

## Response to Referee #1 (essd-2020-100)

We Thank Reviewer for his/her constructive comments.

Responses to the Specific comments:

**General comments:** This study presents high-resolution air quality reanalysis products over China for 2013-2018. The air quality reanalysis assimilated the country-wide surface observations using the regional EnKF data assimilation. The assimilated results were evaluated against the assimilated and independent measurements. The topic of this study is very interesting, and the produced data sets can be useful for various applications. The paper is generally well written. However, because this is the first paper describing the system and data, more careful description of the system and its performance would be useful for readers and future developments.

**Reply:** The authors appreciate the reviewer for his/her constructive and up-to-point comments. We have carefully considered the comments and revised the manuscript accordingly. Please refer to our responses for more details given below.

**Comment 1:** The representativeness error estimation is not clear. How did you estimate  $L_{repr}$  for each station and  $\epsilon^{abs}$  for each species? Urban and rural observations could be (or should be) used in a different way, but this is not mentioned. Were any temporal averages applied to the observations? Temporal variability information could be used a part of representativeness errors. Further explanation is needed.

**Reply:** Thanks for this important suggestion. The representativeness error arises from the different spatial scales that the gridded model results and discrete observations represent, which is parameterized by the formula proposed by Elbern et al. (2007) in this study:

$$r_{repr} = \sqrt{\frac{\Delta x}{L_{repr}}} \times \epsilon^{abs} \quad (1)$$

where  $r_{repr}$  represents the representativeness error,  $\Delta x$  represents the model resolution,  $L_{repr}$  represents the characteristic representativeness length of the observation site and  $\epsilon^{abs}$  represents the error characteristic parameters for different species.

We agree with the reviewer that the  $L_{repr}$  should be treated differently for urban and rural sites since the urban sites usually have smaller representativeness length than the rural sites due to the larger representativeness error. According to Elbern et al. (2007), the representativeness length of urban and rural sites were 2km and 10km. Considering that the observation sites from CNEMC were almost city (urban) sites (>90%), the  $L_{repr}$  was assigned to be 2km in this study for simplicity.

For the estimations of  $\epsilon^{abs}$ , previous studies (Chen et al., 2019; Feng et al., 2018; Jiang et al., 2013; Ma et al., 2019; Pagowski and Grell, 2012; Peng et al., 2017; Werner et al., 2019) usually assigned the  $\epsilon^{abs}$  empirically

to be half of the measurement error following the study by Pagowski et al. (2010). In this study, the  $\varepsilon^{abs}$  was obtained from Li et al. (2019) who estimated the  $\varepsilon^{abs}$  based on a dense observation network in Beijing-Tianjin-Hebei region. In their study, the representativeness error of each species' observation was first estimated by the spatiotemporal averaged standard deviation of the observed values within a 30km×30km grid:

$$r_{repr,i} = \frac{1}{MT} \sum_{m=1}^M \sum_{t=1}^T S_{m,t,i} \quad (2)$$

where  $r_{repr,i}$  represents the representativeness errors of the observations for species  $i$ ,  $S_{m,t,i}$  represents the standard deviation of the observed values of species  $i$  at different sites that are located in a same grid  $m$  at time  $t$ ,  $M$  and  $T$  represents the total number of grid and observation time. After that, the  $\varepsilon_i^{abs}$  for species  $i$  were estimated by a transformation of Eq. (1):

$$\varepsilon_i^{abs} = r_{repr,i} / \sqrt{\frac{\Delta x}{L_{repr}}} \quad (3)$$

where  $\Delta x$  is equal to 30km. Based on the estimated  $L_{repr}$  and the  $\varepsilon_i^{abs}$  for different species, the representativeness errors are estimated using Eq. (1) by specifying the  $\Delta x$  to be 15km. Following the suggestions of the reviewer, we have added more explanation to the estimations of representativeness error in the revised manuscript (*please see lines 223–245 in the revised manuscript*).

**Comment 2:** The assimilated results are compared with the independent observations for PM but with the assimilated observations only for other species (they only demonstrate self-consistency. CAMS is not observation). This provides limited information on the performance of the developed system. The Chi-square diagnostic can be used to see whether the Kalman filtering worked properly. OmF & OmA statistics can also be demonstrated. Given limited validation data, more efforts are required to demonstrate the performance.

**Reply:** Thanks for this important comment. Following the suggestions of reviewer, we have added the analysis of  $\chi^2$  diagnosis and the statistics of observation minus forecast (OmF:  $\mathbf{y}^o - \mathbf{H}(\mathbf{x}^b)$ ) & observation minus analysis (OmA:  $\mathbf{y}^o - \mathbf{H}(\mathbf{x}^a)$ ) in the revised manuscript to demonstrate the performance of our assimilation system (*please see lines 317–369 in the revised manuscript*).

$\chi^2$  diagnosis is a robust criterion for validating the estimated background and observation error covariance in the data assimilation (e.g., Menard et al., 2000; Miyazaki et al., 2015; Miyazaki et al., 2012), which is estimated by comparing the sample covariance of OmF with the sum of estimated background and observation error covariance in the observational space ( $\mathbf{HBH}^T + \mathbf{R}$ ):

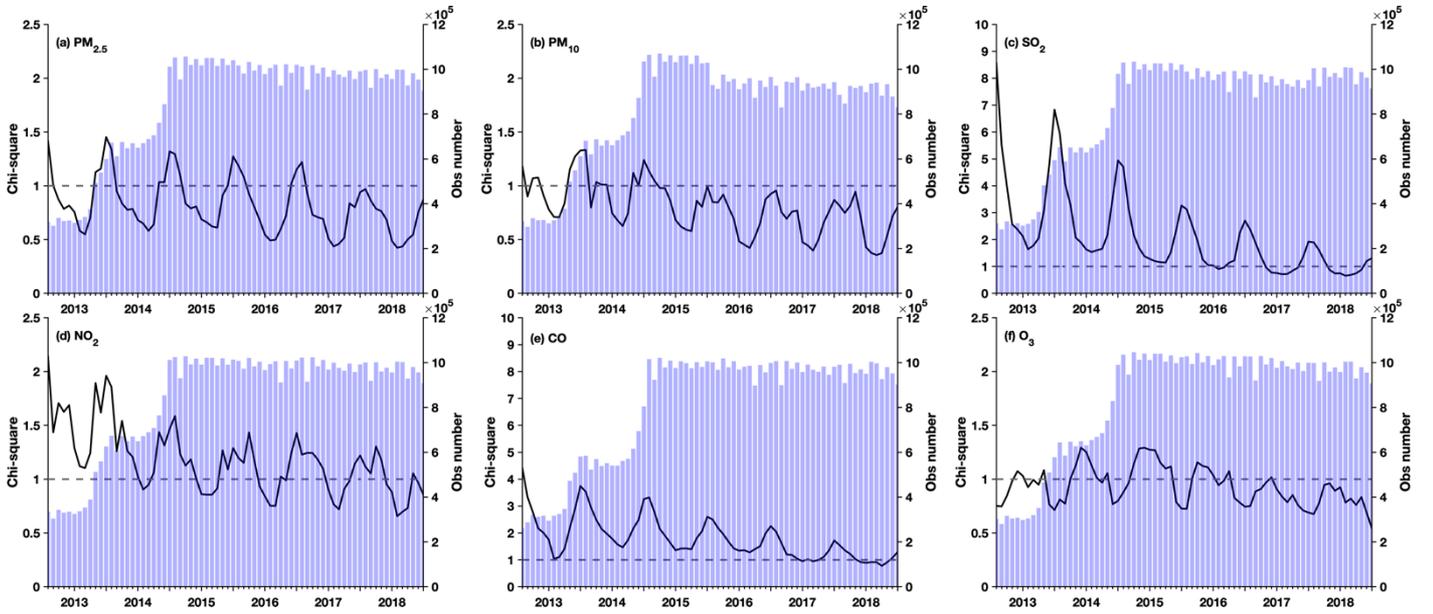
$$\mathbf{Y} = \frac{1}{\sqrt{m}} (\mathbf{HBH}^T + \mathbf{R})^{-\frac{1}{2}} (\mathbf{y}^o - \mathbf{HX}^b) \quad (4)$$

$$\chi^2 = \mathbf{Y}^T \mathbf{Y} \quad (5)$$

where  $m$  is the number of observations. According to the Kalman filtering theory, the mean of  $\chi^2$  should approach

1 if the background and observation error covariances are properly specified, while values greater (lower) than 1 indicates the underestimation (overestimation) of the observation and/or background error covariance.

Figure R1 shows the time series of the monthly  $\chi^2$  values (black lines) for different species as well as the number of assimilated observations per month (blue bars). The mean values of  $\chi^2$  are generally within 50% difference from the ideal value of 1 for PM<sub>2.5</sub>, PM<sub>10</sub>, NO<sub>2</sub> and O<sub>3</sub>, which suggests that the observation and background error covariance are generally well specified in the analysis of these species. Although the  $\chi^2$  values for these species showed pronounced seasonal variations that reflects the different error characteristics in different seasons, the  $\chi^2$  values were roughly stable for PM<sub>2.5</sub> and O<sub>3</sub> throughout the period, and for NO<sub>2</sub> and PM<sub>10</sub> after 2015 when the number of assimilated observations become stable, which generally shows the long-term stability of the performance of data assimilation. The  $\chi^2$  values for SO<sub>2</sub> were nevertheless greater than 1 in most cases, especially before 2017. This would be more relevant to the underestimations of background error covariance of SO<sub>2</sub> as we only specified 12% uncertainty in the SO<sub>2</sub> emissions, suggesting that the emission uncertainty of SO<sub>2</sub> may be underestimated by Zhang et al. (2009). There were also pronounced annual trends in the  $\chi^2$  values of SO<sub>2</sub>, which may be attributed to the increases of observation number from 2013 to 2014 and the substantial decreases of SO<sub>2</sub> observations. Although smaller than the  $\chi^2$  values of SO<sub>2</sub>, the values for CO were greater than 1 in most cases, suggesting the underestimations of the error covariances. Obvious decreasing trend can also be found in the  $\chi^2$  values of CO. The  $\chi^2$  test results suggest that our data assimilation system has relatively poor performance in the analysis of CO and SO<sub>2</sub> concentrations than the other four species, which is consistent with the cross-validation results which showed smaller  $R^2$  values for the reanalysis data of CO and SO<sub>2</sub> concentrations (*Sect.4.2.2 in the revised manuscript*). The annual trend of  $\chi^2$  values in CO and SO<sub>2</sub> also indicates relatively weak stability in the performance of data assimilation system on assimilating CO and SO<sub>2</sub> observations, which may influence the analysis of the annual trends in these two species. Based on these results, we have added discussions on this issue in our revised manuscript to inform the potential users of the problems that they should be aware of (*please see lines 667 – 670 in the revised manuscript*).



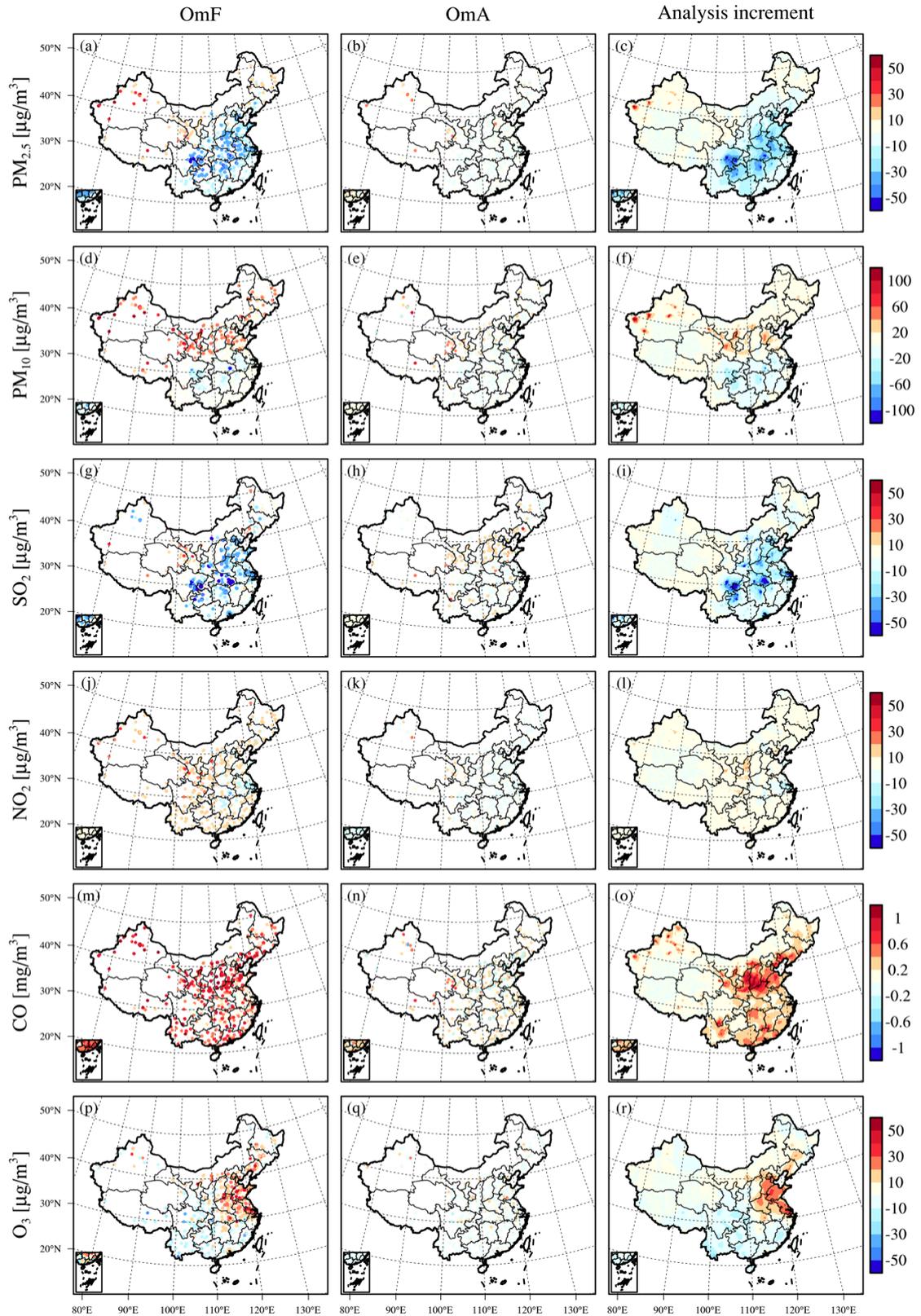
**Figure R1: Time series of the monthly mean  $\chi^2$  values (black lines) and the number of assimilated observations per month (blue bars) for (a)  $PM_{2.5}$ , (b)  $PM_{10}$ , (c)  $SO_2$ , (d)  $NO_2$ , (e)  $CO$  and (f)  $O_3$ .**

Spatial distributions of six-year averaged OmF & OmA values for each species in the observation space were then analyzed to investigate the structure of forecast bias and to measure the improvement in the reanalysis (Fig. R2). The analysis increment, which is estimated from the differences between the analysis and forecast, is also plotted to measure the adjustment made in the model space. The OmF values have showed positive model biases (i.e., negative OmF) in the  $PM_{2.5}$  and  $SO_2$  concentrations in east China, as well as  $PM_{10}$  and  $O_3$  concentrations south China. The negative model biases (i.e., positive OmF) were mainly found in the  $PM_{2.5}$  concentrations in west China, the  $PM_{10}$  concentrations in north China, the  $O_3$  concentrations in central-east China, as well as the concentrations of  $CO$  and  $NO_2$  throughout the whole China.

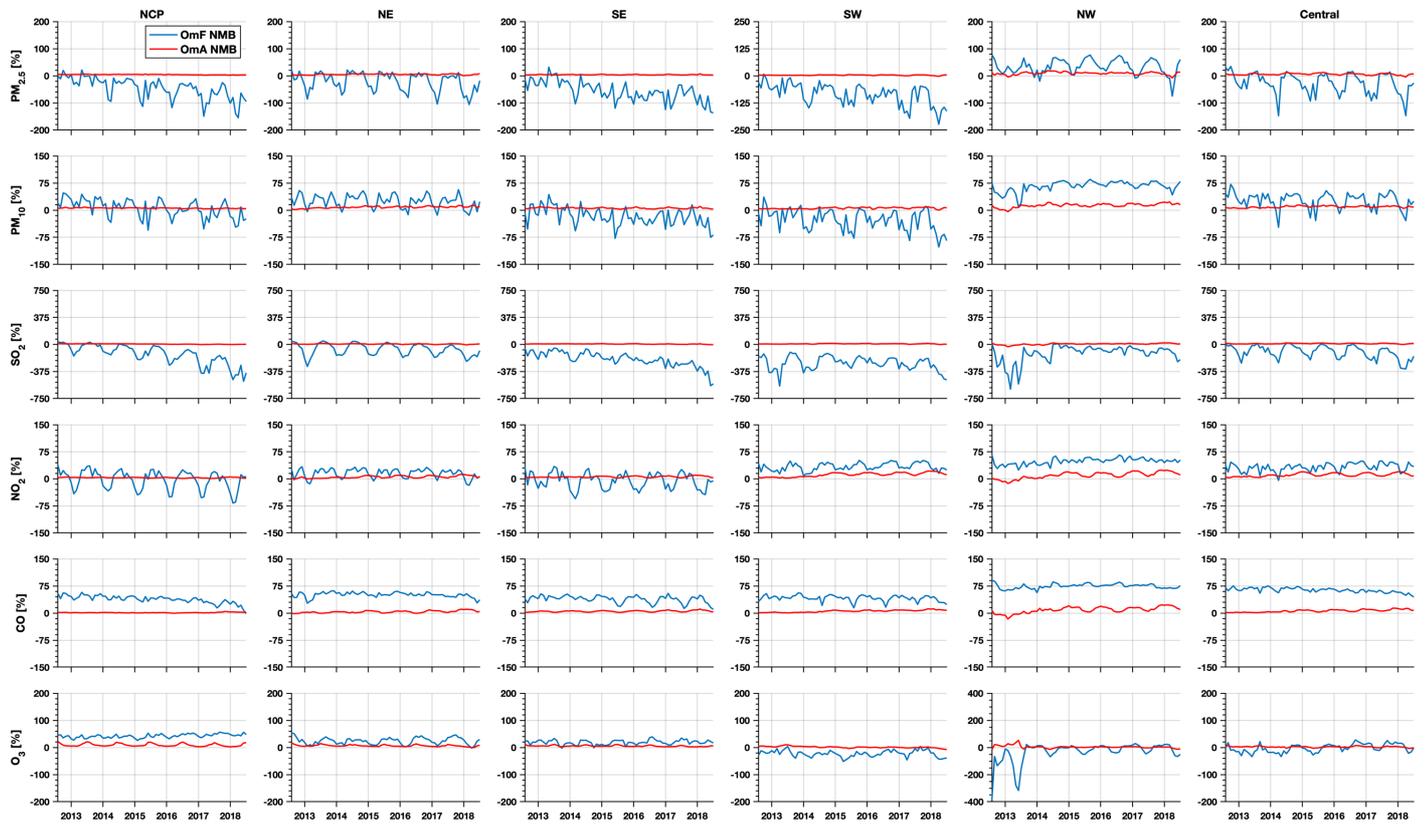
The OmA values suggest that the data assimilation removes most of the model biases for each species, which confirms the good performance of our data assimilation system. According to Fig. R3, the monthly mean OmF biases were almost completely removed in each regions of China because of assimilation, with mean OmF biases reducing by 32–94% for  $PM_{2.5}$ , 33–83% for  $PM_{10}$ , 25–96% for  $SO_2$ , 53–88% for  $NO_2$ , 88–97% for  $CO$  and 54–90% for  $O_3$  concentrations in different regions of China. The mean OmF RMSE were also reduced substantially by 80–93% for  $PM_{2.5}$ , 80–86% for  $PM_{10}$ , 73–96% for  $SO_2$ , 76–91% for  $NO_2$ , 88–96% for  $CO$  and 76–87% for  $O_3$  concentrations in different regions of China (Fig. R4). In addition, despite the mean OmF bias and OmF RMSE exhibit significant annual trend, the OmA bias and OmA RMSE are relatively stable during the assimilation period, which generally confirms the long-term stability of our data assimilation system.

The spatial patterns of analysis increment were in good agreement with those of the OmF values for each species, which generally shows negative (positive) increments for  $PM_{2.5}$  concentrations in east (west) China, negative (positive) increments for  $PM_{10}$  concentrations in south (north) China, negative increments for  $SO_2$

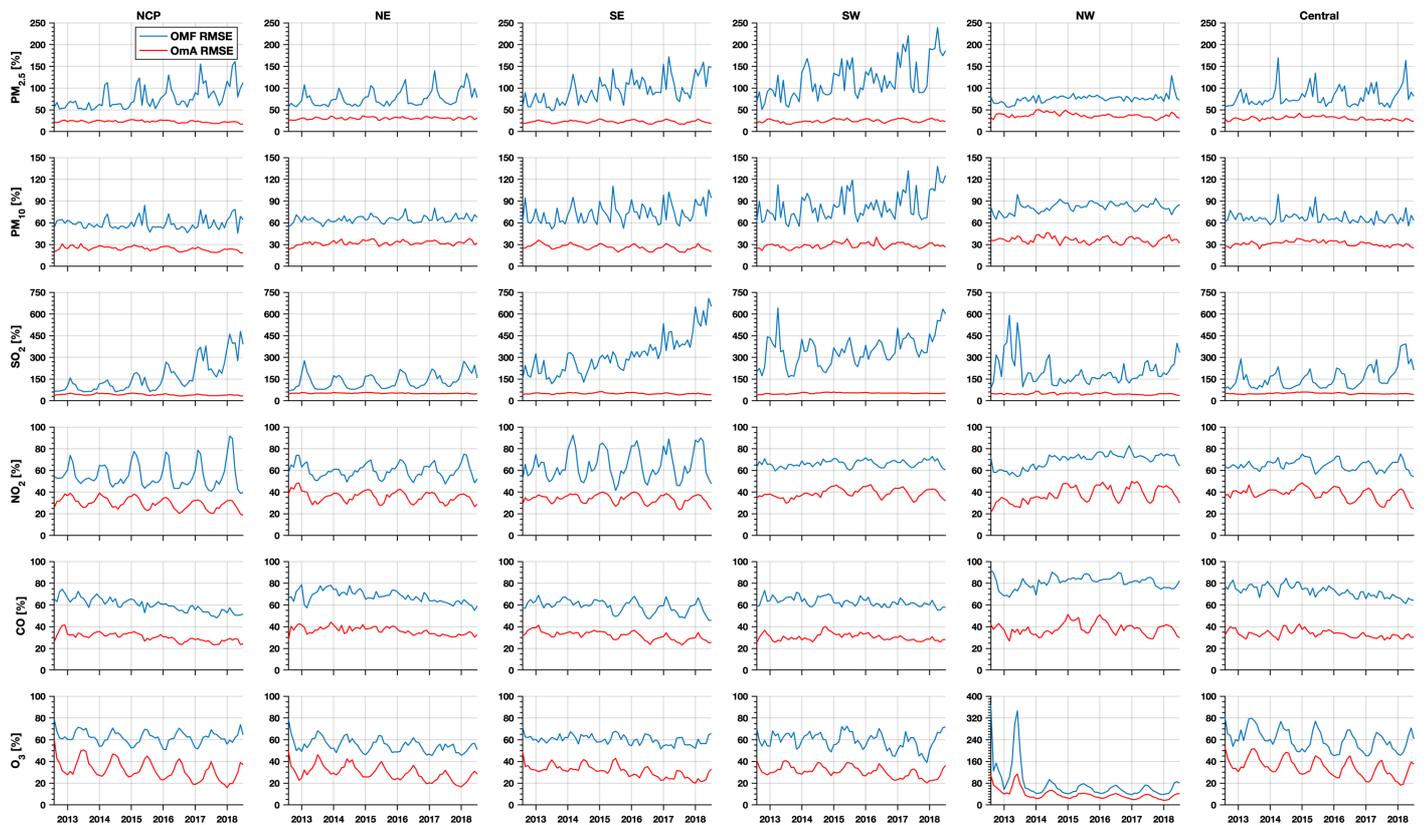
throughout the China, positive increments for CO and NO<sub>2</sub> concentrations throughout the China, and the positive (negative) increments for O<sub>3</sub> concentrations in central-east (south) China. These results confirm that the data assimilation can effectively propagate the observation information into the model state and reduced the model errors.



**Figure R2: Spatial distributions of the six-year mean OmF (left panel), OmA (middle panel) and analysis increment (right panel) for different species in China.**



**Figure R3: Time series of monthly mean OmF and OmA normalized mean bias in different regions of China for different species.**

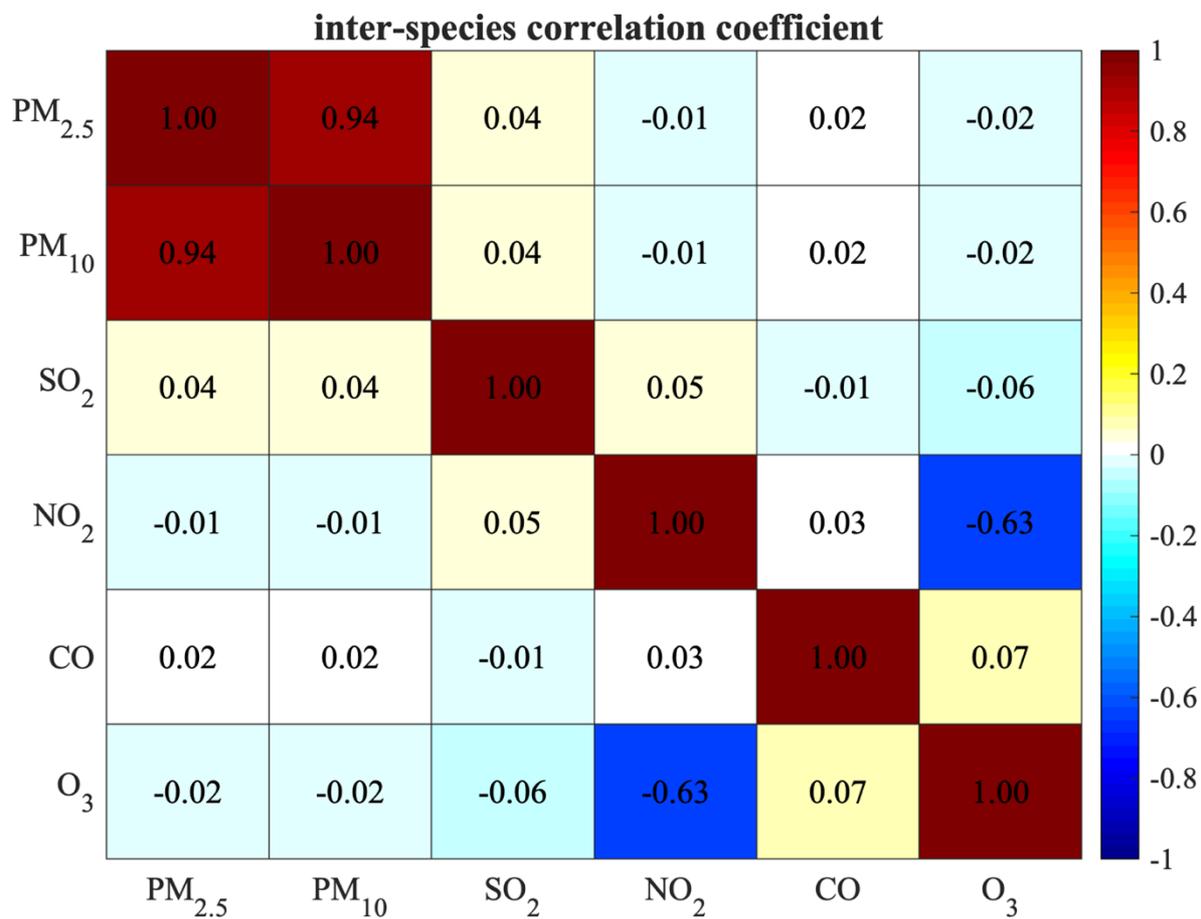


**Figure R4: Time series of monthly mean OmF and OmA normalized root mean square error in different regions of China for different species.**

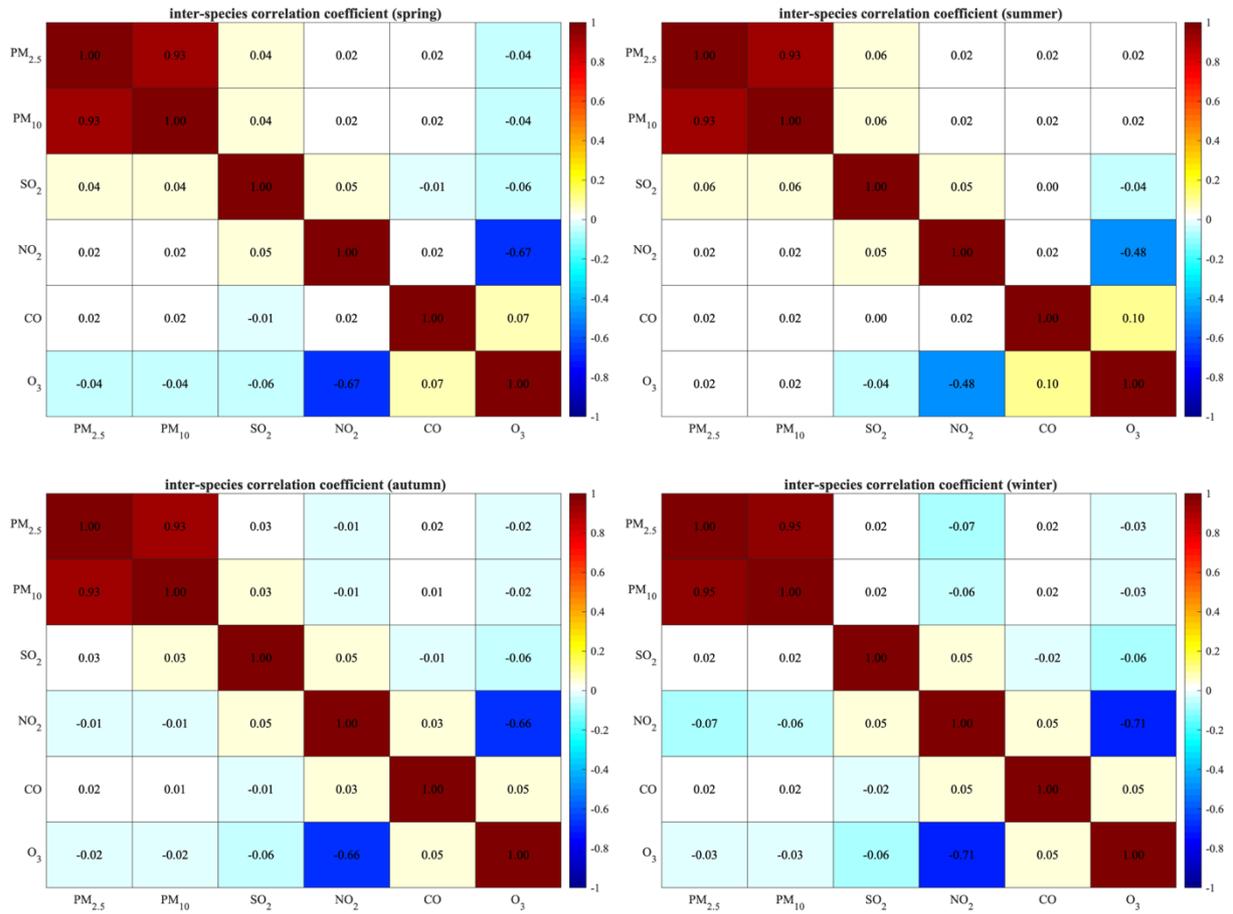
**Comment 3:** Inter-species correlation was totally neglected in background error covariance. This setting is extremely conservative and does not fully utilize the advantages of EnKF data assimilation that produces comprehensive background error patterns. I'm wondering if the authors have tried to implement inter-species correlations. Further discussion is needed (e.g., why it is so conservative, what is the disadvantage of the current setting).

**Reply:** Thanks for this important comment. We agree with the reviewer that including the correlations between the background errors of different chemical species has the capability to improve the assimilation performance as shown in Miyazaki et al. (2012). The reason that we neglected the inter-species correlation in the background error covariance is that we concentrated on the assimilations of primary air pollutants (except of O<sub>3</sub>) whose errors are more related to the errors in their emissions. Since the emission errors of these species were considered to be independent in this study (*Sect. 2.2 in the revised manuscript*), thus the background errors of these species have very weak correlations in most cases as shown in Figs. R5-6. The correlation between background errors of different species were generally near zero for most cases. Thus, we neglected these weak correlations to prevent the spurious correlation between non or weakly related variables in EnKF. In contrast, there are significant positive correlation between the background errors of PM<sub>2.5</sub> and PM<sub>10</sub> and negative correlation between the background errors of NO<sub>2</sub> and O<sub>3</sub>. The high correlation between PM<sub>2.5</sub> and PM<sub>10</sub> is just because PM<sub>2.5</sub> is a part of PM<sub>10</sub>, and there would be redundant information in the observations of PM<sub>2.5</sub> and PM<sub>10</sub> concentrations, thus we did not include the correlation between the PM<sub>2.5</sub> and PM<sub>10</sub> concentrations in the assimilation. The negative correlation between the O<sub>3</sub> and NO<sub>2</sub> is due to the NO<sub>x</sub>-OH-O<sub>3</sub> chemical reactions in the NO<sub>x</sub> saturated conditions that increases of NO<sub>2</sub> concentrations would reduce the O<sub>3</sub> concentrations due to the enhanced NO titration effect. However, the relationship between O<sub>3</sub> and NO<sub>2</sub> concentrations is actually nonlinear depending on the NO<sub>x</sub> limited or saturated conditions (Sillman, 1999), and previous study by Tang et al. (2016) has shown the limitations of the EnKF under strong nonlinear relationships. The cross-variable data assimilations of O<sub>3</sub> and NO<sub>2</sub> may come up with inefficient or even wrong adjustments. Considering the nonlinear relationship between the O<sub>3</sub> and NO<sub>2</sub> concentrations and their unexpected effects on EnKF, we took a conservative way in the assimilations of NO<sub>2</sub> and O<sub>3</sub> by neglecting their error correlations.

We agree with the reviewer that current setting may be too conservative to fully utilize the advantages of EnKF assimilation, however it can avoid possible serious negative influences on the reanalysis data caused by the spurious correlations or nonlinear chemical relationships. The different species can also be assimilated in a consistent way under current settings. Following the suggestions of reviewer, we have clarified the reasons for neglecting the inter-species correlations in the background error covariances in the revised manuscript (*please see lines 274 – 292 in the revised manuscript*).



**Figure R5: Correlations between species in the background error covariance matrix, estimated from the LETKF ensemble averaged from 2013 to 2018. The global mean of the covariance estimated for each station is plotted.**



**Figure R6: Correlations between species in the background error covariance matrix, estimated from the LETKF ensemble averaged in different seasons from 2013 to 2018. The global mean of the covariance estimated for each station is plotted.**

**Comment 4:** Please clarify whether there are any variations in inflation factor and how it was optimized for different species. In most regional ensemble data assimilation systems, fixed lateral boundary condition tends to limit the effectiveness of data assimilation near their boundaries (and also inside when horizontal advection is strong) because of reduced spreads. Did you find any problem with it?

**Reply:** Sorry for the confusion. In this study, the inflation factor was calculated based on Kalman filtering theory which requires that the ensemble and innovation spreads be of similar magnitude (Evensen, 2003; Wang and Bishop, 2003):

$$\langle dd^T \rangle \approx \mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R} \quad (6)$$

$$d = y^o - \mathbf{H}(x^b) \quad (7)$$

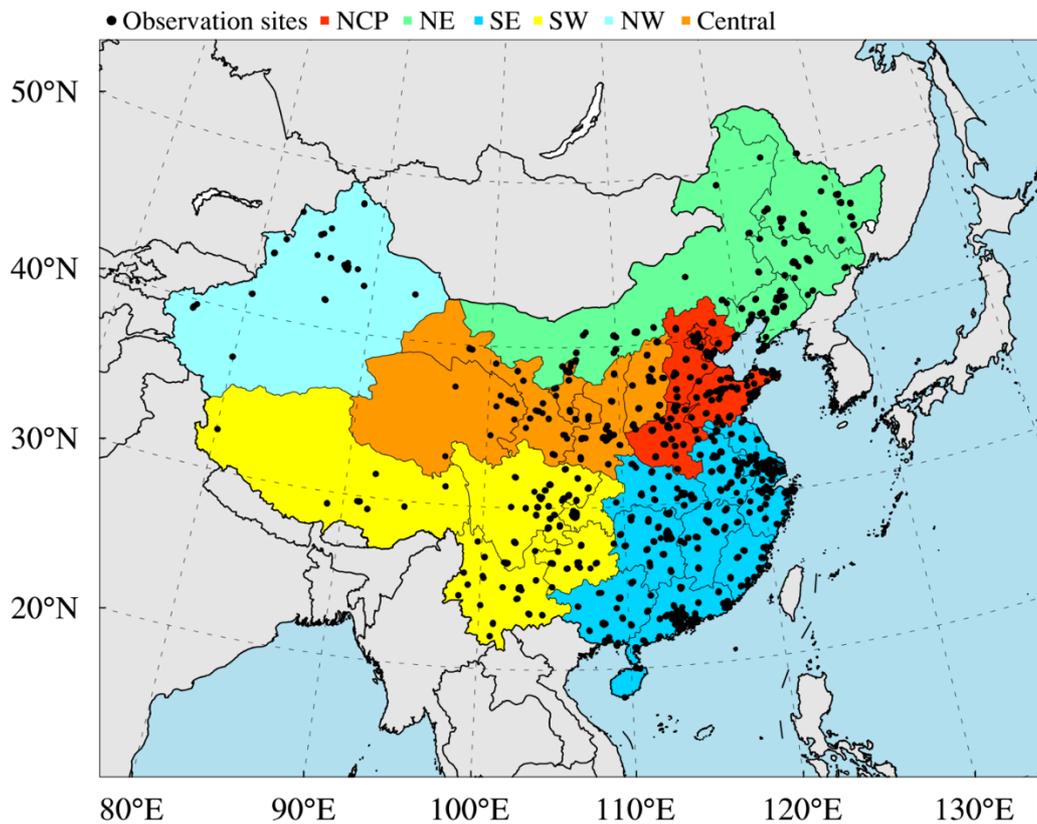
In order to balance the ensemble and innovation spreads, a multiplicative inflation factor for  $\mathbf{B}$  can be approximate by:

$$\lambda = \frac{(\mathbf{R}^{-1/2}d)^T \mathbf{R}^{-1/2}d - p}{\text{trace}\{\mathbf{R}^{-1/2}\mathbf{H}\mathbf{P}^b(\mathbf{R}^{-1/2}\mathbf{H})^T\}} \quad (8)$$

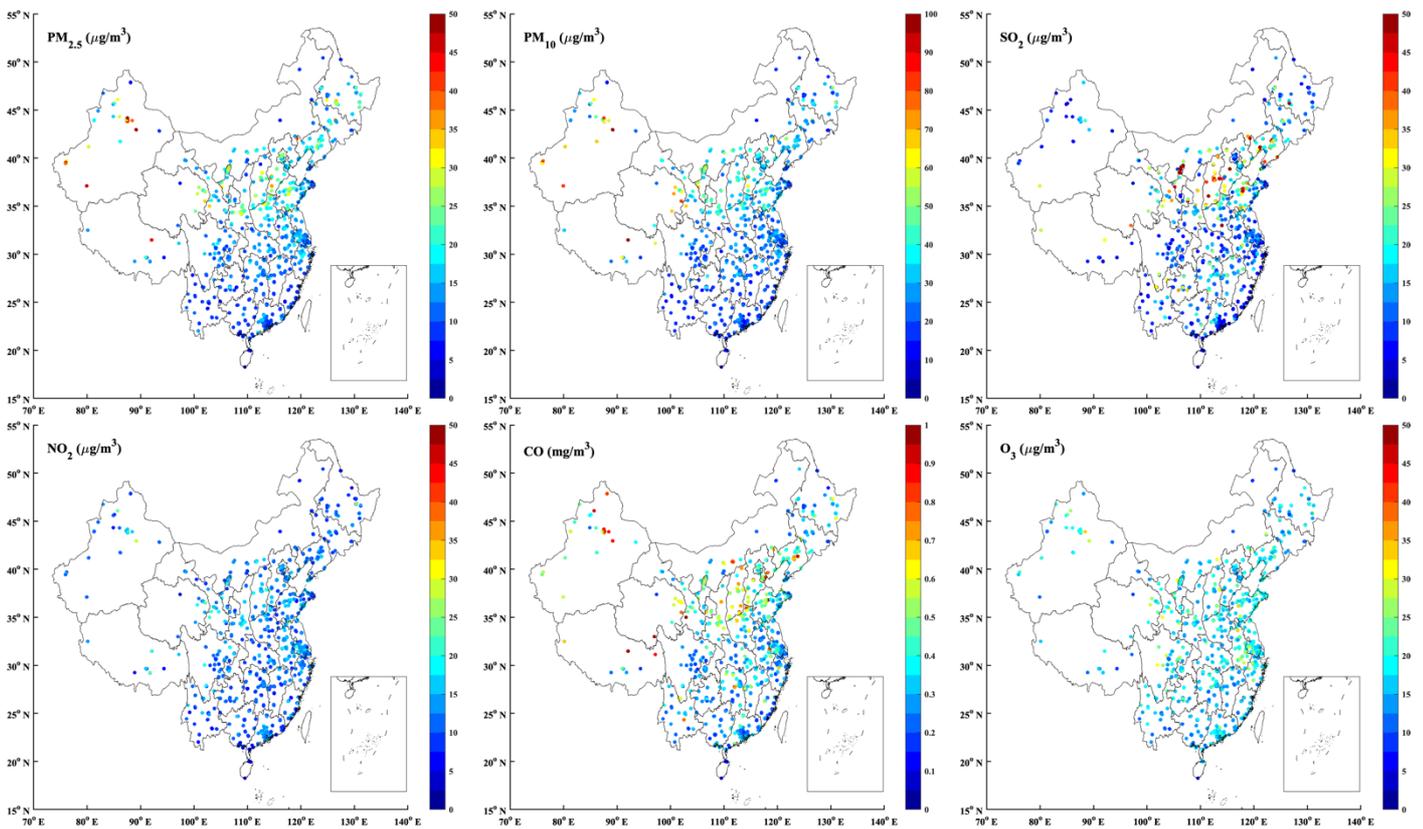
where the trace of the covariance matrix is used to approximate covariance on a globally averaged basis, and  $\langle \cdot \rangle$

denotes the ensemble average. Using Eq (8), the hourly inflation factor was calculated for each species. In addition, the inflation factor was calculated locally in this study. Thus, the inflation factor used in this assimilation is not only species specific, but also varies with time and space, which reflects different error characteristics of different species in different time and places. Following the suggestion of reviewer, we have clarified this issue in the revised manuscript (*please see lines 270 – 273 in the revised manuscript*).

We agree with the review that the use of fixed lateral boundary condition would lead to small ensemble spread near the boundary. Since we only assimilate the surface observations in China which were not near the boundary of the modeling region in most cases (Fig. R7), the effects of fixed boundary condition were small in this study. This can be shown in Fig. R8 which shows that the OmA RMSE values at the sites near the boundary of the China were approximate to those at inland sites. In addition, the inflation technique was also used to inflate the background error covariance, which could reduce the effects of the small ensemble spread on the analysis.



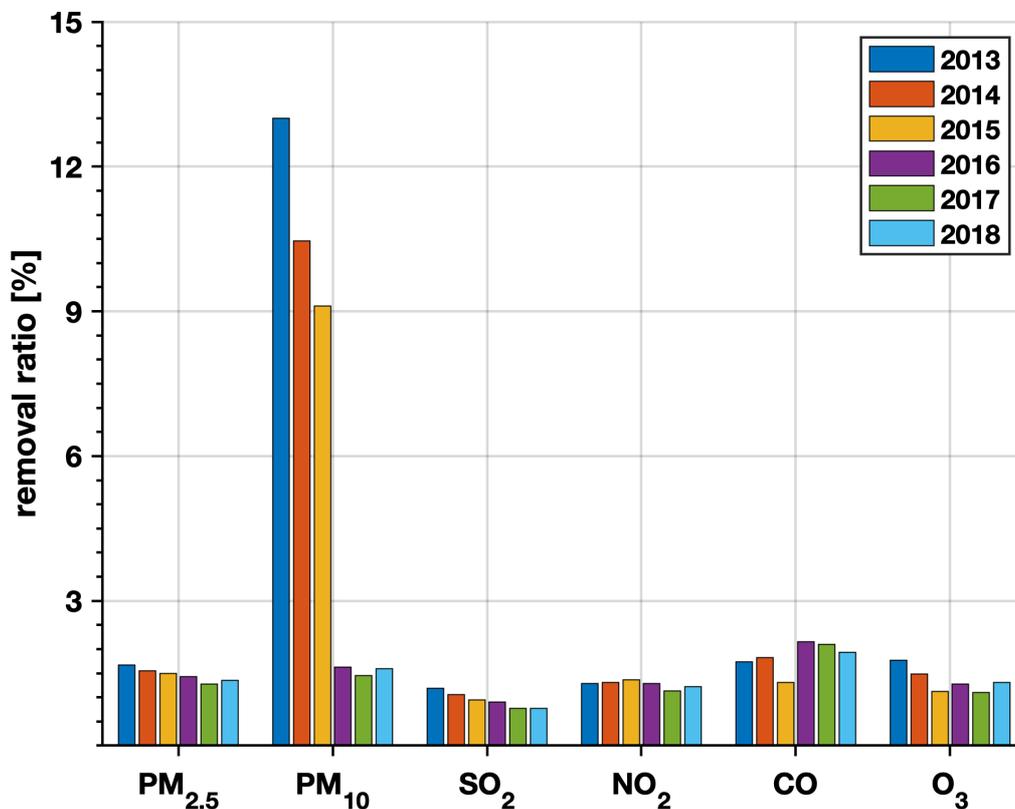
**Figure R7: Modeling domain of the ensemble simulation overlay the distributions of observation sites from CNEMC. Different colours denote the different regions in China, namely North China Plain (NCP), Northeast China (NE), Southwest China (SW), Southeast China (SE), Northwest China (NW) and Central.**



**Figure R8: Spatial distributions of OmA RMSE values for (a)  $PM_{2.5}$ , (b)  $PM_{10}$ , (c)  $SO_2$ , (d)  $NO_2$ , (e) CO and (f)  $O_3$  in China**

**Comment 5:** Using automatic outlier detection method, how much observations were rejected? What was the impact in data assimilation?

**Reply:** Thanks for this comment. Figure R9 shows the removal ratios of the six pollutants from 2013 to 2018, which were less than 1.5% for most air pollutants throughout the assimilation period. The  $PM_{10}$  observations have a high removal ratio (9–13%) during 2013–2015 with most of outliers marked by an observed concentration of  $PM_{2.5}$  higher than that of  $PM_{10}$  at the same hour and same site (Wu et al., 2018). However, there was a sharp decrease in removal ratios of  $PM_{10}$  in 2016 (~1.5%) because of the implementation of a compensation algorithm for the loss of semi-volatile materials in the  $PM_{10}$  measurements (Wu et al., 2018).



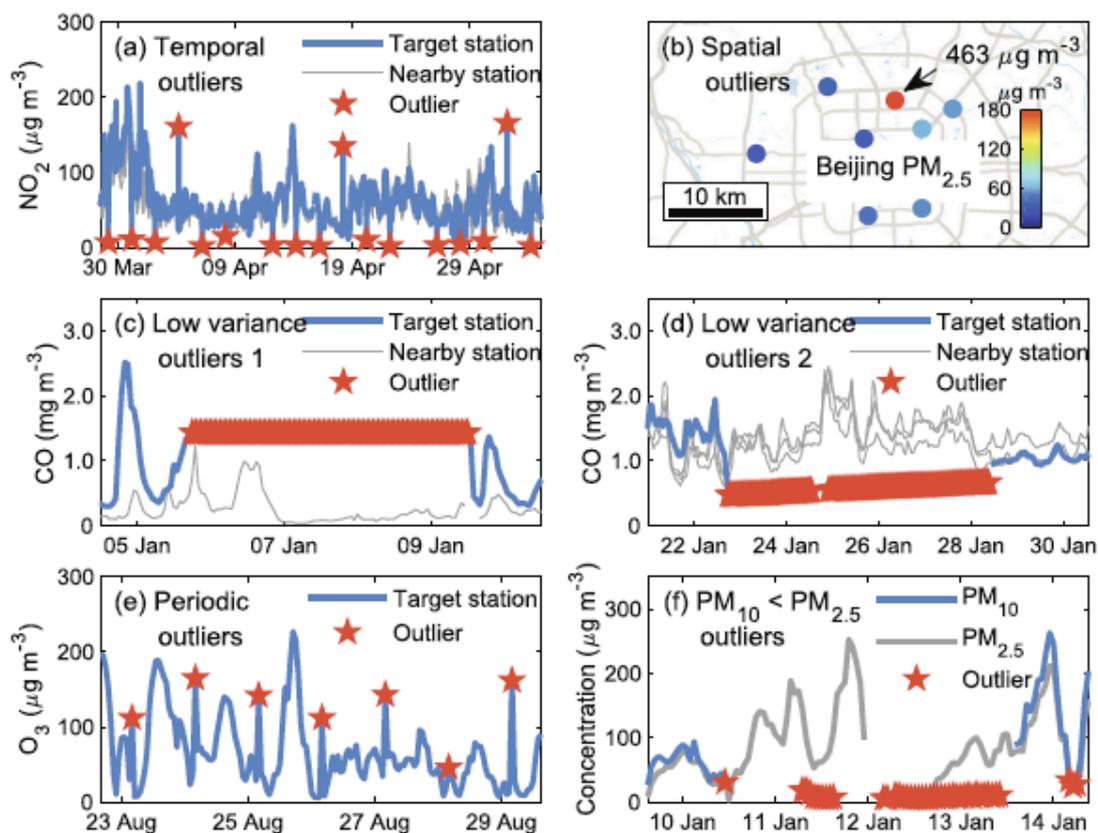
**Figure R9: Removal ratio of all observation sites in China from 2013 to 2018 for different species detected by the automatic outlier detection method.**

The outlier detection method was essential for the assimilations of surface observations due to the existence of outliers in the original observation dataset. The outlier detection method has been applied to detect all four types of outliers in the hourly surface observations of air pollutants, which were characterized by temporal and spatial inconsistency (ST-outliers), instrument-induced low variances (LV-outliers), periodic calibration exceptions (P-outliers) and less PM<sub>10</sub> than PM<sub>2.5</sub> observations (LP-outliers).

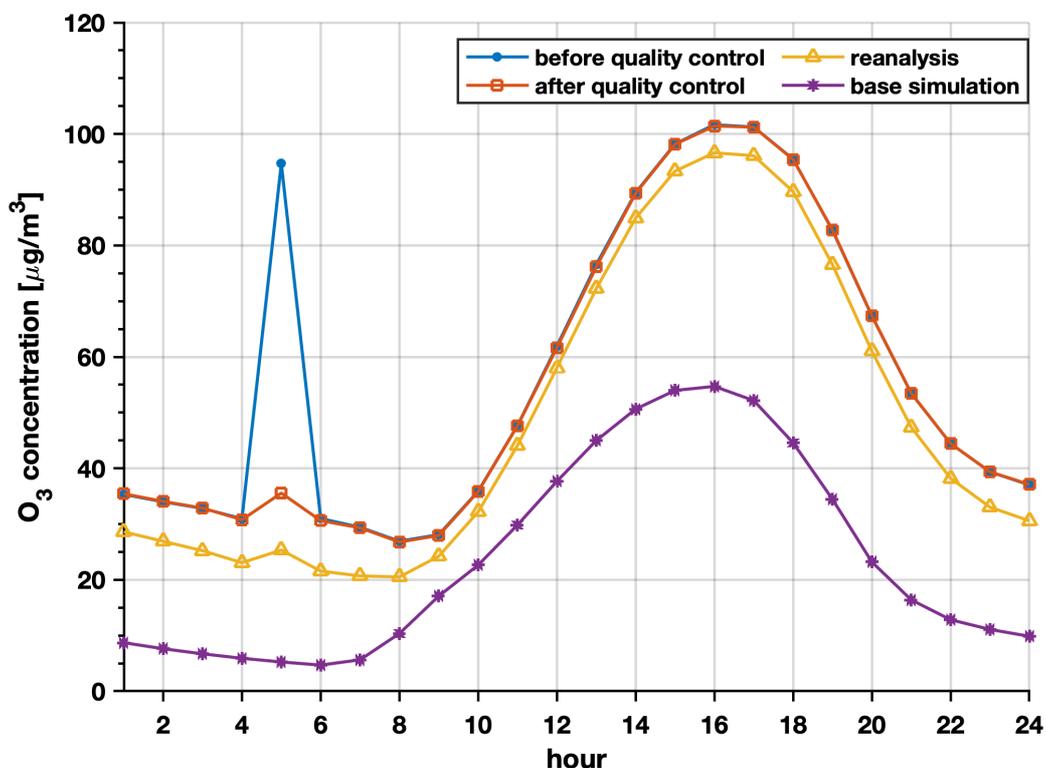
As exemplified by Fig. R10a and Fig. R10b obtained from Wu et al. (2018), the ST-outliers are observations that differ greatly from values observed at adjacent time or those in neighboring areas, such as the abnormally low values in NO<sub>2</sub> observations or the abnormally high values in PM<sub>2.5</sub> observations. The LV-outliers are characterized by a very low variance in time series compared to neighboring sites (Fig. R10d). In cases when the pump of the instruments is stuck, or the filter tape is depleted, the observations even do not change over time (Fig. R10c). The P-outliers are mainly induced by the regular calibration process for the instruments, such as O<sub>3</sub> observation instruments (Fig. R10e), which may interfere with the observations and insert abnormal values into online measurement datasets. The LP-outlier involves PM<sub>2.5</sub> concentrations being higher than PM<sub>10</sub> concentrations observed at the same hour and same site which is mainly caused by the loss of semi-volatile components of particulate matter in the instruments.

The different kinds of outliers emphasized that it is necessary to filter out these outliers before the assimilation,

otherwise these outliers would introduce serious impacts on the quality of reanalysis data both in temporal and spatial consistency, sometimes even lead to wrong assimilation results. For example, as shown in Fig. R11, there is a false  $O_3$  peak in the original observation data due to the P-outliers occurred at 0400 LST. The quality assurance largely reduces this false peak and the observation data after quality assurance show more reasonable diurnal variations of  $O_3$  concentrations, which has guaranteed the quality of reanalysis data. Thus, the outlier detection method used in this study plays an indispensable role in the chemical data assimilations based on surface observations.



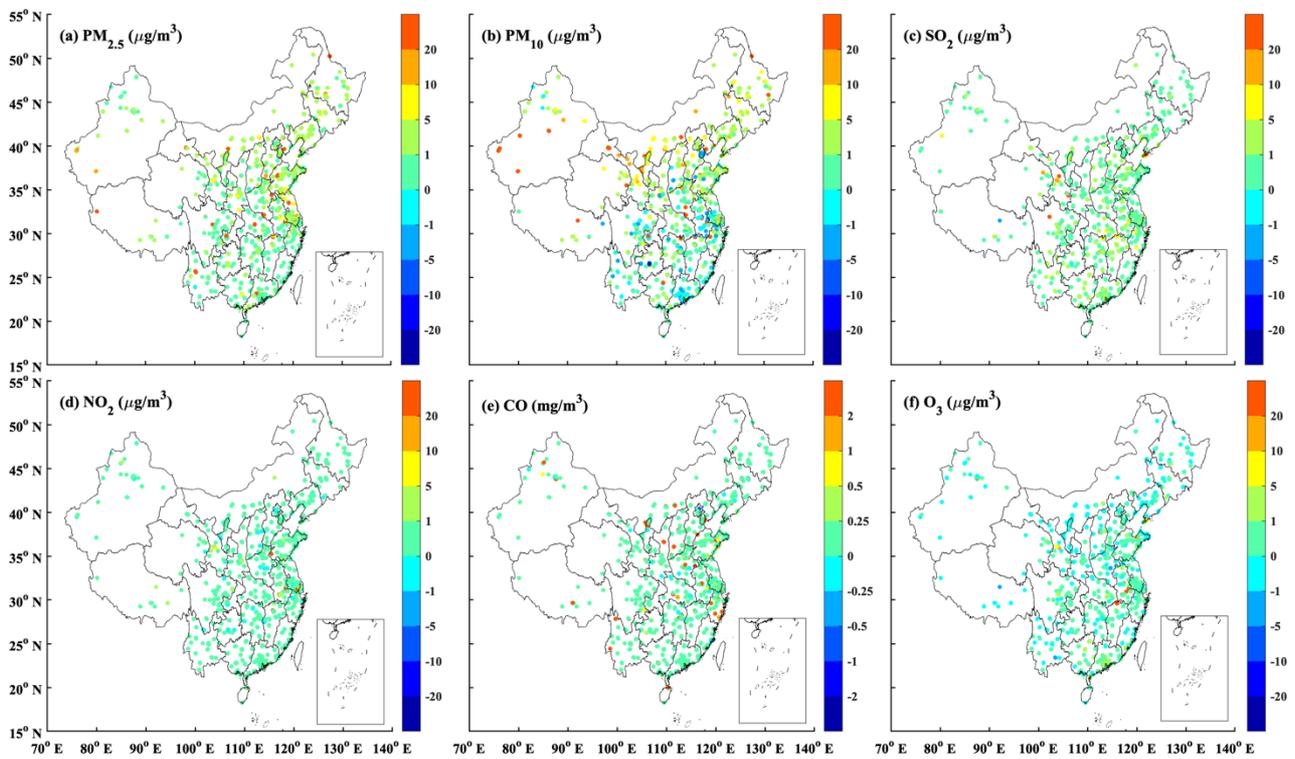
**Figure R10: Examples of classified outliers in surface observations of air pollutants. (a, b) Spatiotemporal outliers have large differences with neighboring observations in time and space. (c, d) Low variance outliers either stay the same or change abnormally slowly in time and differ significantly with observations from nearby sites. (e) Periodic outliers appear periodically, usually every 24 h. (f)  $PM_{10} < PM_{2.5}$  outliers are the  $PM_{10}$  observations that are lower than the  $PM_{2.5}$  observations at the same time and site (taken from figure 1 in Wu et al. (2018)).**



**Figure R11: Six-year averaged diurnal variations of O<sub>3</sub> concentrations in Wuhan, China obtained from observations before and after quality control, reanalysis data and base simulation.**

The differences in annual concentrations caused by quality control were also shown in Fig. R12 to assess the potential impacts of outlier detection on the assimilations. The differences were generally positive for PM<sub>2.5</sub>, SO<sub>2</sub>, NO<sub>2</sub> and CO concentrations, indicating a lower tendency of these species' concentrations due to the use of outlier detection. Negative differences were mainly found in the PM<sub>10</sub> concentrations in south China and the O<sub>3</sub> concentrations throughout China. According to estimation, the impacts of outlier detection were generally small in most stations. The differences were less than 5 µg/m<sup>3</sup> (1 µg/m<sup>3</sup>) for PM<sub>2.5</sub> concentrations over most stations in north (south) China and less than 1 µg/m<sup>3</sup> for the gaseous air pollutants for most stations throughout China. The differences were shown to be relative larger for PM<sub>10</sub> concentrations over northwest China which can be over 20 µg/m<sup>3</sup> in stations around Taklimakan Desert. This would be due to the higher outlier ratios in the observations over the remote areas.

These results suggest that the use of outlier detection is necessary for the assimilations of surface air quality observations, which prevents the negative influences of outliers on the reanalysis and improves its temporal and spatial consistency. The impacts of outlier detection on the estimated concentrations were also small in most stations. Following the suggestion of review, more descriptions about the impacts of outlier detection method on the assimilation were added in the revised manuscript (*please see lines 195 – 216 in the revised manuscript*).



**Figure R12: Spatial distributions of differences in annual concentrations of six air pollutants in China before and after quality control averaged from 2013 to 2018.**

**Comment 6:** Because of the fine-scale variability and large degree of freedoms, the high-region data assimilation would require larger ensembles. I'm wondering if 50 members are sufficient. Further discussion is needed to demonstrate whether the background error is produced properly to propagate observational information in space.

**Reply:** Thanks for this important comment. We agree with the review that high-resolution data assimilation requires larger ensembles due to the fine-scale variability and large degree of freedoms. The ensemble size determines the accuracy to which the background error covariance is approximated. A large ensemble size is essential to capture the proper background error covariance structure, but it is computationally expensive since the cost of EnKF linearly increases with the ensemble size while the accuracy of the covariance estimate improves by its square root (Constantinescu et al., 2007a; Miyazaki et al., 2012). The appropriate ensemble size depends on the specific application and model. The idealized experiments of Constantinescu et al. (2007a) have shown that a 50-member ensemble has significant improvements against smaller ensembles which is also computationally affordable given the computational resources. In a realistic chemical data assimilation application with horizontal resolution of  $\sim 2.8^\circ$ , Miyazaki et al. (2012) has shown that the analysis is improved significantly by increasing the ensemble size from 16 to 32 and is further somewhat improved by increasing it from 32 to 48. However, the impact was much less significant by increasing it from 48 to 64. An ensemble size of 48 was thus recommended. Ensemble size of 50 members are also typical in numerical weather prediction which are thought to provide a good balance between accuracy and computational efficiency (Constantinescu et al., 2007b).

Thus, the ensemble size was chosen to be 50 in this study based on the previous publications (Constantinescu et al., 2007a, b; Miyazaki et al., 2012) and our previous high-resolution (~9km) regional assimilation work (Tang et al., 2016; Tang et al., 2011; Tang et al., 2013) which showed that a 50-member ensemble keeps good balance between computational efficiency and assimilation performance. Several measures were also conducted to deal with the large degree of freedoms in our high-resolution assimilation work. First, we assumed that the emission errors were spatially correlated when we perturbed the emissions. An isotropic gaussian correlation model with a decorrelation length of 150km was used in the error covariance of emissions, which was written as

$$\rho(i, j) = \exp\left\{-\frac{1}{2}\left[\frac{h(i, j)}{l}\right]^2\right\} \quad (9)$$

where  $\rho(i, j)$  represents the correlation between grid  $i$  and  $j$ ,  $h(i, j)$  represents the distance between these two points and  $l$  represents the decorrelation length. This would reduce the degree of freedoms in the state vector and alleviate the impacts of limited ensembles on high-resolution assimilation applications. Secondly, we adopted an adaptive inflation method to prevent the underestimations of the background error covariance due to the limited ensemble sizes. Thirdly, the local analysis scheme has been used in our study to deal with the rank problems and spurious correlation caused by the limited ensemble size. These measures enable our applications of the EnKF with limited ensemble size on the high-resolution data assimilation at affordable computational cost. As shown in Fig R2, the spatial patterns of analysis increment were in good agreement with those of the OmF residuals for each species, this suggests that estimated background error covariance can effectively propagate the observation information into the model state and reduced the model errors.

Therefore, given the expensive computational cost in the high-resolution ensemble simulations, the 50-member ensemble was used in this study as a trade-off between assimilation performance and computational efficiency. However, better assimilation performance is expected when a larger ensemble size is used. Following the suggestions of review, we added more discussions on the choice of ensemble size in the revised manuscript (*please see lines 159 – 190 and lines 662 – 667 in the revised manuscript*).

## References

- Chen, D., Liu, Z., Ban, J., Zhao, P., and Chen, M.: Retrospective analysis of 2015–2017 wintertime PM<sub>2.5</sub> in China: response to emission regulations and the role of meteorology, *Atmos. Chem. Phys.*, 19, 7409–7427, <https://doi.org/10.5194/acp-19-7409-2019>, 2019.
- Constantinescu, E. M., Sandu, A., Chai, T. F., and Carmichael, G. R.: Assessment of ensemble-based chemical data assimilation in an idealized setting, *Atmos. Environ.*, 41, 18–36, <https://doi.org/10.1016/j.atmosenv.2006.08.006>, 2007a.
- Constantinescu, E. M., Sandu, A., Chai, T. F., and Carmichael, G. R.: Ensemble-based chemical data assimilation. I: General approach, *Q. J. R. Meteorol. Soc.*, 133, 1229–1243, <https://doi.org/10.1002/qj.76>, 2007b.
- 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, <https://doi.org/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 statistics, *J. Geophys. Res.-Oceans*, 99, 10143–10162, <https://doi.org/10.1029/94JC00572>, 1994.
- Feng, S. Z., Jiang, F., Jiang, Z. Q., Wang, H. M., Cai, Z., and Zhang, L.: Impact of 3DVAR assimilation of surface PM<sub>2.5</sub> observations on PM<sub>2.5</sub> forecasts over China during wintertime, *Atmos. Environ.*, 187, 34–49, <https://doi.org/10.1016/j.atmosenv.2018.05.049>, 2018.
- Jiang, Z. Q., Liu, Z. Q., Wang, T. J., Schwartz, C. S., Lin, H. C., and Jiang, F.: Probing into the impact of 3DVAR assimilation of surface PM<sub>10</sub> observations over China using process analysis, *J. Geophys. Res.-Atmos.*, 118, 6738–6749, <https://doi.org/10.1002/jgrd.50495>, 2013.
- Li, F., Tang, X., Wang, Z., Zhu, L., Wang, X., Wu, H., Lu, M., Li, J., and Zhu, J.: Estimation of Representative Errors of Surface Observations of Air Pollutant Concentrations Based on High-Density Observation Network over Beijing-Tianjin-Hebei Region, *Chinese Journal of Atmospheric Sciences*, 43, 277–284, [https://doi.org/1006-9895\(2019\)43:2<277>2.0.TX;2-S](https://doi.org/1006-9895(2019)43:2<277>2.0.TX;2-S), 2019.
- Ma, C. Q., Wang, T. J., Mizzi, A. P., Anderson, J. L., Zhuang, B. L., Xie, M., and Wu, R. S.: Multiconstituent Data Assimilation With WRF-Chem/DART: Potential for Adjusting Anthropogenic Emissions and Improving Air Quality Forecasts Over Eastern China, *J. Geophys. Res.-Atmos.*, 124, 7393–7412, <https://doi.org/10.1029/2019jd030421>, 2019.
- Menard, R. and Chang, L. P.: Assimilation of stratospheric chemical tracer observations using a Kalman filter. Part II: chi(2)-validated results and analysis of variance and correlation dynamics, *Mon. Weather Rev.*, 128, 2672–2686, [https://doi.org/10.1175/1520-0493\(2000\)128<2672:Aoscto>2.0.Co;2](https://doi.org/10.1175/1520-0493(2000)128<2672:Aoscto>2.0.Co;2), 2000.
- Miyazaki, K., Eskes, H. J., and Sudo, K.: A tropospheric chemistry reanalysis for the years 2005–2012 based on an assimilation of OMI, MLS, TES, and MOPITT satellite data, *Atmos. Chem. Phys.*, 15, 8315–8348, <https://doi.org/10.5194/acp-15-8315-2015>, 2015.
- Miyazaki, K., Eskes, H. J., Sudo, K., Takigawa, M., van Weele, M., and Boersma, K. F.: Simultaneous assimilation of satellite NO<sub>2</sub>, O<sub>3</sub>, CO, and HNO<sub>3</sub> data for the analysis of tropospheric chemical composition and emissions, *Atmos. Chem. Phys.*, 12, 9545–9579, <https://doi.org/10.5194/acp-12-9545-2012>, 2012.
- Pagowski, M. and Grell, G. A.: Experiments with the assimilation of fine aerosols using an ensemble Kalman filter, *J. Geophys. Res.-Atmos.*, 117, 15, <https://doi.org/10.1029/2012jd018333>, 2012.
- 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. R. Meteorol. Soc.*, 136, 2013–2024, <https://doi.org/10.1002/qj.700>, 2010.
- Peng, Z., Liu, Z., Chen, D., and Ban, J.: Improving PM<sub>2.5</sub> forecast over China by the joint adjustment of initial conditions and source emissions with an ensemble Kalman filter, *Atmos. Chem. Phys.*, 17, 4837–4855, <https://doi.org/10.5194/acp-17-4837-2017>, 2017.
- Sillman, S.: The relation between ozone, NO<sub>x</sub> and hydrocarbons in urban and polluted rural environments, *Atmos. Environ.*, 33, 1821–1845, [https://doi.org/10.1016/s1352-2310\(98\)00345-8](https://doi.org/10.1016/s1352-2310(98)00345-8), 1999.
- 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,

- <https://doi.org/10.5194/acp-11-12901-2011>, 2011.
- Tang, X., Zhu, J., Wang, Z. F., Wang, M., Gbaguidi, A., Li, J., Shao, M., Tang, G. Q., and Ji, D. S.: Inversion of CO emissions over Beijing and its surrounding areas with ensemble Kalman filter, *Atmos. Environ.*, 81, 676-686, <https://doi.org/10.1016/j.atmosenv.2013.08.051>, 2013.
- Tang, X., Zhu, J., Wang, Z., Gbaguidi, A., Lin, C., Xin, J., Song, T., and Hu, B.: Limitations of ozone data assimilation with adjustment of NO<sub>x</sub> emissions: mixed effects on NO<sub>2</sub> forecasts over Beijing and surrounding areas, *Atmos. Chem. Phys.*, 16, 6395-6405, <https://doi.org/10.5194/acp-16-6395-2016>, 2016.
- Wang, X. G. and Bishop, C. H.: A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes, *J. Atmos. Sci.*, 60, 1140-1158, [https://doi.org/10.1175/1520-0469\(2003\)060<1140:Acobae>2.0.Co;2](https://doi.org/10.1175/1520-0469(2003)060<1140:Acobae>2.0.Co;2), 2003.
- Werner, M., Kryza, M., Pagowski, M., and Guzikowski, J.: Assimilation of PM<sub>2.5</sub> ground base observations to two chemical schemes in WRF-Chem - The results for the winter and summer period, *Atmos. Environ.*, 200, 178-189, <https://doi.org/10.1016/j.atmosenv.2018.12.016>, 2019.
- Wu, H. J., Tang, X., Wang, Z. F., Wu, L., Lu, M. M., Wei, L. F., and Zhu, J.: Probabilistic Automatic Outlier Detection for Surface Air Quality Measurements from the China National Environmental Monitoring Network, *Adv. Atmos. Sci.*, 35, 1522-1532, <https://doi.org/10.1007/s00376-018-8067-9>, 2018.
- 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, <https://doi.org/10.5194/acp-9-5131-2009>, 2009.