Response to Reviewer #1's comments:

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

Summary and general comments:

This manuscript investigated the application of ensemble Kalman filter (EnKF) for constraining the atmospheric chemical species including PM_{10}, $PM_{2.5}$, SO_2, NO_2, O_3 and CO. The simultaneous assimilation of various surface air quality measurements improved the representation of the initial conditions and emission factors of aforementioned species, as well as their 72-hours forecasts. This investigation on the assimilation of various air quality observations for a severe haze pollution event provides a promising case study for the regional air-quality modeling. I would recommend the minor revision with the considerations of several issues as listed below.

List of minor comments:

1) Section 2.1: Which dataset (reanalysis) did you use for the meteorological initial and boundary conditions? Were the perturbations also added to the meteorology? If not, please add one or two sentences to mention that the uncertainty of the meteorology forecasts is not considered in this study.

The meteorological initial and boundary conditions were provided by the National Centers for Environmental Prediction Global Forecast System (GFS). The temperature, water vapor, velocity, geopotential height and dry surface pressure fields of the meteorological initial and boundary conditions were perturbed by adding Gaussian random noise with a zero mean and static background error covariances (Torn et al., 2006) to generated the 50 ensemble members by WRFDA. We have added these sentences in Line 274-278, Page 10.

2) L107-108: Are emission scaling factors λ spatially varying?

Yes, the emission scaling factors $\boldsymbol{\lambda}$ here are spatially varying. In our system, we use the ensemble forecast chemical fields $\mathbf{C}_{i,t}^{\mathrm{f}}$ and the previous DA cycles' analysis scaling factors $\boldsymbol{\lambda}_{i,t-3}^{\mathrm{a}}$, $\boldsymbol{\lambda}_{i,t-2}^{\mathrm{a}}$, $\boldsymbol{\lambda}_{i,t-1}^{\mathrm{a}}$ to evaluate the emission scaling factors $\boldsymbol{\lambda}_{i,t}^{\mathrm{f}}$. Since $\mathbf{C}_{i,t}^{\mathrm{f}}$ were spatially varying, the ensemble concentration ratios $\boldsymbol{\kappa}_{i,t} = \mathbf{C}_{i,t}^{\mathrm{f}}/\overline{\mathbf{C}_t^{\mathrm{f}}}$ were spatially varying too. Thus, $\boldsymbol{\lambda}_{i,t}^{\mathrm{f}} = \frac{1}{4}\left(\boldsymbol{\lambda}_{i,t-3}^{\mathrm{a}} + \boldsymbol{\lambda}_{i,t-2}^{\mathrm{a}} + \boldsymbol{\lambda}_{i,t-1}^{\mathrm{a}} + \boldsymbol{\lambda}_{i,t}^{\mathrm{p}}\right) = \frac{1}{4}\left(\boldsymbol{\lambda}_{i,t-3}^{\mathrm{a}} + \boldsymbol{\lambda}_{i,t-2}^{\mathrm{a}} + \boldsymbol{\lambda}_{i,t-1}^{\mathrm{a}} + (\boldsymbol{\kappa}_{i,t})_{\mathrm{inf}}\right) = \frac{1}{4}\left(\boldsymbol{\lambda}_{i,t-3}^{\mathrm{a}} + \boldsymbol{\lambda}_{i,t-2}^{\mathrm{a}} + \boldsymbol{\lambda}_{i,t-1}^{\mathrm{a}} + \beta\left(\boldsymbol{\kappa}_{i,t} - \overline{\boldsymbol{\kappa}_t}\right) + \overline{\boldsymbol{\kappa}_t}\right)$ were spatially varying.

We have added these sentences in Line 156-158, Page6.

3) L154-156: Why the inflation factors for the chemical species β are different among the variables? Could you please provide the strategy you took to find these values?

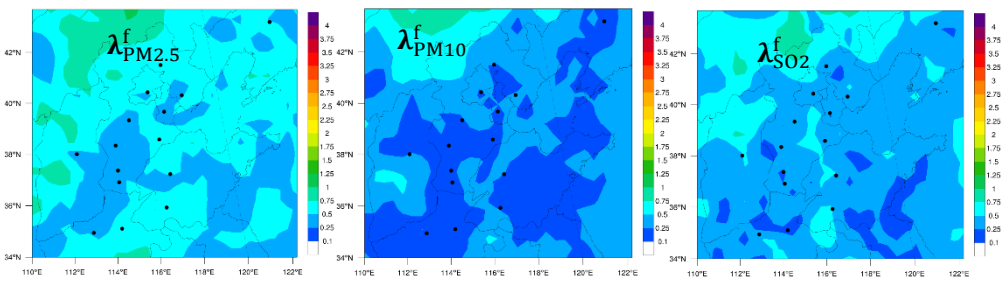
Peng et al. (2015) first used the forecast model of scaling operator $\mathbf{M}_{\mathrm{SF}}$ to prepare the ensemble emission scaling factors $\boldsymbol{\lambda}^{\mathrm{f}}$ in order to optimize all CO_2 fluxes as a whole at grid scale. In Peng et al. (2015), the ensemble spread of $\boldsymbol{\kappa}_{i,t} = \mathbf{C}_{i,t}^{\mathrm{f}}/\overline{\mathbf{C}_t^{\mathrm{f}}}$ 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 1.25 for β =60, 70, 75, 80. And the CO_2 DA system worked comparatively well for β =60, 70, 75, 80. The assimilated CO2 fluxes deviated markedly from the "true" CO2 fluxes when the ensemble spread of $\lambda_{\mathrm{i,t}}^{\mathrm{a}}$ were too small for β =10, 50 or when the ensemble spread of $\lambda_{\mathrm{i,t}}^{\mathrm{a}}$ were too large for β =100. 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.

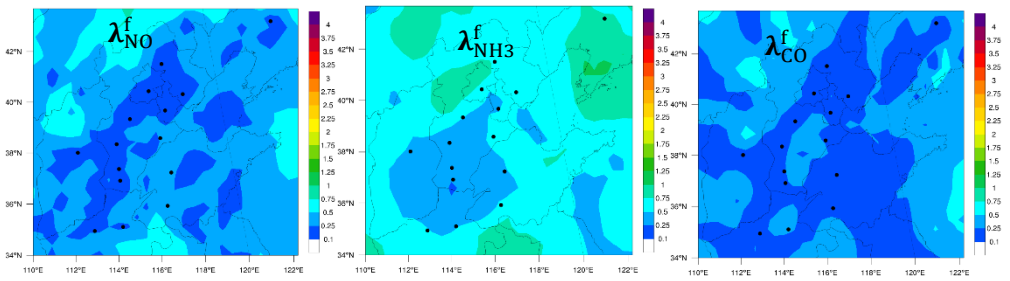
In Peng et al. (2017), four emission scaling factors, $\boldsymbol{\lambda}^{\mathrm{f}}_{\mathrm{PM2.5}}$, $\boldsymbol{\lambda}^{\mathrm{f}}_{\mathrm{SO2}}$, $\boldsymbol{\lambda}^{\mathrm{f}}_{\mathrm{NO}}$ and $\boldsymbol{\lambda}^{\mathrm{f}}_{\mathrm{NH3}}$, are optimized in Peng et al. (2017) when the pure surface $PM_{2.5}$ observations are assimilated. We use the same inflation factor β to keep the ensemble spreads of $\boldsymbol{\lambda}^{\mathrm{f}}_{\mathrm{PM2.5}}$, $\boldsymbol{\lambda}^{\mathrm{f}}_{\mathrm{SO2}}$, $\boldsymbol{\lambda}^{\mathrm{f}}_{\mathrm{NO}}$ and $\boldsymbol{\lambda}^{\mathrm{f}}_{\mathrm{NH3}}$ at a certain level. Several sensitive experiments were performed to investigate β (1.2, 1.5, 1.8, 2, 2.5). It is seemed that reasonable results can be obtained when the ensemble spread of the emission scaling factors $\boldsymbol{\lambda}^{\mathrm{f}}_{\mathrm{PM2.5}}$ ranged from 0.1 to 1. Finally, $\beta = 1.5$ was chosen in Peng et al. (2017). The area-averaged ensemble spreads of $\boldsymbol{\lambda}^{\mathrm{f}}_{\mathrm{PM2.5}}$, $\boldsymbol{\lambda}^{\mathrm{f}}_{\mathrm{SO2}}$, $\boldsymbol{\lambda}^{\mathrm{f}}_{\mathrm{NO}}$ and $\boldsymbol{\lambda}^{\mathrm{f}}_{\mathrm{NH3}}$ were stably distributed around 0.5, 1.0, 1.5 and 0.8 respectively over the three sub-regions: Beijing–Tianjin–Hebei region, Yangtze River delta and Pearl River delta. It is apparent that the ensemble spread of $\lambda^{\mathrm{f}}_{\mathrm{SO2}}$ and $\lambda^{\mathrm{f}}_{\mathrm{NO}}$ is a little large due to the same β.

Therefore, it is better to choose different inflation factors for different emission scaling factors. We have performed several sensitive experiments to determine the value of β over a 2-day period before the experiments written in the manuscript. The criterion we choose β is to keep the ensemble spread of the scaling factors ranging from 0.1 to 1 in most model area. Finally, β is chosen as 1.3, 1.4, 1.3, 1.2, 1.2, and 1.4 for $\boldsymbol{\lambda}^{\mathrm{f}}_{\mathrm{PM2.5}}$, $\boldsymbol{\lambda}^{\mathrm{f}}_{\mathrm{PM10}}$, $\boldsymbol{\lambda}^{\mathrm{f}}_{\mathrm{SO2}}$, $\boldsymbol{\lambda}^{\mathrm{f}}_{\mathrm{NO}}$, $\boldsymbol{\lambda}^{\mathrm{f}}_{\mathrm{NH3}}$ and $\boldsymbol{\lambda}^{\mathrm{f}}_{\mathrm{CO}}$ (See ReFig. 1)

Perhaps there are very few negative values for $(\boldsymbol{\kappa}_{\mathrm{i,t}})_{\mathrm{inf}}$ after inflation. A quality control procedure is performed for $(\boldsymbol{\kappa}_{i,t})_{\mathrm{inf}}$ before further appliance. All these negative data were set as 0 in this work. Then $(\boldsymbol{\kappa}_{i,t})_{\mathrm{inf}}$ were re-centered to ensure the ensemble mean values of $(\boldsymbol{\kappa}_{i,t})_{\mathrm{inf}}$ were all 1. Then, another quality control procedure is performed for $\boldsymbol{\lambda}^{\mathrm{a}}_{i,t}$ to keep them positive. Thus, all $\boldsymbol{\lambda}^{\mathrm{f}}_{i,t}$ and $\boldsymbol{\lambda}^{\mathrm{a}}_{i,t}$ could be positive.

We have added these sentences in Line 158-163, 166-169, Page 6.





ReFig. 1. Spatial distribution of the ensemble spread for $\boldsymbol{\lambda}^{f}_{PM2.5}$, $\boldsymbol{\lambda}^{f}_{PM10}$, $\boldsymbol{\lambda}^{f}_{SO2}$, $\boldsymbol{\lambda}^{f}_{NO}$, $\boldsymbol{\lambda}^{f}_{NH3}$ and $\boldsymbol{\lambda}^{f}_{CO}$at the lowest model level at 0000 UTC 6 October 2014 in the NCP region.

4) **L257-259: How did you perturb the initial conditions, lateral boundary conditions and emissions? In other words, please provide how you estimated the background uncertainty and spatial correlations (i.e. background covariance structures) for the chemical state variables in adding perturbations?**

Before the first DA cycle, a 50-member ensemble of four-day spin-up forecasts was performed, with perturbed meteorological initial conditions (ICs), lateral boundary conditions (LBCs) and emissions, from 0000 UTC 1 October to 2300 UTC 4 October 2014. The perturbed meteorological ICs and LBCs are created by adding Gaussian random noise (Torn et al., 2006) to the temperature, water vapor, velocity, geopotential height and dry surface pressure fields of the products of the National Centers for Environmental Prediction Global Forecast System (GFS) by WRFDA. The perturbed emissions were generated also by adding Gaussian random noise with a standard deviation of 10 percent of the corresponding anthropogenic emissions. The aerosol ICs were zero and the aerosol LBCs were idealized profiles embedded within the WRF/Chem model. They are the same as in Peng et al. (2017). It is noted that the perturbed emissions were only used in the initial part.

In the DA part, the ICs were the analysis of the previous DA cycle, the meteorological LBCs were the perturbed LBCs. The anthropogenic emissions, $\mathbf{E}^{f}_{PM2.5}$, $\mathbf{E}^{f}_{PM10}$, $\mathbf{E}^{f}_{SO2}$, $\mathbf{E}^{f}_{NO}$, $\mathbf{E}^{f}_{NH3}$, $\mathbf{E}^{f}_{CO}$, sulfate $\mathbf{E}^{f}_{SO4}$ and nitrate $\mathbf{E}^{f}_{NO3}$ are calculated by using the forecast emission scaling factors. Other species, such as the organic compounds $\mathbf{E}_{org}$ and elemental compounds $\mathbf{E}_{BC}$, are perturbed by adding Gaussian

random noise. Since the emissions are calculated by EQ. (1), their background uncertainties and the spatial correlations are completely dependent on those of the corresponding emission factors. The forecast scaling factors are calculated by EQ. (2) ~ (5). And no other perturbations are added to the scaling factors; no other correlations are assumed for the scaling factors.

The experimental design is the same as in Peng et al (2017). We have rewritten briefly in Section 4 to avoid the repetition (Line 272-293, Page 10-11).

5) L275-279 and Figure 2: This is very promising. I would imagine that the impacts of other sources of uncertainties in air-quality forecast that were not directly considered in this study (such as chemical schemes and parameterizations in forecast model, and meteorology) were indirectly considered through the well-calibrated inflations of state variables. Could you please make a comment about the impacts of these other uncertainty sources in discussion section? I believe it would be helpful for the future readers of this manuscript.

It is true that the impacts of other sources of uncertainties in air-quality forecast (such as chemical schemes and parameterizations in forecast model, and meteorology) were not directly considered through the well-calibrated inflations of state variables. EnKF assimilation is influenced greatly by model errors and observation errors. But it is very difficult to accurately evaluate the uncertainties of models, though the covariance inflation technique was simply applied for all state variables to roughly compensate for model errors. Therefore, we can only obtain suboptimal results through EnKF assimilation.

We have added the above paragraph in Lines 476-482, Page 17.

6) Figure 4: It is not clear to me what "The shaded backgrounds indicate the distribution of the observations, where the top edge represented the 90th percentile and the bottom edge the 10th percentile" means. Does this distribution represent the observation values of individual sites in the Beijing–Tianjin–Hebei (BTH) region? Are other (red, black, pink, blue and light green) thick lines

average of all sites in BTH region? The purpose to show these two values together is unclear to me, since the grey shaded line and other thick lines do not seem to be comparable each other. I would recommend to add more explanations about this figure, or to remove the grey shaded lines.

Yes. the grey shaded line represent the distribution of the observation values of individual sites in the Beijing–Tianjin–Hebei (BTH) region. Other (red, black, pink, blue and light green) thick lines represent the average values of all sites in BTH region. No more information could be obtained from the grey shaded line since the average values of observations (red line) were shown. Thus we remove the grey shaded lined in Figure 4.

List of specific comments:

1) L174: Please change "chose" to "chosen".

We have revised the word in Line 188, Page 6.

2) L296: I think "was able to" better fits with this context than "could".

We have changed the word in Line 328, Page 12.

3) Figure 4: The acronyms of "an" and "ct" is not described (although they can be guessed from the figure caption). Could you please add the explanation of those acronyms in the figure captions, such as "the analysis (referred to as "an", pink line)"?

We have changed theese in Line 848-855, Page 32.

4) Figure 11: Please add the explanation of grey shaded lines in the top panels.

We remove the grey shaded lined in Figure 11, similar to Figure 4.