{i,t}} = \overline{\mathbf{C}_{i,t}^{\mathrm{f}}/\overline{\mathbf{C}_{t}^{\mathrm{f}}}} = \overline{\mathbf{C}_{i,t}^{\mathrm{f}}}/\overline{\mathbf{C}_{t}^{\mathrm{f}}} =$

$\overline{C_t^f}/\overline{C_t^f}=1$. 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 $\lambda_{i,t}^f$. For example, in our manuscript, M = 4. Thus, $\lambda_{i,t}^p$, $\lambda_{i,t-1}^a$, $\lambda_{i,t-2}^a$, and $\lambda_{i,t-3}^a$ 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 ($E_{PM2.5i}$, $E_{PM2.5j}$, E_{SO4i}, E_{SO4j}, E_{NO3i}, E_{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) $\lambda_{PM2.5}^f$, λ_{SO2}^f, λ_{NO}^f and λ_{NH3}^f 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 μg m^{-3} and 243 (around 1.5%) were discarded due to the ensemble mean of the first guess departure exceeding 100 μg 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.
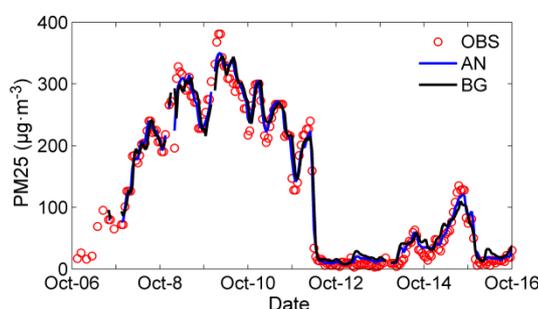
ReFig1. Time series of the hourly PM$_{2.5}$ obtained from observations (red circle), the ensemble mean of the analysis (blue line) and the ensemble mean of the background (the ensemble mean of the background, black line) in Beijing.

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

expJ and expC. The figure of the PM$_{2.5}$ mass differences was removed to save space. It is very clear that the analysis of expJ and that of expC are much different from the simulation of the control run.

Then we rewrote paragraph 2 in Section 5.2 in Lines 517, Page 19 to Lines 533, Page 20.

14: Line 349, We can find adjustments over the SE Asia and India where you have no PM2.5 observation.

The analysis increments (i.e. $\bar{\mathbf{x}}^a - \bar{\mathbf{x}}^b$) indicate the direct impact of assimilating PM$_{2.5}$ observation. They are determined by both the observation increments and the relative magnitudes of the forecast error and the observation error. From Figure 4 (e) and (f), we can see the increments of both assimilation experiments are distributed around the locations of observations as expected. However, the impact of assimilating PM$_{2.5}$ observations is not limited to the areas where observations were located, observations information is also transported to other areas through the WRF-Chem forecast. Besides, the ensemble forecasts also partly contributed to the PM$_{2.5}$ mass differences (assimilation minus control). Therefore, the spatial distributions of the PM$_{2.5}$ mass in both assimilation experiments were significantly different from the control run. Thus we can find adjustments over the SE Asia and India where no PM2.5 observation is available.

We have added the above explanations in Lines 517 to 528, Page 19.

15: Figure 5, Overlaying of a priori emissions (it will be flat lines) in Figure 5 may emphasize that the assimilation can generate the temporal variations in the emissions.

I have overlaid a priori emissions (the dash dot line) in Figure 5.

16: Line 375, Is the burning of crop residues limited in the JJJ region? Li et al. shows that the northern part of YRD also has large emissions from the burning.

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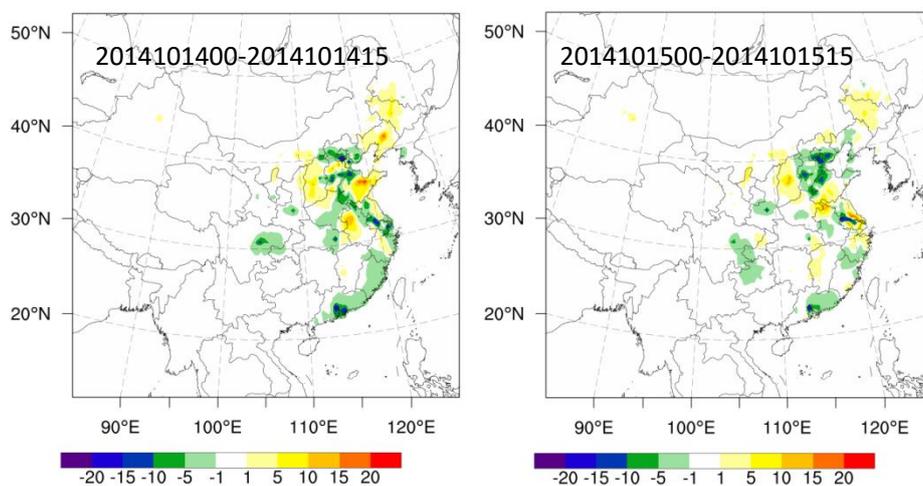
ReFig2 Spatial distribution of the mean differences between the ensemble mean analysis and the prior emissions of the unspeciated primary sources of $PM_{2.5}$ at the lowest model level