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 ε^{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:

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

where r_{repr} represents the representativeness error, Δx represents the model resolution, L_{repr} represents the characteristic representativeness length of the observation site and ε^{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 ε^{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 ε^{abs} empirically to be half of the measurement error following the study by Pagowski et al. (2010). In this study, the ε^{abs} was obtained from Li et al. (2019) who estimated the ε^{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 ε_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 Δx is equal to 30km. Based on the estimated L_{repr} and the ε_i^{abs} for different species, the representativeness errors are estimated using Eq. (1) by specifying the Δx 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*).

Changes in the manuscript: lines 223–245

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 χ^2 diagnosis and the statistics of observation minus forecast (OmF: $\boldsymbol{y^o} - \boldsymbol{H}(\boldsymbol{x^b})$) & observation minus analysis (OmA: $\boldsymbol{y^o} - \boldsymbol{H}(\boldsymbol{x^a})$ in the revised manuscript to demonstrate the performance of our assimilation system (*please see lines 317–369 in the revised manuscript*).

χ^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 ($\mathbf{HBH^T} + \mathbf{R}$):

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

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

where *m* is the number of observations. According to the Kalman filtering theory, the mean of χ^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 χ^2 values (black lines) for different species as well as the number of assimilated observations per month (blue bars). The mean values of χ^2 are generally within 50% difference from the ideal value of 1 for $PM_{2.5}$, PM_{10}, NO_2 and O_3, which suggests that the observation and background error covariance are generally well specified in the analysis of these species. Although the χ^2 values for these species showed pronounced seasonal variations that reflects the different error characteristics in different seasons, the χ^2 values were roughly stable for $PM_{2.5}$ and O_3 throughout the period, and for NO_2 and PM_{10} after 2015 when the number of assimilated observations become stable, which generally shows the long-term stability of the performance of data assimilation. The χ^2 values for SO_2 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_2 as we only specified 12% uncertainty in the SO_2 emissions. suggesting that the emission uncertainty of SO_2 may be underestimated by Zhang et al. (2009). There were also pronounced annual trends in the χ^2 values of SO_2, which may be attributed to the increases of observation number from 2013 to 2014 and the substantial decreases of SO_2 observations. Although smaller than the χ^2 values of SO_2, 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 χ^2 values of CO. The χ^2 test results suggest that our data assimilation system has relatively poor performance in the analysis of CO and SO_2 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_2 concentrations (*Sect.4.2.2 in the revised manuscript*). The annual trend of χ^2 values in CO and SO_2 also indicates relatively weak stability in the performance of data assimilation system on assimilating CO and SO_2 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*).

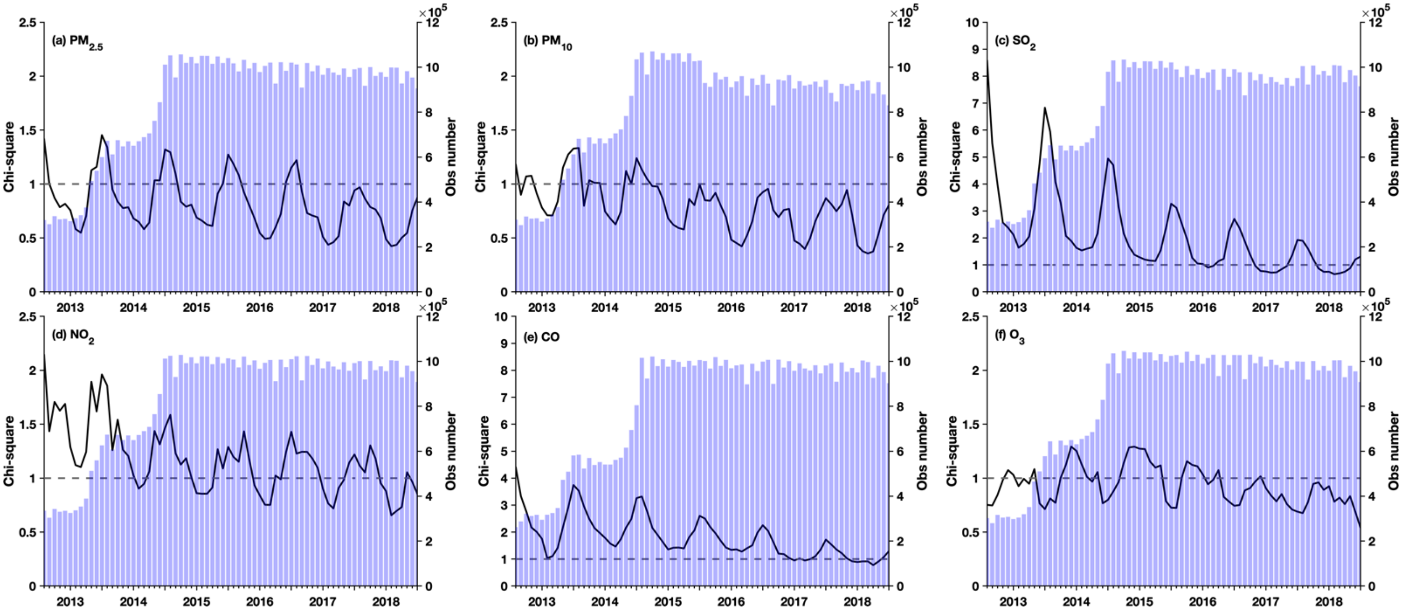


Figure R1: Time series of the monthly mean χ^2 values (black lines) and the number of assimilated observations per month (blue bars) for (a) $PM_{2.5}$, (b) PM_{10}, (c) SO_2, (d) NO_2, (e) CO and (f) O_3.

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_2 concentrations throughout the China, and the positive (negative) increments for O_3 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.

Changes in the manuscript: lines 317–319, lines 667–670, Figure 3 and Figure 4.

Changes in the supplementary: Figure S5 and Figure S6.

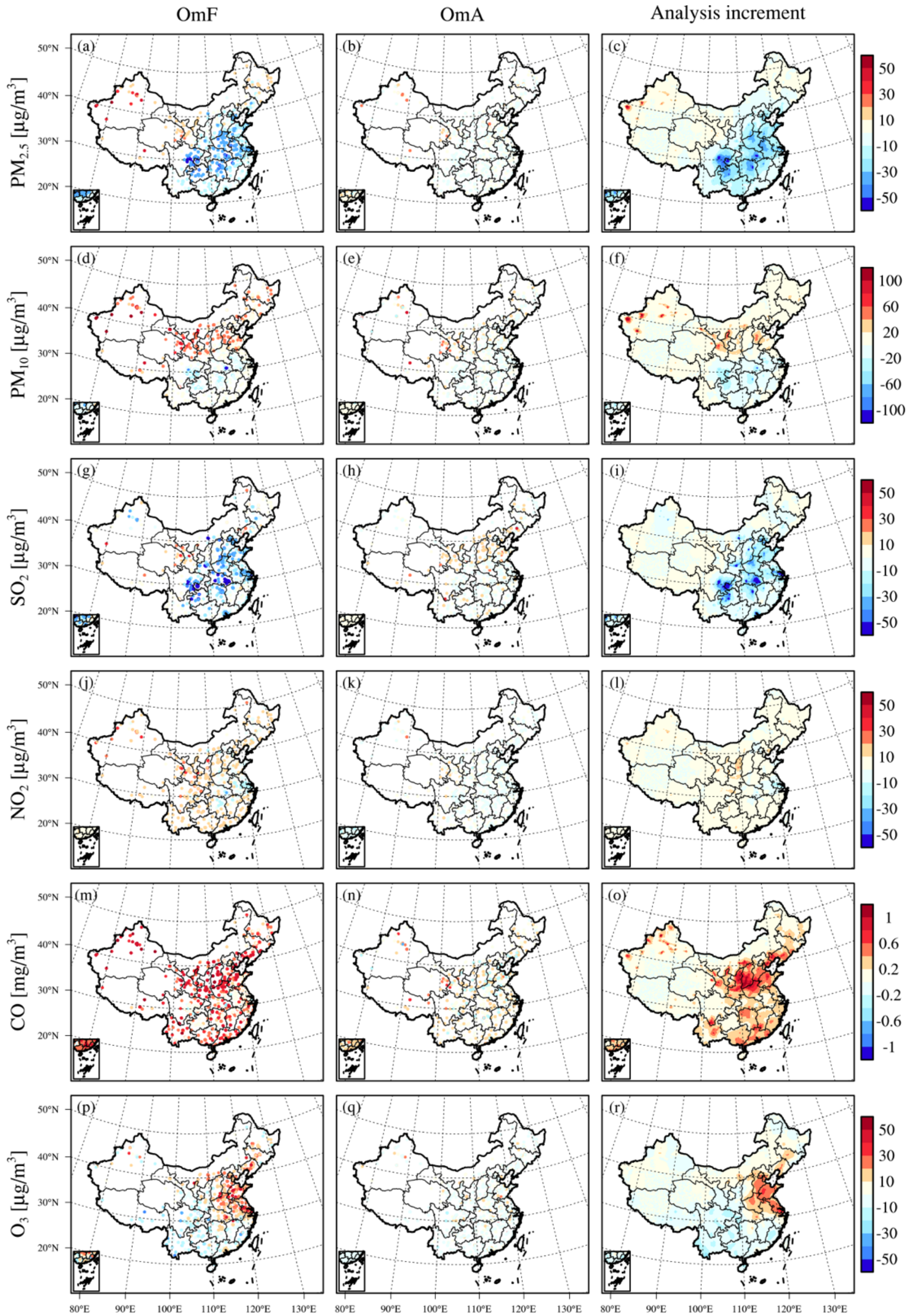


Figure R2: Spatial distributions of the six-year mean OmF (left panel), OmA (middle panel) and analysis increment (right panel) for different species in China.

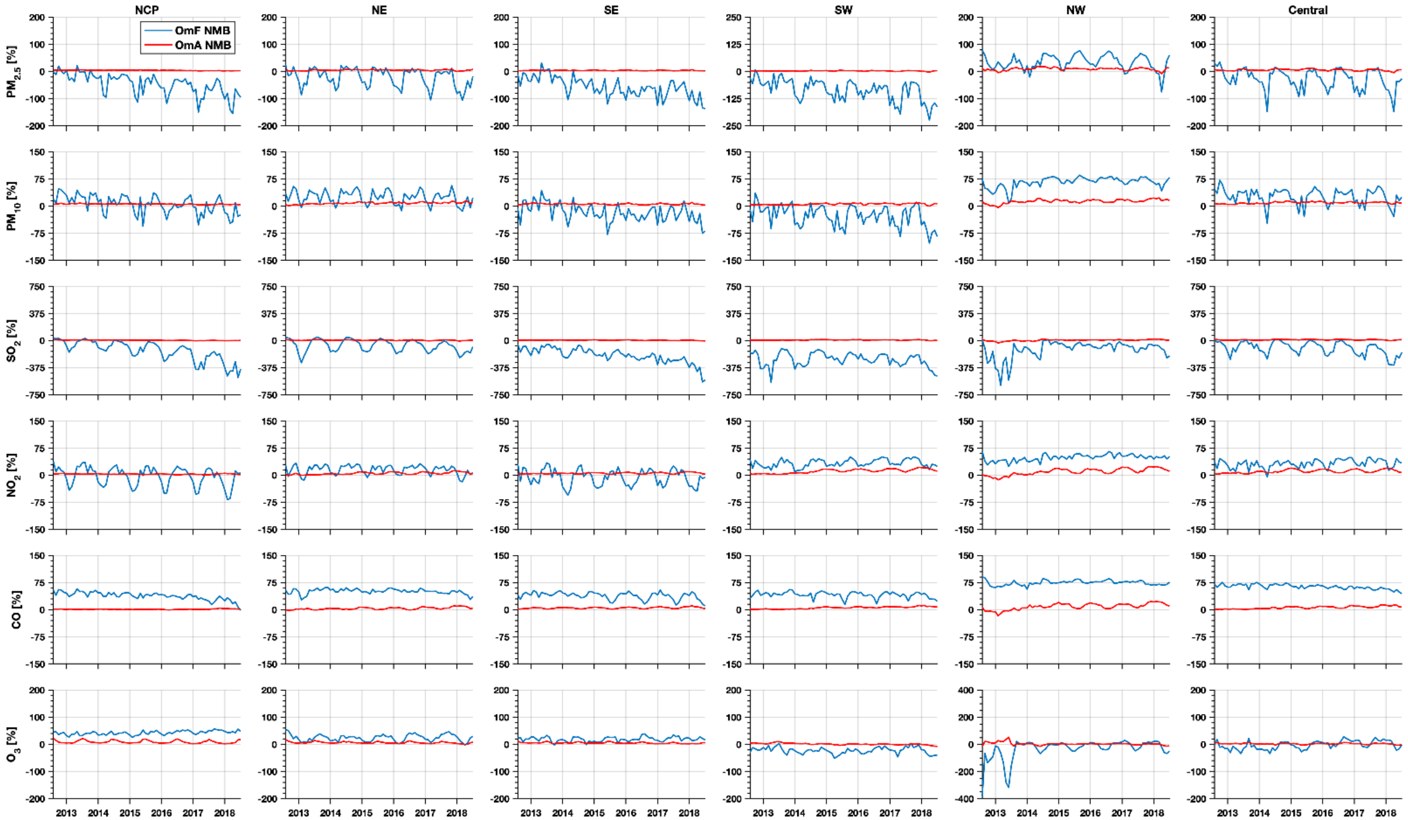


Figure R3: Time series of monthly mean OmF and OmA normalized mean bias in different regions of China for different species.

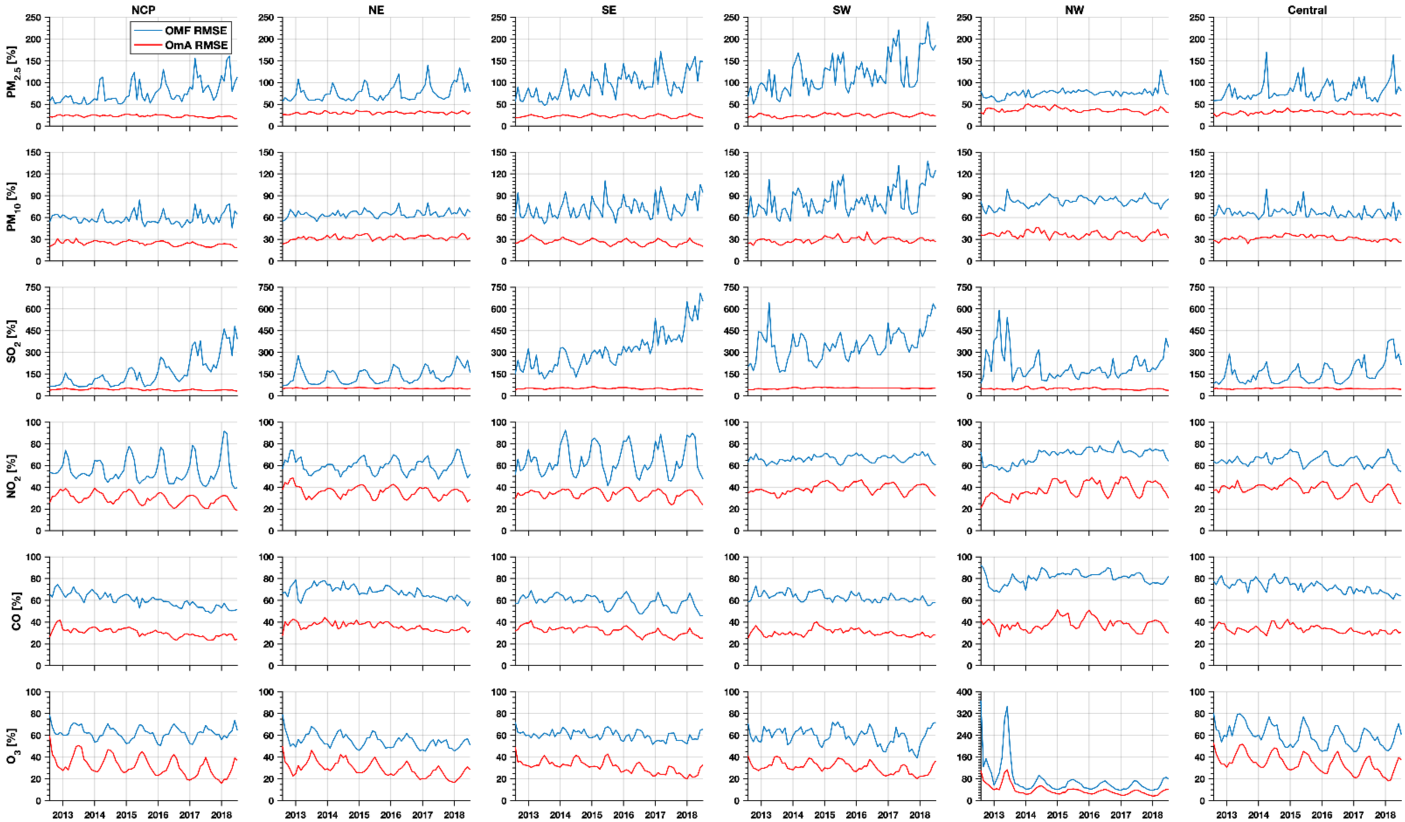


Figure R4: Time series of monthly mean OmF and OmA normalized root mean square error in different regions of China for different species.

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_3) 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 $PM_{2.5}$ and PM_{10} and negative correlation between the background errors of NO_2 and O_3. The high correlation between $PM_{2.5}$ and PM_{10} is just because $PM_{2.5}$ is a part of PM_{10}, and there would be redundant information in the observations of $PM_{2.5}$ and PM_{10} concentrations, thus we did not include the correlation between the $PM_{2.5}$ and PM_{10} concentrations in the assimilation. The negative correlation between the O_3 and NO_2 is due to the NO_x-OH-O_3 chemical reactions in the NO_x saturated conditions that increases of NO_2 concentrations would reduce the O_3 concentrations due to the enhanced NO titration effect. However, the relationship between O_3 and NO_2 concentrations is actually nonlinear depending on the NO_x 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_3 and NO_2 may come up with inefficient or even wrong adjustments. Considering the nonlinear relationship between the O_3 and NO_2 concentrations and their unexpected effects on EnKF, we took a conservative way in the assimilations of NO_2 and O_3 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*).

Changes in the manuscript: lines 274–292.

Changes in the supplementary: Figure S3 and Figure S4.

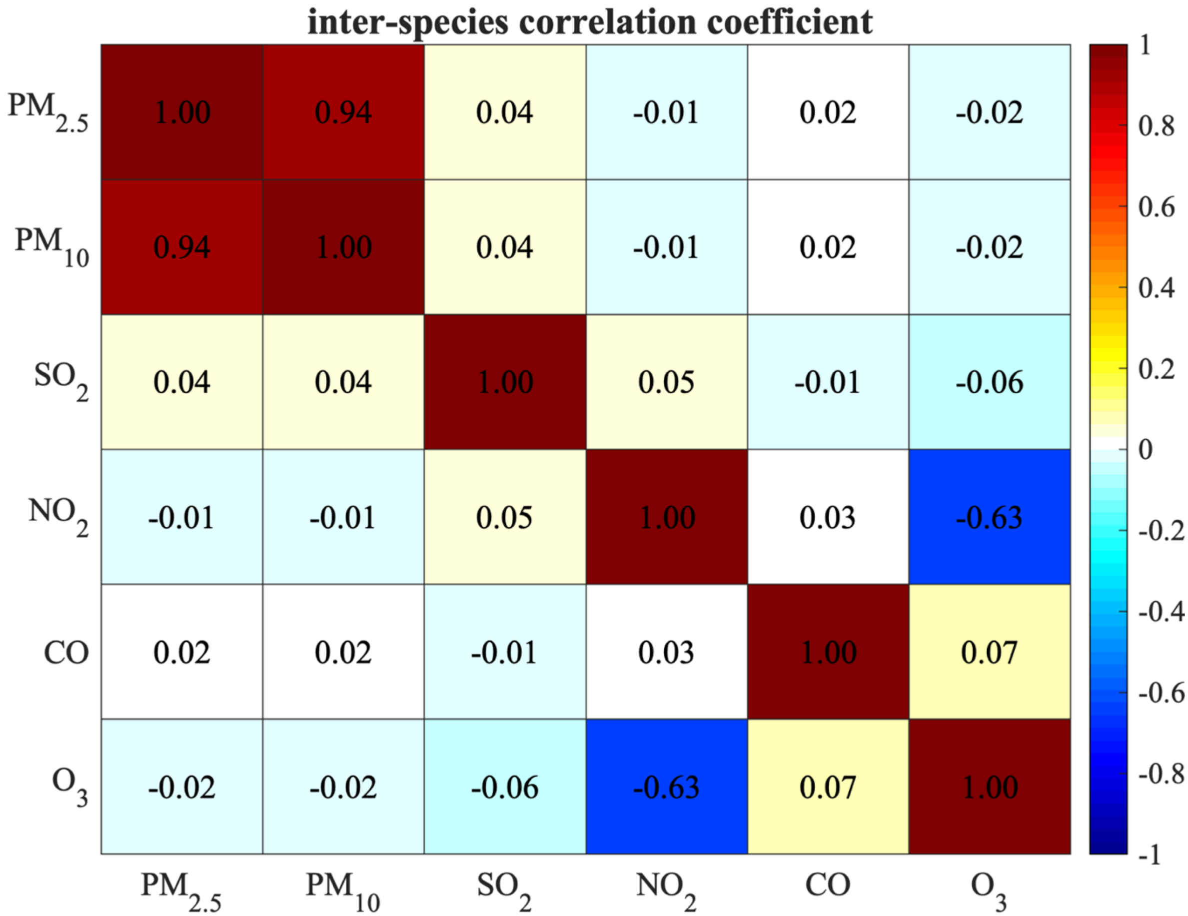


Figure R5: Correlations between species in the background error covariance matrix, estimated from the LETKF ensemble averaged from 2013 to 2018. The global mean of the covariance estimated for each station is plotted.

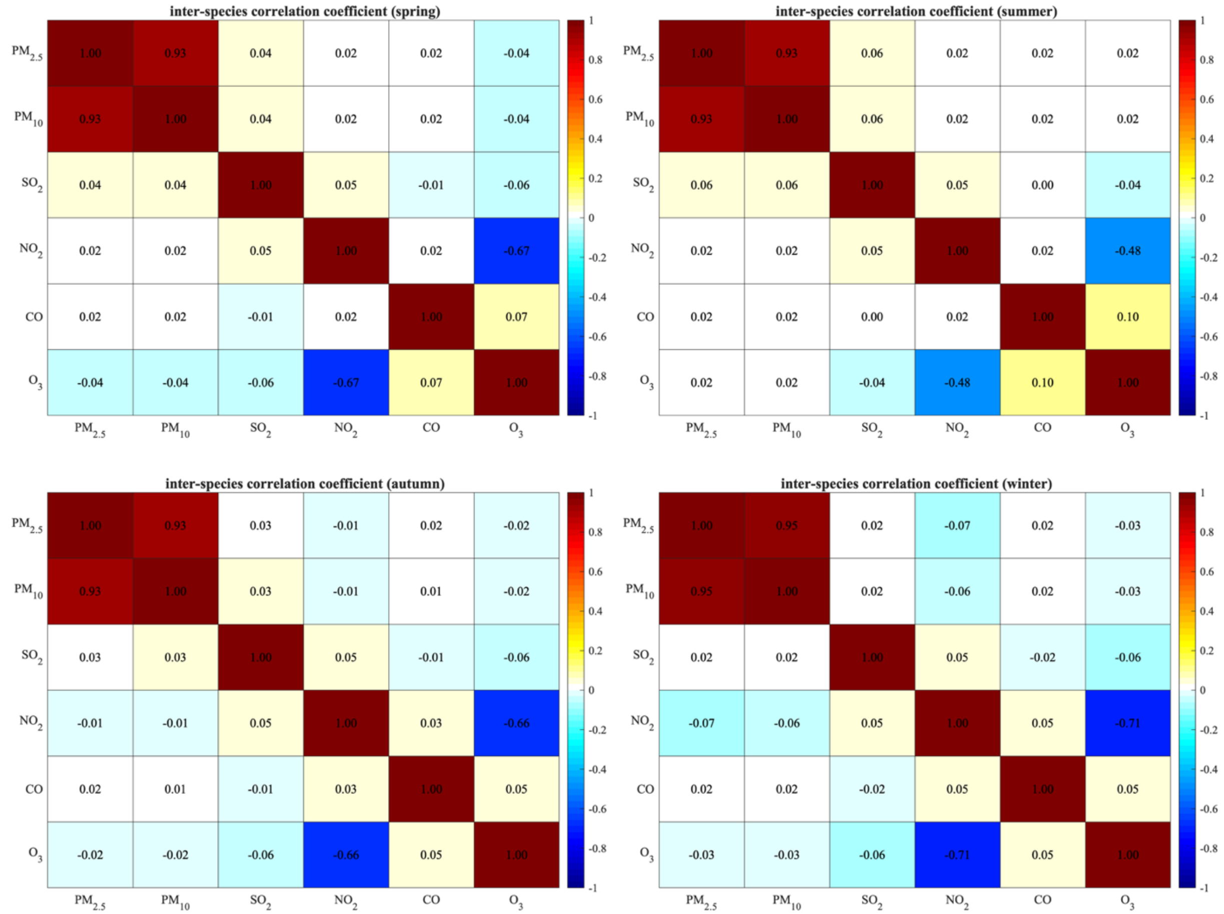


Figure R6: Correlations between species in the background error covariance matrix, estimated from the LETKF ensemble averaged in different seasons from 2013 to 2018. The global mean of the covariance estimated for each station is plotted.

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):

$$\langle dd^T \rangle \approx \mathrm{HBH^T} + \mathrm{R} \tag{6}$$

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

In order to balance the ensemble and innovation spreads, a multiplicative inflation factor for **B** can be approximate by:

$$\lambda = \frac{\left(\mathbf{R}^{-1/2}\boldsymbol{d}\right)^{\mathrm{T}}\mathbf{R}^{-1/2}\boldsymbol{d} - p}{trace\left\{\mathbf{R}^{-1/2}\mathbf{H}\mathbf{P}^{b}\left(\mathbf{R}^{-1/2}\mathbf{H}\right)^{\mathrm{T}}\right\}} \tag{8}$$

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.

Changes in the manuscript: lines 270–273.

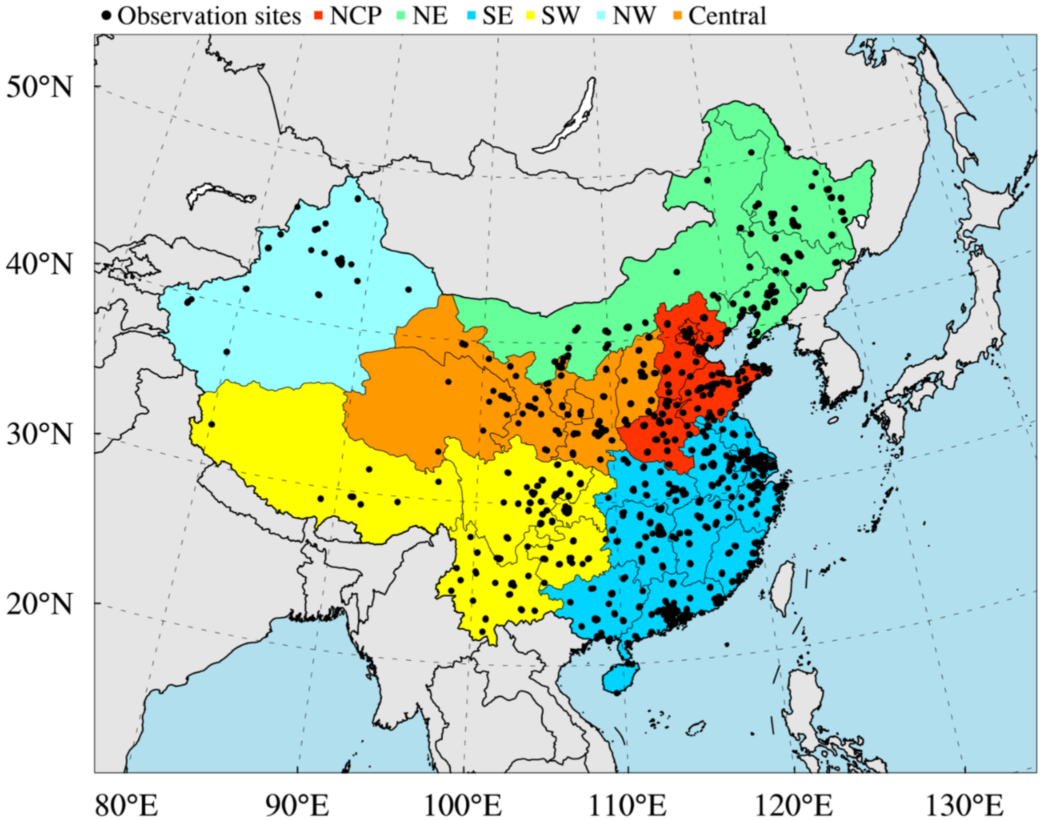


Figure R7: Modeling domain of the ensemble simulation overlay the distributions of observation sites from CNEMC. Different colours denote the different regions in China, namely North China Plain (NCP), Northeast China (NE), Southwest China (SW), Southeast China (SE), Northwest China (NW) and Central.

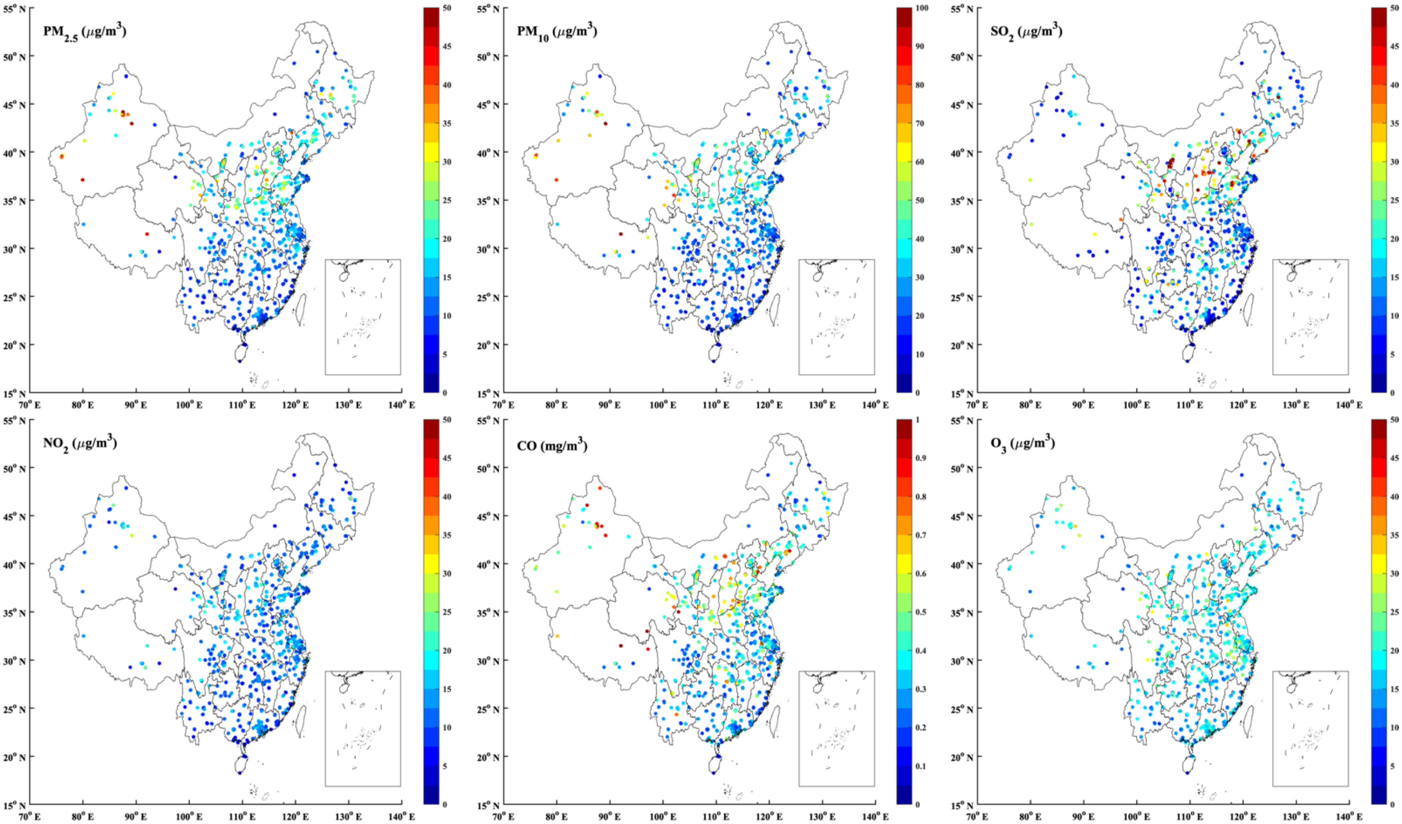


Figure R8: Spatial distributions of OmA RMSE values for (a) $PM_{2.5}$, (b) PM_{10}, (c) SO_2, (d) NO_2, (e) CO and (f) O_3 in China

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).

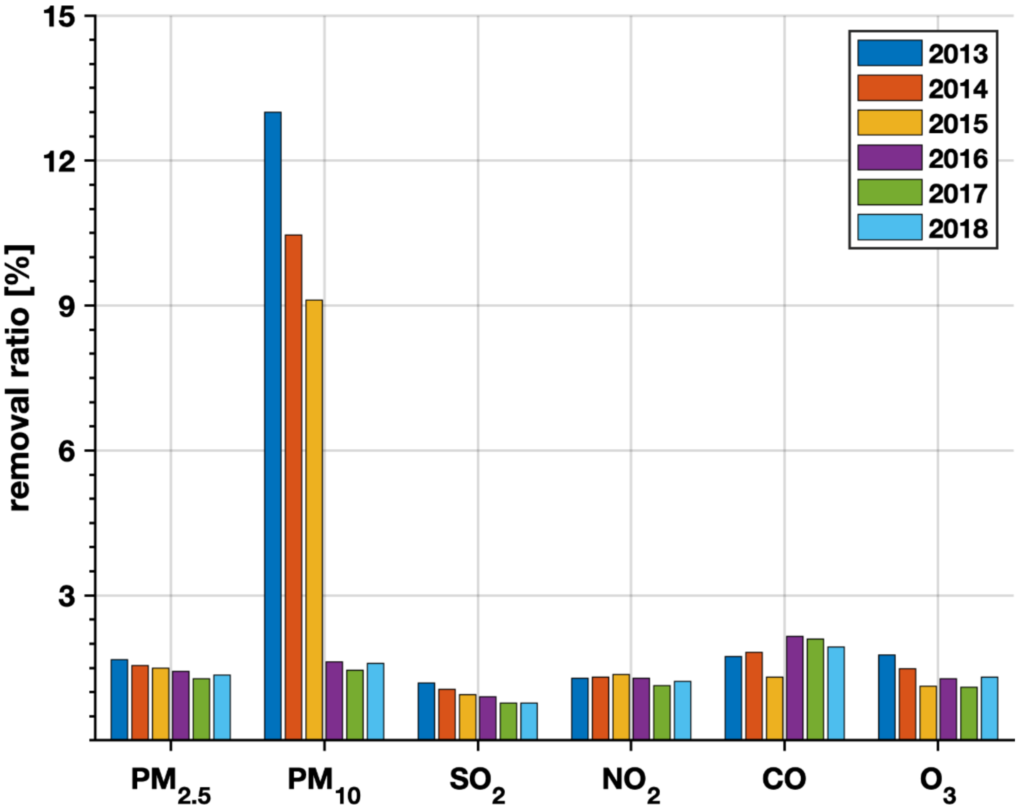


Figure R9: Removal ratio of all observation sites in China from 2013 to 2018 for different species detected by the automatic outlier detection method.

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_{10} than $PM_{2.5}$ 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_2 observations or the abnormally high values in $PM_{2.5}$ 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_3 observation instruments (Fig. R10e), which may interfere with the observations and insert abnormal values into online measurement datasets. The LP-outlier involves $PM_{2.5}$ concentrations being higher than PM_{10} 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.

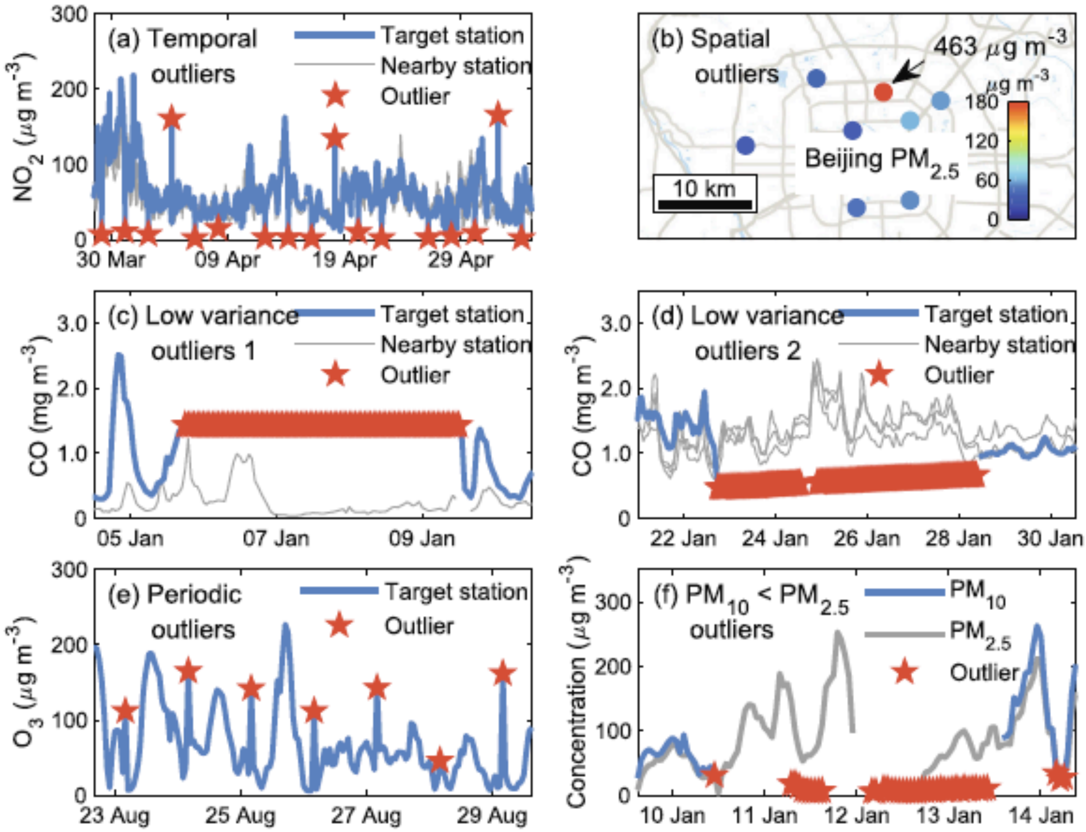


Figure R10: Examples of classified outliers in surface observations of air pollutants. (a, b) Spatiotemporal outliers have large differences with neighboring observations in time and space. (c, d) Low variance outliers either stay the same or change abnormally slowly in time and differ significantly 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)).

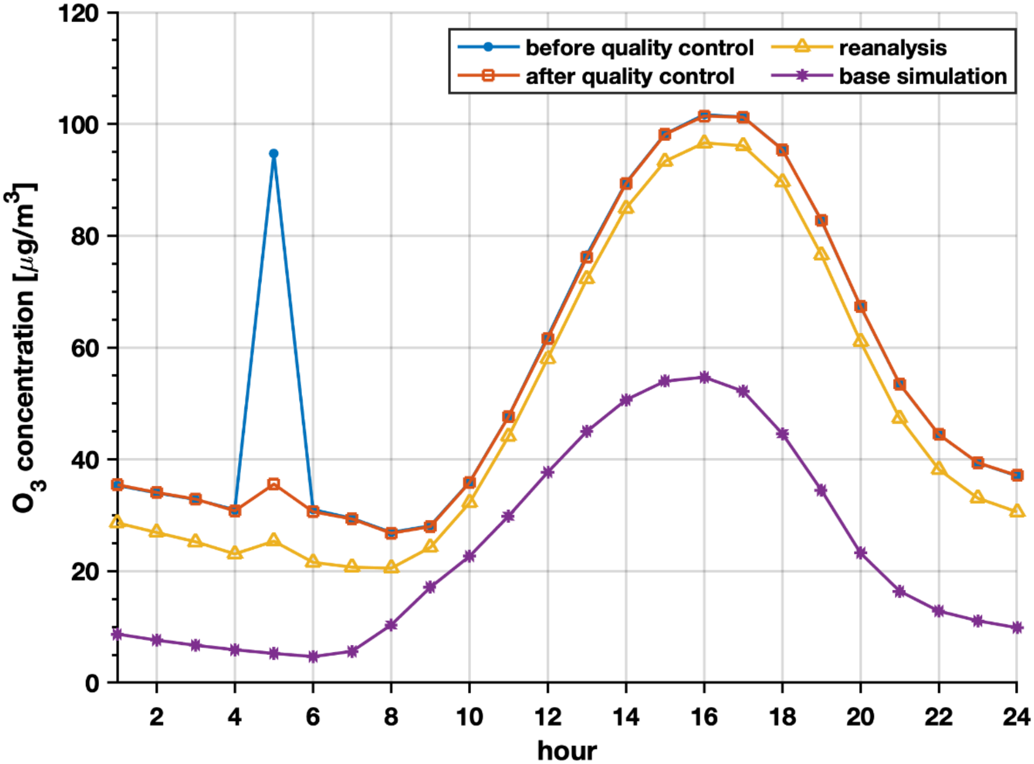


Figure R11: Six-year averaged diurnal variations of O_3 concentrations in Wuhan, China obtained from observations before and after quality control, reanalysis data and base simulation.

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 $PM_{2.5}$, SO_2, NO_2 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_{10} concentrations in south China and the O_3 concentrations throughout China. According to estimation, the impacts of outlier detection were generally small in most stations. The differences were less than 5 $\mu g/m^3$ (1 $\mu g/m^3$) for $PM_{2.5}$ concentrations over most stations in north (south) China and less than 1 $\mu g/m^3$ for the gaseous air pollutants for most stations throughout China. The differences were shown to be relative larger for PM_{10} concentrations over northwest China which can be over 20 $\mu g/m^3$ 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*).

Changes in the manuscript: lines 195–216.

Changes in the supplementary: Figure S1 and Figure S2.

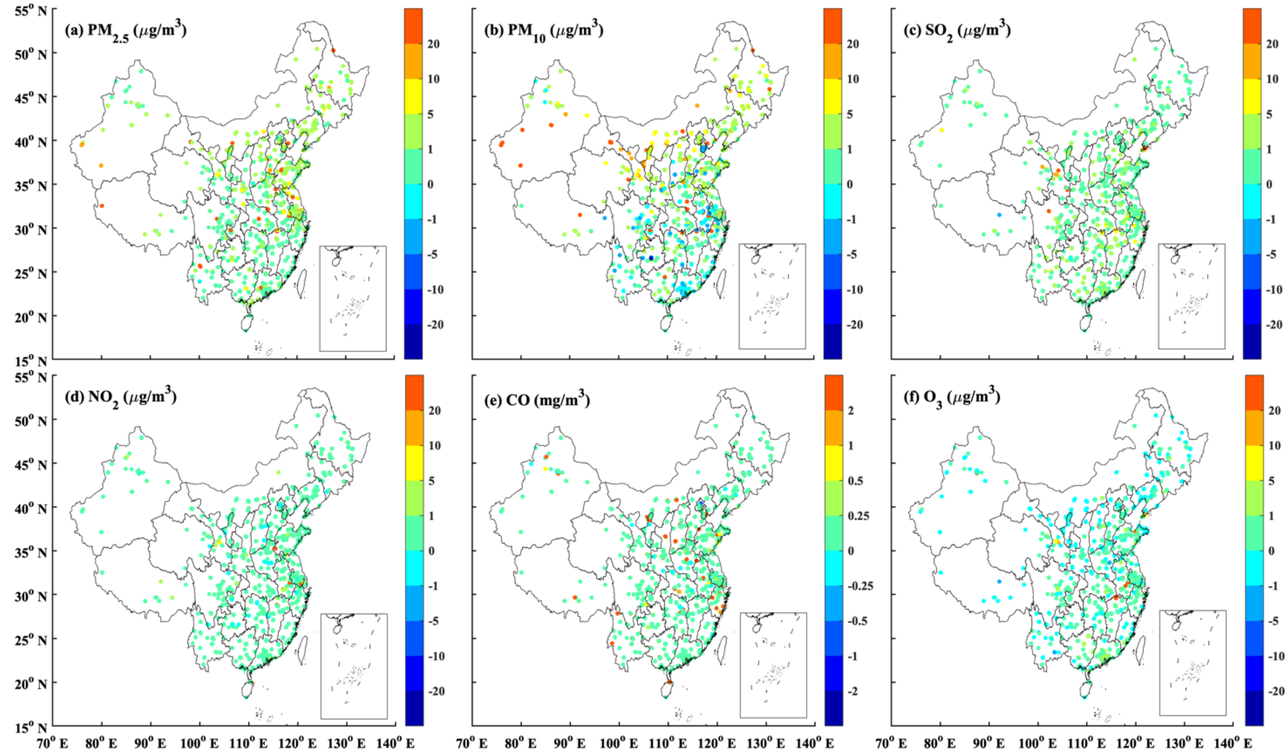


Figure R12: Spatial distributions of differences in annual concentrations of six air pollutants in China before and after quality control averaged from 2013 to 2018.

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 ~2.8°, 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 assimiation 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

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

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 adapative 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 supurious correlation caused by the limited ensemble size. This 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*).

Changes in the manuscript: lines 158–189 and lines 661–666.

References

Chen, D., Liu, Z., Ban, J., Zhao, P., and Chen, M.: Retrospective analysis of 2015–2017 wintertime $PM_{2.5}$ 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 $PM_{2.5}$ observations on $PM_{2.5}$ 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_{10} 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, 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 Changs, 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, 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_2, O_3, CO, and HNO_3 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 PM2. 5 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_x and hydrocarbons in urban and polluted rural environments, Atmos. Environ., 33, 1821-1845, 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_x emissions: mixed effects on NO_2 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, 2003.

Werner, M., Kryza, M., Pagowski, M., and Guzikowski, J.: Assimilation of $PM_{2.5}$ 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.

Response to Referee #2 (essd-2020-100)

We Thank Reviewer for his/her constructive comments.

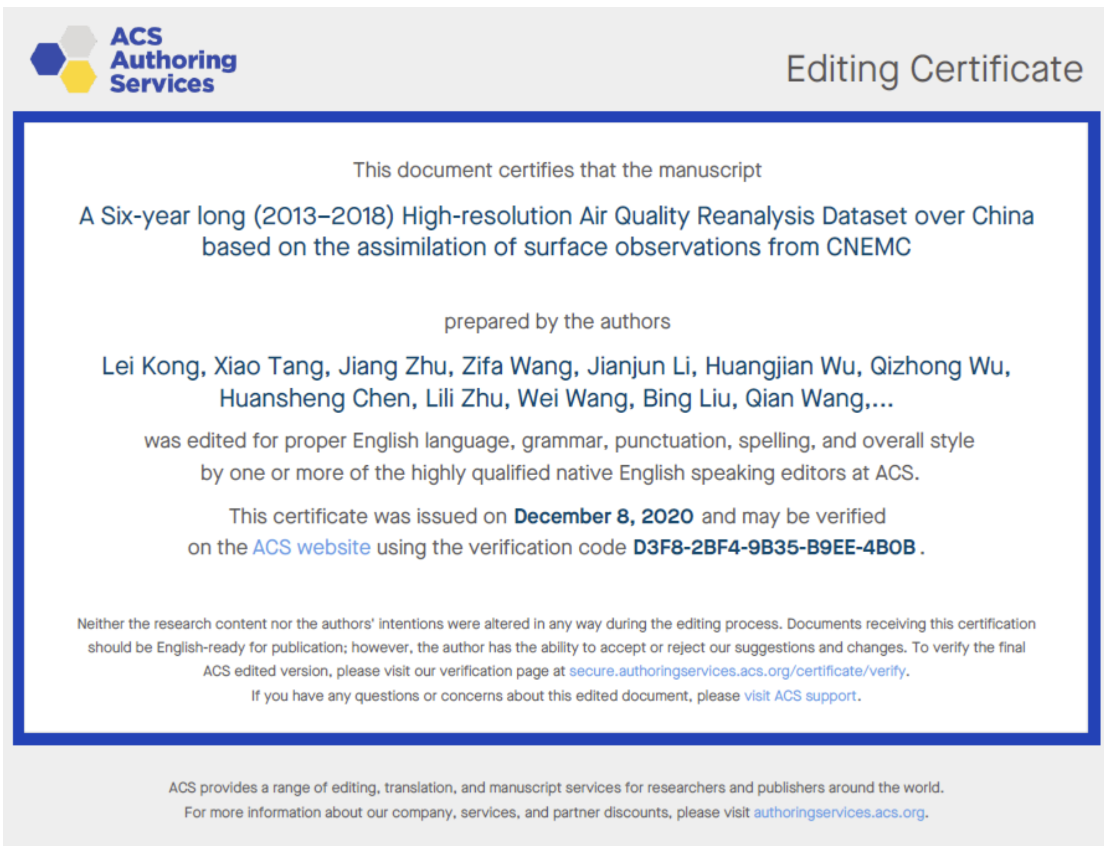
Responses to the Specific comments:

General comments: This paper presents a six-year reanalysis of air quality in China, providing data for six criteria air pollutants at 15 km spatial resolution and hourly temporal resolution. This is a unique data set and the methods for producing it appear to be sound. I presume that there will be some demand for the data set, although it would have been good to get a more specific identification of stakeholders from the paper. The main problem I have with the paper is that the writing is atrocious. The title is ungrammatical (should 'base" be "based"?), the first sentence of the abstract is vapid, and it doesn't improve from there. The paper is basically unreadable. I don't know to what extent this matters for an ESSD publication, but as a reader I would feel that if the paper is that bad then the dataset must be bad too. It's not clear to me what to do about this except to request that the authors rewrite their paper to abide by grammatical, clear, and concise language - I will leave it to the Editor to decide whether this is an appropriate request.

Reply: We sincerely apologize for the poor presentation quality. A thorough revision has been made to improve the language of the paper. The paper was edited for grammar, phrasing, and punctuation. In addition, many edits were made to further improve the flow and readability of the text. Specifically, A variety of edits were made to ensure smooth transitions between sentences and to link related thoughts. Sentence flow is improved in the revised manuscript by ensuring the appropriate use of conjunctions and introductory words and phrases. Certain edits were made to remove redundant, repetitive or unnecessary phrasing and to present the information in a more straightforward manner. Some edits were also made to improve conciseness by trimming unnecessary words and streamlining the flow of the manuscript. The manuscript was also been polished by highly qualified native English speaking editors at American Chemical Society (Fig. R1). After these revisions, we believe the language of the revised manuscript can meet the requirement of the publication in ESSD.

Following the suggestion of reviewer, the potential usages of our reanalysis data are also emphasized in the revised manuscript to get a more specific identification of stakeholders from the paper. For example, the dataset can be used in the retrospective air quality analysis in China, health and environmental impact assessment of air pollution at fine scales, model evaluation and satellite calibration, optimization of monitoring sites and provision of basic training datasets for statistical or artificial intelligence (AI)-based forecasting (*please see lines 41–45 and lines 113–116*).

Changes in the manuscript: lines 41–45, lines 113–116 and language edits throughout the paper.


ACS Authoring Services

Editing Certificate

This document certifies that the manuscript

A Six-year long (2013–2018) High-resolution Air Quality Reanalysis Dataset over China based on the assimilation of surface observations from CNEMC

prepared by the authors

Lei Kong, Xiao Tang, Jiang Zhu, Zifa Wang, Jianjun Li, Huangjian Wu, Qizhong Wu, Huansheng Chen, Lili Zhu, Wei Wang, Bing Liu, Qian Wang,...

was edited for proper English language, grammar, punctuation, spelling, and overall style by one or more of the highly qualified native English speaking editors at ACS.

This certificate was issued on **December 8, 2020** and may be verified on the ACS website using the verification code **D3F8-2BF4-9B35-B9EE-4B0B**.

Neither the research content nor the authors' intentions were altered in any way during the editing process. Documents receiving this certification should be English-ready for publication; however, the author has the ability to accept or reject our suggestions and changes. To verify the final ACS edited version, please visit our verification page at secure.authoringservices.acs.org/certificate/verify.
If you have any questions or concerns about this edited document, please visit ACS support.

ACS provides a range of editing, translation, and manuscript services for researchers and publishers around the world.
For more information about our company, services, and partner discounts, please visit authoringservices.acs.org.

Figure R1: The editing certificate of the revise manuscript provided by ACS.

A six-year-long (2013–2018) high-resolution air quality reanalysis dataset in China based on the assimilation of surface observations from CNEMC

Lei Kong[1,2], Xiao Tang[1,2], Jiang Zhu[1,2], Zifa Wang[1,2,3], Jianjun Li[4], Huangjian Wu[1,5], Qizhong Wu[6], Huansheng Chen[1], Lili Zhu[4], Wei Wang[4], Bing Liu[4], Qian Wang[7], Duohong Chen[8], Yuepeng Pan[1,2], Tao Song[1,2], Fei Li[1], Haitao Zheng[9], Guanglin Jia[10], Miaomiao Lu[11], Lin Wu[1,2], Gregory R. Carmichael[12]

[1]LAPC & ICCES, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, 100029, China
[2]College of Earth and Planetary Sciences, University of Chinese Academy of Sciences, Beijing, 100049, China
[3]Center for Excellence in Regional Atmospheric Environment, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen 361021, China
[4]China National Environmental Monitoring Centre, Beijing, 100012, China
[5]Guanghua School of Management, Peking University, Beijing 100871, China
[6]College of Global Change and Earth System Science, Beijing Normal University, Beijing 100875, China
[7]Shanghai Environmental Monitoring Center, Shanghai, 200030, China
[8]State Environmental Protection Key Laboratory of Regional Air Quality Monitoring, Guangdong Environmental Monitoring Center, Guangzhou, 510308, China
[9]Key Lab of Environmental Optics and Technology, Anhui Institute of Optics and Fine Mechanics, Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei, 230031, China
[10]School of Environment and Energy, South China University of Technology, University Town, Guangzhou 510006, China
[11]State Environmental Protection Key Laboratory of Urban Ambient Air Particulate Matter Pollution Prevention and Control, College of Environmental Science and Engineering, Nankai University, Tianjin 300350, China
[12]Center for Global and Regional Environmental Research, University of Iowa, Iowa City, IA, 52242, USA

Correspondence to: Xiao Tang (tangxiao@mail.iap.ac.cn) and Jiang Zhu (jzhu@mail.iap.ac.cn)

Abstract

A six-year-long high-resolution Chinese air quality reanalysis (CAQRA) dataset is presented in this study obtained from the assimilation of surface observations from China National Environmental Monitoring Centre (CNEMC) using the ensemble Kalman filter (EnKF) and Nested Air Quality Prediction Modeling System (NAQPMS). This dataset contains surface fields of six conventional air pollutants in China (i.e., $PM_{2.5}$, PM_{10}, SO_2, NO_2, CO and O_3) for period 2013–2018 at high spatial (15 km×15 km) and temporal (1 hour) resolutions. This paper aims to document this dataset by providing detailed descriptions of the assimilation system and the first validation results for the above reanalysis dataset. The fivefold cross-validation (CV) method is adopted to demonstrate the quality of the reanalysis. The CV results show that the CAQRA yields an excellent performance in reproducing the magnitude and variability of surface air pollutants in China from 2013 to 2018 (CV R^2 = 0.52–0.81, CV root mean square error (RMSE) = 0.54 mg/m^3 for CO and CV RMSE = 16.4–39.3 μg/m^3 for the other pollutants at the hourly scale). Through comparison to the Copernicus Atmosphere Monitoring Service reanalysis (CAMSRA) dataset produced by the European Centre for Medium-Range Weather Forecasts (ECWMF), we show that CAQRA attains a high

accuracy in representing surface gaseous air pollutants in China due to the assimilation of surface observations. The fine horizontal resolution of CAQRA also makes it more suitable for air quality studies at the regional scale. The $PM_{2.5}$ reanalysis dataset is further validated against the independent datasets from the U.S. Department State Air Quality Monitoring Program over China, which exhibits a good agreement with the independent observations (R^2 = 0.74–0.86 and RMSE =16.8–33.6 $\mu g/m^3$ in different cities). Besides, through the comparison to satellite estimated $PM_{2.5}$ concentrations, we show that the accuracy of the $PM_{2.5}$ reanalysis is higher than that of most satellite estimates. The CAQRA is the first high-resolution air quality reanalysis dataset in China that simultaneously provides the surface concentrations of six conventional air pollutants, which is of great value for many studies, such as health impact assessment of air pollution, investigation of air quality changes in China, model evaluation and satellite calibration, optimization of monitoring sites and provision of training data for statistical or artificial intelligence (AI)-based forecasting. All datasets are freely available at https://doi.org/10.11922/sciencedb.00053 (Tang et al., 2020a), and a prototype product containing the monthly and annual means of the CAQRA dataset has also been released at https://doi.org/10.11922/sciencedb.00092 (Tang et al., 2020b) to facilitate the potential evaluation of the CAQRA dataset by users.

1 Introduction

Air pollution is a critical environmental issue that adversely affects human health and is closely connected to climate change (von Schneidemesser et al., 2015). Exposure to ambient air pollution has been confirmed by many epidemiological studies to be a leading contributor to the global disease burden, which increases both morbidity and mortality (Cohen et al., 2017). China, as the largest developing country, has achieved great economic development since the 1980s. This large-scale economic expansion, however, is accompanied by a dramatic increase in air pollutant emissions, leading to severe air pollution in China (Kan et al., 2012). Since 2012, the Chinese government has established a nationwide ground-based air quality monitoring network (Fig. 1) to monitor the surface concentrations of six conventional air pollutants in China, i.e., particles with an aerodynamic diameter of 2.5 μm or smaller ($PM_{2.5}$), particles with an aerodynamic diameter of 10 μm or smaller (PM_{10}), sulfur dioxide (SO_2), nitrogen dioxide (NO_2), carbon monoxide (CO) and Ozone (O_3), which plays an irreplaceable role in understanding the air pollution in China. In addition, since the implementation of Action Plan for the Prevention and Control of Air Pollution in 2013, a series of aggressive control measures has been applied in China to reduce the emissions of air pollutants. According to the estimates of Zheng et al. (2018b), the Chinese anthropogenic emissions has decreased by 59% for SO_2, 21% for NO_x, 23% for CO, 36% for PM_{10} and 35% for $PM_{2.5}$ from 2013 to 2017. Concurrently, the air quality in China has changed dramatically over the past six years (Silver et al., 2018; Zheng et al., 2017). Such large changes in the Chinese air quality and their effects on human health and the environment have become an increasingly hot topic in many scientific fields (e.g., Xue et al., 2019; Zheng et al., 2017) which requires a long-term air quality dataset in China with high accuracy and spatiotemporal resolutions.

Ground-based observations can provide accurate information on the spatial and temporal distributions of air pollutants in China, but they are sparsely and unevenly distributed in space. Satellite observations exhibit the advantages of a high spatial coverage and have widely been applied in air pollution monitoring over large domains. A series of satellite retrievals related to air quality has been developed over the past two decades, such as the observations of NO_2, SO_2 and O_3 columns from the Ozone Monitoring Instrument (OMI; Levelt et al., 2006), CO column observations from the Measurement of Pollution in the Troposphere (MOPITT; Deeter et al., 2003) and aerosol optical depth (AOD) observations from the Moderate Resolution Imaging Spectroradiometer (MODIS; Barnes et al., 1998). The satellite column measurements have also been used to estimate surface concentrations using different methods, such as chemical transport models (CTMs) (e.g., van Donkelaar et al., 2016; van Donkelaar et al., 2010), advanced statistical methods (e.g., Ma et al., 2014; Ma et al., 2016; Xue et al., 2019; Zou et al., 2017) and semi-empirical models (e.g., Lin et al., 2015; Lin et al., 2018), which have been proven to be an effective way to acquire wide-coverage distributions of surface air pollutant with a good accuracy (Chu et al., 2016; Shin et al., 2019). However, challenges remain in satellite-based concentration estimates due to missing values related to cloud contamination, uncertainties in satellite measurements, and difficulties in modelling the complex relationship between surface concentrations and column measurements (Shin et al., 2019; van Donkelaar et al., 2016; Xue et al., 2019). In addition, most satellite-based estimates of surface concentrations exhibit low temporal resolutions (daily or even longer), which limits their application in fine-scale studies, such as the assessment of the acute health effects of the air quality. To our knowledge, a nationwide long-term estimate of the surface concentrations of all conventional air pollutants in China at the hourly scale have not yet been reported in previous satellite estimates.

A long-term air quality reanalysis dataset of critical air pollutants can provide constrained estimates of their concentrations at all locations and times, which optimally combines the accuracy of observations and the physical information and spatial continuity of CTMs through advanced data assimilation techniques. Reanalysis datasets are uniform, continuous and state-of-science best-estimate data products that have been adopted by a vast number of research communities. For example, several long-term meteorological reanalysis datasets have been developed by various weather centres in different regions/countries, such as the ERA-Interim reanalysis developed by the European Centre for Medium-Range Weather Forecasts (ECMWF; Dee et al., 2011), the National Center for Atmospheric Research (NCAR)/National Centers for Environmental Protection (NCEP) reanalysis developed by the NCEP (Saha et al., 2010), the Modern-Era Retrospective Analysis for Research and Applications (MERRA) developed by the NASA Global Modeling and Assimilation Office (NASA-GMAO; Rienecker et al., 2011), the Japanese 55-year Reanalysis (JRA-55) developed by the Japan Meteorological Agency (Kobayashi et al., 2015) and the China Meteorological Administration's global atmospheric Reanalysis (CRA-40) developed by the China Meteorological Administration (CMA). The use of data assimilation in atmospheric chemistry reanalysis is more recent, and certain reanalysis datasets for the atmospheric composition have been produced over the past decades, for example the Monitoring Atmospheric Composition and Climate (MACC), Copernicus Atmosphere Monitoring Service (CAMS) interim reanalysis (CIRA), and CAMS reanalysis (CAMSRA) produced by the ECWMF (Flemming et al., 2017; Inness et al., 2019; Inness et al., 2013), the MERRA-2 aerosol reanalysis produced by the NASA-GMAO (Randles et al., 2017), the tropospheric chemistry reanalysis

(TCR) from 2005–2012 produced by Miyazaki et al. (2015) and its latest version TCR-2 (Miyazaki et al., 2020), the global reanalysis of carbon monoxide produced by Gaubert et al. (2016), the multi-sensor total ozone reanalysis from 1970–2012 produced by van der A et al. (2015) and the Japanese Reanalysis for Aerosols (JRAero) from 2011–2015 produced by Yumimoto et al. (2017). These reanalysis datasets promote our understanding of the atmospheric composition and also facilitate the air quality research. However, these datasets are all global datasets with coarse horizonal resolutions (> 50 km), which may be insufficient to capture the high spatial variability of air pollutants at the regional scale. In addition, some of these reanalysis datasets only provide air quality data prior to 2012 and only focus on specific species. There is still no high-resolution air quality reanalysis dataset in China capturing its dramatic air quality change during recent years.

In view of these discrepancies, in this study, we develop a high-resolution regional air quality reanalysis dataset in China from 2013 to 2018 (which will be extended in the future on a yearly basis) by assimilating surface observations from China National Environmental Monitoring Centre (CNEMC). The developed reanalysis dataset may help mitigate the lack of high-resolution air quality datasets in China by providing surface concentration fields of all six conventional air pollutants in China at high spatial (15 km×15 km) and temporal (hourly) resolutions, which is of great value to (1) retrospective air quality analysis in China, (2) health and environmental impact assessment of air pollution at fine scales, (3) model evaluation and satellite calibration, (4) optimization of monitoring sites and (5) provision of basic training datasets for statistical or artificial intelligence (AI)-based forecasting.

2 Description of the chemical data assimilation system

The Chinese air quality reanalysis (CAQRA) dataset was produced with the chemical data assimilation system (ChemDAS) developed by the Institute of Atmospheric Physics, Chinese Academy of Sciences (IAP, CAS) (Tang et al., 2011). This system consists of (i) a three-dimensional CTM called the Nested Air Quality Prediction Modeling System (NAQPMS) developed by Wang et al. (2000), (ii) an ensemble Kalman filter (EnKF) assimilation algorithm, and (iii) surface observations from CNEMC with the automatic outlier detection method. We adopted an offline analysis scheme in this study since there are no previous experiences with online chemical data assimilation at such a high horizontal resolution. The lessons learned from this offline analysis application could also facilitate future implementation of online analysis. In the offline analysis scheme, a free ensemble simulation was first conducted, and the observations were then assimilated using the EnKF. A similar offline analysis scheme has also been applied in previous reanalysis studies, such as Candiani et al. (2013) and Kumar et al. (2012). Detailed descriptions of the ensemble simulation, observations and data assimilation algorithm used in this study are presented below.

2.1 Air pollution prediction model

The NAQPMS model was used as the forecast model to represent the atmospheric chemistry, which has been applied in previous assimilation studies (Tang et al., 2011; Tang et al., 2013). The model is driven by the hourly meteorological fields produced by the Weather Research and Forecasting (WRF) model (Skamarock, 2008). Gas phase chemistry is simulated with the carbon bond mechanism Z (CBM-Z) developed by Zaveri and Peters (1999). Aqueous-phase chemistry and wet deposition are simulated based on the Regional Acid Deposition Model (RADM) mechanism in the Community Multi-scale Air Quality (CMAQ) model version 4.6. In regard to aerosol processes, the thermodynamic model ISORROPIA 1.7 (Nenes et al., 1998) is applied for the simulations of inorganic atmospheric aerosols. Six secondary organic aerosols (SOAs) are explicitly treated in the NAQPMS model based on Li et al. (2011). To simulate the interactions between particles and gases, 28 heterogeneous reactions involving sulfate, soot, dust and sea salt particles are included based on previous studies (Li et al., 2015; Li et al., 2012). Size-resolved mineral dust emissions are calculated online as a function of the relative humidity, frictional velocity, mineral particle size distribution and surface roughness (Li et al., 2012). Sea salt emissions are calculated with the scheme of Athanasopoulou et al. (2008). The dry deposition of gases and aerosols is modelled based on the scheme of Wesely (1989), and advection is simulated with the accurate mass conservation algorithm of Walcek and Aleksic (1998).

Figure 1 shows the modelling domain of this study, which covers most parts of East Asia with a fine horizontal resolution of 15 km. The vertical coordinate system consists of 20 terrain-following levels with the model top reaching up to 20000 m and the first layer at approximately 50 m. Nine vertical layers are set within 2 km of the surface to better characterize the vertical mixing process within the boundary layers. The emissions of air pollutants considered in this study include the monthly anthropogenic emissions retrieved from the Hemispheric Transport of Air Pollution (HTAP) v2.2 emission inventory with a base year of 2010 (Janssens-Maenhout et al., 2015), biomass burning emissions retrieved from the Global Fire Emissions Database (GFED) version 4 (Randerson et al., 2017; van der Werf et al., 2010), biogenic volatile organic compound (BVOC) emissions retrieved from the Model of Emissions of Gases and Aerosols from Nature (MEGAN)-MACC (Sindelarova et al., 2014), marine VOC emissions retrieved from the POET database (Granier et al., 2005), soil NO_x emissions retrieved from the Regional Emission Inventory in Asia (Yan et al., 2003) and lightning NOx emissions retrieved from Price et al., 1997. Clean initial conditions are used in the air quality simulations with a two-week free run of the NAQPMS model as the spin-up time. The top and boundary conditions are provided by the Model for Ozone and Related Chemical Tracers (MOZART; Brasseur et al., 1998; Hauglustaine et al., 1998) model, and the meteorological fields are provided by the WRF model. In each daily meteorology simulation, a 36-h free run of the WRF model is conducted with the first 12-h simulation period as the spin-up run and the remaining 24-h period providing the meteorologic inputs for the NAQPMS model. The initial and boundary conditions for the meteorology simulations are provided by the NCAR/NCEP 1° × 1° reanalysis data.

2.2 Generation of ensemble simulation

The EnKF uses an ensemble of model simulation to represent the forecast uncertainty which should include the most model uncertain aspects. Considering that the emissions are a major source of uncertainty in air quality prediction (Carmichael et al., 2008; Hanna et al., 1998; Li et al., 2017), in this study the ensemble were generated by perturbing the emissions based on their error probability distribution functions (PDFs) which were assumed to be Gaussian distributions. Table 1 lists the perturbed species considered in this study as well as their corresponding emission uncertainties obtained from previous studies.

The perturbed emissions were parameterized by multiplying the base emissions with a perturbation factor $\boldsymbol{\beta}$, as expressed in Eq. (1):

$$\boldsymbol{E_i} = \boldsymbol{E} \circ \boldsymbol{\beta_i}, i = 1, 2, \cdots, N \tag{1}$$

where $\boldsymbol{E}$ denotes the vector of base emissions, $\circ$ denotes the Schur product and N denotes the ensemble size. The performance of the EnKF is strongly related to the ensemble size which determines the accuracy to which the background error covariance is approximated (Constantinescu et al., 2007; Miyazaki et al., 2012). A large ensemble size is important in capturing the proper background error covariance structure, especially in high-resolution data assimilation application due to the fine-scale variability and large degree of freedoms. However, a large ensemble is computationally expensive as the cost of EnKF linearly increases with ensemble size while the accuracy of covariance estimate improves by its square root (Constantinescu et al., 2007). Thus, an appropriate ensemble should keep a good balance between accuracy and computational cost. Constantinescu et al. (2007) in their ideal experiments showed that a 50-member ensemble has significant improvement against smaller ensembles, and Miyazaki et al. (2012) in their real chemical assimilation experiments showed that the improvement was much less significantly by further increasing the ensemble size from 48 to 64. Thus, the ensemble size was chosen as 50 in this study by referencing pervious publications and also our previous high-resolution regional assimilation work (Tang et al., 2011; Tang et al., 2013; Tang et al., 2016) which showed that a 50-member ensemble keeps good balance between assimilation performance and computational efficiency. However, it should be noted that our application has higher horizontal resolution than that of Constantinescu et al. (2007) and Miyazaki et al. (2012), which may require larger ensemble size due to the larger degree of freedoms in our application. Thus, to reduce the degree of freedoms in our high-resolution data assimilation work, we assumed that the emission errors were spatially correlated, and an isotropic correlation model was assumed in the covariance of the emission errors, which is written as:

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

where $\rho(i,j)$ represents the correlation between grids i and j, $h(i,j)$ is the distance between these two points and l is the decorrelation length, which was specified as 150 km in this study. According to the PDF of the emission errors, $\boldsymbol{\beta}$ follows the same Gaussian distribution as that of the emission errors except that its mean equals 1. Using the method of Evensen (1994), fifty smooth pseudorandom perturbation fields of $\boldsymbol{\beta}$ were generated for each perturbed species. In addition, the emission perturbations were kept independent from each other to prevent pseudo-correlation among the different species.

2.3 Observations

Surface observations of the hourly ambient $PM_{2.5}$, PM_{10}, SO_2, NO_2, CO and O_3 concentrations retrieved from the CNEMC were used in this study. The number of observation sites was approximately 510 in 2013 and increased to 1436 in 2015. Real-time observations of these six air pollutants at each monitoring site are routinely gathered by the CNEMC and released to the public (available at http://www.cnemc.cn/; last accessed: 17 April 2020) at hourly intervals. A challenge that should be overcome in the assimilations of surface observation is that there are occasional outliers occurring in these observations due to the instrument malfunctions, influences of harsh environments and limitation of measurement method. Filtering out these outliers is necessary before the assimilation, otherwise these outliers may cause unrealistic spatial and temporal variations in the reanalysis. To address this issue, a fully automatic outlier detection method was developed by Wu et al. (2018) to filter out the observation outliers. An automatic outlier detection method is very important in chemical data assimilation since there is a large amount of observation data on multiple species. Four types of outliers characterized by temporal and spatial inconsistencies, instrument-induced low variances, periodic calibration exceptions and lower PM_{10} concentrations than those of $PM_{2.5}$ were detected and removed before the assimilation. Figure S1 shows the removal ratios of the six air pollutants from 2013 to 2018, which are generally around 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 these outliers marked by lower PM_{10} concentrations than those of $PM_{2.5}$. 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). To assess the potential impacts of outlier detection on the assimilations, the differences in annual concentrations caused by quality control are shown in Fig. S2. The differences were generally positive for $PM_{2.5}$, SO_2, NO_2 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_{10} concentrations in south China and the O_3 concentrations throughout China. According to estimation, the impacts of outlier detection were generally small in most stations. The differences were less than 5 $\mu g/m^3$ (1 $\mu g/m^3$) for $PM_{2.5}$ concentrations over most stations in north (south) China and less than 1 $\mu g/m^3$ for the gaseous air pollutants for most stations throughout China. The differences were shown to be relative larger for PM_{10} concentrations over northwest China which can be over 20 $\mu g/m^3$ in stations around Taklimakan Desert. This would be due to the higher outlier ratios in the observations over the remote areas. More details on the outlier detection method were available in Wu et al. (2018).

A proper estimate of the observation error is important in regard to the filter performance since the observation and background errors determine the relative weights of the observation and background values in the analysis. The observation error includes measurement and representativeness errors. For each species, the measurement error was given by their respective instruments, namely, 5% for $PM_{2.5}$ and PM_{10}, 2% for SO_2, NO_2 and CO, and 4% for O_3 according to officially released documents of the Chinese Ministry of Ecology and Environmental Protection (HJ 193–2013 and HJ 654–2013, available at http://www.cnemc.cn/jcgf/dqhj/; last accessed: 17 April 2020). 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:

$$r_{repr} = \sqrt{\frac{\Delta x}{L_{repr}}} \times \epsilon^{abs} \tag{3}$$

where r_{repr} represents the representativeness error, Δx represents the model resolution, L_{repr} represents the characteristic representativeness length of the observation site and ε^{abs} represents the error characteristic parameters for different species.

The estimation of L_{repr} is dependent on the types of observation sites with urban sites usually having smaller representative length than the rural sites have due to the larger representativeness errors. Considering that the observation sites from CNEMC were almost city (urban) sites (>90%), the L_{repr} was assigned to be 2km in this study according to Elbern et al., 2007.

For the estimations of ε^{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 ε^{abs} empirically to be half of the measurement error following the study by Pagowski et al. (2010). In this study, the ε^{abs} was obtained from Li et al. (2019) who estimated the ε^{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:

$$r_{repr,i} = \frac{1}{MT}\sum_{m=1}^{M}\sum_{t=1}^{T} S_{m,t,i} \tag{4}$$

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 the estimations of $r_{repr,i}$, the ε_i^{abs} for species i were estimated by a transformation of Eq. (3):

$$\varepsilon_i^{abs} = r_{repr,i} / \sqrt{\frac{\Delta \mathrm{x}}{L_{repr}}} \tag{5}$$

where Δx is equal to 30km. Based on the estimated L_{repr} and the ε_i^{abs} for different species, the representativeness errors are estimated using Eq. (3) by specifying the Δx to be 15km.

2.4 Data assimilation algorithm

We used a variant of the EnKF approach, i.e., the local ensemble transform Kalman filter (LETKF; Hunt et al., 2007), to assimilate the observations into the model state. The LETKF has several advantageous over the original EnKF (e.g., Miyazaki et al., 2012). As a kind of deterministic filter, it does not need to perturb the observations, which avoids introducing additional sampling errors. In addition, the LETKF performs the analysis locally in space and time, which not only alleviate the rank problem of the EnKF method but also suppress the long-distance spurious correlation caused by the limited ensemble size. The formulation of the LETKF can be written as:

$$\overline{\boldsymbol{x}^{a}} = \overline{\boldsymbol{x}^{b}} + \mathbf{X}^{\mathrm{b}}\overline{\boldsymbol{w}}^{a} \tag{6}$$

$$\overline{\boldsymbol{w}}^{a} = \widetilde{\mathbf{P}}^{\mathrm{a}}\left(\mathbf{H}\mathbf{X}^{\mathrm{b}}\right)^{\mathbf{T}}\mathbf{R}^{-\mathbf{1}}\left(\boldsymbol{y}^{o} - \mathbf{H}\overline{\boldsymbol{x}^{b}}\right) \tag{7}$$

$$\widetilde{\mathbf{P}}^{\mathrm{a}} = \left[\frac{(N_{ens}-1)\mathbf{I}}{1+\lambda} + \left(\mathbf{H}\mathbf{X}^{\mathrm{b}}\right)^{\mathbf{T}}\mathbf{R}^{-\mathbf{1}}\left(\mathbf{H}\mathbf{X}^{\mathrm{b}}\right)\right]^{-1} \tag{8}$$

$$\overline{\boldsymbol{x}^{b}} = \frac{1}{N_{ens}}\sum_{i=1}^{N_{ens}}\boldsymbol{x}_i^{b}\,;\, \mathbf{X}_{\mathbf{i}}^{\mathbf{b}} = \frac{\mathbf{1}}{\sqrt{\boldsymbol{N}-\mathbf{1}}}\left(\boldsymbol{x}_i^{b} - \overline{\boldsymbol{x}^{b}}\right) \tag{9}$$

where $\overline{\boldsymbol{x}^{\boldsymbol{a}}}$ is the analysis state, $\overline{\boldsymbol{x}^{\boldsymbol{b}}}$ is the background state, $\mathbf{X}^{\mathbf{b}}$ represents the background perturbations, $\overline{\boldsymbol{w}}^{\boldsymbol{a}}$ is the analysis in the ensemble space spanned by $\mathbf{X}^{\mathbf{b}}$, $\widetilde{\mathbf{P}}^{\mathbf{a}}$ is the analysis error covariance in the ensemble space with dimensions of $N_{ens} \times N_{ens}$, $\mathbf{y}^{\mathbf{o}}$ is the vector of observations used in the analysis of this grid, $\mathbf{R}$ is the observation error covariance matrix, and $\mathbf{H}$ is the linear observational operator that maps the model space to the observation space. The scalar λ in Eq. (8) denotes the inflation factor for the background covariance matrix, which was estimated with the algorithm proposed by Wang and Bishop (2003):

$$\lambda = \frac{\left(\mathbf{R}^{-1/2}\boldsymbol{d}\right)^{\mathrm{T}}\mathbf{R}^{-1/2}\boldsymbol{d} - p}{trace\left\{\mathbf{R}^{-1/2}\mathbf{H}\mathbf{P}^{\boldsymbol{b}}\left(\mathbf{R}^{-1/2}\mathbf{H}\right)^{\mathrm{T}}\right\}} \tag{10}$$

$$\boldsymbol{d} = \boldsymbol{y}^{\boldsymbol{o}} - \mathbf{H}\overline{\boldsymbol{x}^{\boldsymbol{b}}} \tag{11}$$

$$\mathbf{P}^{\boldsymbol{b}} = \mathbf{X}^{\mathbf{b}}\left(\mathbf{X}^{\mathbf{b}}\right)^{\mathbf{T}} \tag{12}$$

where $\boldsymbol{d}$ represents the residuals, p is the number of observations, $\mathbf{P}^{\boldsymbol{b}}$ is the ensemble-estimated background error covariance matrix, and the trace of the covariance matrix is used to approximate covariance on a globally averaged basis. The inflation is necessary for the ensemble-based assimilation algorithm since the ensemble-estimated background error covariance is very likely to underestimate the true background error covariance due to the limited ensemble size and occurrence of the model error (Liang et al., 2012). Without any treatment to prevent background error covariance underestimation, the model forecast would be overconfident and eventually result in filter divergence. Using Eq. (10), 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 the different species in different time and places.

Besides, the inter-species correlation was neglected in the background error covariance similar to previous chemical data assimilation studies (e.g., Inness et al., 2015; Inness et al., 2019; Ma et al., 2019) although Miyazaki et al, (2012) has shown the benefits of including correlations between the background errors of different chemical species. This is, on the one hand, to avoid the effects of the spurious correlation between non or weakly related variables. On the other hand, different from Miyazaki et al., 2012, this study concentrated on the assimilations of primary air pollutants (except of O_3) 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), thus the correlation between background errors of different species were generally near zero for most cases as shown in Figs. S3-4. The high correlations only occur in background errors of $PM_{2.5}$ and PM_{10} as well as the NO_2 and O_3. The high positive correlation between $PM_{2.5}$ and PM_{10} is just because $PM_{2.5}$ is a part of PM_{10}, and there would be redundant information in the observations of $PM_{2.5}$ and PM_{10} concentrations, thus we did not include the correlation between the $PM_{2.5}$ and PM_{10} concentrations in the assimilation. The negative correlation between the O_3 and NO_2 is due to the NO_x-OH-O_3 chemical reactions in the NO_x saturated conditions that the increases of NO_2 concentrations would reduce the O_3 concentrations due to the enhanced NO titration effect. However, the relationship between O_3 and NO_2 concentrations is actually nonlinear depending on the NO_x 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_3 and NO_2 may come up with inefficient or even wrong adjustments. Considering the nonlinear relationship between the O_3 and NO_2 concentrations and their unexpected effects on EnKF, we took a conservative way in the assimilations of NO_2 and O_3 by neglecting their error correlations. This would also make different species be assimilated in a consistent way. Therefore, in this study each air pollutant is assimilated independently by only using the observations of this pollutant.

Figure 2 shows the local scheme we used in the assimilation, where the plus and dot symbols indicate the centres of the model grids and locations of the observation sites, respectively. In each model grid, only the observation sites located within a $(2l + 1)$ by $(2l + 1)$ rectangular area centred at this model grid were considered in the calculations of its analysis. The cut-off radius l was chosen as 12 model grids, approximately 180 km at the 15-km horizontal resolution. The use of a cut-off radius, however, could cause analysis discontinuities when an observation enters or leaves the local domain when moving from one model grid to another (Sakov and Bertino, 2011). To increase the smoothness of the analysis state, following Hunt et al. (2007), we artificially reduced the impact of the observations close to the boundary of the local domain by multiplying the entries in $\mathbf{R^{-1}}$ by a factor decaying from one to zero with increasing distance of the observation from the central model grid. The decay factors used in this study are calculated by:

$$\rho(i) = exp\left\{-\frac{h(i)^2}{2L^2}\right\} \tag{13}$$

where $\rho(i)$ is the decay factor for observation i, $h(i)$ is the distance between observation i and the central model grid point, and L is the decorrelation length chosen as 80 km, smaller than the cut-off radius, to increase the smoothness of the analysis state. Typically, only the state of the central model grid is updated and used to construct the global analysis field. However, experience has shown that an observable discontinuity remains in the analysis over certain regions. To address this issue, following the method of Ott et al. (2004), we simultaneously updated the state of a small patch (l =1) around the central model grid (the updated region in Fig. 2) at each local analysis step. The final analysis of a given model grid was then obtained as the weighted mean of all the analysis values of this model grid. A weighted mean was necessary since the analysis of the different patches adopted different decay factors for the observation error. The weight of each analysis value in model grid i is calculated by Eq. (14):

$$W_{i,j} = \frac{\exp\left(-\frac{h(i,j)^2}{L^2}\right)}{\sum_{j=1}^{m} \exp\left(-\frac{h(i,j)^2}{L^2}\right)} \tag{14}$$

where $h(i,j)$ is the distance of model grid i to the central model grid of the patch generating the jth analysis value of this grid, m is the number of patches containing this model grid and L is the decorrelation length, which was chosen as 80 km in this study.

3 Data assimilation statistics

3.1 χ^2 diagnosis

We first applied the $\boldsymbol{\chi}^2$ test to demonstrate the performance of our data assimilation system, which is important in evaluating the reanalysis (Miyazaki et al., 2015). The χ^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 observation minus forecast (OmF) with the sum of estimated background and observation error covariance in the observational space ($\mathbf{HBH^T + R}$):

$$\boldsymbol{Y} = \frac{1}{\sqrt{m}}(\mathbf{HBH^T + R})^{-\frac{1}{2}}(\boldsymbol{y^o} - \boldsymbol{HX^b}) \tag{15}$$

$$\chi^2 = \boldsymbol{Y^T Y} \tag{16}$$

where m is the number of observations. According to the Kalman filtering theory, the mean of χ^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 3 shows the time series of the monthly χ^2 values (black lines) for different species as well as the number of assimilated observations per month (blue bars). The mean values of χ^2 are generally within 50% difference from the ideal value of 1 for $PM_{2.5}$, PM_{10}, NO_2 and O_3, which suggests that the observation and background error covariance are generally well specified in the analysis of these species. Although the χ^2 values for these species showed pronounced seasonal variations that reflect the different error characteristics in different seasons, the χ^2 values were roughly stable for $PM_{2.5}$ and O_3 throughout the assimilation periods, and for NO_2 and PM_{10} after 2015 when the number of assimilated observations become stable, which generally shows the long-term stability of the performance of data assimilation. The χ^2 values for SO_2 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_2 as we only specified 12% uncertainty in the SO_2 emissions, suggesting that the emission uncertainty of SO_2 may be underestimated by Zhang et al. (2009). There were also pronounced annual trends in the χ^2 values of SO_2, which may be attributed to the increases of observation number from 2013 to 2014 and the substantial decreases of SO_2 observations from 2013 to 2018. Although smaller than the χ^2 values of SO_2, the values for CO were greater than 1 in most cases, suggesting the underestimations of the error covariances. Similar to the χ^2 values of SO_2, obvious decreasing trend can also be found in the χ^2 values of CO. These results suggest that our data assimilation system has relatively poor performance in the analysis of CO and SO_2 concentrations than the other four species, which is consistent with the cross-validation results (Sect. 4.2.2) that showed smaller R^2 values for the reanalysis data of CO and SO_2 concentrations. The annual trend of χ^2 values in CO and SO_2 also indicates relatively weak stability in the performance of data assimilation system on assimilating CO and SO_2 observations, which may influence the analysis of the annual trends in these two species.

3.2 OmF & OmA analysis

Spatial distributions of six-year averaged OmF and observation minus analysis (OmA) for each species in the observation space were then analysed to investigate the structure of forecast bias and to measure the improvement in the reanalysis (Fig. 4). The analysis increment, which is estimated from the differences between the analysis and forecast, is also plotted to measure the adjustments made in the model space. The OmF values have showed persistent 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 in 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. S5, the monthly mean OmF biases were almost completely removed in each regions of China because of the 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 root mean square error (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. S6). 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_2 concentrations throughout the China, and the positive (negative) increments for O_3 concentrations in central-east (south) China. These results confirm that the data assimilation can effectively propagate the observation information into the model state and reduce the model errors.

4 Evaluation Results

In this section, we present the fields of the CAQRA dataset and compare them to the observations. It aims to provide a brief introduction to the CAQRA dataset and gives a first assessment of the quality of this dataset. The cross-validation (CV) method was applied in the assessment of the CAQRA dataset, in which a proportion of the observation data was withheld from the data assimilation process and adopted as a validation dataset. We conducted five CV experiments by randomly dividing the observation sites of the CNEMC into five groups (with 20% of the observation sites in each group). In each experiment, the analysis was performed with one group of the observation data omitted in the assimilation process. Analysis results at the validation sites, i.e., the observation sites not used in the assimilation process, were then collected and used to validate the assimilation results. For convenience, the analysis results at the validation sites of the five CV experiments were combined and comprised a validation dataset containing all observation sites (the CV run). This dataset was then evaluated against the observations to assess the quality of the CAQRA dataset. In addition, independent $PM_{2.5}$ observations retrieved from the U.S. Department State Air Quality Monitoring Program over China were also employed in the assessment of the $PM_{2.5}$ reanalysis field. The quality of the CAQRA dataset was assessed at different spatial and temporal scales to better understand the CAQRA dataset. Additionally, the validation results of the ensemble mean of the simulations without assimilation (the base simulation) are provided to highlight the impacts of assimilation.

4.1 Particulate matter (PM)

4.1.1 Spatial distribution of the PM reanalysis data over China

We first present the reanalysis fields of the PM concentrations ($PM_{2.5}$ and PM_{10}) in China. Figure 5 shows the six-year mean (2013–2018) spatial distribution of the $PM_{2.5}$ concentration in China obtained from the CAQRA dataset, base simulation and observations. The CAQRA dataset provides a continuous map of the $PM_{2.5}$ concentration in China and suitably reproduces the observed magnitude of the $PM_{2.5}$ concentration in China. The highest $PM_{2.5}$ concentrations were observed in the NCP region due to its intensive industrial activities and the associated high emissions of $PM_{2.5}$ and its precursors (Qi et al., 2017). High $PM_{2.5}$ concentrations were also found in the SE region, where the $PM_{2.5}$ concentration is influenced by both local emissions and the long-range transport of air pollutants from northern China (Lu et al., 2017). In the NW region, in addition to hotspots exhibiting high $PM_{2.5}$ concentrations in large cities, high $PM_{2.5}$ concentrations were also observed in the Taklimakan Desert due to the influences of dust emissions. The observed magnitude and spatial variability of the PM_{10} concentration were also represented well by the PM_{10} reanalysis field. In general, the spatial distributions of the PM_{10} reanalysis were similar to those of the $PM_{2.5}$ reanalysis except in Gansu and Ningxia provinces, where high PM_{10} concentrations and relatively low $PM_{2.5}$ concentrations occurred. This may be related to the large contributions of dust emissions in these areas. The base simulation notably overestimated the $PM_{2.5}$ and PM_{10} concentrations in China. This may occur due to the systematic biases in the emission inventory (Kong et al., 2019) and because negative trends of PM and its precursor emissions were not considered in our simulations. In addition, the $PM_{2.5}$ concentration hotspots in the NW region and Tibetan Plateau were not captured in the base simulation, possibly due to the absence of emissions in these remote regions.

Seasonal maps of the $PM_{2.5}$ and PM_{10} concentrations are shown in Figs. S7–8 in the Supplement, which reveal profound seasonal variations. Both the $PM_{2.5}$ and PM_{10} concentrations exhibit maximum values in winter in most regions of China due to the increased anthropogenic emissions related to enhanced power generation, industrial activities and fossil fuel burning for heating purposes (Li et al., 2017). Unfavourable meteorological conditions with stable boundary conditions also contribute to the high PM concentrations in winter. In contrast, due to the low emission rate and intense mixing processes, the PM concentrations are the lowest in summer. The PM concentrations in the Taklimakan Desert exhibit a different seasonality, with the highest PM concentrations occurring in spring and the lowest levels occurring in winter. This occurs because the major PM sources in the Taklimakan Desert are not anthropogenic emissions but dust emissions, which are usually the highest in spring due to the frequent strong dust storms. Figure 6 further shows an example of the hourly PM reanalysis results, including a year-round time series of the site mean hourly PM concentrations in Beijing. This figure shows that PM reanalysis suitably captures the hourly evolution of the PM concentrations. Both the heavy haze episodes during the wintertime and the strong dust storms during the springtime are represented well in PM reanalysis.

4.1.2 Assessment of the PM reanalysis data over China

The CV method was used to assess the quality of the PM reanalysis data over China. Table 2 summarizes the site-based CV results for the reanalysis data from 2013 to 2018 at the different temporal scales. It should be mentioned that these sites are all validation sites not used in the data assimilation process. The validation results indicated that by assimilating the surface PM concentrations, the reanalysis data exhibit a relatively high performance in reproducing the magnitude and variability of the surface PM concentrations in China. The CV R^2 values were up to 0.81 and 0.72 in regard to the hourly $PM_{2.5}$ and PM_{10} concentrations, respectively, which were much higher than the values of 0.26 and 0.17, respectively, in the base simulation. The bias was substantially reduced in the $PM_{2.5}$ and PM_{10} reanalysis data with CV mean bias (MBE) values of approximately -2.6 μg/m^3 (-4.9%) and -6.8 μg/m^3 (-8.7%), respectively, at the hourly scale, much smaller than the large bias in the base simulation. The CV RMSE values were only approximately half of the base simulation RMSE values, which were approximately 17.6 and 39.3 μg/m^3 for the hourly $PM_{2.5}$ and PM_{10} concentrations, respectively. The reanalysis data showed a good performance at the daily, monthly and yearly scales, with CV RMSE values ranging from 9.0 to 15.1 μg/m^3 for the $PM_{2.5}$ concentration and from 19.1 to 28.8 μg/m^3 for the PM_{10} concentration.

The quality of the $PM_{2.5}$ and PM_{10} reanalysis data in the different regions of China is further summarized in Table S1-2. At the hourly scale, small negative biases of the $PM_{2.5}$ reanalysis data were found in the NCP (-4.8%), NE (-5.8%), SE (-3.8%) and SW (-3.4%) regions. The biases in the NW and central regions were relatively large, with CV normalized mean bias (CV NMB) values of approximately -7.3% and -8.2%, respectively. Two reasons might explain the large biases in these two regions. First, the observation sites are sparse in the NW and central regions. As a result, the $PM_{2.5}$ concentration is not suitably constrained at certain sites in the CV method. Second, the emissions of $PM_{2.5}$ and its precursors might be very low in these two regions, leading to underestimation of the background errors since we only considered the emission uncertainty in the ensemble simulations. Although this problem was alleviated by using the inflation technique to compensate for the missing errors, the overconfident model results still degraded the assimilation performance to a certain extent, making the analysis less influenced by the observations. The errors of the $PM_{2.5}$ reanalysis data exhibited apparent spatial differences (Table S1). The CV RMSE values were the smallest in the SE (14.9 μg/m^3) and SW (16.5 μg/m^3) regions and increased to ~25 μg/m^3 in the NCP, NE and central regions. Consistent with the bias distributions, the largest CV RMSE value was found in the NW region, which reached 52.1 μg/m^3 but was still much smaller than the RMSE value of the base simulation (73.0 μg/m^3). The errors of the $PM_{2.5}$ reanalysis data were small at the daily, monthly and yearly scales, with CV RMSE values of approximately 10.6–39.4 μg/m^3 at the daily scale, 7.4–26.9 μg/m^3 at the monthly scale and 6.1–23.5 μg/m^3 at the yearly scale. In terms of the hourly PM_{10} reanalysis data, the CV results (Table S2) indicated that small negative biases occurred in the NCP, NE, SE and

SW regions, ranging from -9.6% (NE region) to -5.9% (SE region). The biases were larger in the NW and central regions, with the CV NBM values increasing to approximately 18.0% and 14.1%, respectively. The errors of the PM_{10} reanalysis data also exhibited a spatial heterogeneity. The CV RMSE value was the smallest in the SE (26.0 $\mu g/m^3$) and SW (30.2 $\mu g/m^3$) regions and increased to approximately 39.8 and 43.7 $\mu g/m^3$ in the NE and NCP regions, respectively. The largest errors were found in the central and NW regions, with CV RMSE values of approximately 105.5 and 57.3 $\mu g/m^3$, respectively. The PM_{10} reanalysis data revealed small errors at the daily, monthly and yearly scales, with CV RMSE values of approximately 18.6–85.5 $\mu g/m^3$ at the daily scale, 13.7–64.0 $\mu g/m^3$ at the monthly scale and 12.3–55.8 $\mu g/m^3$ at the yearly scale.

4.1.3 Trend study of the PM reanalysis data over China

A realistic representation of the observed interannual change is another important aspect of the reanalysis dataset. The performance of the reanalysis data in representing the observed interannual changes in the $PM_{2.5}$ and PM_{10} concentrations was thus evaluated nationwide and in the different regions of China. Figures 7–8 show time series of the monthly mean $PM_{2.5}$ and PM_{10} concentrations nationwide and in the different regions. The observed national $PM_{2.5}$ concentration revealed a profound seasonal cycle with the highest concentration in winter and the lowest level in summer. The annual trends of the $PM_{2.5}$ and PM_{10} concentrations were also calculated using the Mann-Kendall (M-K) trend test and the Theil-Sen trend estimation method, which are summarized in Table 3. A significant negative trend was observed in the $PM_{2.5}$ concentration nationwide, with a calculated annual trend of approximately -5.8 ($p<0.05$) $\mu g \cdot m^{-3} \cdot yr^{-1}$. The NE and NCP regions exhibited the highest negative trends among the six regions, with calculated trends of approximately -7.5 ($p<0.05$) and -7.0 ($p<0.05$) $\mu g \cdot m^{-3} \cdot yr^{-1}$, respectively. In the other regions, the negative trends ranged from -6.3 to -5.2 $\mu g \cdot m^{-3} \cdot yr^{-1}$. The base simulation suitably reproduced the observed seasonal cycle of the $PM_{2.5}$ concentration in all regions. The magnitude of the $PM_{2.5}$ concentration in 2013 was also captured well in the different regions, suggesting that the emission inventories of 2010 were generally reasonable for the simulation of the $PM_{2.5}$ concentration in 2013. However, starting from 2014, the base simulation tended to overestimate the observations in the NCP, SE and SW regions, indicating that the emission inventory of 2010 may be too high for the simulation of the $PM_{2.5}$ concentration in these regions after 2014. In contrast, the base simulation significantly underestimated the $PM_{2.5}$ concentration in the NW region. The model performance of the base simulation was relatively good in the NE and central regions throughout the six years. Although the base simulation captured the negative trends of the observed $PM_{2.5}$ concentration in China and the different regions, the simulated trends were much lower than those indicated by the observations. Since we adopted the same emission inventory in the simulations of the air pollutants in the different years, the simulated trends in the base simulation were only driven by the variations in meteorological conditions. This suggests that the change in meteorological conditions only explained a small proportion of the negative trends in the $PM_{2.5}$ concentration in China and that emission reductions contributed more to the decline in the $PM_{2.5}$ concentration. The CV run agreed better with the observations. The observed trends of the $PM_{2.5}$ concentration in China and each subregion were all suitably captured by the reanalysis in the CV run. Similar results were obtained for the analysis of the trend of the PM_{10} concentration, as shown in Fig. 8. The observed PM_{10} concentration also exhibited significant negative trends, which were captured well by the PM_{10} reanalysis in the CV run. The base simulation attained a better performance in reproducing the PM_{10} concentration in China than in reproducing the $PM_{2.5}$ concentration, while significant underestimations of the PM_{10} concentration occurred in the NW and central regions. The calculated negative trends of the base simulation were still lower than those indicated by the observations. This again highlights the large contributions of emission reduction to the improvement of the air quality in China in these years.

4.1.4 Independent validation of the $PM_{2.5}$ reanalysis data

In addition to the CV method, the $PM_{2.5}$ reanalysis data were further validated against an independent dataset acquired from the U.S. Department State Air Quality Monitoring Program over China (http://www.stateair.net/; last accessed: 17 April 2020), which contains the hourly $PM_{2.5}$ concentration in Beijing, Chengdu, Guangzhou, Shanghai and Shenyang cities. Table 4 presents a comparison of the observed $PM_{2.5}$ concentrations to those obtained from the CAQRA dataset and base simulation. The results indicated that the magnitude and variability of the $PM_{2.5}$ reanalysis data agreed better with those of the observed $PM_{2.5}$ concentrations in all cities. Both the MBE and RMSE values were greatly reduced in the CAQRA dataset, which only ranged from -7.1 to -0.3 $\mu g \cdot m^{-3}$ and from 16.8 to 33.6 $\mu g \cdot m^{-3}$, respectively, in these cities. The correlation coefficient was also greatly improved in CAQRA (R^2 = 0.74–0.86) over the base simulation (R^2 = 0.09–0.38). These results confirm that the CAQRA dataset attains a high quality performance in representing the $PM_{2.5}$ pollution in China in these years.

4.1.5 Comparison to the satellite-estimated $PM_{2.5}$ concentration

Previous studies have shown that estimating the ground-based $PM_{2.5}$ concentration from the satellite-derived AOD is an effective way to map the $PM_{2.5}$ concentration with a good accuracy. To further demonstrate the accuracy of our $PM_{2.5}$ reanalysis data, we also compared the accuracy to that of satellite-estimated $PM_{2.5}$ concentrations. Table 5 summarizes several representative studies focusing on the estimation of the ground-based $PM_{2.5}$ concentration in China at the national level using different kinds of methods. Most of these studies estimated the ground-based $PM_{2.5}$ concentration at the daily scale since they employed polar-orbiting satellite data (e.g., MODIS) that only provide daily AOD observations. The estimation conducted by Liu et al. (2019) was an exception which exhibited an hourly resolution due to the use of AOD measurements from a geostationary satellite (Himawari-8). The horizontal resolution in these studies was mainly approximately 10 km except that of Lin et al. (2018), which revealed the finest horizontal resolution (1 km), and that of Zhan et al., 2017, which revealed the coarsest horizontal resolution (0.5°). Few studies have provided long-term $PM_{2.5}$ data covering recent years. In comparison, our $PM_{2.5}$ reanalysis data provide long-term data in China at a fine temporal resolution (1 h) and a high accuracy. A fine temporal resolution is important for epidemiological studies, especially for the assessment of the acute health effects of air pollution. Furthermore, the accuracy of our reanalysis data (CV R^2 = 0.86 and CV RMSE = 15.1 $\mu g \cdot m^{-3}$) was also higher than that of most of these satellite estimates (CV R^2 = 0.56–0.86 and CV RMSE = 15.0–20.2 $\mu g \cdot m^{-3}$).

4.2 Gases

4.2.1 Spatial distribution of the reanalysis data of gaseous air pollutants over China

Next, we present the reanalysis fields for gaseous air pollutants in China, namely, SO_2, CO, NO_2 and O_3. Figure 9 shows the spatial distribution of the six-year average SO_2 and CO concentrations in China obtained from the CAQRA dataset, base simulation and observations. The SO_2 reanalysis data captured the magnitude and spatial distribution of the SO_2 concentration in China well, while the base simulation greatly overestimated the SO_2 concentration due to the positive biases of the SO_2 emissions in the simulations. Consistent with the observations, the SO_2 reanalysis data exhibited high spatial heterogeneity, with the highest values located in the NCP region, especially in Shandong, Shanxi and Hebei provinces. Several SO_2 concentration hotspots were also found in the NE region. SO_2 is mainly emitted from fossil fuel consumption, especially coal burning (Lu et al., 2010). Shandong, Shanxi, Inner Mongolia and Hebei provinces are the four largest consumers of coal in China according to the China Energy Statistical Yearbook (NBSC 2017a, b), which explains the high SO_2 concentrations in these provinces. The spatial distribution of the CO reanalysis data was similar to that of the SO_2 reanalysis data and agreed well with the observed spatial distribution. In contrast, the base simulation highly underestimated the CO concentration, especially in the NCP region. In addition, both the observations and reanalysis data showed CO concentration hotspots in the NW region and Xizang Province, while these hotspots were largely underestimated or even missing in the base simulation. According to previous studies, such underestimation might be related to underestimated CO emissions in China (Kong et al., 2020; Tang et al., 2013). In regard to NO_2 (Fig. 10), both the reanalysis data and base simulation captured the observed magnitude and spatial distribution of the NO_2 concentration in China. High NO_2 concentrations generally occurred in the NCP region and the major city clusters in China. However, the base simulation generally revealed an underestimated NO_2 concentration in China. The spatial distribution of the O_3 concentration (Fig. 10) demonstrated a lower spatial heterogeneity than that of the other gases. The O_3 reanalysis data suitably captured the observed magnitude and spatial distribution of the O_3 concentration in China, while the base simulation generally underestimated the O_3 concentration in China. Figures S9–12 further show seasonal maps of the reanalysis fields of these gases. All gases exhibited a profound seasonal cycle, with maximum values observed in winter and the lowest values in summer except O_3, which demonstrated the opposite seasonal cycle. The highest SO_2, CO and NO_2 concentrations in winter could occur due to the increased anthropogenic emissions and the more stable atmospheric conditions during this season. Regarding O_3, the highest value in summer was closely related to the enhanced photochemical reactions in summer associated with the high temperature and solar radiance.

4.2.2 Assessment of the gas reanalysis data over China

Evaluation results of the above gas reanalysis data are provided in Table 2. The table indicates that the reanalysis data attain an excellent performance in representing the magnitude and variability of these gaseous air pollutants in China, with CV R^2 values ranging from 0.51 for SO_2 to 0.76 for O_3 and CV MBE (CV NMB) values of approximately -2.0 $\mu g \cdot m^{-3}$ (-8.5%), -2.3 $\mu g \cdot m^{-3}$ (-6.9%), -0.06 $mg \cdot m^{-3}$ (-6.1%) and -2.3 $\mu g \cdot m^{-3}$ (-4.0%) for the hourly SO_2, NO_2, CO and O_3 reanalysis data, respectively. Compared to the base simulation, the errors were reduced by approximately half in the reanalysis data with CV RMSE values of approximately 24.9 $\mu g \cdot m^{-3}$, 16.4 $\mu g \cdot m^{-3}$, 0.54 $mg \cdot m^{-3}$ and 21.9 $\mu g \cdot m^{-3}$ for the hourly SO_2, NO_2, CO and O_3 reanalysis data, respectively. The reanalysis data achieved a good performance at the daily, monthly and yearly scales. The CV RMSE values of the daily SO_2 and NO_2 reanalysis data were also smaller than those of the SO_2 and NO_2 concentration datasets in China previously developed by Zhan et al. (2018) and Zhang et al. (2019), respectively, based on the random-forest-spatiotemporal-kriging model wherein the RMSE values of the daily SO_2 and NO_2 concentrations were estimated to be 19.5 and 13.3 $\mu g \cdot m^{-3}$, respectively.

In terms of the different regions (Tables S3–6), the hourly SO_2 reanalysis data indicated small negative biases (approximately 2–10%) in all regions except in the central region, where the negative bias was relatively large (17.0%). The smallest CV RMSE values of the SO_2 reanalysis data were observed in the SE, SW and NW regions (smaller than 25 $\mu g \cdot m^{-3}$), while in the other regions, the CV RMSE values exceeded 30 $\mu g \cdot m^{-3}$. The hourly NO_2 reanalysis data showed small negative biases in all regions, which were relatively small in the NE, NCP and SE regions (ranging from -5.9 to -3.5%) and were relatively large in the SW, NW and central regions (ranging from -15.1 to -12.9%). The CV RMSE for the hourly NO_2 reanalysis data was approximately 15 $\mu g \cdot m^{-3}$ in all regions except in the NW (24.3 $\mu g \cdot m^{-3}$) and central (20.5 $\mu g \cdot m^{-3}$) regions. The hourly CO reanalysis data exhibited small negative biases in all regions. The largest biases were still found in the NW region, which reached approximately 15.0%, while in the other regions, the biases ranged from -11.2% to -2.5%. The CV RMSE values for the hourly CO reanalysis data were the smallest in South China (approximately 0.39 and 0.46 $mg \cdot m^{-3}$ in the SE and SW regions, respectively) and increased to 0.64 and 0.59 $mg \cdot m^{-3}$ in the NCP and NE regions, respectively. The largest CV RMSE was observed in the NW region, which amounted to approximately 1.13 $mg \cdot m^{-3}$. The biases of the hourly O_3 reanalysis data were uniformly distributed in the different regions, with the CV NMB value ranging from -6.1% to 1.4%. Similarly, the CV RMSE value of the O_3 reanalysis data was approximately 20 $\mu g \cdot m^{-3}$ in all regions except in the NW region (28.3 $\mu g \cdot m^{-3}$).

4.2.3 Trend study of the gas reanalysis data over China

Figure 11 shows time series of the monthly mean SO_2 concentration in China obtained from the CV run, base simulation and observations. Additionally, time series of the monthly mean SO_2 concentration in the different regions are shown. The observed SO_2 concentrations showed significant negative trends ($P<0.05$) in China (-6.2 $\mu g \cdot m^{-3} \cdot yr^{-1}$, Table 6) and in all regions (ranging from -2.3 to -9.5 $\mu g \cdot m^{-3} \cdot yr^{-1}$, Table 6) due to the large reductions in SO_2 emissions across China. During the 11th-13rd Five-Year Plans (FYPs) and the Air Pollution Prevention and Control Plan, the Chinese government invested great efforts to reduce SO_2 emissions, such as the installation of flue-gas desulfurization (FGD) and selective catalytic reduction systems, construction of large units, decommissioning of small units and replacement of coal with cleaner energies (Li et al., 2017; Zheng et al., 2018b). As a result, the SO_2 emissions substantially decreased in China, especially in the industrial and power sectors. The base simulation significantly overestimated the SO_2 concentration in all regions, especially after 2013.

The negative trends of the SO_2 concentration were also largely underestimated in the base simulation. In contrast, the SO_2 reanalysis data captured the magnitude and negative trends of the observed SO_2 concentrations in China and in all regions well. The NO_2 observations showed negative trends in China as well (Fig. 12). However, the negative trend was not significant except in the NE region (Table 6). This is consistent with the small reductions in NO_x emissions (21%) in China due to the small changes in the emissions originating from the transportation sector, accounting for almost one-third of the NO_x emissions in China. The pollution controls applied in the transportation section were exactly offset by the growing emissions related to vehicle growth (Zheng et al., 2018b). The base simulation generally underestimated the NO_2 concentration during the wintertime, and the observed negative trends of the NO_2 concentration were also underestimated in all regions. By assimilating the observed NO_2 concentrations, the reanalysis data agreed better with the observations both in regard to the magnitude and negative trends. The CO observations exhibited significant negative trends in all regions except in the NW region (Fig. 13), with calculated negative trends ranging from -0.18 to -0.06 $\mu g \cdot m^{-3} \cdot yr^{-1}$. Such negative trends have also been observed in satellite measurements, such as MOPITT observations (Zheng et al., 2018a), which are mainly attributed to the reduced anthropogenic emissions in China, as suggested by both bottom-up and top-down methods (Zheng et al., 2019). The base simulation largely underestimated the CO concentration in all regions. In addition, the negative trends of the CO concentration were also notably underestimated in the base simulation, which highlights the major contribution of emission reduction to the decreased CO concentration in these regions. The CO reanalysis data agreed well with the observations and captured the negative trends of the CO concentration in all regions. The O_3 concentration exhibited the opposite trend to that exhibited by the other air pollutants (Fig. 14), which revealed significant positive trends in all regions, ranging from 2.3 to 5.4 $\mu g \cdot m^{-3} \cdot yr^{-1}$ and indicating enhanced photochemical pollution in China. This phenomenon has been observed and investigated by Li et al. (2019), who suggested that the rapid decrease in the $PM_{2.5}$ concentration and the resultant reduction in the aerosol sink of hydroperoxyl (HO_2) radicals were important factors contributing to the enhanced O_3 concentration in China. The base simulation generally captured the magnitude of the O_3 concentration in the SE, SW, NW and central regions but underestimated the O_3 concentration in the NCP and NE regions, especially in spring and summer. In addition, the base simulation underestimated the observed positive trends of the O_3 concentration in all regions, which suggests that meteorological variability only contributed a small proportion of the observed O_3 trend in China. Again, the O_3 reanalysis data are substantially better than the base simulation and suitably reproduce the observed trends of the O_3 concentration in all regions.

4.2.4 Comparison to the CAMS reanalysis data

To further evaluate the accuracy of our reanalysis dataset for gaseous air pollutants, the CAMSRA dataset produced by the ECMWF (Inness et al., 2019) was employed as a reference in a comparison to our reanalysis dataset. The CAMSRA dataset is the latest global reanalysis dataset of the atmospheric composition, which assimilates satellite retrievals of O_3, CO, NO_2 and AOD. Three-hour reanalysis data of the SO_2, NO_2, CO and O_3 concentrations at the surface model level from 2013 to 2018 were adopted in this study, which were downloaded from https://atmosphere.copernicus.eu/copernicus-releases-new-global-reanalysis-data-set-atmospheric-composition (last accessed: 17 April 2020) at a resolution of 1 degree by 1 degree. Here, we only focus on a comparison of the gaseous pollutants since the CAMSRA dataset does not provide $PM_{2.5}$ and PM_{10} concentrations.

Figure 15 shows the spatial distribution of the six-year average concentration of these gaseous air pollutants in China obtained from the CAMSRA dataset. Compared to the spatial distributions determined with the CAQRA dataset and observations (Figs. 9–10), the CAMSRA dataset greatly overestimates the surface SO_2 and O_3 concentrations in China. In addition, due to the higher spatial resolution (15 km) of the CAQRA dataset than that of the CAMSRA dataset (approximately 50 km), our products provide more detailed spatial patterns of the surface air pollutants in China, which are better suited for air quality studies at the regional scale. Table 7 quantitatively compares the accuracy of the CAQRA dataset to that of the CAMSRA dataset in the estimation of the surface concentrations of gaseous air pollutants in China. Compared to CAMSRA (R^2 = 0.00–0.23), CAQRA attains a much better performance in capturing the spatiotemporal variability in the surface concentrations of gaseous air pollutants in China, with R^2 values ranging from 0.53 to 0.77. The MBE and RMSE values are also smaller in the CAQRA dataset than those in the CAMSRA dataset, especially for the SO_2 and O_3 concentrations. This is attributed to the assimilation of surface observations in CAQRA, while CAMSRA only assimilates satellite retrievals. These results suggest that the CAQRA dataset provides surface air quality datasets in China of a higher quality than the air quality datasets provided by the CAMSRA dataset, which is especially valuable for future relevant studies with high demands in spatiotemporal resolution and accuracy.

5 Conclusions

A high-resolution CAQRA dataset was produced in this study by assimilating surface observations of the $PM_{2.5}$, PM_{10}, SO_2, NO_2, CO and O_3 concentrations retrieved from the CNEMC. This dataset provides time-consistent concentration fields of $PM_{2.5}$, PM_{10}, SO_2, NO_2, CO and O_3 in China from 2013 to 2018 (will be extended in the future on a yearly basis) at high spatial (15 km) and temporal (1 hour) resolutions. The CAQRA dataset was produced with the ChemDAS, which applied the NAQPMS model as the forecast model, and the LETKF to assimilate the observations in the postprocessing mode. The background error covariance was calculated from ensemble simulations, which considered the emission uncertainties of the major air pollutants. An inflation technique was also applied to dynamically inflate the background error to prevent underestimation of the true background error covariance.

The fivefold CV method was employed to validate the reanalysis dataset, which provided us with the first indication of the quality of the CAQRA dataset. The validation results suggested that the CAQRA dataset attains an excellent performance in representing the spatiotemporal variability of surface air pollutants in China, with CV R^2 values ranging from 0.52 for the hourly SO_2 concentration to 0.81 for the hourly $PM_{2.5}$ concentration. The CV MBE values of the reanalysis data were -2.6 $\mu g \cdot m^{-3}$, -6.8 $\mu g \cdot m^{-3}$, -2.0 $\mu g \cdot m^{-3}$, -2.3 $\mu g \cdot m^{-3}$, -0.06 $mg \cdot m^{-3}$ and -2.3 $\mu g \cdot m^{-3}$ for the hourly concentrations of $PM_{2.5}$, PM_{10}, SO_2, NO_2, CO and O_3, respectively. The CV RMSE values of the reanalysis data for these air pollutants were estimated to be approximately 21.3 $\mu g \cdot m^{-3}$, 39.3 $\mu g \cdot m^{-3}$, 24.9 $\mu g \cdot m^{-3}$, 16.4 $\mu g \cdot m^{-3}$, 0.54 $mg \cdot m^{-3}$ and 21.9 $\mu g \cdot m^{-3}$, respectively. In the different regions of China, the NW and central regions exhibited relatively large biases and errors, which mainly occurred due to the relatively sparse observations and underestimated background errors. The Chinese air quality has substantially changed over the last six years. The observations indicate significant decreasing trends for all air pollutants except O_3, which shows an increasing trend over the last six years. The reanalysis data reveal an excellent performance in representing the trends of all air pollutants in China, suggesting the suitability of the reanalysis data for air pollutant trend analysis in China.

In addition to the CV method, the $PM_{2.5}$ reanalysis data were also evaluated against independent observations retrieved from the U.S. Department State Air Quality Monitoring Program over China. The results suggested that the reanalysis data suitably reproduce the magnitude and variability of the observed $PM_{2.5}$ concentration in all cities, with the MBE and RMSE values only ranging from -7.1 to -0.3 $\mu g \cdot m^{-3}$ and from 16.8 to 33.6 $\mu g \cdot m^{-3}$, respectively. The reanalysis data of the gaseous air pollutants were also compared to the latest global reanalysis data contained in the CAMSRA dataset produced by the ECMWF. The CAMSRA dataset is of great value in providing three-dimensional distributions of multiple chemical species globally. As a regional dataset, our products attain a higher spatial resolution than does the CAMSRA dataset, which could better suit air quality studies at the regional scale. Although our products only provide the surface concentrations of six conventional air pollutants in China, the accuracy of the CAQRA dataset was estimated to be higher than that of the CAMSRA dataset due to the assimilation of surface observations. Hence, our products exhibit their own value in regional air quality studies with high demands in spatiotemporal resolution and accuracy. We also compared our $PM_{2.5}$ reanalysis data to previous satellite estimates of the surface $PM_{2.5}$ concentration, which revealed that the $PM_{2.5}$ reanalysis data are more accurate than most satellite estimates and exhibit a relatively fine temporal resolution.

As the first version of the CAQRA dataset, certain limitations remain that potential users should be aware of. First, the discontinuities in the availability and coverage of assimilated observations will affect the reanalysis quality and the estimated interannual trends. As shown in Sect.3.1, there has been a consistent increase in the number of assimilated observations from 2013 to 2015 due to the increases of observation sites. The smaller number of assimilated observations in 2013 and 2014 would provide less constrains on the background state and thus degrade the reanalysis in these two years. This may cause spurious interannual changes and trends from 2013 to 2018. Thus, cautions are needed when using the reanalysis for long-term air quality change from 2013 to 2018. However, this problem would be not serious after 2015 when the number of assimilated observations become stable. In addition, the observation sites used in the assimilation are mainly urban or suburban sites that do not provide enough information on the air pollution in rural areas, which may influence the quality of CAQRA in rural areas. Secondly, we only perturbed the emissions to represent the forecast uncertainty in this study, which may underestimate the forecast uncertainty due to the omitting of other error sources, such as the uncertainty in poorly parameterized physical or chemical processes, and the uncertainty in meteorological simulation. The limited ensemble size would also lead to underestimation of the forecast error especially in the high-resolution assimilation applications. Although the inflation method is used to compensate for the missing errors, the underestimated forecast uncertainty would still degrad the assimilation performance to a certain extent as exemplified by the larger biases in the reanalysis over NW and Central regions. Thirdly, we did not consider the annual trend of emissions in the ensemble simulation. This would lead to temporal changes in the statistics of innovation due to the substantial changes of observations, which would influence the long stability of the data assimilation as suggested by the χ^2 test although the OmA statistics generally confirms a passable stability in our assimilation system. Last but not least, the current CAQRA only contains the surface concentrations of the air pollutants in China which cannot provide the information on the vertical structure of the air pollutants. to further improve the accuracy of our air quality reanalysis dataset, in the future, an online EnKF run could be conducted to simultaneously correct the emissions and concentrations. More observation types, such as observation data of the $PM_{2.5}$ composition, could also be assimilated to provide $PM_{2.5}$ composition fields in China, which could support both epidemiological studies and climate research.

Data availability

The whole CAQRA reanalysis dataset can be freely downloaded at https://doi.org/10.11922/sciencedb.00053 (Tang et al., 2020a), and the prototype product, which contains the monthly and annual means of the CAQRA dataset, is available at https://doi.org/10.11922/sciencedb.00092 (Tang et al., 2020b).

Author contributions

X.T., J.Z., and Z.W. conceived and designed the project; H.W., L.K., X.T., and L.W. established the data assimilation system; Q.W. and L.K. performed the meteorology simulations; X.T., L.K., H.C., H.W., H.Z., G.J. and M.L. conducted the ensemble simulations with the NAQPMS model; J.L., L.Z., W.W., B.L., Q.W., D.C. and T.S. provided the air quality monitoring data; W.H. executed the quality control of the observation data; F.L. estimated the representativeness error of the observations; and L.K. carried out the CAQRA calculations, generated the figures and wrote the paper with comments provided by G.C.

Competing interests

The authors declare that they have no conflicts of interest.

Acknowledgements

This study was funded by the National Natural Science Foundation (Grant Nos. 91644216, 41575128, and 41875164), the CAS Strategic Priority Research Program (Grant No. XDA19040201), and the CAS Information Technology Program (Grant No. XXH13506-302).

Tables

Table 1: Uncertainties in the emissions of the different species

Species	SO_2[a]	NO_x[a]	CO[a]	Non-methane volatile organic compounds (NMVOCs)[a]	NH_3[b]	PM_{10}[a]	$PM_{2.5}$[a]	Black carbon (BC)[a]	Organic carbon (OC)[a]
Emission Uncertainty	12%	31%	70%	68%	53%	132%	130%	208%	258%

[a] Emission uncertainty obtained from Zhang et al. (2009)

[b] Emission uncertainty obtained from Streets et al. (2003)

Table 2: Site-based cross-validation results for the reanalysis data (outside brackets) and base simulation (inside brackets) from 2013 to 2018 at the different temporal scales

	$PM_{2.5}$ (μg/m³)				PM_{10} (μg/m³)			
	R^2	MBE	NMB (%)	RMSE	R^2	MBE	NMB (%)	RMSE
Hourly	0.81 (0.26)	-2.6 (17.6)	-4.9 (34.7)	21.3 (54.1)	0.72 (0.17)	-6.8 (-7.6)	-7.8 (-8.7)	39.3 (75.7)
Daily	0.86 (0.32)	-2.5 (17.4)	-4.9 (34.3)	15.1 (46.4)	0.81 (0.22)	-6.7 (-7.0)	-7.7 (-8.1)	28.8 (64.1)
Monthly	0.88 (0.40)	-2.5 (17.4)	-5.0 (34.1)	10.3 (33.6)	0.83 (0.28)	-6.7 (-7.3)	-7.7 (-8.4)	21.1 (44.4)
Yearly	0.86 (0.37)	-3.0 (15.2)	-5.6 (28.7)	9.0 (28.9)	0.79 (0.27)	-7.5 (-10.2)	-8.3 (-11.3)	19.1 (38.2)
	SO_2 (μg/m³)				NO_2 (μg/m³)			
	R^2	MBE	NMB (%)	RMSE	R^2	MBE	NMB (%)	RMSE
Hourly	0.52 (0.03)	-2.0 (25.5)	-8.5 (106.6)	24.9 (67.2)	0.61 (0.22)	-2.3 (-5.0)	-6.9 (-14.8)	16.4 (24.9)
Daily	0.67 (0.04)	-2.0 (25.6)	-8.5 (106.9)	17.5 (59.3)	0.67 (0.27)	-2.3 (-5.0)	-6.8 (-14.8)	12.3 (19.9)
Monthly	0.74 (0.04)	-2.1 (25.4)	-8.6 (105.7)	13.2 (52.0)	0.67 (0.34)	-2.3 (-5.0)	-6.8 (-14.8)	10.0 (15.9)
Yearly	0.71 (0.04)	-2.6 (23.1)	-9.9 (87.2)	12.0 (47.5)	0.62 (0.42)	-2.5 (-5.9)	-7.3 (-17.3)	9.1 (13.6)
	CO (mg/m³)				O_3 (μg/m³)			
	R^2	MBE	NMB (%)	RMSE	R^2	MBE	NMB (%)	RMSE
Hourly	0.55 (0.17)	-0.06 (-0.47)	-6.1 (-44.7)	0.54 (0.87)	0.76 (0.35)	-2.3 (-10.5)	-4.0 (-17.8)	21.9 (38.3)
Daily	0.61 (0.20)	-0.06 (-0.47)	-5.8 (-44.6)	0.44 (0.77)	0.74 (0.25)	-2.3 (-10.4)	-3.9 (-17.8)	16.6 (31.3)
Monthly	0.62 (0.21)	-0.06 (-0.47)	-6.0 (-44.7)	0.36 (0.69)	0.74 (0.28)	-2.3 (-10.4)	-3.9 (-17.8)	13.1 (25.3)
Yearly	0.52 (0.09)	-0.08 (-0.51)	-6.9 (-46.7)	0.37 (0.72)	0.53 (0.03)	-2.2 (-9.8)	-3.8 (-17.2)	10.4 (21.2)

Table 3: Calculated annual trends of the $PM_{2.5}$ and PM_{10} concentrations in China

	$PM_{2.5}$ (μg/m^3)			PM_{10} (μg/m^3)		
	Observation	Cross-validation	Base simulation	Observation	Cross-validation	Base simulation
China	**-5.8 (-13.4, -3.5)**[a]	**-5.0 (-12.6, -3.1)**	**-2.0 (-3.6, -0.7)**	**-7.2 (-18.4, -3.2)**	**-6.0 (-17.0, -2.9)**	**-2.5 (-3.6, -0.7)**
NCP	**-7.0 (-15.7, -5.5)**	**-6.6 (-14.5, -4.8)**	**-3.5 (-4.7, -1.9)**	**-8.3 (-20.4, -5.1)**	**-7.6 (-19.2, -4.4)**	**-4.2 (-4.7, -1.9)**
NE	**-7.5 (-11.0, -3.9)**	**-6.7 (-10.0, -3.5)**	**-3.2 (-5.8, -1.2)**	**-11.2 (-17.4, -4.7)**	**-10.4 (-16.4, -4.7)**	**-3.7 (-5.8, -1.2)**
SE	**-5.2 (-11.3, -2.8)**	**-4.9 (-10.6, -2.7)**	-0.9 (-3.1, 1.3)	**-6.0 (-14.9, -2.4)**	**-5.8 (-13.2, -1.9)**	-1.6 (-3.1, 1.3)
SW	**-6.3 (-12.8, -2.6)**	**-4.9 (-12.2, -2.4)**	-1.4 (-7.5, 0.4)	**-7.9 (-19.9, -2.2)**	**-5.5 (-17.5, -2.1)**	**-1.3 (-7.5, 0.4)**
NW	-5.7 (-11.6, 2.1)[b]	-3.3 (-10.7, 1.8)	-1.3 (-4.9, 2.9)	-0.5 (-14.4, 1.6)	-2.2 (-8.5, 3.4)	-2.3 (-4.9, 2.9)
Central	**-5.8 (-19.8, -0.8)**	-3.6 (-17.7, 0.2)	-0.6 (-5.9, 0.9)	-8.9 (-28.5, 0.2)	-6.8 (-26.9, 0.5)	-2.0 (-5.9, 0.9)

[a] The bold font denotes that the calculated trend is significant at the 0.05 significance level, and the values in brackets denote the 95% confidence interval.

Table 4: Independent validation results of the CAQRA dataset (outside brackets) and base simulation (inside brackets) against the observation data retrieved from the U.S. Department State Air Quality Monitoring Program over China

	R^2	MBE (μg/m^3)	NMB (%)	RMSE (μg/m^3)
Beijing	0.86 (0.37)	-0.3 (11.4)	-0.3 (13.2)	33.6 (75.6)
Shanghai	0.86 (0.34)	5.5 (39.6)	10.9 (78.3)	17.1 (64.8)
Chengdu	0.85 (0.19)	-7.1 (59.3)	-8.9 (74.7)	23.1 (91.5)
Guangzhou	0.74 (0.09)	-3.3 (11.1)	-7.5 (25.1)	16.8 (38.8)
Shenyang	0.85 (0.29)	-2.2 (16.8)	-3.2 (24.3)	24.8 (59.1)

Table 5 Comparison of the accuracy of our $PM_{2.5}$ reanalysis data to that of satellite estimates

Reference	Spatial resolution	Temporal resolution	Temporal coverage	CV R^2	CV RMSE	Method
Ma et al. (2016)	0.1° ×0.1°	daily	2004–2013	0.79	27.4	LME + GAM
Xue et al. (2019)	0.1° ×0.1°	daily	2000–2016	0.56	30.2	CTM + HD-expansion + GAM
Xue et al. (2017)	0.1° ×0.1°	daily	2014	0.72	23.0	CTM + LME + spatiotemporal kriging
Chen et al. (2018)	0.1° ×0.1°	daily	2005–2016	0.83	18.1	RF
Lin et al. (2018)	1 km× 1km	daily	2001 – 2015	0.78[a]	19.3[a]	Semi-empirical
Chen et al. (2019)	3 km×3 km	daily	2014 – 2015	0.86	15.0	XGBoost + NELRM
Yao et al. (2019)	6 km×6 km	daily	2014	0.60	21.8	TEFR + GWR
You et al. (2016)	0.1° ×0.1°	daily	2014	0.79	18.6	GWR
Zhan et al. (2017)	0.5° ×0.5°	daily	2014	0.76	23.0	GW-GBM
Li et al. (2017b)	0.1° ×0.1°	daily	2015	0.82	16.4	Geoi-DBN
Liu et al. (2019)	0.125° ×0.125°	hourly	2016	0.86	17.3	RF
This study	15 km× 15km	hourly	2013–2018	0.81	21.3	EnKF
		daily	2013–2018	0.86	15.1	EnKF

[a] The accuracy of the $PM_{2.5}$ estimates of Lin et al. (2018) was assessed at the monthly scale.

LME: Linear mixed-effect model

GWR: Geographically weighted regression model

GAM: Generalized additive model

HD-expansion: High-dimensional expansion

RF: Random forest

XGBoost: Extreme gradient boosting

NELRM: Non-linear exposure-lag-response model

TEFR: Time fixed-effects regression model

GW-GBM: Geographically weighted gradient boosting machine

Geoi-DBN: Geographical deep belief network

Table 6: Calculated annual trends of the SO_2, NO_2, CO and O_3 concentrations in China

	SO_2 (μg/m^3)			NO_2 (μg/m^3)		
	Observation	Cross-validation	Base simulation	Observation	Cross-validation	Base simulation
China	**-6.2 (-12.0, -3.9)** [a]	**-4.9 (-10.3, -3.0)**	**-1.7 (-6.2, -0.8)**	-2.6 (-5.9, 0.1)	-2.1 (-5.9, 0.1)	**-0.9 (-3.0, -0.3)**
NCP	**-9.5 (-16.5, -7.2)**	**-8.1 (-14.5, -5.9)**	**-1.7 (-4.1, -1.4)**	-2.0 (-5.9, 0.0)	-2.1 (-5.6, 0.1)	**-0.6 (-1.6, -0.3)**
NE	**-6.8 (-14.6, -4.9)**	**-5.9 (-12.1, -4.1)**	**-1.8 (-7.6, -0.6)**	**-3.0 (-4.9, -1.1)**	**-3.3 (-5.4, -1.2)**	**-1.3 (-3.8, -0.3)**
SE	**-4.4 (-6.7, -2.5)**	**-3.7 (-5.6, -2.0)**	**-1.0 (-2.9, -0.1)**	-2.4 (-5.3, 0.1)	-2.5 (-5.1, 0.1)	**-1.0 (-1.8, -0.3)**
SW	**-4.2 (-8.8, -1.9)**	**-2.8 (-7.6, -1.3)**	**-3.4 (-15.6, -1.9)**	-1.8 (-6.2, 0.3)	-1.6 (-6.5, 0.2)	**-0.7 (-3.9, -0.2)**
NW	**-2.3 (-11.1, 0.6)**	**-4.2 (-7.7, -1.1)**	-1.9 (-13.7, 1.0)	-3.4 (-8.4, 2.3)	-1.7 (-9.5, 1.3)	-1.0 (-6.5, 0.3)
Central	**-7.9 (-17.5, -3.3)**	**-5.5 (-15.7, -2.3)**	-0.6 (-10.2, 0.0)	-2.0 (-6.6, 1.9)	-1.0 (-8.0, 2.2)	-0.5 (-3.8, 0.1)

	CO (mg/m^3)			O_3 (μg/m^3)		
	Observation	Cross-validation	Base simulation	Observation	Cross-validation	Base simulation
China	**-0.12 (-0.17, -0.06)**	**-0.12 (-0.18, -0.07)**	**-0.02 (-0.05 -0.01)**	**3.5 (2.1, 5.0)**	**3.8 (2.1, 5.0)**	**2.0 (0.1, 5.9)**
NCP	**-0.18 (-0.25, -0.11)**	**-0.17 (-0.24, -0.11)**	**-0.03 (-0.05, -0.02)**	**5.3 (2.5, 8.7)**	**5.5 (2.4, 8.8)**	1.4 (-0.5, 5.0)
NE	**-0.13 (-0.21, -0.05)**	**-0.13 (-0.20, -0.06)**	**-0.03 (-0.07, -0.01)**	**4.8 (1.5, 10.0)**	**4.6 (1.4, 9.5)**	2.8 (-0.4, 8.0)
SE	**-0.06 (-0.09, -0.04)**	**-0.06 (-0.08, -0.04)**	**-0.01 (-0.02, -0.01)**	**2.3 (0.3, 3.4)**	**2.6 (0.8, 3.5)**	**1.7 (0.3, 3.0)**
SW	**-0.11 (-0.19, -0.04)**	**-0.09 (-0.21, -0.04)**	**-0.02 (-0.06, -0.01)**	**3.2 (1.2, 5.0)**	**3.5 (1.8, 5.4)**	2.7 (-0.9, 7.1)
NW	-0.14 (-0.46, 0.04)	-0.14 (-0.30, 0.04)	-0.03 (-0.06, 0.00)	**5.4 (1.6, 9.8)**	**4.0 (1.4, 10.1)**	2.6 (-0.2, 8.8)
Central	**-0.16 (-0.27, -0.09)**	**-0.17 (-0.25, -0.10)**	-0.01 (-0.06, 0.00)	**5.3 (2.3, 9.2)**	**4.5 (1.4, 7.8)**	2.2 (-0.3, 7.7)

[a] The bold font denotes that the calculated trend is significant at the 0.05 significance level, and the values in brackets denote the 95% confidence interval.

Table 7: Comparison of the data accuracy of CAQRA and CAMSRA in China

	CAQRA				CAMSRA			
	SO_2 (μg/m^3)	NO_2 (μg/m^3)	CO (mg/m^3)	O_3 (μg/m^3)	SO_2 (μg/m^3)	NO_2 (μg/m^3)	CO (mg/m^3)	O_3 (μg/m^3)
R^2	0.53	0.61	0.55	0.77	0.04	0.23	0.13	0.00
MBE	-2.0	-2.3	-0.1	-2.3	19.4	1.7	-0.2	30.6
NMB (%)	-8.5	-6.9	-6.1	-4.0	81.2	5.2	-17.5	52.1
RMSE	24.8	16.4	0.5	21.9	54.5	27.3	0.9	55.2

Figures

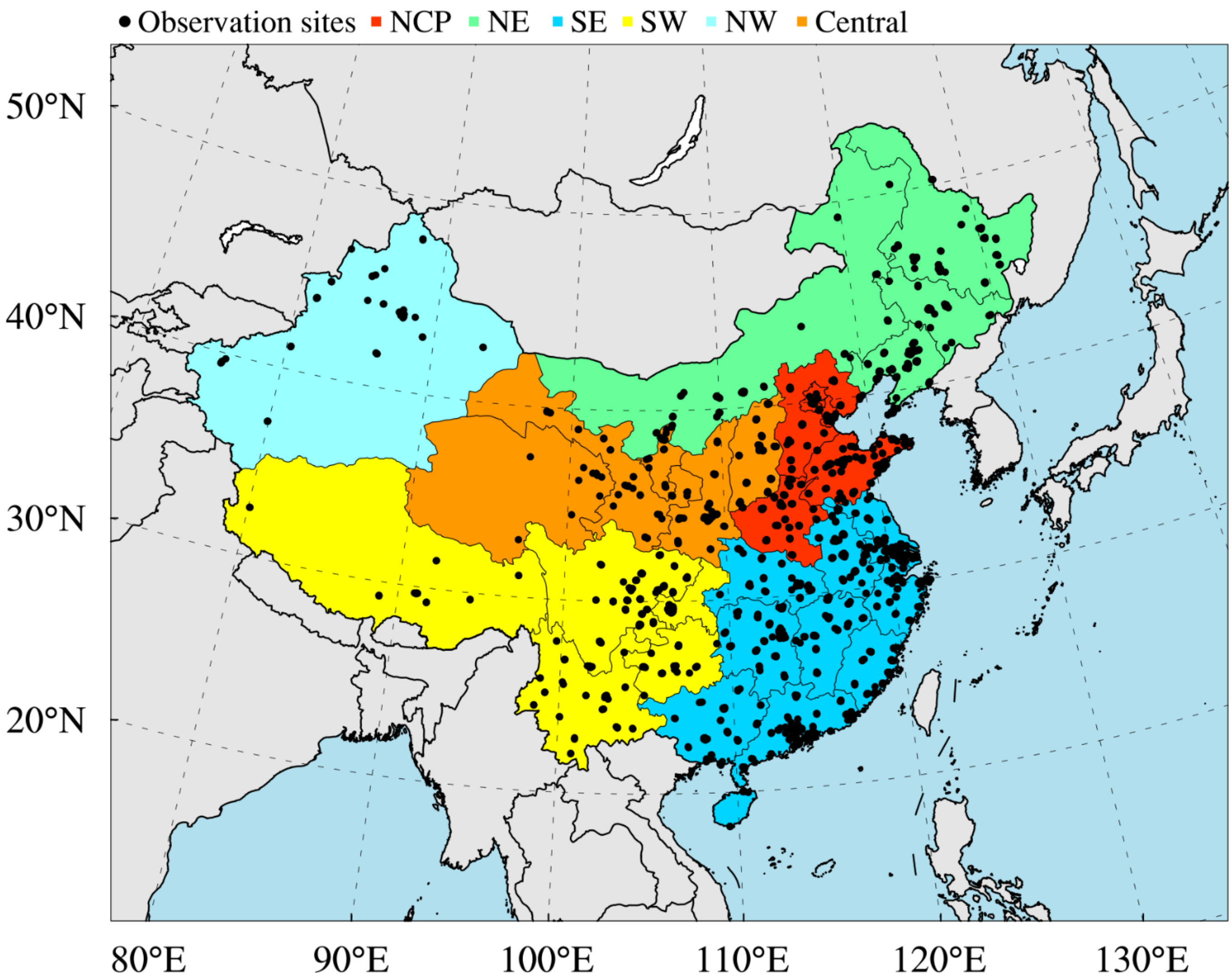


Figure 1: Modelling domain of the ensemble simulation overlain on the distribution of the observation sites of the CNEMC. The different colours denote the different regions in China, namely, the North China Plain (NCP), Northeast China (NE), Southwest China (SW), Southeast China (SE), Northwest China (NW) and Central China.

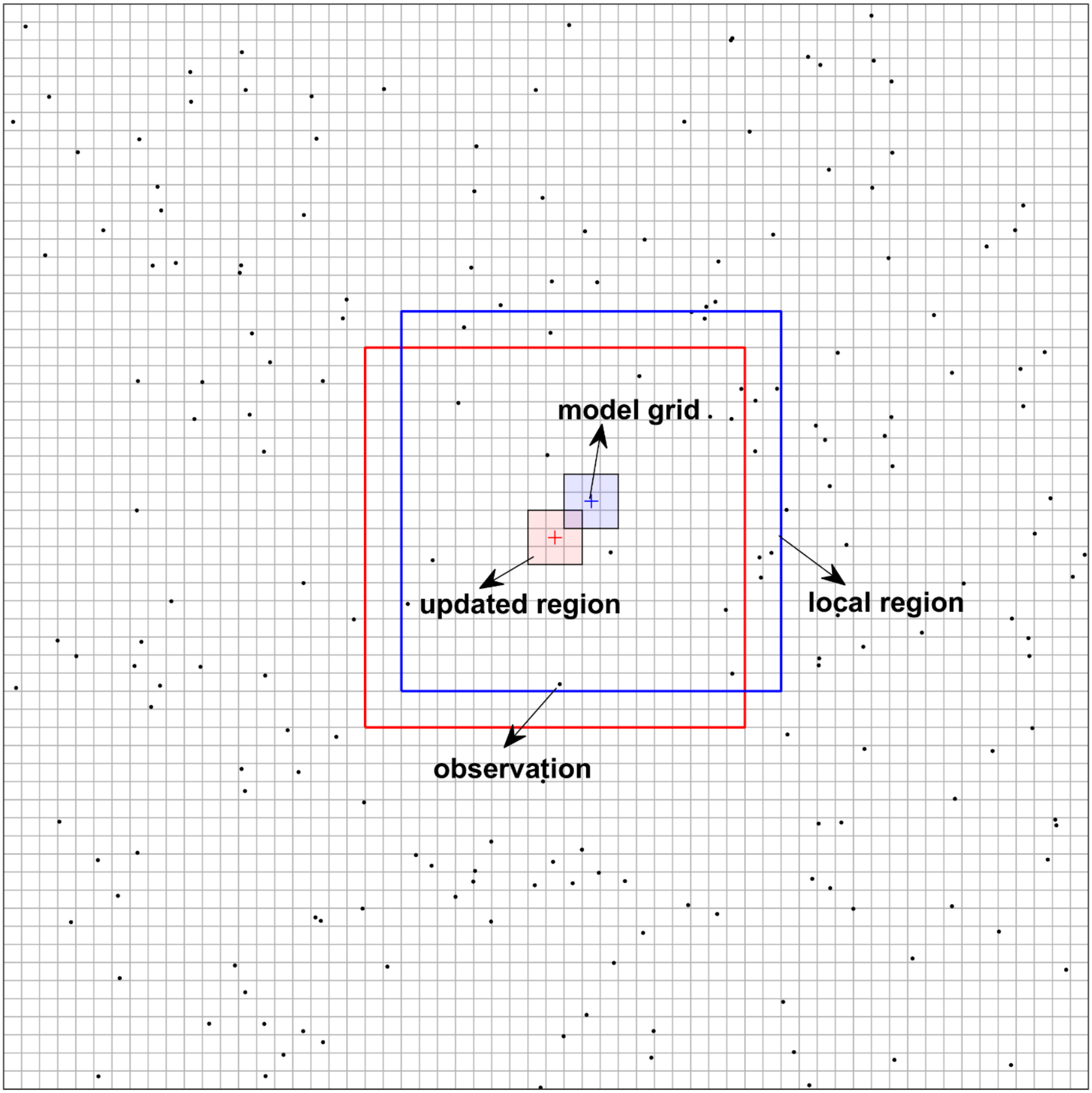


Figure 2: Illustration of the local analysis scheme used in the assimilation. The plus and dot symbols denote the centres of the model grids and the location of the observation sites, respectively. The large rectangular region denotes the local region, and the shaded region denotes the updated region.

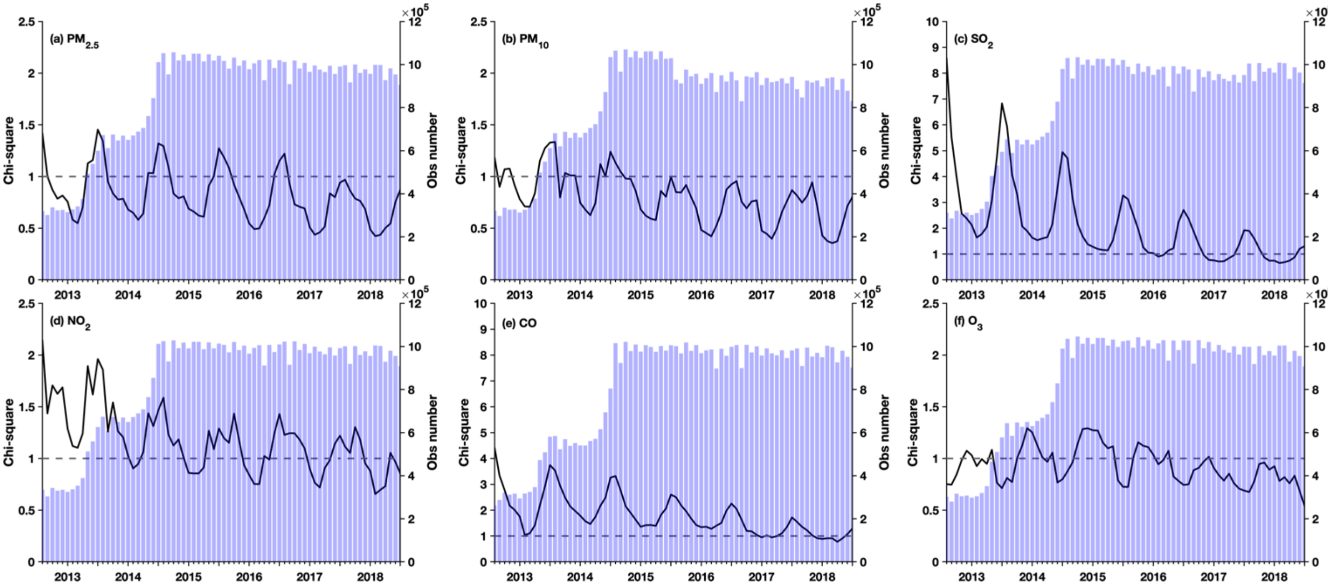


Figure 3: Time series of the monthly mean χ^2 values (black line) and the number of assimilated observations per month (blue bars) for (a) $PM_{2.5}$, (b) PM_{10}, (c) SO_2, (d) NO_2, (e) CO and (f) O_3.

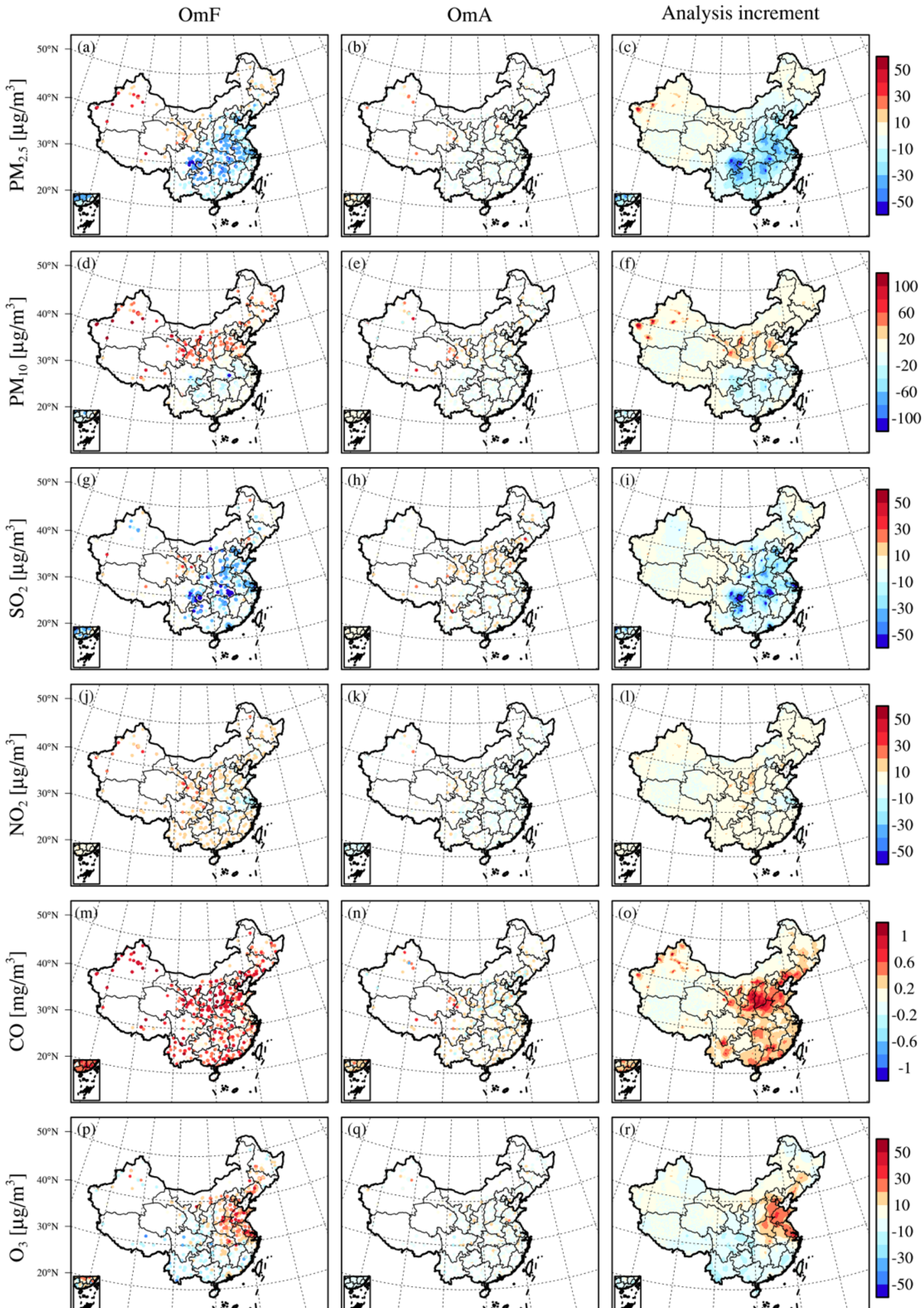


Figure 4: Spatial distributions of the six-year mean OmF (left panel), OmA (middle panel) and analysis increment (right panel) for different species in China.

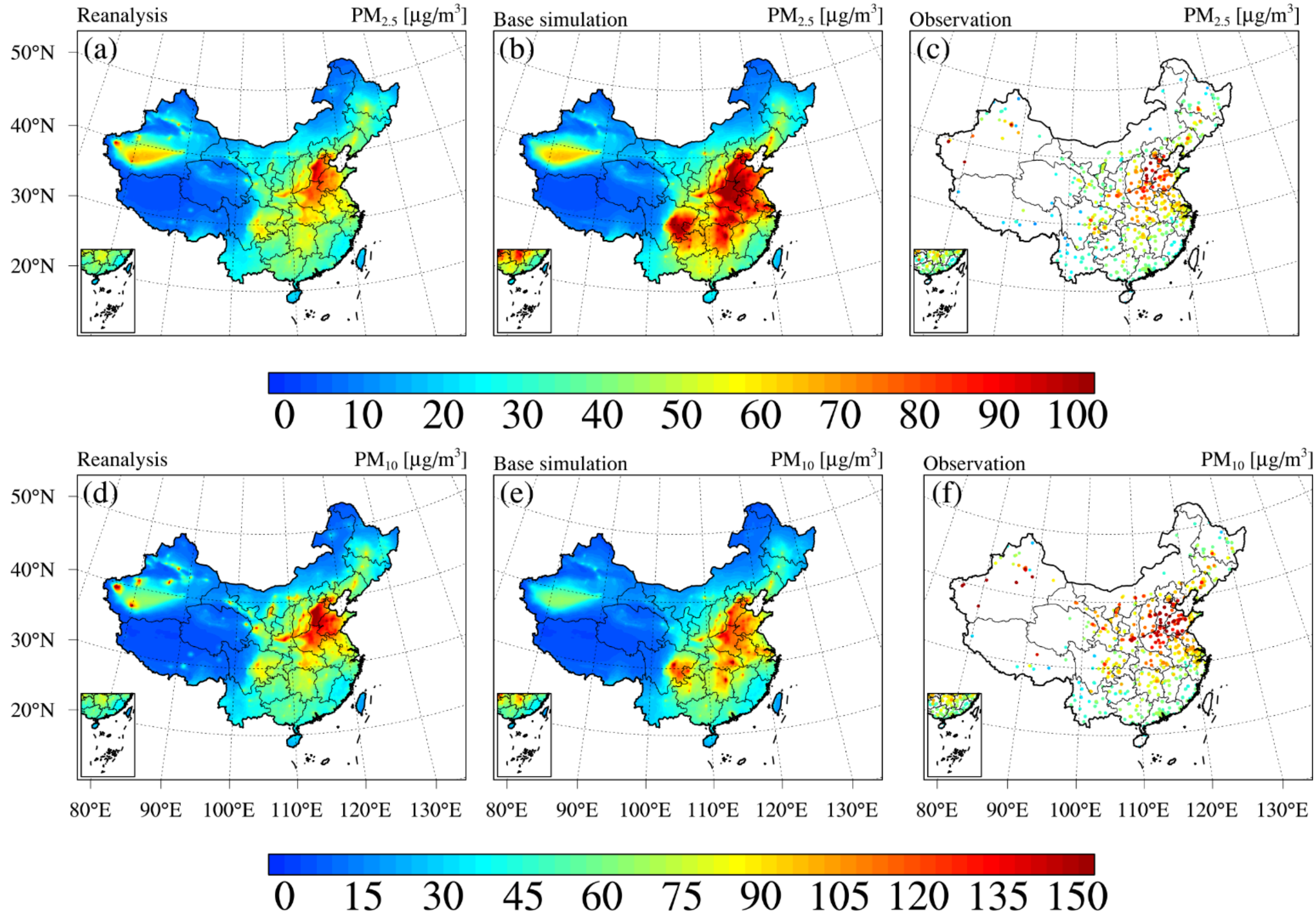


Figure 5: Spatial distributions of the (a–c) $PM_{2.5}$ and (d–f) PM_{10} concentrations in China from (a, d) CAQRA, (b, e) base simulation and (c, f) observations averaged from 2013 to 2018.

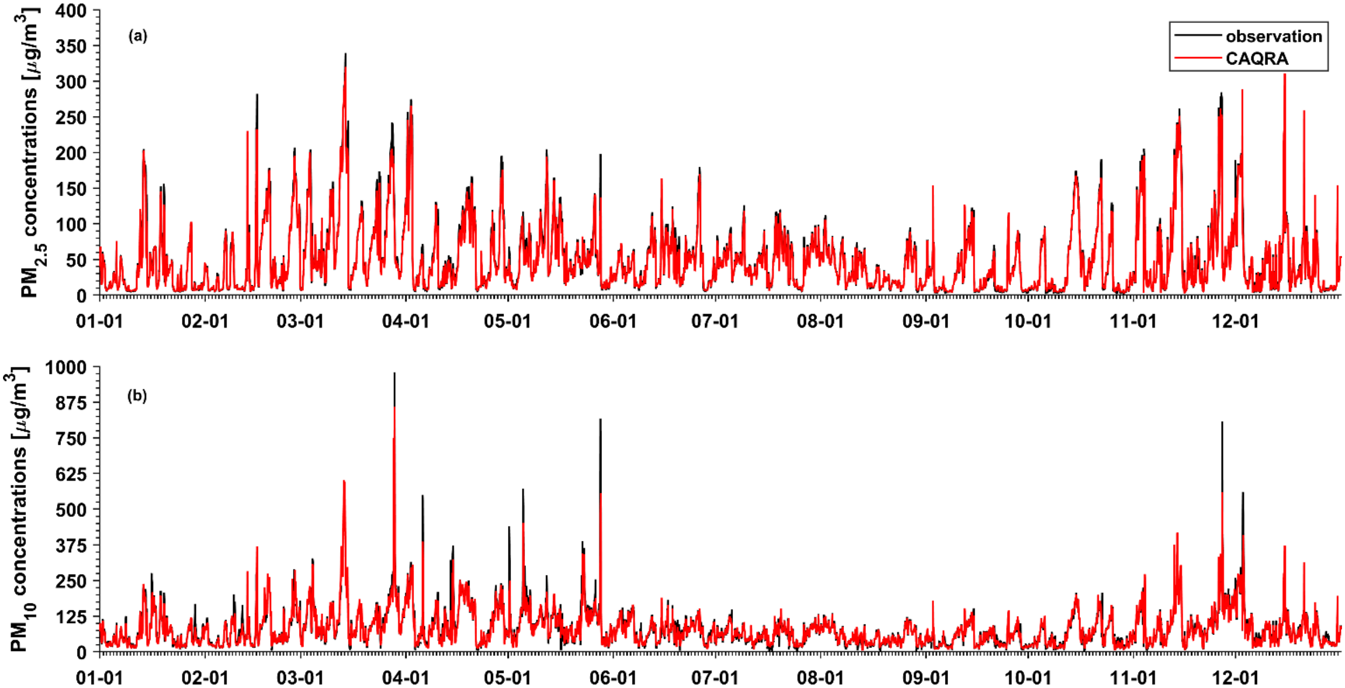


Figure 6: Time series of the site mean hourly (a) $PM_{2.5}$ and (b) PM_{10} concentrations in Beijing obtained from the observations and CAQRA.

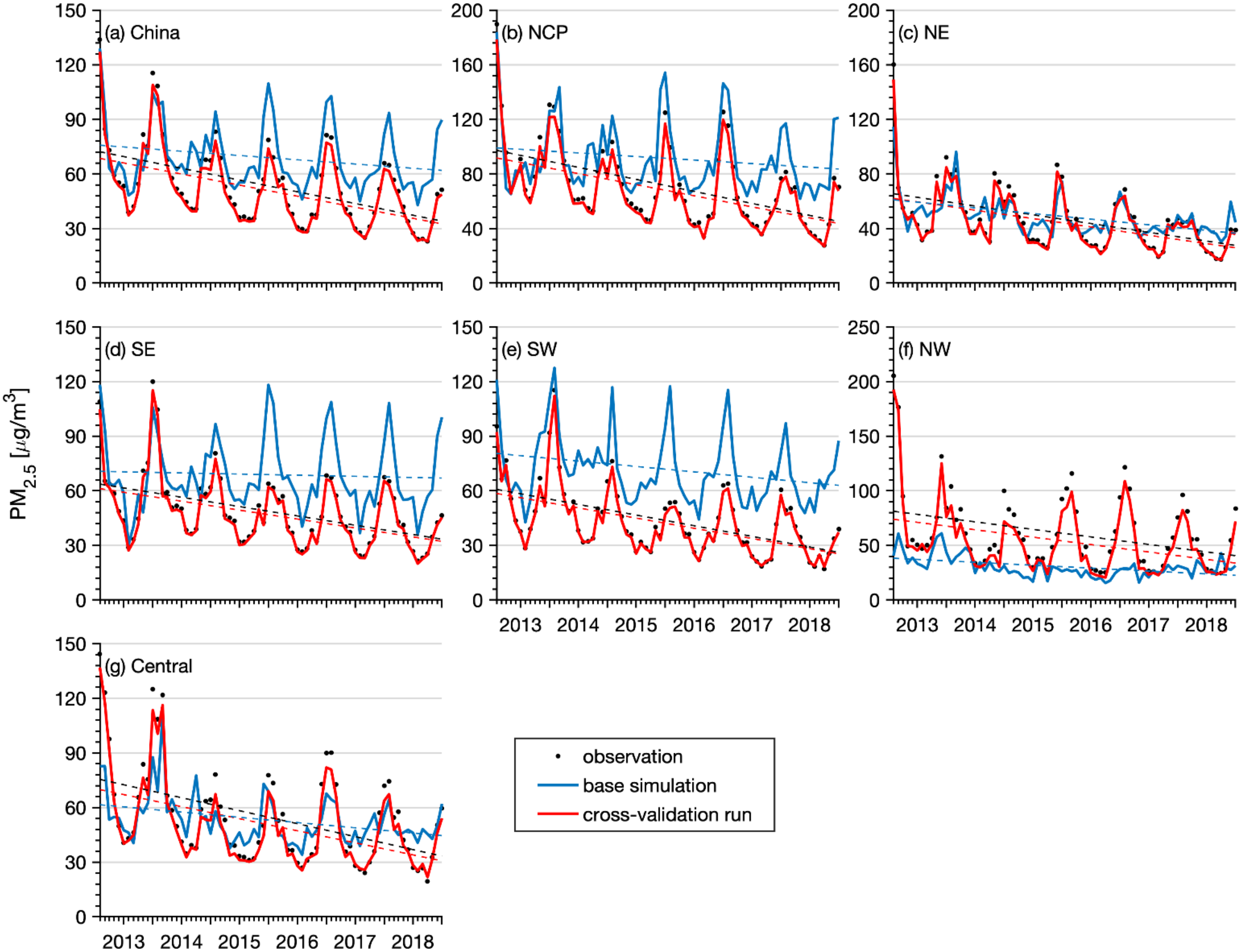
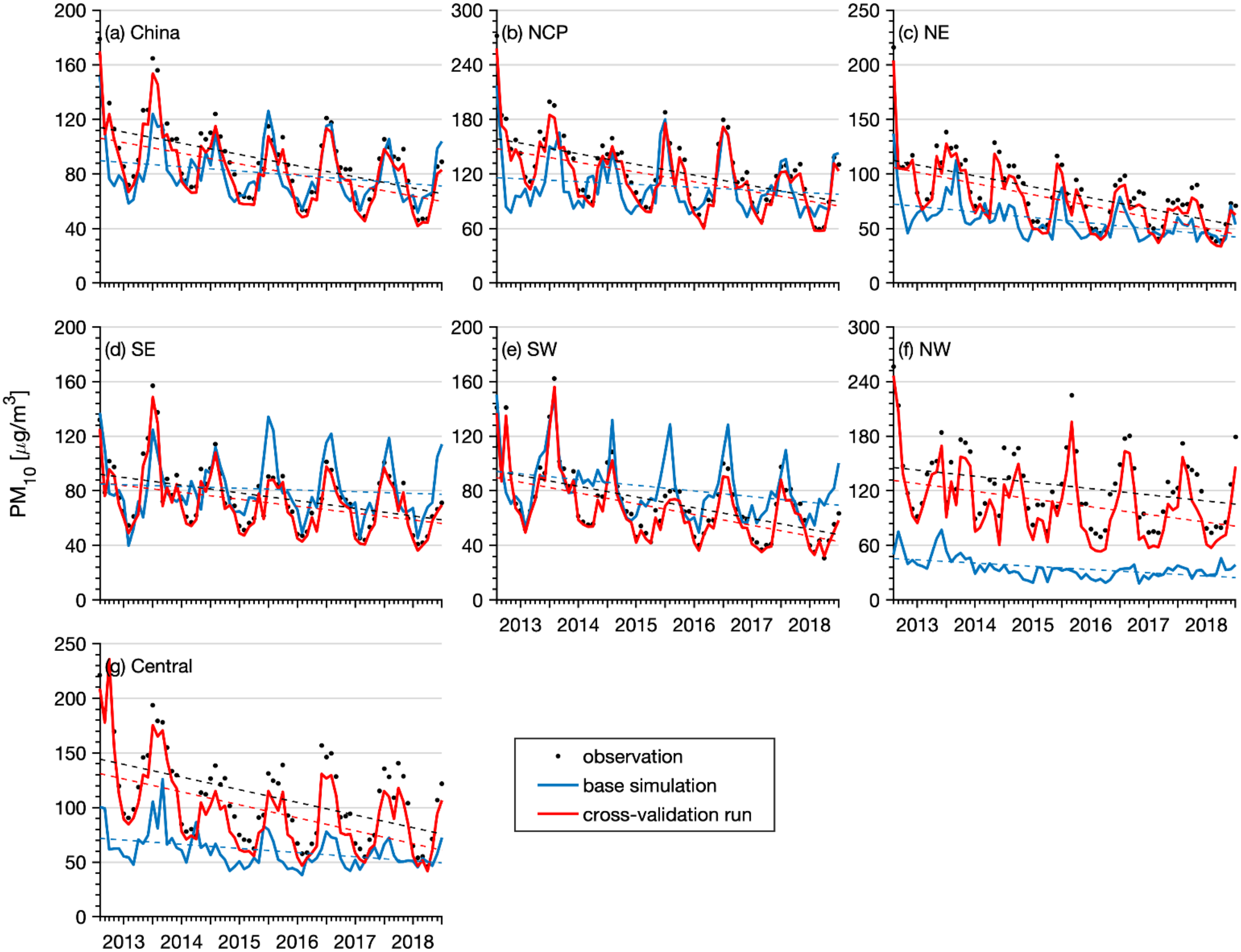


Figure 7: Time series of the monthly mean $PM_{2.5}$ concentrations in (a) China, (b) NCP, (c) NE, (d) SE, (e) SW, (f) NW and (f) central regions obtained from the cross-validation run (red line), base simulation (blue line) and observations (black dots).



817

818 **Figure 8: Same as Fig. 7 but for the PM_{10} concentration.**

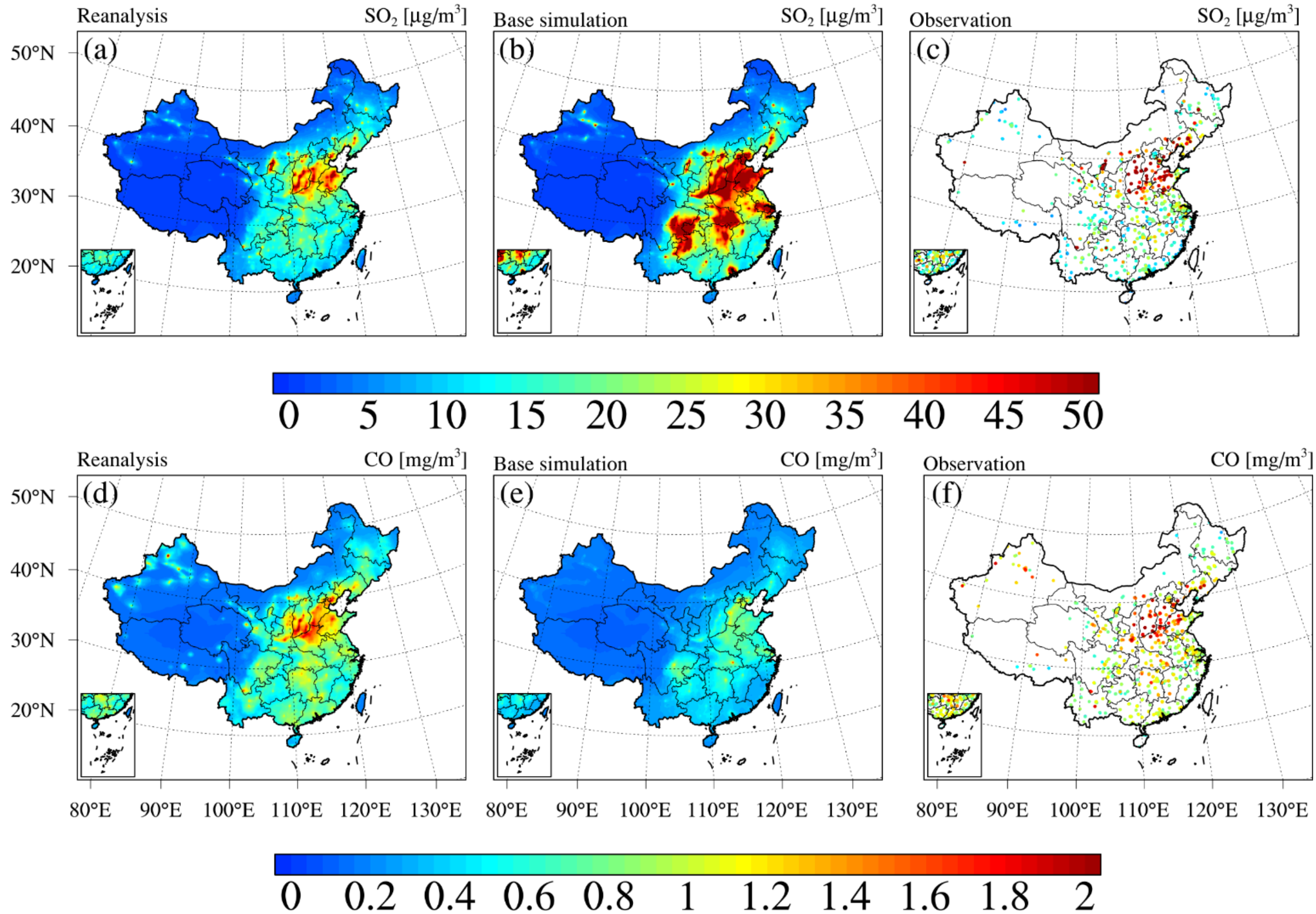


Figure 9: Same as Fig. 5 but for the SO_2 and CO concentrations.

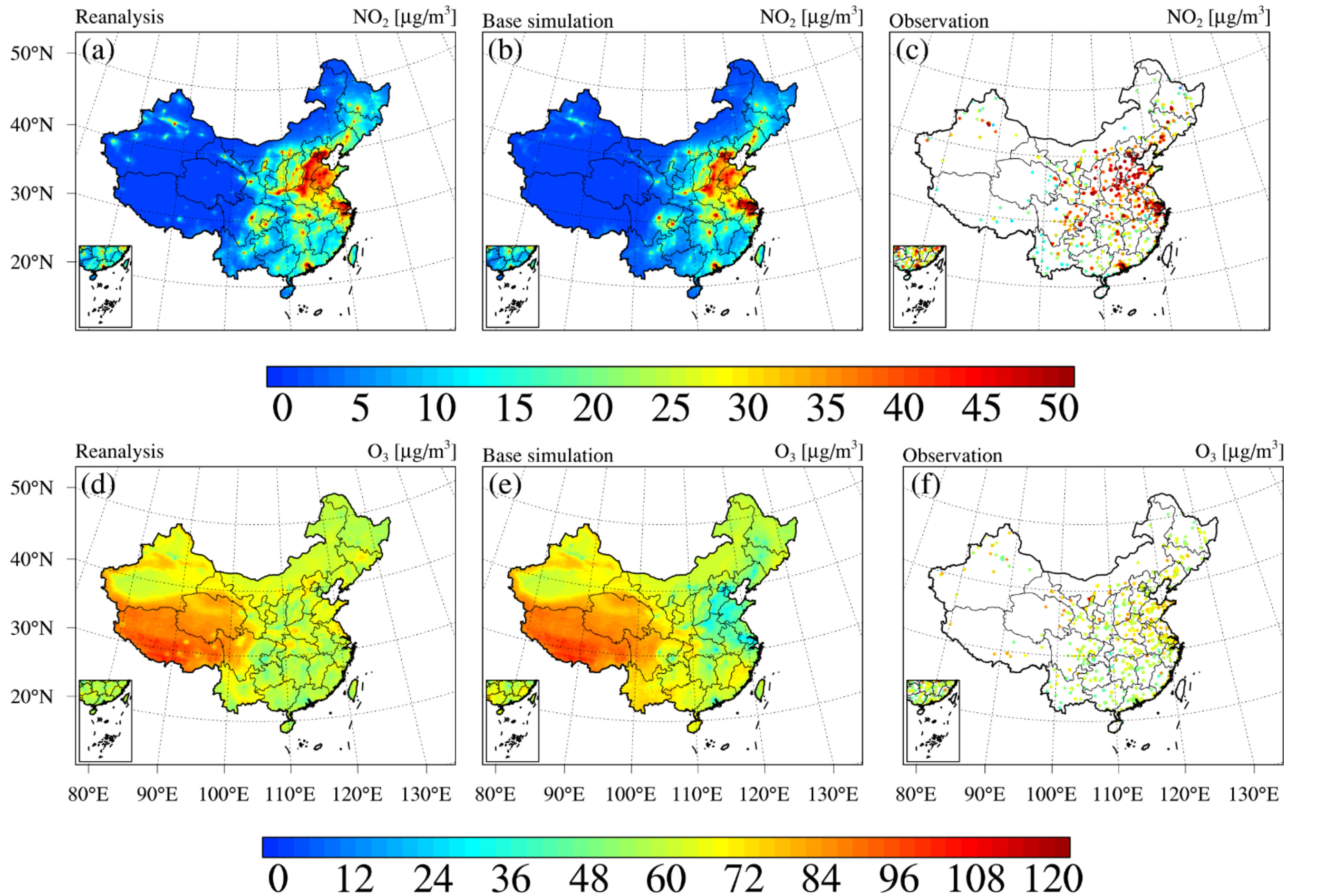


Figure 10: Same as Fig. 5 but for NO_2 and O_3.

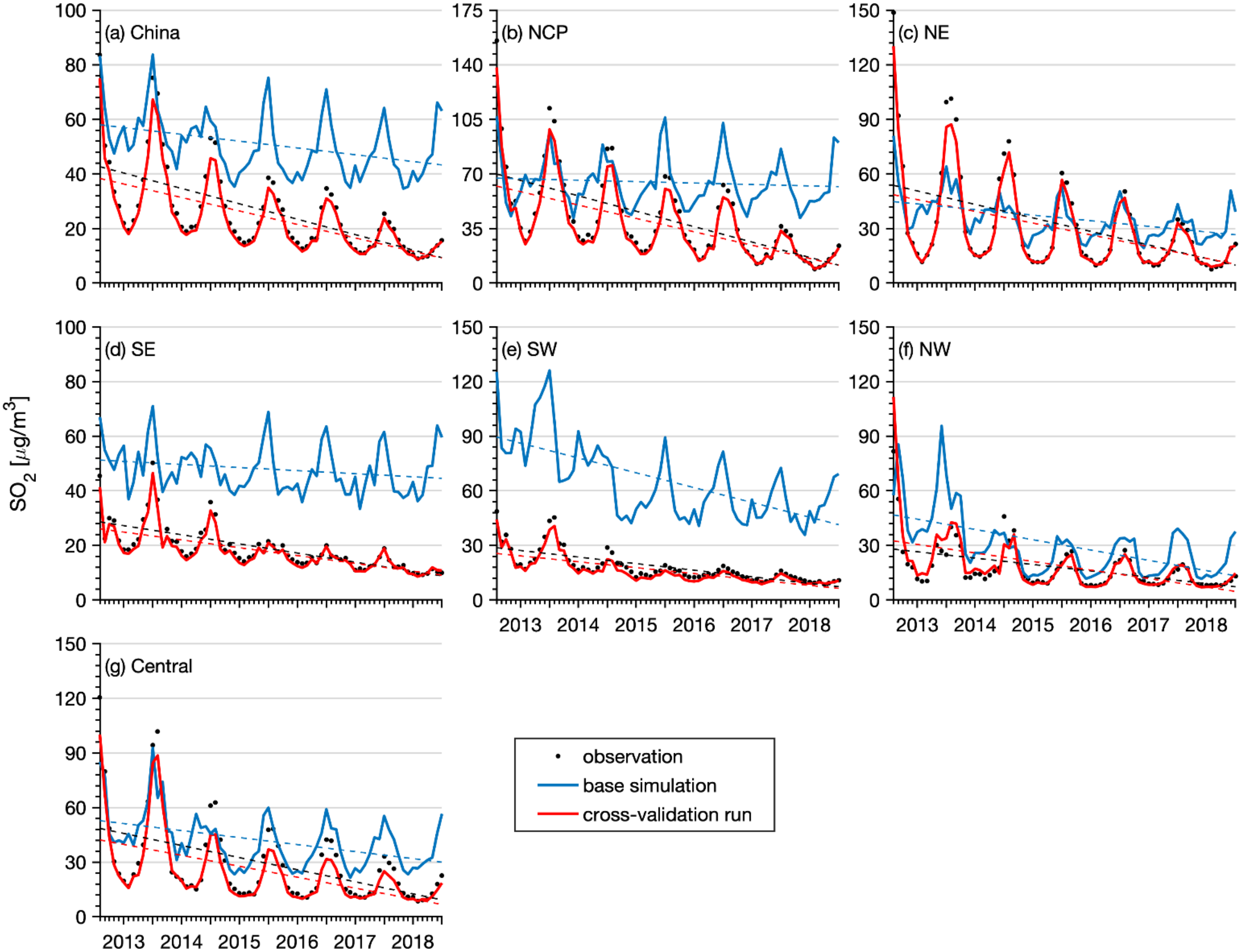


Figure 11: Same as Fig. 7 but for the SO_2 concentration.

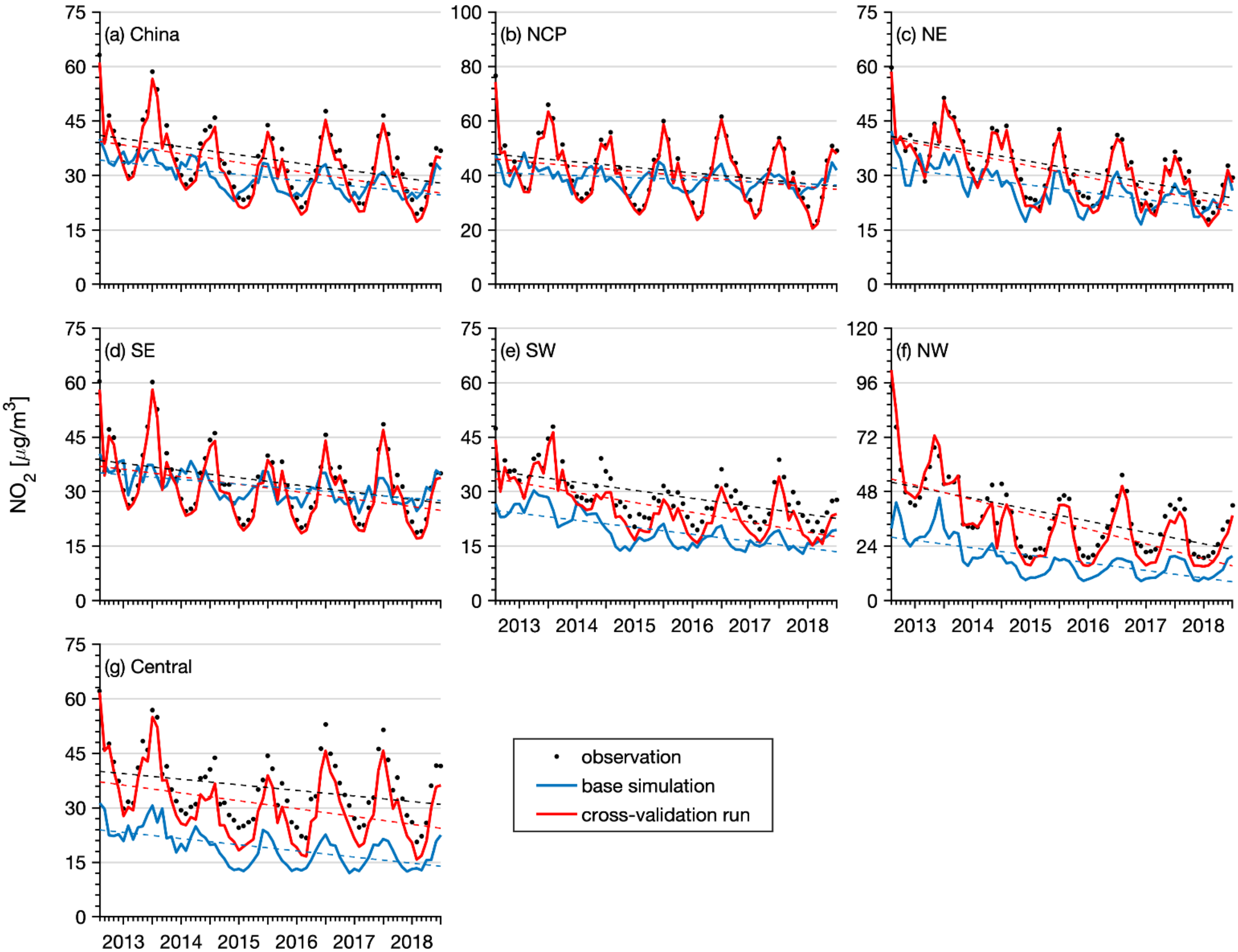


Figure 12: Same as Fig. 7 but for the NO_2 concentration.

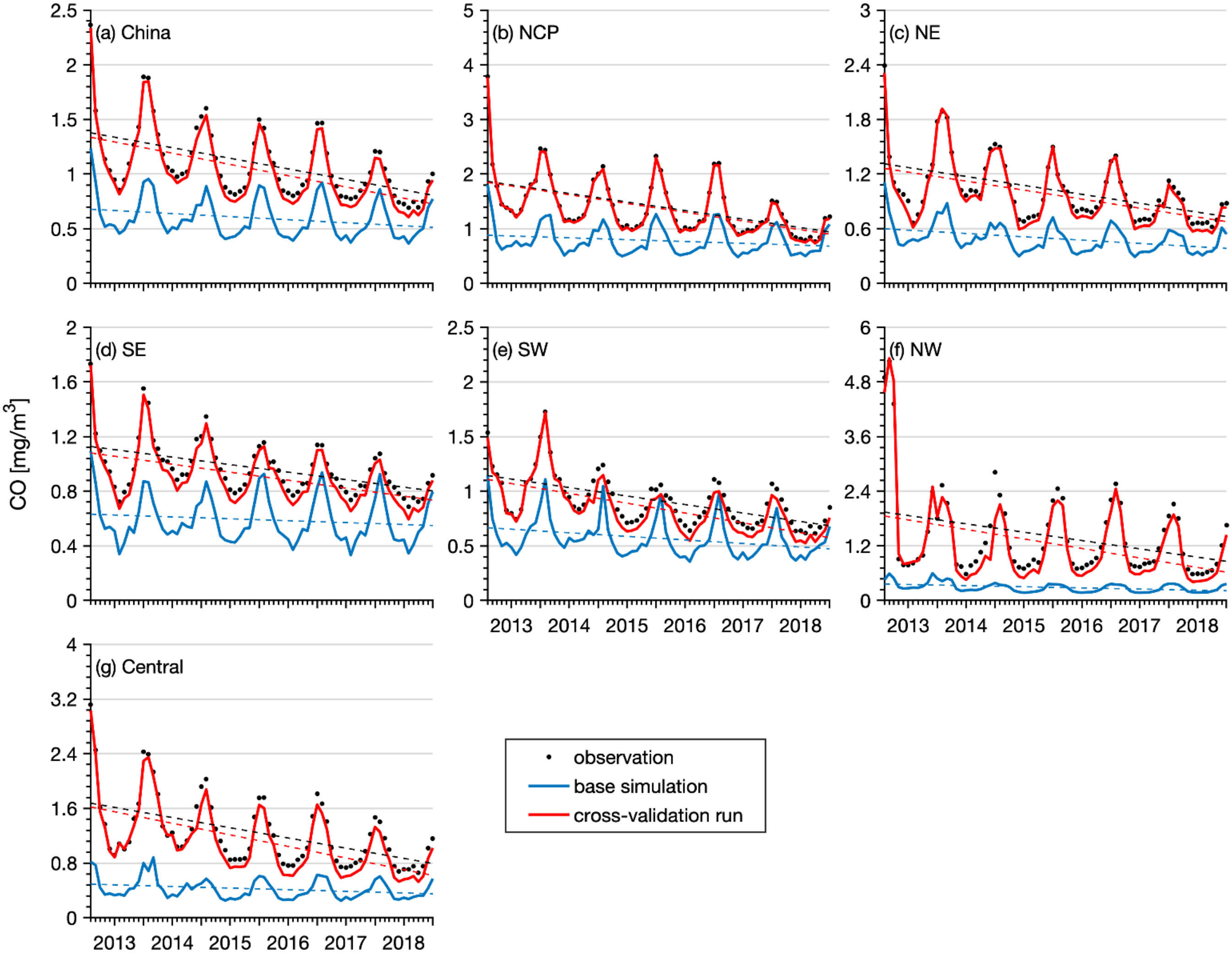


Figure 13: Same as Fig. 7 but for the CO concentration.

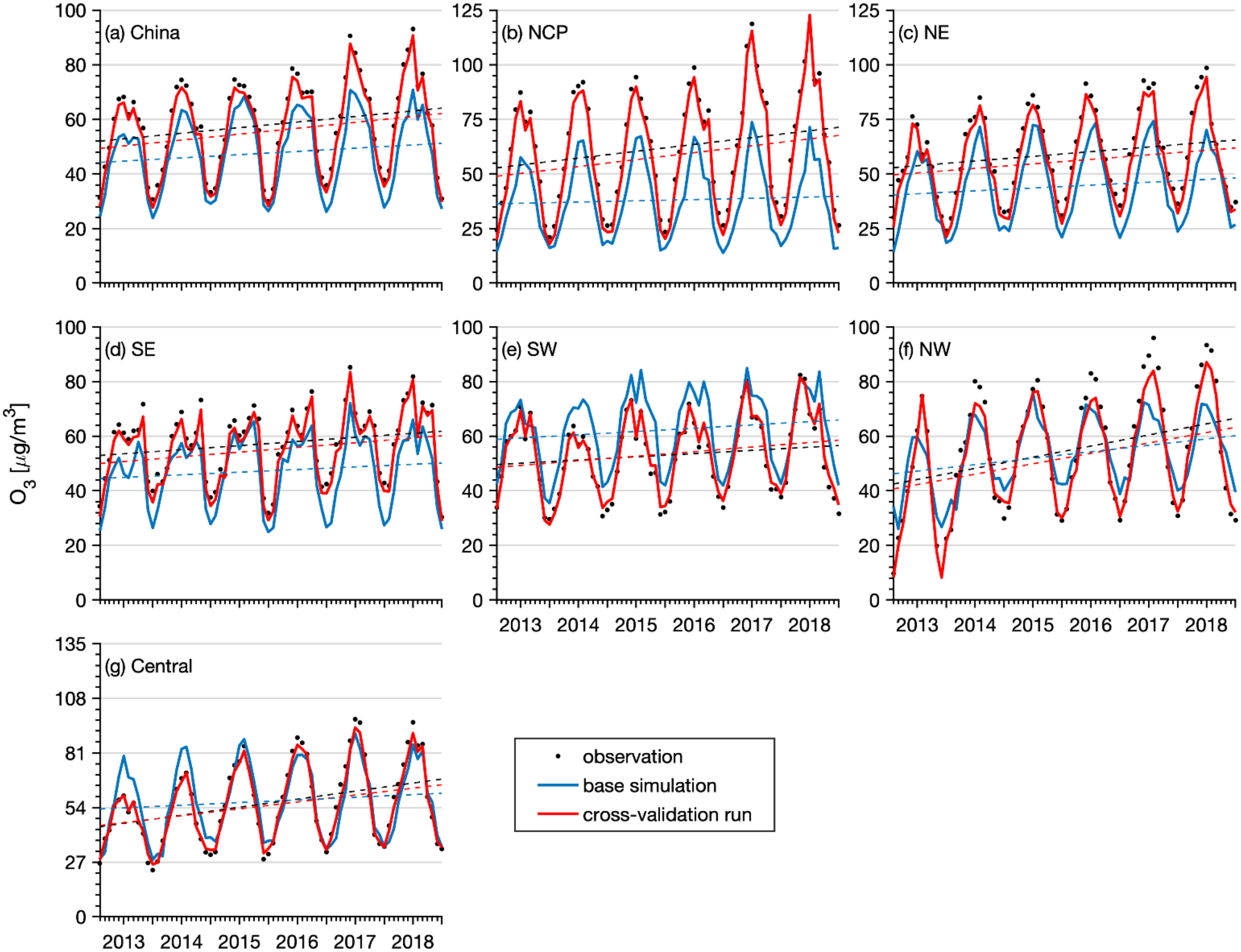


Figure 14: Same as Fig. 7 but for the O_3 concentration.

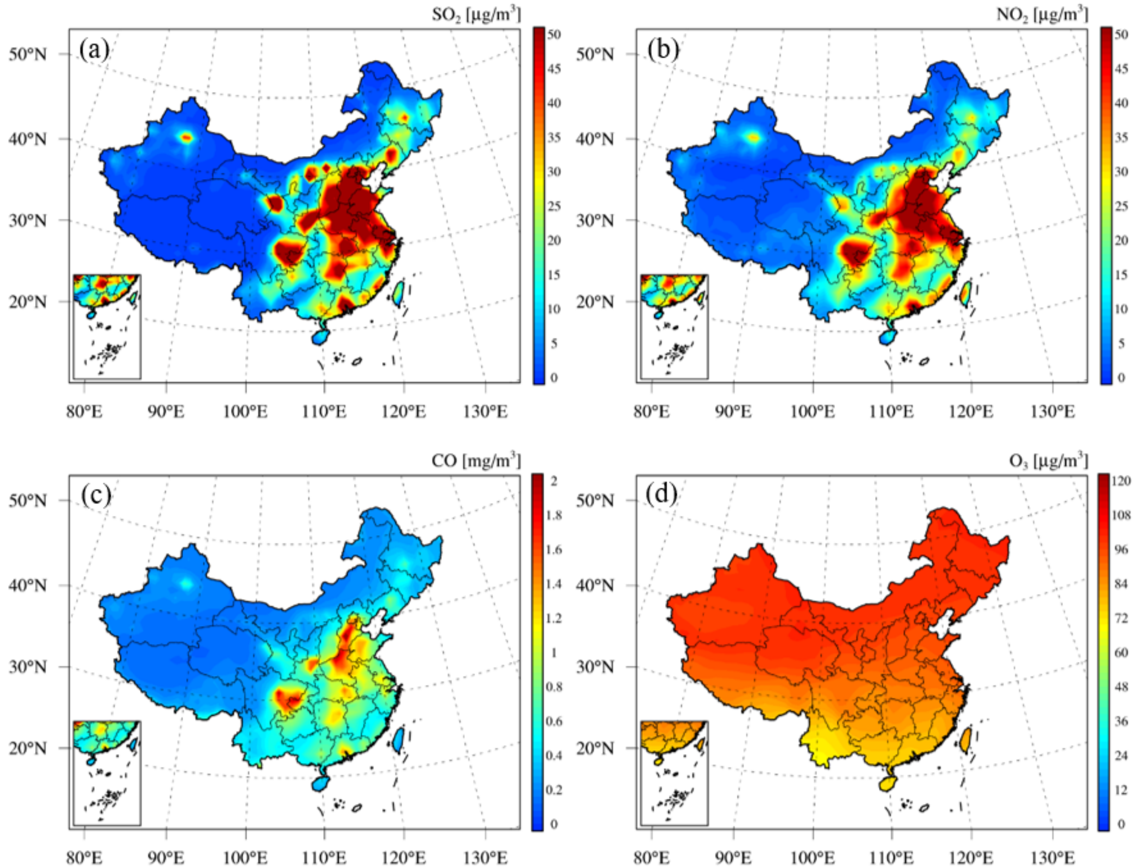


Figure 15: Spatial distributions of the multiyear average concentrations of (a) SO_2, (b) NO_2, (c) CO and (d) O_3 from 2013 to 2018 obtained from CAMSRA.

References

Athanasopoulou, E., Tombrou, M., Pandis, S. N., and Russell, A. G.: The role of sea-salt emissions and heterogeneous chemistry in the air quality of polluted coastal areas, Atmos. Chem. Phys., 8, 5755-5769, https://doi.org/10.5194/acp-8-5755-2008, 2008.

Barnes, W. L., Pagano, T. S., and Salomonson, V. V.: Prelaunch characteristics of the Moderate Resolution Imaging Spectroradiometer (MODIS) on EOS-AM1, IEEE Trans. Geosci. Remote Sensing, 36, 1088-1100, https://doi.org/10.1109/36.700993, 1998.

Brasseur, G. P., Hauglustaine, D. A., Walters, S., Rasch, P. J., Muller, J. F., Granier, C., and Tie, X. X.: MOZART, a global chemical transport model for ozone and related chemical tracers 1. Model description, J. Geophys. Res.-Atmos., 103, 28265-28289, https://doi.org/10.1029/98jd02397, 1998.

Candiani, G., Carnevale, C., Finzi, G., Pisoni, E., and Volta, M.: A comparison of reanalysis techniques: Applying optimal interpolation and Ensemble Kalman Filtering to improve air quality monitoring at mesoscale, Sci. Total Environ., 458, 7-14, https://doi.org/10.1016/j.scitotenv.2013.03.089, 2013.

Carmichael, G., Sakurai, T., Streets, D., Hozumi, Y., Ueda, H., Park, S., Fung, C., Han, Z., Kajino, M., and Engardt, M.: MICS-Asia II: The model intercomparison study for Asia Phase II methodology and overview of findings, Atmos. Environ., 42, 3468-3490, https://doi.org/10.1016/j.atmosenv.2007.04.007, 2008.

Chen, D., Liu, Z., Ban, J., Zhao, P., and Chen, M.: Retrospective analysis of 2015–2017 wintertime $PM_{2.5}$ 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.

Chen, G. B., Li, S. S., Knibbs, L. D., Hamm, N. A. S., Cao, W., Li, T. T., Guo, J. P., Ren, H. Y., Abramson, M. J., and Guo, Y. M.: A machine learning method to estimate $PM_{2.5}$ concentrations across China with remote sensing, meteorological and land use information, Sci. Total Environ., 636, 52-60, https://doi.org/10.1016/j.scitotenv.2018.04.251, 2018.

Chen, Z. Y., Zhang, T. H., Zhang, R., Zhu, Z. M., Yang, J., Chen, P. Y., Ou, C. Q., and Guo, Y. M.: Extreme gradient boosting model to estimate PM2.5 concentrations with missing-filled satellite data in China, Atmos. Environ., 202, 180-189, https://doi.org/10.1016/j.atmosenv.2019.01.027, 2019.

Chu, Y. Y., Liu, Y. S., Li, X. Y., Liu, Z. Y., Lu, H. S., Lu, Y. A., Mao, Z. F., Chen, X., Li, N., Ren, M., Liu, F. F., Tian, L. Q., Zhu, Z. M., and Xiang, H.: A Review on Predicting Ground $PM_{2.5}$ Concentration Using Satellite Aerosol Optical Depth, Atmosphere, 7, 25, https://doi.org/ 10.3390/atmos7100129, 2016.

Cohen, A. J., Brauer, M., Burnett, R., Anderson, H. R., Frostad, J., Estep, K., Balakrishnan, K., Brunekreef, B., Dandona, L., Dandona, R., Feigin, V., Freedman, G., Hubbell, B., Jobling, A., Kan, H., Knibbs, L., Liu, Y., Martin, R., Morawska, L., Pope, C. A., Shin, H., Straif, K., Shaddick, G., Thomas, M., van Dingenen, R., van Donkelaar, A., Vos, T., Murray, C. J. L., and Forouzanfar, M. H.: Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015, Lancet, 389, 1907-1918, https://doi.org/10.1016/s0140-6736(17)30505-6, 2017.

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, 2007.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Holm, E. V., Isaksen, L., Kallberg, P., Kohler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J. J., Park, B. K., Peubey, C., de Rosnay, P., Tavolato, C., Thepaut, J. N., and Vitart, F.: The ERA-Interim reanalysis:

configuration and performance of the data assimilation system, Q. J. R. Meteorol. Soc., 137, 553-597, https://doi.org/10.1002/qj.828, 2011.

Deeter, M. N., Emmons, L. K., Francis, G. L., Edwards, D. P., Gille, J. C., Warner, J. X., Khattatov, B., Ziskin, D., Lamarque, J. F., Ho, S. P., Yudin, V., Attie, J. L., Packman, D., Chen, J., Mao, D., and Drummond, J. R.: Operational carbon monoxide retrieval algorithm and selected results for the MOPITT instrument, J. Geophys. Res.-Atmos., 108, 4399, https://doi.org/10.1029/2002JD003186, 2003.

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 $PM_{2.5}$ observations on $PM_{2.5}$ forecasts over China during wintertime, Atmos. Environ., 187, 34-49, https://doi.org/ 10.1016/j.atmosenv.2018.05.049, 2018.

Flemming, J., Benedetti, A., Inness, A., Engelen, R. J., Jones, L., Huijnen, V., Remy, S., Parrington, M., Suttie, M., Bozzo, A., Peuch, V. H., Akritidis, D., and Katragkou, E.: The CAMS interim Reanalysis of Carbon Monoxide, Ozone and Aerosol for 2003–2015, Atmos. Chem. Phys., 17, 1945-1983, https://doi.org/10.5194/acp-17-1945-2017, 2017.

Gaubert, B., Arellano, A. F., Barre, J., Worden, H. M., Emmons, L. K., Tilmes, S., Buchholz, R. R., Vitt, F., Raeder, K., Collins, N., Anderson, J. L., Wiedinmyer, C., Alonso, S. M., Edwards, D. P., Andreae, M. O., Hannigan, J. W., Petri, C., Strong, K., and Jones, N.: Toward a chemical reanalysis in a coupled chemistry-climate model: An evaluation of MOPITT CO assimilation and its impact on tropospheric composition, J. Geophys. Res.-Atmos., 121, 7310-7343, https://doi.org/10.1002/2016jd024863, 2016.

Granier, C., Lamarque, J., Mieville, A., Muller, J., Olivier, J., Orlando, J., Peters, J., Petron, G., Tyndall, G., and Wallens, S.: POET, a database of surface emissions of ozone precursors. 2005.

Hanna, S. R., Chang, J. C., and Fernau, M. E.: Monte Carlo estimates of uncertainties in predictions by a photochemical grid model (UAM-IV) due to uncertainties in input variables, Atmos. Environ., 32, 3619-3628, https://doi.org/10.1016/s1352-2310(97)00419-6, 1998.

Hauglustaine, D. A., Brasseur, G. P., Walters, S., Rasch, P. J., Muller, J. F., Emmons, L. K., and Carroll, C. A.: MOZART, a global chemical transport model for ozone and related chemical tracers 2. Model results and evaluation, J. Geophys. Res.-Atmos., 103, 28291-28335, https://doi.org/10.1029/98jd02398, 1998.

Hunt, B. R., Kostelich, E. J., and Szunyogh, I.: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter, Physica D, 230, 112-126, https://doi.org/10.1016/j.physd.2006.11.008, 2007.

Inness, A., Ades, M., Agusti-Panareda, A., Barre, J., Benedictow, A., Blechschmidt, A. M., Dominguez, J. J., Engelen, R., Eskes, H., Flemming, J., Huijnen, V., Jones, L., Kipling, Z., Massart, S., Parrington, M., Pench, V. H., Razinger, M., Remy, S., Schulz, M., and Suttie, M.: The CAMS reanalysis of atmospheric composition, Atmos. Chem. Phys., 19, 3515-3556, https://doi.org/10.5194/acp-2018-1078, 2019.

Inness, A., Baier, F., Benedetti, A., Bouarar, I., Chabrillat, S., Clark, H., Clerbaux, C., Coheur, P., Engelen, R. J., Errera, Q., Flemming, J., George, M., Granier, C., Hadji-Lazaro, J., Huijnen, V., Hurtmans, D., Jones, L., Kaiser, J. W., Kapsomenakis, J., Lefever, K., Leitao, J., Razinger, M., Richter, A., Schultz, M. G., Simmons, A. J., Suttie, M., Stein, O., Thepaut, J. N., Thouret, V., Vrekoussis, M., Zerefos, C., and Team, M.: The MACC reanalysis: an 8 yr data set of atmospheric composition, Atmos. Chem. Phys., 13, 4073-4109, https://doi.org/10.5194/acp-13-4073-2013, 2013.

Inness, A., Blechschmidt, A. M., Bouarar, I., Chabrillat, S., Crepulja, M., Engelen, R. J., Eskes, H., Flemming, J., Gaudel, A., Hendrick, F., Huijnen, V., Jones, L., Kapsomenakis, J., Katragkou, E., Keppens, A., Langerock, B., de Maziere, M., Melas, D., Parrington, M., Peuch,

V. H., Razinger, M., Richter, A., Schultz, M. G., Suttie, M., Thouret, V., Vrekoussis, M., Wagner, A., and Zerefos, C.: Data assimilation of satellite-retrieved ozone, carbon monoxide and nitrogen dioxide with ECMWF's Composition-IFS, Atmos. Chem. Phys., 15, 5275-5303, https://doi.org/10.5194/acp-15-5275-2015, 2015.

Janssens-Maenhout, G., Crippa, M., Guizzardi, D., Dentener, F., Muntean, M., Pouliot, G., Keating, T., Zhang, Q., Kurokawa, J., Wankmuller, R., van der Gon, H. D., Kuenen, J. J. P., Klimont, Z., Frost, G., Darras, S., Koffi, B., and Li, M.: HTAP_v2.2: a mosaic of regional and global emission grid maps for 2008 and 2010 to study hemispheric transport of air pollution, Atmos. Chem. Phys., 15, 11411-11432, https://doi.org/10.5194/acp-15-11411-2015, 2015.

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_{10} observations over China using process analysis, J. Geophys. Res.-Atmos., 118, 6738-6749, https://doi.org/10.1002/jgrd.50495, 2013.

Kan, H., Chen, R., and Tong, S.: Ambient air pollution, climate change, and population health in China, Environ. Int., 42, 10-19, https://doi.org/10.1016/j.envint.2011.03.003, 2012.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K., and Takahashi, K.: The JRA-55 Reanalysis: General Specifications and Basic Characteristics, J. Meteorol. Soc. Jpn., 93, 5-48, https://doi.org/10.2151/jmsj.2015-001, 2015.

Kong, L., Tang, X., Zhu, J., Wang, Z., Fu, J. S., Wang, X., Itahashi, S., Yamaji, K., Nagashima, T., Lee, H. J., Kim, C. H., Lin, C. Y., Chen, L., Zhang, M., Tao, Z., Li, J., Kajino, M., Liao, H., Wang, Z., Sudo, K., Wang, Y., Pan, Y., Tang, G., Li, M., Wu, Q., Ge, B., and Carmichael, G. R.: Evaluation and uncertainty investigation of the NO_2, CO and NH_3 modeling over China under the framework of MICS-Asia III, Atmos. Chem. Phys., 20, 181-202, https://doi.org/10.5194/acp-2018-1158, 2020.

Kumar, U., De Ridder, K., Lefebvre, W., and Janssen, S.: Data assimilation of surface air pollutants (O_3 and NO_2) in the regional-scale air quality model AURORA, Atmos. Environ., 60, 99-108, https://doi.org/10.1016/j.atmosenv.2012.06.005, 2012.

Levelt, P. F., Van den Oord, G. H. J., Dobber, M. R., Malkki, A., Visser, H., de Vries, J., Stammes, P., Lundell, J. O. V., and Saari, H.: The Ozone Monitoring Instrument, IEEE Trans. Geosci. Remote Sensing, 44, 1093-1101, https://doi.org/10.1109/TGRS.2006.872333, 2006.

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, 2019.

Li, J., Dong, H. B., Zeng, L. M., Zhang, Y. H., Shao, M., Wang, Z. F., Sun, Y. L., and Fu, P. Q.: Exploring Possible Missing Sinks of Nitrate and Its Precursors in Current Air Quality Models-A Case Simulation in the Pearl River Delta, China, Using an Observation-Based Box Model, Sola, 11, 124-128, https://doi.org/10.2151/sola.2015-029, 2015.

Li, J., Wang, Z., Wang, X., Yamaji, K., Takigawa, M., Kanaya, Y., Pochanart, P., Liu, Y., Irie, H., Hu, B., Tanimoto, H., and Akimoto, H.: Impacts of aerosols on summertime tropospheric photolysis frequencies and photochemistry over Central Eastern China, Atmos. Environ., 45, 1817-1829, https://doi.org/10.1016/j.atmosenv.2011.01.016, 2011.

Li, J., Wang, Z., Zhuang, G., Luo, G., Sun, Y., and Wang, Q.: Mixing of Asian mineral dust with anthropogenic pollutants over East Asia: a model case study of a super-duststorm in March 2010, Atmos. Chem. Phys., 12, 7591-7607, https://doi.org/10.1016/j.atmosenv.2011.01.016, 2012.

Li, K., Jacob, D. J., Liao, H., Shen, L., Zhang, Q., and Bates, K. H.: Anthropogenic drivers of 2013-2017 trends in summer surface ozone in China, Proc. Natl. Acad. Sci. U.S.A., 116, 422-427, https://doi.org/10.1073/pnas.1812168116, 2019.

Li, M., Liu, H., Geng, G. N., Hong, C. P., Liu, F., Song, Y., Tong, D., Zheng, B., Cui, H. Y., Man, H. Y., Zhang, Q., and He, K. B.: Anthropogenic emission inventories in China: a review, Natl. Sci. Rev., 4, 834-866, https://doi.org/10.1093/nsr/nwx150, 2017.

Li, T. W., Shen, H. F., Yuan, Q. Q., Zhang, X. C., and Zhang, L. P.: Estimating Ground-Level $PM_{2.5}$ by Fusing Satellite and Station Observations: A Geo-Intelligent Deep Learning Approach, Geophys. Res. Lett., 44, 11985-11993, https://doi.org/10.1002/2017gl075710, 2017.

Liang, X., Zheng, X. G., Zhang, S. P., Wu, G. C., Dai, Y. J., and Li, Y.: Maximum likelihood estimation of inflation factors on error covariance matrices for ensemble Kalman filter assimilation, Q. J. R. Meteorol. Soc., 138, 263-273, https://doi.org/10.1002/qj.912, 2012.

Lin, C. Q., Li, Y., Yuan, Z. B., Lau, A. K. H., Li, C. C., and Fung, J. C. H.: Using satellite remote sensing data to estimate the high-resolution distribution of ground-level $PM_{2.5}$, Remote Sens. Environ., 156, 117-128, https://doi.org/10.1016/j.rse.2014.09.015, 2015.

Lin, C. Q., Liu, G., Lau, A. K. H., Li, Y., Li, C. C., Fung, J. C. H., and Lao, X. Q.: High-resolution satellite remote sensing of provincial PM2.5 trends in China from 2001 to 2015, Atmos. Environ., 180, 110-116, https://doi.org/10.1016/j.atmosenv.2018.02.045, 2018.

Liu, J. J., Weng, F. Z., and Li, Z. Q.: Satellite-based $PM_{2.5}$ estimation directly from reflectance at the top of the atmosphere using a machine learning algorithm, Atmos. Environ., 208, 113-122, https://doi.org/10.1016/j.atmosenv.2019.04.002, 2019.

Lu, M. M., Tang, X., Wang, Z. F., Gbaguidi, A., Liang, S. W., Hu, K., Wu, L., Wu, H. J., Huang, Z., and Shen, L. J.: Source tagging modeling study of heavy haze episodes under complex regional transport processes over Wuhan megacity, Central China, Environ. Pollut., 231, 612-621, https://doi.org/10.1016/j.envpol.2017.08.046, 2017.

Lu, Z., Streets, D. G., Zhang, Q., Wang, S., Carmichael, G. R., Cheng, Y. F., Wei, C., Chin, M., Diehl, T., and Tan, Q.: Sulfur dioxide emissions in China and sulfur trends in East Asia since 2000, Atmos. Chem. Phys., 10, 6311-6331, https://doi.org/10.5194/acp-10-6311-2010, 2010.

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.

Ma, Z. W., Hu, X. F., Huang, L., Bi, J., and Liu, Y.: Estimating Ground-Level $PM_{2.5}$ in China Using Satellite Remote Sensing, Environ. Sci. Technol., 48, 7436-7444, https://doi.org/10.1021/es5009399, 2014.

Ma, Z. W., Hu, X. F., Sayer, A. M., Levy, R., Zhang, Q., Xue, Y. G., Tong, S. L., Bi, J., Huang, L., and Liu, Y.: Satellite-Based Spatiotemporal Trends in PM2.5 Concentrations: China, 2004-2013, Environ. Health Perspect., 124, 184-192, https://doi.org/10.1289/ehp.1409481, 2016.

Menard, R. and Changs, 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, 2000.

Miyazaki, K., Bowman, K., Sekiya, T., Eskes, H., Boersma, F., Worden, H., Livesey, N., Payne, V. H., Sudo, K., Kanaya, Y., Takigawa, M., and Ogochi, K.: Updated tropospheric chemistry reanalysis and emission estimates, TCR-2, for 2005–2018, Earth Syst. Sci. Data., 12, 2223-2259, https://doi.org/10.5194/essd-12-2223-2020, 2020.

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_2, O_3, CO, and HNO_3 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.

NBSC, China energy statistical Yearbook, 2017a, http://tongji.cnki.net/kns55/Navi/HomePage.aspx?id=N2017110016&name=YCXME&floor=1, Accessed date: 17 April 2020.

NBSC, China statistical Yearbook on environment, 2017b, http://www.stats.gov.cn/ztjc/ztsj/hjtjzl/, Accessed date: 17 April 2020.

Nenes, A., Pandis, S. N., and Pilinis, C.: ISORROPIA: A new thermodynamic equilibrium model for multiphase multicomponent inorganic aerosols, Aquat. Geochem., 4, 123-152, https://doi.org/10.1023/a:1009604003981, 1998.

Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E. J., Corazza, M., Kalnay, E., Patil, D. J., and Yorke, J. A.: A local ensemble Kalman filter for atmospheric data assimilation, Tellus Ser. A-Dyn. Meteorol. Oceanol., 56, 415-428, https://doi.org/10.1111/j.1600-0870.2004.00076.x, 2004.

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 PM2. 5 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.

Price, C., Penner, J., and Prather, M.: NO_x from lightning .1. Global distribution based on lightning physics, J. Geophys. Res.-Atmos., 102, 5929-5941, https://doi.org/10.1029/96jd03504, 1997.

Qi, J., Zheng, B., Li, M., Yu, F., Chen, C. C., Liu, F., Zhou, X. F., Yuan, J., Zhang, Q., and He, K. B.: A high-resolution air pollutants emission inventory in 2013 for the Beijing-Tianjin-Hebei region, China, Atmos. Environ., 170, 156-168, https://doi.org/10.1016/j.atmosenv.2017.09.039, 2017.

Randerson, J. T., Van Der Werf, G. R., Giglio, L., Collatz, G. J., and Kasibhatla, P. S.: Global Fire Emissions Database, Version 4.1 (GFEDv4). ORNL Distributed Active Archive Center, 2017.

Randles, C. A., da Silva, A. M., Buchard, V., Colarco, P. R., Darmenov, A., Govindaraju, R., Smirnov, A., Holben, B., Ferrare, R., Hair, J., Shinozuka, Y., and Flynn, C. J.: The MERRA-2 Aerosol Reanalysis, 1980 Onward. Part I: System Description and Data Assimilation Evaluation, J. Clim., 30, 6823-6850, https://doi.org/10.1175/jcli-d-16-0609.1, 2017.

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G. K., Bloom, S., Chen, J. Y., Collins, D., Conaty, A., Da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C. R., Reichle, R., Robertson, F. R., Ruddick, A. G., Sienkiewicz, M., and Woollen, J.: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications, J. Clim., 24, 3624-3648, https://doi.org/10.1175/jcli-d-11-00015.1, 2011.

Saha, S., Moorthi, S., Pan, H. L., Wu, X. R., Wang, J. D., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H. X., Stokes, D., Grumbine, R., Gayno, G., Wang, J., Hou, Y. T., Chuang, H. Y., Juang, H. M. H., Sela, J., Iredell, M., Treadon, R., Kleist, D., Van Delst, P., Keyser, D., Derber, J., Ek, M., Meng, J., Wei, H. L., Yang, R. Q., Lord, S., Van den Dool, H., Kumar, A., Wang, W. Q., Long, C., Chelliah, M., Xue, Y., Huang, B. Y., Schemm, J. K., Ebisuzaki, W., Lin, R., Xie, P. P., Chen, M. Y., Zhou, S. T., Higgins, W., Zou, C. Z., Liu, Q. H., Chen, Y., Han, Y., Cucurull, L., Reynolds, R. W., Rutledge, G., and Goldberg, M.: THE NCEP CLIMATE

FORECAST SYSTEM REANALYSIS, Bulletin of the American Meteorological Society, 91, 1015-1057, https://doi.org/10.1175/2010BAMS3001.1, 2010.

Sakov, P. and Bertino, L.: Relation between two common localisation methods for the EnKF, Comput. Geosci., 15, 225-237, https://doi.org/10.1007/s10596-010-9202-6, 2011.

Shin, M., Kang, Y., Park, S., Im, J., Yoo, C., and Quackenbush, L. J.: Estimating ground-level particulate matter concentrations using satellite-based data: a review, GISci. Remote Sens., https://doi.org/10.1080/15481603.2019.1703288, 2019. 1-16, 2019.

Sillman, S.: The relation between ozone, NO_x and hydrocarbons in urban and polluted rural environments, Atmos. Environ., 33, 1821-1845, https://doi.org/10.1016/s1352-2310(98)00345-8, 1999.

Silver, B., Reddington, C. L., Arnold, S. R., and Spracklen, D. V.: Substantial changes in air pollution across China during 2015-2017, Environ. Res. Lett., 13, 8, https://doi.org/10.1088/1748-9326/aae718, 2018.

Sindelarova, K., Granier, C., Bouarar, I., Guenther, A., Tilmes, S., Stavrakou, T., Muller, J. F., Kuhn, U., Stefani, P., and Knorr, W.: Global data set of biogenic VOC emissions calculated by the MEGAN model over the last 30 years, Atmos. Chem. Phys., 14, 9317-9341, https://doi.org/10.5194/acp-14-9317-2014, 2014.

Skamarock, W. C.: A description of the advanced research WRF version 3, Ncar Technical, 113, 7--25, 2008.

Streets, D. G., Bond, T. C., Carmichael, G. R., Fernandes, S. D., Fu, Q., He, D., Klimont, Z., Nelson, S. M., Tsai, N. Y., Wang, M. Q., Woo, J. H., and Yarber, K. F.: An inventory of gaseous and primary aerosol emissions in Asia in the year 2000, J. Geophys. Res.-Atmos., 108, 23, https://doi.org/10.1029/2002jd003093, 2003.

Tang, X., Kong, L., Zhu, J., Wang, Z. F. , Li, J. J., Wu, H. J., Wu, Q. Z., Chen, H. S., Zhu, L. L., Wang, W., Liu, B., Wang, Q., Chen D. H., Pan Y. P., Song, T., Li, F., Zheng, H. T., Jia, G. L., Lu, M. M., Wu, L., Carmichael, G. R.: A Six-year long High-resolution Air Quality Reanalysis Dataset over China from 2013 to 2018. V2. Science Data Bank, https://doi.org/10.11922/sciencedb.00053, 2020a.

Tang, X., Kong, L., Zhu, J., Wang, Z. F. , Li, J. J., Wu, H. J., Wu, Q. Z., Chen, H. S., Zhu, L. L., Wang, W., Liu, B., Wang, Q., Chen D. H., Pan Y. P., Song, T., Li, F., Zheng, H. T., Jia, G. L., Lu, M. M., Wu, L., Carmichael, G. R.: A Six-year long High-resolution Air Quality Reanalysis Dataset over China from 2013 to 2018 (monthly and annual version). V1. Science Data Bank, https://doi.org/10.11922/sciencedb.00092, 2020b.

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_x emissions: mixed effects on NO_2 forecasts over Beijing and surrounding areas, Atmos. Chem. Phys., 16, 6395-6405, https://doi.org/10.5194/acp-16-6395-2016, 2016.

van der A, R. J., Allaart, M. A. F., and Eskes, H. J.: Extended and refined multi sensor reanalysis of total ozone for the period 1970-2012, Atmos. Meas. Tech., 8, 3021-3035, https://doi.org/10.5194/amt-8-3021-2015, 2015.

van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Mu, M., Kasibhatla, P. S., Morton, D. C., DeFries, R. S., Jin, Y., and van Leeuwen, T. T.: Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009), Atmos. Chem. Phys., 10, 11707-11735, https://doi.org/10.5194/acp-10-11707-2010, 2010.

van Donkelaar, A., Martin, R. V., Brauer, M., Hsu, N. C., Kahn, R. A., Levy, R. C., Lyapustin, A., Sayer, A. M., and Winker, D. M.: Global Estimates of Fine Particulate Matter using a Combined Geophysical-Statistical Method with Information from Satellites, Models, and Monitors, Environ. Sci. Technol., 50, 3762-3772, https://doi.org/10.1021/acs.est.5b05833, 2016.

van Donkelaar, A., Martin, R. V., Brauer, M., Kahn, R., Levy, R., Verduzco, C., and Villeneuve, P. J.: Global Estimates of Ambient Fine Particulate Matter Concentrations from Satellite-Based Aerosol Optical Depth: Development and Application, Environ. Health Perspect., 118, 847-855, https://doi.org/10.1289/ehp.0901623, 2010.

von Schneidemesser, E., Monks, P. S., Allan, J. D., Bruhwiler, L., Forster, P., Fowler, D., Lauer, A., Morgan, W. T., Paasonen, P., Righi, M., Sindelarova, K., and Sutton, M. A.: Chemistry and the Linkages between Air Quality and Climate Change, Chem. Rev., 115, 3856-3897, https://doi.org/10.1021/acs.chemrev.5b00089, 2015.

Walcek, C. J. and Aleksic, N. M.: A simple but accurate mass conservative, peak-preserving, mixing ratio bounded advection algorithm with FORTRAN code, Atmos. Environ., 32, 3863-3880, https://doi.org/10.1016/S1352-2310(98)00099-5, 1998.

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, 2003.

Wang, Z. F., Sha, W. M., and Ueda, H.: Numerical modeling of pollutant transport and chemistry during a high-ozone event in northern Taiwan, Tellus Ser. B-Chem. Phys. Meteorol., 52, 1189-1205, https://doi.org/10.1034/j.1600-0889.2000.01064.x, 2000.

Werner, M., Kryza, M., Pagowski, M., and Guzikowski, J.: Assimilation of $PM_{2.5}$ 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.

Wesely, M. L.: Parameterization of surface resistances to gaseous dry deposition in regional-scale numerical models, Atmos. Environ., 23, 1293-1304, https://doi.org/10.1016/0004-6981(89)90153-4 1989.

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.

Xue, T., Zheng, Y., Tong, D., Zheng, B., Li, X., Zhu, T., and Zhang, Q.: Spatiotemporal continuous estimates of PM2.5 concentrations in China, 2000–2016: A machine learning method with inputs from satellites, chemical transport model, and ground observations, Environ. Int., 123, 345-357, https://doi.org/10.1016/j.envint.2018.11.075, 2019.

Xue, T., Zheng, Y. X., Geng, G. N., Zheng, B., Jiang, X. J., Zhang, Q., and He, K. B.: Fusing Observational, Satellite Remote Sensing and Air Quality Model Simulated Data to Estimate Spatiotemporal Variations of $PM_{2.5}$ Exposure in China, Remote Sens., 9, 19, https://doi.org/10.3390/rs9030221, 2017.

Yan, X. Y., Akimoto, H., and Ohara, T.: Estimation of nitrous oxide, nitric oxide and ammonia emissions from croplands in East, Southeast and South Asia, Glob. Change Biol., 9, 1080-1096, https://doi.org/10.1046/j.1365-2486.2003.00649.x, 2003.

Yao, F., Wu, J., Li, W., and Peng, J.: A spatially structured adaptive two-stage model for retrieving ground-level $PM_{2.5}$ concentrations from VIIRS AOD in China, ISPRS Journal of Photogrammetry and Remote Sensing, 151, 263-276, https://doi.org/10.1016/j.isprsjprs.2019.03.011, 2019.

You, W., Zang, Z. L., Zhang, L. F., Li, Y., Pan, X. B., and Wang, W. Q.: National-Scale Estimates of Ground-Level $PM_{2.5}$ Concentration in China Using Geographically Weighted Regression Based on 3 km Resolution MODIS AOD, Remote Sens., 8, 13, https://doi.org/10.3390/rs8030184, 2016.

Yumimoto, K., Tanaka, T. Y., Oshima, N., and Maki, T.: JRAero: the Japanese Reanalysis for Aerosol v1.0, Geosci. Model Dev., 10, 3225-3253, https://doi.org/10.5194/gmd-10-3225-2017, 2017.

Zaveri, R. A. and Peters, L. K.: A new lumped structure photochemical mechanism for large-scale applications, J. Geophys. Res.-Atmos., 104, 30387-30415, https://doi.org/10.1029/1999jd900876, 1999.

Zhan, Y., Luo, Y. Z., Deng, X. F., Chen, H. J., Grieneisen, M. L., Shen, X. Y., Zhu, L. Z., and Zhang, M. H.: Spatiotemporal prediction of continuous daily PM2.5 concentrations across China using a spatially explicit machine learning algorithm, Atmos. Environ., 155, 129-139, https://doi.org/10.1016/j.atmosenv.2017.02.023, 2017.

Zhan, Y., Luo, Y. Z., Deng, X. F., Zhang, K. S., Zhang, M. H., Grieneisen, M. L., and Di, B. F.: Satellite-Based Estimates of Daily NO2 Exposure in China Using Hybrid Random Forest and Spatiotemporal Kriging Model, Environ. Sci. Technol., 52, 4180-4189, https://doi.org/10.1021/acs.est.7b05669, 2018.

Zhang, H. Y., Di, B. F., Liu, D. R., Li, J. R., and Zhan, Y.: Spatiotemporal distributions of ambient SO_2 across China based on satellite retrievals and ground observations: Substantial decrease in human exposure during 2013-2016, Environ. Res., 179, 9, https://doi.org/10.1016/j.envres.2019.108795, 2019.

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.

Zheng, B., Chevallier, F., Ciais, P., Yin, Y., Deeter, M. N., Worden, H. M., Wang, Y. L., Zhang, Q., and He, K. B.: Rapid decline in carbon monoxide emissions and export from East Asia between years 2005 and 2016, Environ. Res. Lett., 13, 9, https://doi.org/10.1088/1748-9326/aab2b3, 2018a.

Zheng, B., Chevallier, F., Yin, Y., Ciais, P., Fortems-Cheiney, A., Deeter, M. N., Parker, R. J., Wang, Y. L., Worden, H. M., and Zhao, Y. H.: Global atmospheric carbon monoxide budget 2000-2017 inferred from multi-species atmospheric inversions, Earth Syst. Sci. Data, 11, 1411-1436, https://doi.org/10.5194/essd-11-1411-2019, 2019.

Zheng, B., Tong, D., Li, M., Liu, F., Hong, C. P., Geng, G. N., Li, H. Y., Li, X., Peng, L. Q., Qi, J., Yan, L., Zhang, Y. X., Zhao, H. Y., Zheng, Y. X., He, K. B., and Zhang, Q.: Trends in China's anthropogenic emissions since 2010 as the consequence of clean air actions, Atmos. Chem. Phys., 18, 14095-14111, https://doi.org/10.5194/acp-18-14095-2018, 2018b.

Zheng, Y. X., Xue, T., Zhang, Q., Geng, G. N., Tong, D., Li, X., and He, K. B.: Air quality improvements and health benefits from China's clean air action since 2013, Environ. Res. Lett., 12, 9, https://doi.org/10.1088/1748-9326/aa8a32, 2017.

Zou, B., Chen, J. W., Zhai, L., Fang, X., and Zheng, Z.: Satellite Based Mapping of Ground $PM_{2.5}$ Concentration Using Generalized Additive Modeling, Remote Sens., 9, 16, https://doi.org/10.3390/rs9010001, 2017.

Supplementary

Tables

Table S1: CV results of the reanalysis (outside bracket) and base simulation (in bracket) for $PM_{2.5}$ concentrations in different regions of China at different temporal scales

$PM_{2.5}$ ($\mu g/m^3$)	NCP				NE			
	R^2	MBE	NMB(%)	RMSE	R^2	MBE	NMB (%)	RMSE
Hourly	0.85 (0.33)	-3.3 (22.4)	-4.8 (32.8)	25.1 (62.6)	0.77 (0.25)	-2.6 (2.8)	-5.8 (6.5)	22.6 (44.5)
Daily	0.90 (0.44)	-3.4 (22.3)	-4.9 (32.4)	17.5 (51.2)	0.86 (0.32)	-2.6 (2.6)	-5.9 (6.0)	14.7 (35.1)
Monthly	0.92 (0.56)	-3.4 (22.2)	-4.9 (32.4)	11.4 (34.1)	0.86 (0.38)	-2.6 (2.7)	-5.9 (6.0)	9.7 (21.4)
Yearly	0.92 (0.56)	-3.6 (20.8)	-5.0 (29.2)	8.7 (27.3)	0.79 (0.35)	-3.1 (0.4)	-6.6 (0.8)	8.8 (16.7)
	SE				SW			
	R^2	MBE	NMB(%)	RMSE	R^2	MBE	NMB (%)	RMSE
Hourly	0.85 (0.25)	-1.8 (22.2)	-3.8 (47.6)	14.9 (51.5)	0.79 (0.22)	-1.4 (30.3)	-3.4 (74.7)	16.5 (57.4)
Daily	0.90 (0.31)	-1.8 (22.2)	-3.8 (47.4)	10.6 (45.4)	0.86 (0.29)	-1.4 (30.0)	-3.4 (74.2)	12.1 (51.6)
Monthly	0.92 (0.45)	-1.8 (22.1)	-3.8 (47.2)	7.4 (33.7)	0.86 (0.49)	-1.5 (29.8)	-3.7 (73.3)	9.7 (42.8)
Yearly	0.90 (0.37)	-2.0 (20.5)	-4.0 (42.0)	6.1 (29.3)	0.79 (0.47)	-2.2 (27.2)	-5.0 (63.2)	9.5 (38.8)
	NW				Central			
	R^2	MBE	NMB(%)	RMSE	R^2	MBE	NMB (%)	RMSE
Hourly	0.52 (0.11)	-7.3 (-28.7)	-13.1 (-51.1)	52.1 (73.0)	0.72 (0.23)	-4.1 (0.8)	-8.2 (1.6)	26.6 (47.5)
Daily	0.66 (0.15)	-7.5 (-29.0)	-13.2 (-51.3)	39.4 (66.0)	0.83 (0.30)	-4.2 (0.7)	-8.3 (1.4)	19.1 (39.9)
Monthly	0.72 (0.28)	-7.4 (-28.9)	-13.1 (-51.3)	26.9 (50.3)	0.85 (0.42)	-4.2 (0.7)	-8.2 (1.4)	13.1 (26.1)
Yearly	0.64 (0.40)	-9.8 (-33.5)	-16.1 (-54.9)	23.5 (43.1)	0.77 (0.31)	-5.4 (-3.6)	-10.1 (-6.7)	12.5 (24.3)

Table S2: CV results of the reanalysis (outside bracket) and base simulation (in bracket) for PM_{10} concentrations in different regions of China at different temporal scales

PM_{10} ($\mu g/m^3$)	NCP				NE			
	R^2	MBE	NMB(%)	RMSE	R^2	MBE	NMB (%)	RMSE
Hourly	0.79 (0.23)	-7.7 (-14.6)	-6.4 (-12.1)	43.7 (88.3)	0.71 (0.18)	-7.6 (-23.6)	-9.6 (-29.8)	39.8 (70.8)
Daily	0.86 (0.31)	-7.6 (-14.2)	-6.3 (-11.7)	30.9 (71.8)	0.79 (0.25)	-7.6 (-23.6)	-9.7 (-30.0)	27.1 (56.8)
Monthly	0.86 (0.38)	-7.6 (-14.2)	-6.3 (-11.8)	21.4 (44.9)	0.76 (0.29)	-7.7 (-23.6)	-9.8 (-30.0)	19.4 (39.6)
Yearly	0.85 (0.46)	-7.6 (-15.8)	-6.2 (-12.8)	17.6 (33.0)	0.67 (0.31)	-8.3 (-26.5)	-10.3 (-32.6)	18.4 (36.2)
	SE				SW			
	R^2	MBE	NMB(%)	RMSE	R^2	MBE	NMB (%)	RMSE
Hourly	0.77 (0.18)	-4.4 (6.9)	-5.9 (9.4)	26.0 (61.2)	0.69 (0.15)	-5.1 (13.0)	-7.5 (19.1)	30.2 (66.2)
Daily	0.85 (0.23)	-4.1 (8.1)	-5.6 (11.1)	18.6 (52.0)	0.77 (0.21)	-5.0 (13.1)	-7.4 (19.6)	22.4 (56.5)
Monthly	0.85 (0.38)	-4.2 (7.5)	-5.7 (10.2)	13.7 (33.3)	0.76 (0.38)	-5.2 (12.5)	-7.8 (18.5)	18.7 (41.4)
Yearly	0.81 (0.36)	-4.7 (4.9)	-6.1 (6.5)	12.3 (26.3)	0.62 (0.38)	-6.8 (8.7)	-9.6 (12.2)	19.3 (35.7)
	NW				Central			
	R^2	MBE	NMB(%)	RMSE	R^2	MBE	NMB (%)	RMSE
Hourly	0.46 (0.08)	-21.5 (-88.5)	-18.0 (-74.1)	105.5 (150.2)	0.61 (0.11)	-14.6 (-45.6)	-14.1 (-43.9)	57.3 (96.4)
Daily	0.56 (0.11)	-21.5 (-89.3)	-17.9 (-74.1)	85.5 (141.6)	0.72 (0.14)	-14.6 (-45.5)	-14.1 (-43.8)	42.1 (84.6)
Monthly	0.59 (0.17)	-20.8 (-89.5)	-17.2 (-74.0)	64.0 (118.9)	0.74 (0.28)	-14.6 (-45.3)	-14.1 (-43.8)	30.2 (62.5)
Yearly	0.58 (0.23)	-23.8 (-92.3)	-19.3 (-74.7)	55.8 (110.2)	0.67 (0.25)	-16.4 (-50.1)	-15.4 (-46.8)	28.0 (60.4)

Table S3: CV results of the reanalysis (outside bracket) and base simulation (in bracket) for SO_2 concentrations in different regions of China at different temporal scales

SO_2 ($\mu g/m^3$)	NCP				NE			
	R^2	MBE	NMB(%)	RMSE	R^2	MBE	NMB (%)	RMSE
Hourly	0.62 (0.10)	-3.6 (26.4)	-9.4 (69.4)	31.5 (63.1)	0.46 (0.08)	-2.0 (5.1)	-6.9 (17.5)	34.8 (53.2)
Daily	0.74 (0.16)	-3.6 (26.4)	-9.4 (69.6)	22.8 (52.7)	0.62 (0.13)	-2.0 (5.1)	-7.0 (17.6)	23.8 (42.2)
Monthly	0.79 (0.19)	-3.7 (26.2)	-9.6 (68.4)	17.1 (43.6)	0.71 (0.14)	-2.0 (5.0)	-6.9 (17.3)	17.9 (34.9)
Yearly	0.81 (0.18)	-4.2 (23.7)	-10.2 (56.9)	13.3 (36.1)	0.56 (0.14)	-2.4 (2.7)	-7.6 (8.7)	15.9 (27.7)
	SE				SW			
	R^2	MBE	NMB(%)	RMSE	R^2	MBE	NMB (%)	RMSE
Hourly	0.42 (0.01)	-1.0 (29.8)	-5.7 (169.6)	14.6 (69.3)	0.27 (0.01)	-1.9 (44.2)	-12.1 (277.2)	16.7 (88.1)
Daily	0.55 (0.01)	-1.0 (29.9)	-5.7 (170.2)	10.5 (63.3)	0.38 (0.01)	-1.9 (44.1)	-12.2 (276.5)	11.8 (80.3)
Monthly	0.61 (0.01)	-1.0 (29.7)	-5.7 (168.6)	7.8 (55.8)	0.46 (0.02)	-2.0 (43.9)	-12.4 (273.7)	9.1 (73.7)
Yearly	0.66 (0.01)	-1.4 (28.0)	-7.1 (144.5)	7.9 (52.7)	0.53 (0.01)	-2.7 (41.2)	-15.2 (231.3)	9.5 (68.3)
	NW				Central			
	R^2	MBE	NMB(%)	RMSE	R^2	MBE	NMB (%)	RMSE
Hourly	0.31 (0.01)	-0.3 (9.4)	-2.3 (61.6)	22.7 (40.4)	0.30 (0.02)	-4.4 (13.2)	-17.0 (51.3)	36.0 (58.9)
Daily	0.42 (0.01)	-0.3 (9.4)	-1.8 (62.2)	17.8 (36.2)	0.49 (0.03)	-4.4 (13.2)	-17.0 (51.5)	23.6 (49.1)
Monthly	0.48 (0.03)	-0.3 (9.3)	-2.2 (61.1)	13.4 (30.3)	0.59 (0.03)	-4.4 (13.1)	-17.0 (51.0)	18.2 (43.2)
Yearly	0.29 (0.00)	-1.9 (6.6)	-10.5 (35.9)	15.8 (28.0)	0.50 (0.00)	-5.6 (8.6)	-19.0 (29.3)	18.6 (40.2)

Table S4: CV results of the reanalysis (outside bracket) and base simulation (in bracket) for NO_2 concentrations in different regions of China at different temporal scales

NO_2 ($\mu g/m^3$)	NCP				NE			
	R^2	MBE	NMB(%)	RMSE	R^2	MBE	NMB (%)	RMSE
Hourly	0.67 (0.20)	-1.4 (-3.0)	-3.5 (-7.1)	16.8 (26.5)	0.61 (0.27)	-1.6 (-5.9)	-5.0 (-19.1)	15.8 (22.4)
Daily	0.72 (0.22)	-1.4 (-2.9)	-3.3 (-7.1)	12.4 (20.8)	0.66 (0.34)	-1.5 (-5.9)	-4.9 (-19.0)	11.7 (17.2)
Monthly	0.72 (0.24)	-1.4 (-2.9)	-3.3 (-7.1)	9.3 (15.5)	0.64 (0.37)	-1.5 (-5.9)	-5.0 (-19.1)	9.3 (13.7)
Yearly	0.67 (0.36)	-1.4 (-3.8)	-3.3 (-9.0)	7.5 (11.0)	0.64 (0.45)	-1.5 (-6.4)	-4.8 (-20.3)	7.8 (11.7)
	SE				SW			
	R^2	MBE	NMB(%)	RMSE	R^2	MBE	NMB (%)	RMSE
Hourly	0.64 (0.23)	-1.9 (-1.3)	-5.9 (-4.0)	14.9 (24.1)	0.49 (0.19)	-3.9 (-9.9)	-14.0 (-35.7)	16.4 (23.2)
Daily	0.71 (0.28)	-1.8 (-1.3)	-5.8 (-4.0)	11.2 (19.2)	0.55 (0.28)	-3.9 (-9.9)	-14.0 (-35.7)	12.8 (18.7)
Monthly	0.72 (0.36)	-1.8 (-1.2)	-5.8 (-3.9)	8.8 (14.7)	0.48 (0.32)	-4.0 (-10.0)	-14.4 (-36.0)	12.6 (17.2)
Yearly	0.66 (0.49)	-1.9 (-2.2)	-6.0 (-6.6)	7.8 (11.7)	0.46 (0.37)	-4.6 (-11.0)	-16.1 (-38.7)	11.8 (16.4)
	NW				Central			
	R^2	MBE	NMB(%)	RMSE	R^2	MBE	NMB (%)	RMSE
Hourly	0.46 (0.20)	-4.3 (-18.0)	-12.9 (-54.4)	24.3 (31.5)	0.50 (0.20)	-5.2 (-16.7)	-15.1 (-48.3)	20.5 (29.2)
Daily	0.55 (0.27)	-4.1 (-18.0)	-12.5 (-54.4)	18.3 (27.0)	0.58 (0.28)	-5.2 (-16.7)	-15.0 (-48.3)	15.4 (24.7)
Monthly	0.59 (0.40)	-4.2 (-18.0)	-12.7 (-54.3)	15.3 (23.8)	0.61 (0.40)	-5.2 (-16.6)	-15.1 (-48.3)	12.8 (21.6)
Yearly	0.40 (0.36)	-6.0 (-19.7)	-17.3 (-56.3)	16.5 (23.8)	0.55 (0.36)	-5.6 (-17.6)	-16.0 (-50.1)	12.2 (21.4)

Table S5: CV results of the reanalysis (outside bracket) and base simulation (in bracket) for CO concentrations in different regions of China at different temporal scales

CO (mg/m^3)	NCP				NE			
	R^2	MBE	NMB(%)	RMSE	R^2	MBE	NMB (%)	RMSE
Hourly	0.67 (0.25)	-0.03 (-0.59)	-2.49 (-43.4)	0.64 (1.13)	0.50 (0.20)	-0.05 (-0.51)	-5.3 (-51.9)	0.59 (0.88)
Daily	0.72 (0.31)	-0.03 (-0.59)	-2.15 (-43.3)	0.50 (0.99)	0.56 (0.25)	-0.05 (-0.51)	-4.9 (-51.7)	0.46 (0.78)
Monthly	0.74 (0.34)	-0.03 (-0.59)	-2.24 (-43.5)	0.38 (0.85)	0.59 (0.25)	-0.05 (-0.51)	-5.2 (-52.0)	0.37 (0.70)
Yearly	0.71 (0.14)	-0.04 (-0.64)	-2.75 (-45.1)	0.32 (0.85)	0.55 (0.14)	-0.06 (-0.56)	-5.9 (-54.0)	0.35 (0.74)
	SE				SW			
	R^2	MBE	NMB(%)	RMSE	R^2	MBE	NMB (%)	RMSE
Hourly	0.42 (0.13)	-0.06 (-0.36)	-6.4 (-38.2)	0.39 (0.62)	0.36 (0.07)	-0.08 (-0.32)	-9.4 (-36.4)	0.46 (0.65)
Daily	0.45 (0.15)	-0.06 (-0.36)	-6.1 (-38.0)	0.34 (0.57)	0.40 (0.08)	-0.08 (-0.31)	-9.1 (-36.3)	0.39 (0.59)
Monthly	0.44 (0.14)	-0.06 (-0.36)	-6.2 (-38.1)	0.28 (0.51)	0.40 (0.08)	-0.08 (-0.32)	-9.4 (-36.7)	0.34 (0.54)
Yearly	0.38 (0.05)	-0.06 (-0.38)	-6.5 (-39.3)	0.25 (0.50)	0.36 (0.01)	-0.09 (-0.36)	-10.1 (-39.1)	0.36 (0.57)
	NW				Central			
	R^2	MBE	NMB(%)	RMSE	R^2	MBE	NMB (%)	RMSE
Hourly	0.38 (0.12)	-0.19 (-1.02)	-15.0 (-79.3)	1.13 (1.55)	0.44 (0.22)	-0.13 (-0.76)	-11.2 (-65.2)	0.73 (1.11)
Daily	0.45 (0.18)	-0.19 (-1.01)	-14.6 (-79.2)	0.92 (1.43)	0.49 (0.27)	-0.13 (-0.76)	-10.8 (-65.1)	0.62 (1.04)
Monthly	0.50 (0.29)	-0.19 (-1.02)	-15.1 (-79.3)	0.75 (1.32)	0.53 (0.32)	-0.13 (-0.76)	-11.1 (-65.2)	0.52 (0.97)
Yearly	0.13 (0.12)	-0.31 (-1.18)	-21.1 (-80.8)	0.85 (1.35)	0.19 (0.08)	-0.17 (-0.84)	-13.3 (-67.3)	0.69 (1.08)

Table S6: CV results of the reanalysis (outside bracket) and base simulation (in bracket) for O_3 concentrations in different regions of China at different temporal scales

O_3 ($\mu g/m^3$)	NCP				NE			
	R^2	MBE	NMB(%)	RMSE	R^2	MBE	NMB (%)	RMSE
Hourly	0.83 (0.50)	-3.9 (-24.5)	-6.1 (-39.1)	22.1 (44.3)	0.76 (0.38)	-3.6 (-15.4)	-6.0 (-25.6)	21.0 (36.7)
Daily	0.83 (0.48)	-3.8 (-24.5)	-6.0 (-39.1)	16.3 (37.0)	0.76 (0.34)	-3.6 (-15.3)	-5.9 (-25.5)	16.4 (30.8)
Monthly	0.85 (0.62)	-3.8 (-24.4)	-6.1 (-39.1)	12.6 (31.6)	0.76 (0.42)	-3.6 (-15.2)	-5.9 (-25.4)	13.1 (25.2)
Yearly	0.72 (0.29)	-3.7 (-23.1)	-6.2 (-38.6)	9.2 (26.8)	0.62 (0.18)	-3.5 (-14.4)	-6.1 (-25.0)	10.0 (20.4)
	SE				SW			
	R^2	MBE	NMB(%)	RMSE	R^2	MBE	NMB (%)	RMSE
Hourly	0.77 (0.57)	-2.3 (-10.0)	-3.9 (-17.4)	21.1 (38.2)	0.69 (0.40)	0.8 (9.6)	1.4 (18.0)	22.2 (33.6)
Daily	0.69 (0.44)	-2.2 (-10.0)	-3.8 (-17.3)	15.8 (31.0)	0.64 (0.34)	0.8 (9.7)	1.5 (18.1)	17.4 (26.7)
Monthly	0.69 (0.43)	-2.2 (-10.0)	-3.9 (-17.3)	12.4 (24.2)	0.64 (0.44)	0.8 (9.7)	1.6 (18.1)	14.4 (21.1)
Yearly	0.45 (0.07)	-2.4 (-10.1)	-4.2 (-17.7)	10.0 (20.7)	0.42 (0.28)	1.3 (10.0)	2.6 (19.4)	12.0 (17.7)
	NW				Central			
	R^2	MBE	NMB(%)	RMSE	R^2	MBE	NMB (%)	RMSE
Hourly	0.52 (0.31)	-2.7 (-2.2)	-4.6 (-3.8)	28.3 (33.2)	0.71 (0.45)	-1.5 (-0.8)	-2.5 (-1.3)	23.9 (32.5)
Daily	0.50 (0.31)	-2.6 (-2.1)	-4.5 (-3.6)	22.9 (26.6)	0.67 (0.42)	-1.4 (-0.7)	-2.4 (-1.1)	17.8 (23.9)
Monthly	0.58 (0.42)	-2.6 (-2.1)	-4.5 (-3.6)	19.1 (22.1)	0.72 (0.56)	-1.4 (-0.7)	-2.4 (-1.2)	13.9 (17.6)
Yearly	0.37 (0.24)	-1.6 (-0.6)	-2.9 (-1.1)	15.7 (17.1)	0.53 (0.30)	-0.8 (0.2)	-1.4 (0.4)	11.6 (14.1)

Figures

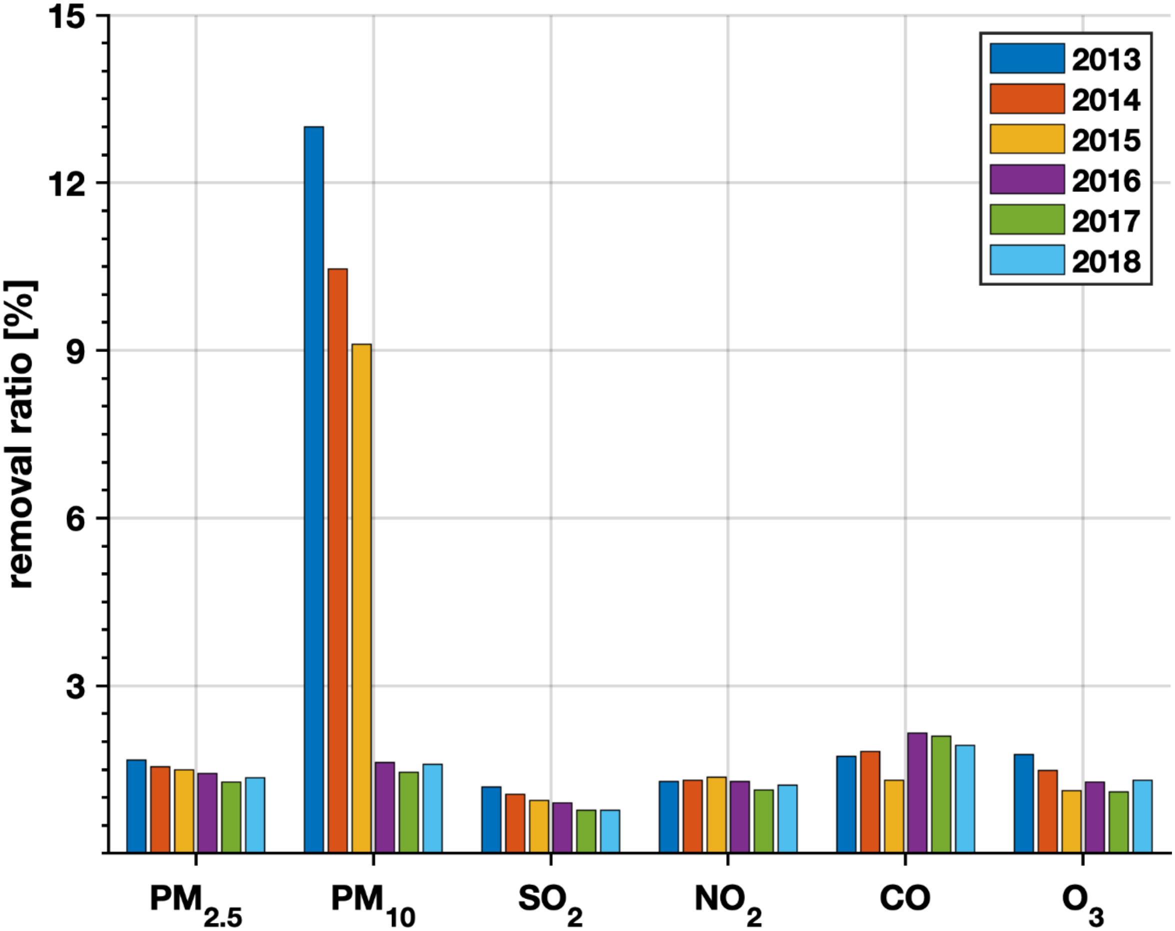


Figure S1: Removal ratio of all observation sites in China from 2013 to 2018 for different species detected by the automatic outlier detection method.

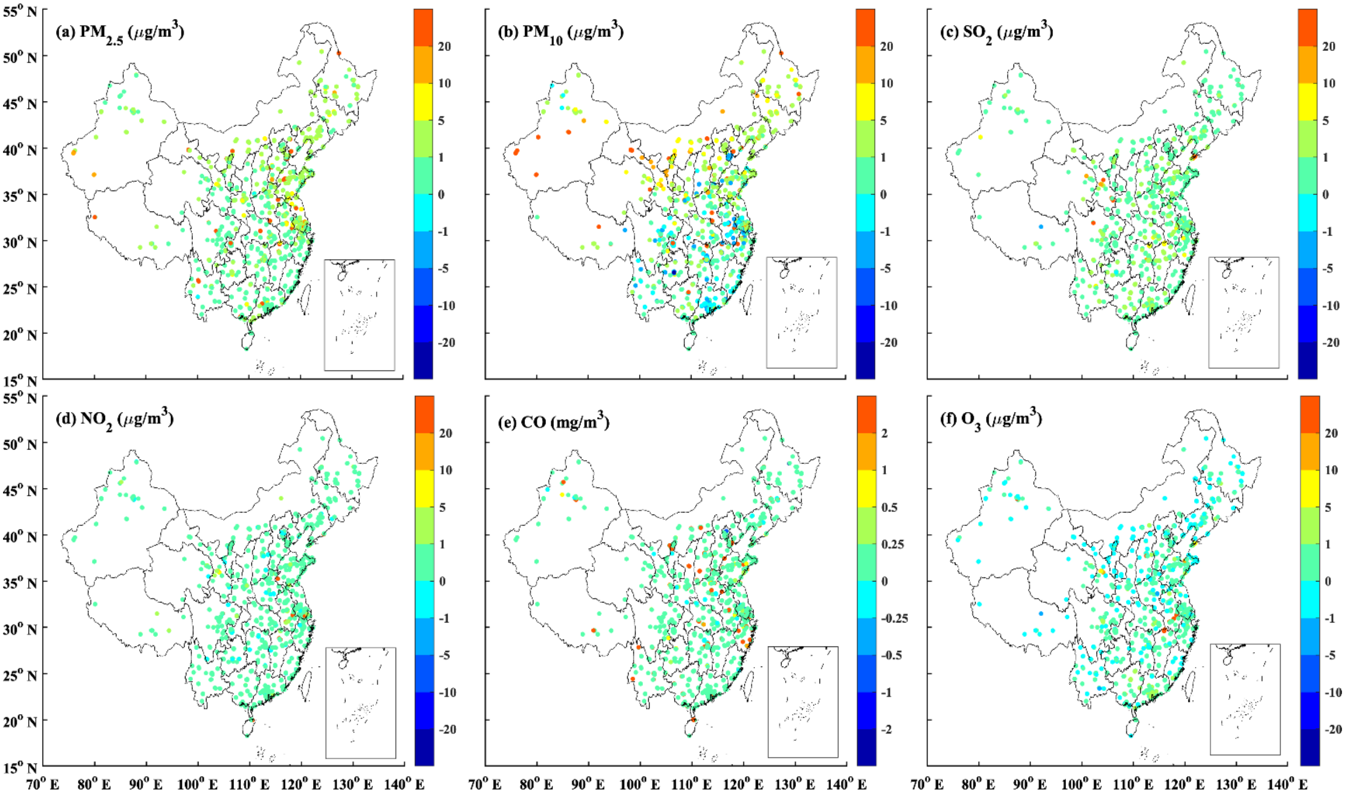


Figure S2: Spatial distributions of differences in annual concentrations of six air pollutants in China before and after quality control averaged from 2013 to 2018.

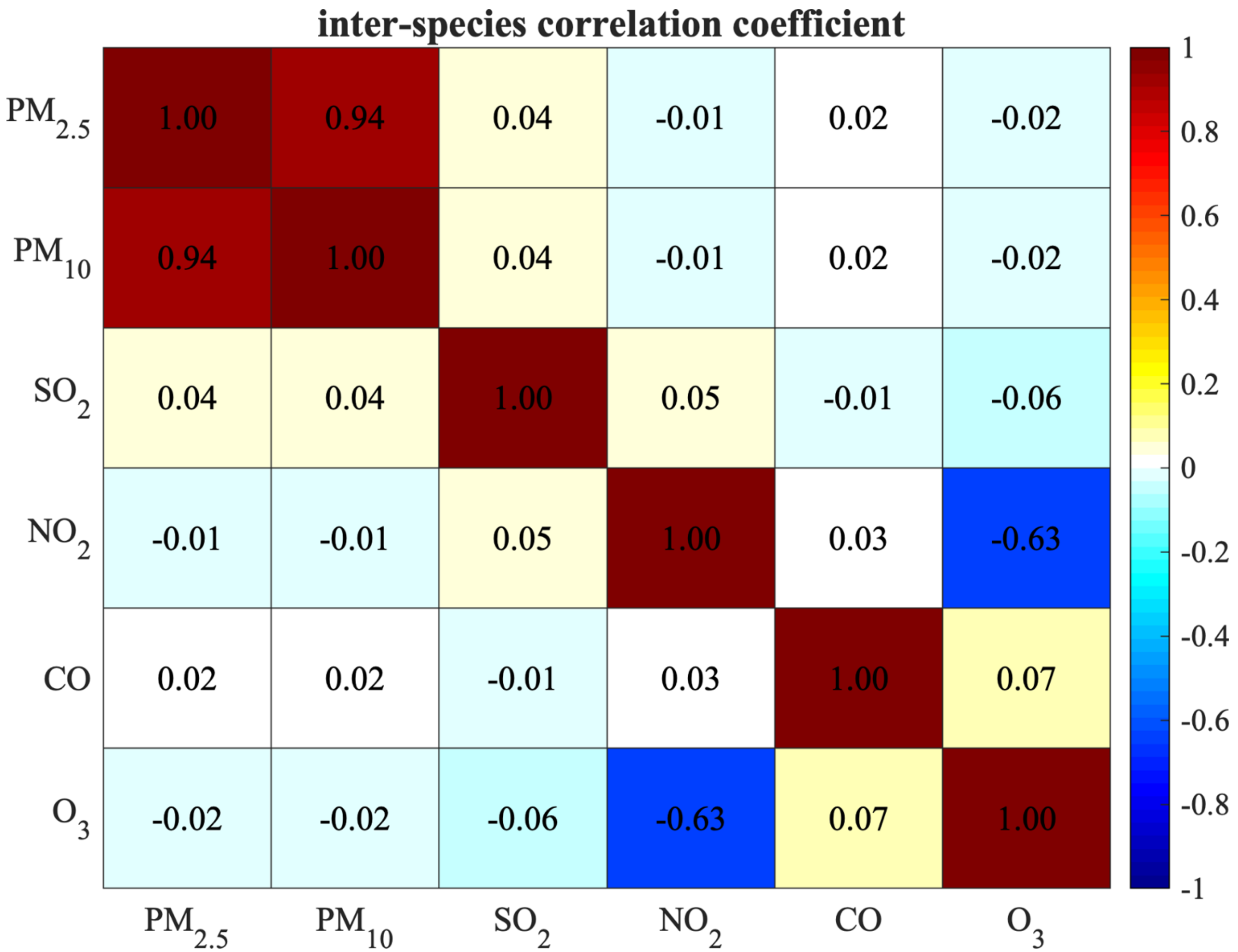


Figure S3: Correlations between species in the background error covariance matrix, estimated from the LETKF ensemble averaged from 2013 to 2018. The global mean of the covariance estimated for each station is plotted.

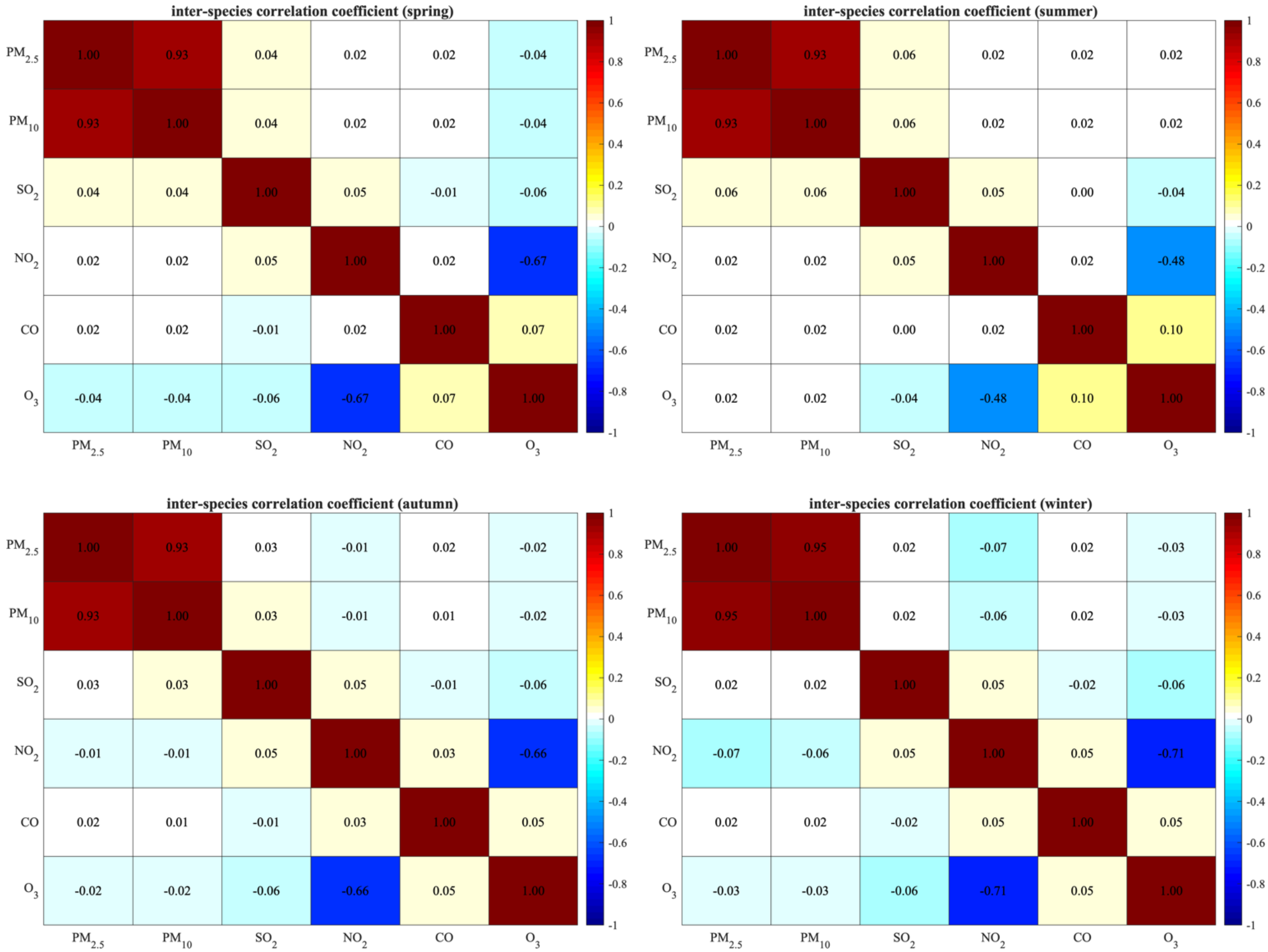


Figure S4: Correlations between species in the background error covariance matrix, estimated from the LETKF ensemble averaged in different seasons from 2013 to 2018. The global mean of the covariance estimated for each station is plotted.

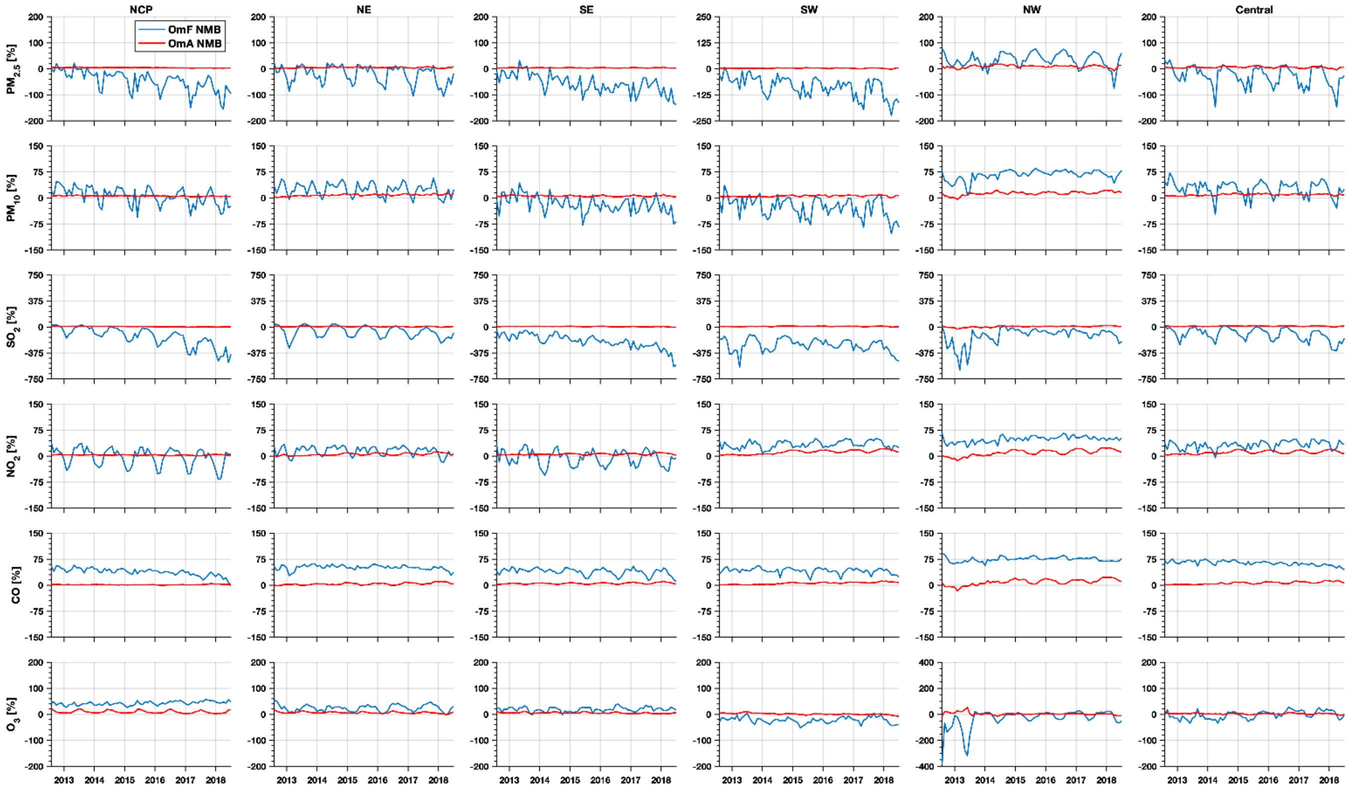


Figure S5: Time series of monthly mean OmF and OmA normalized mean bias in different regions of China for different species.

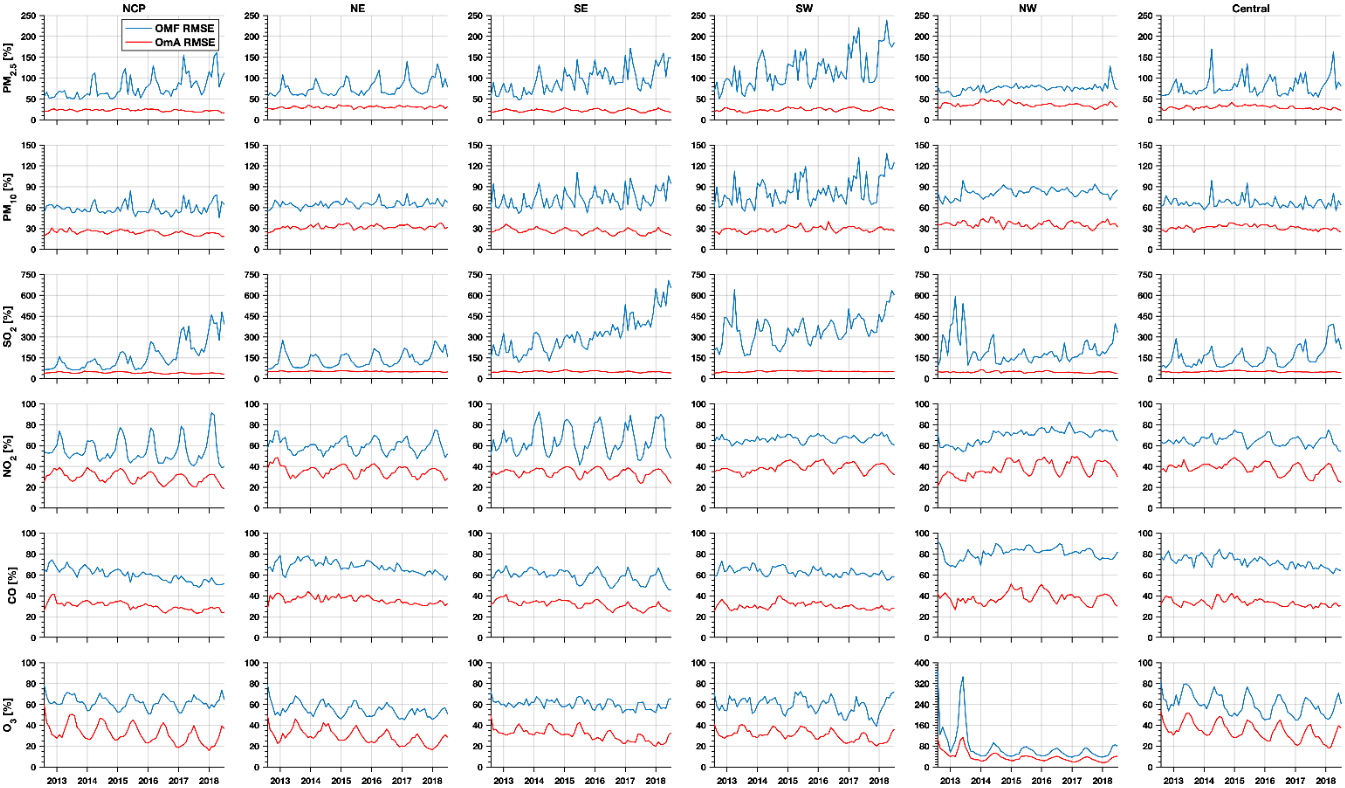


Figure S6: Time series of monthly mean OmF and OmA normalized root mean square error in different regions of China for different species.

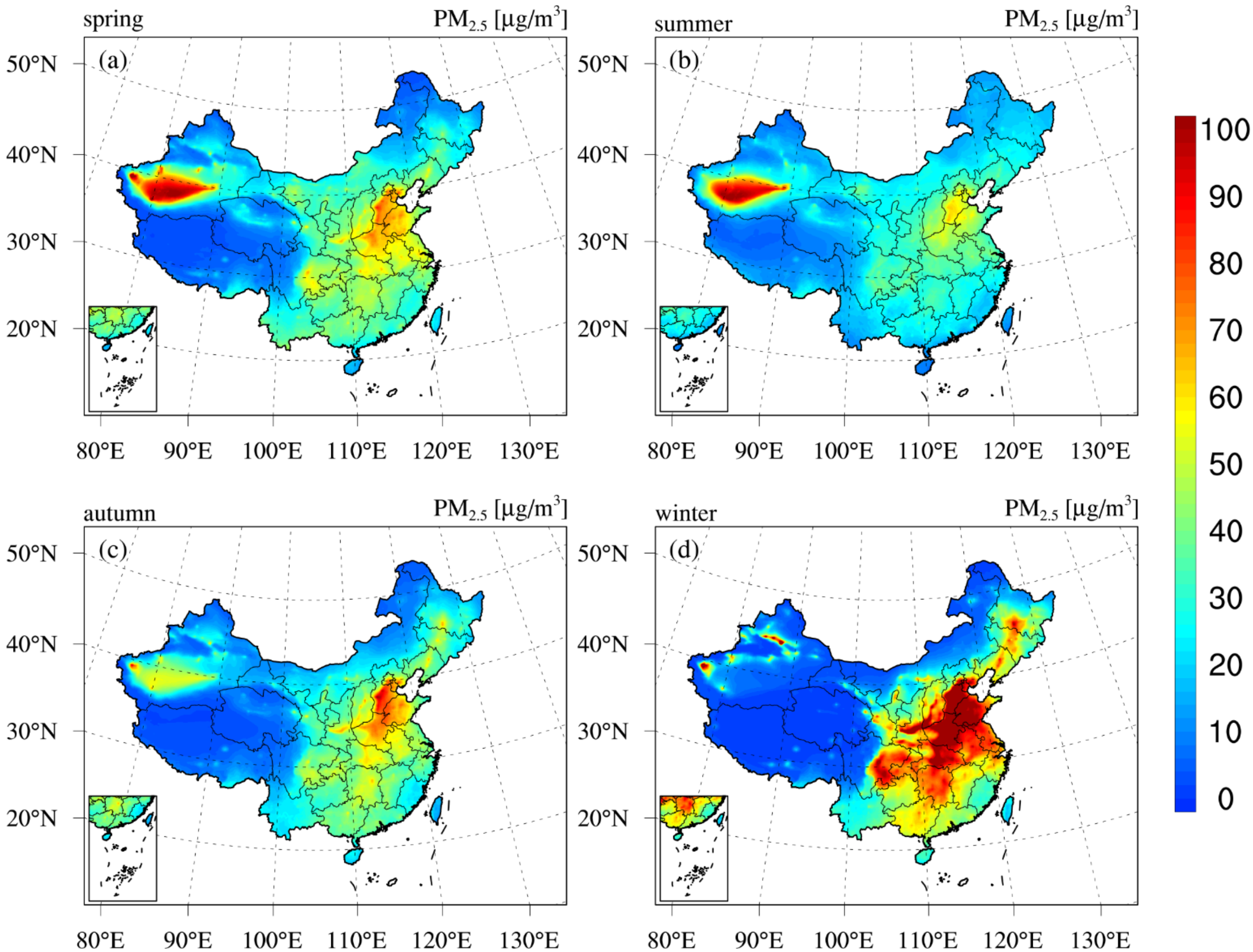


Figure S7: Spatial distributions of the $PM_{2.5}$ concentrations in China during (a) spring, (b) summer, (c) autumn and (d) winter averaged from 2013 to 2018.

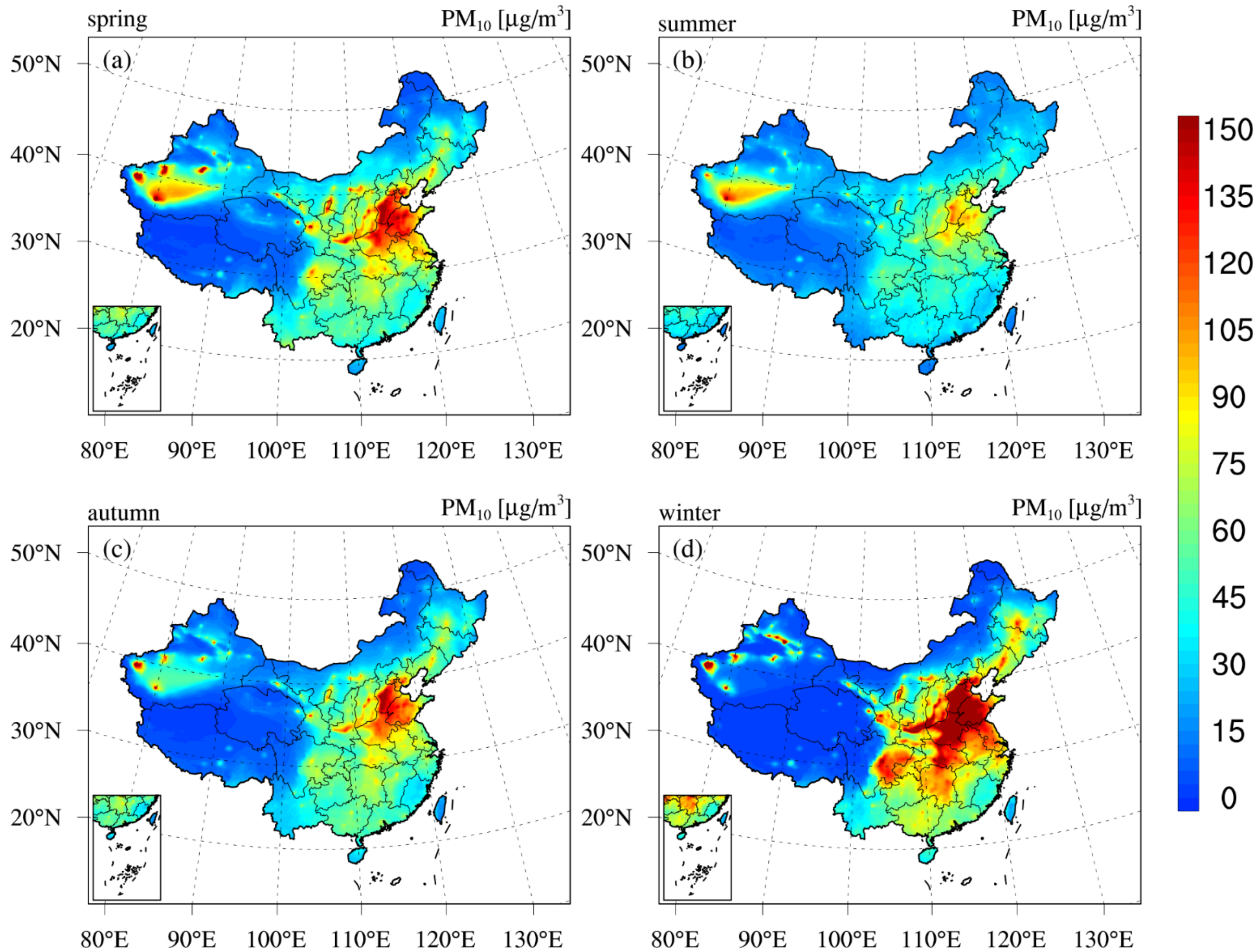


Figure S8: Same as Fig. S7 but for PM$_{10}$ concentrations.

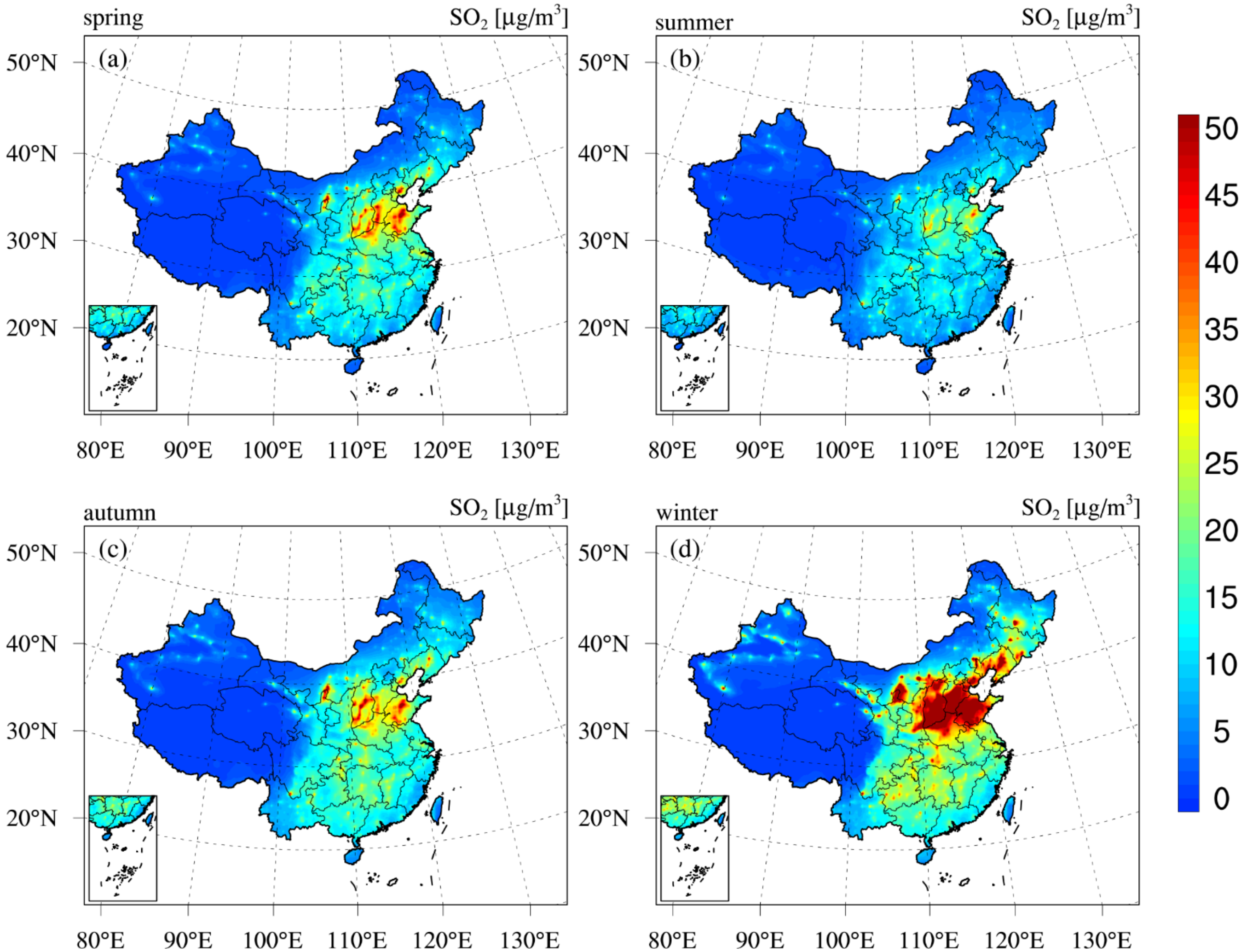


Figure S9: Same as Fig. S7 but for SO_2 concentrations.

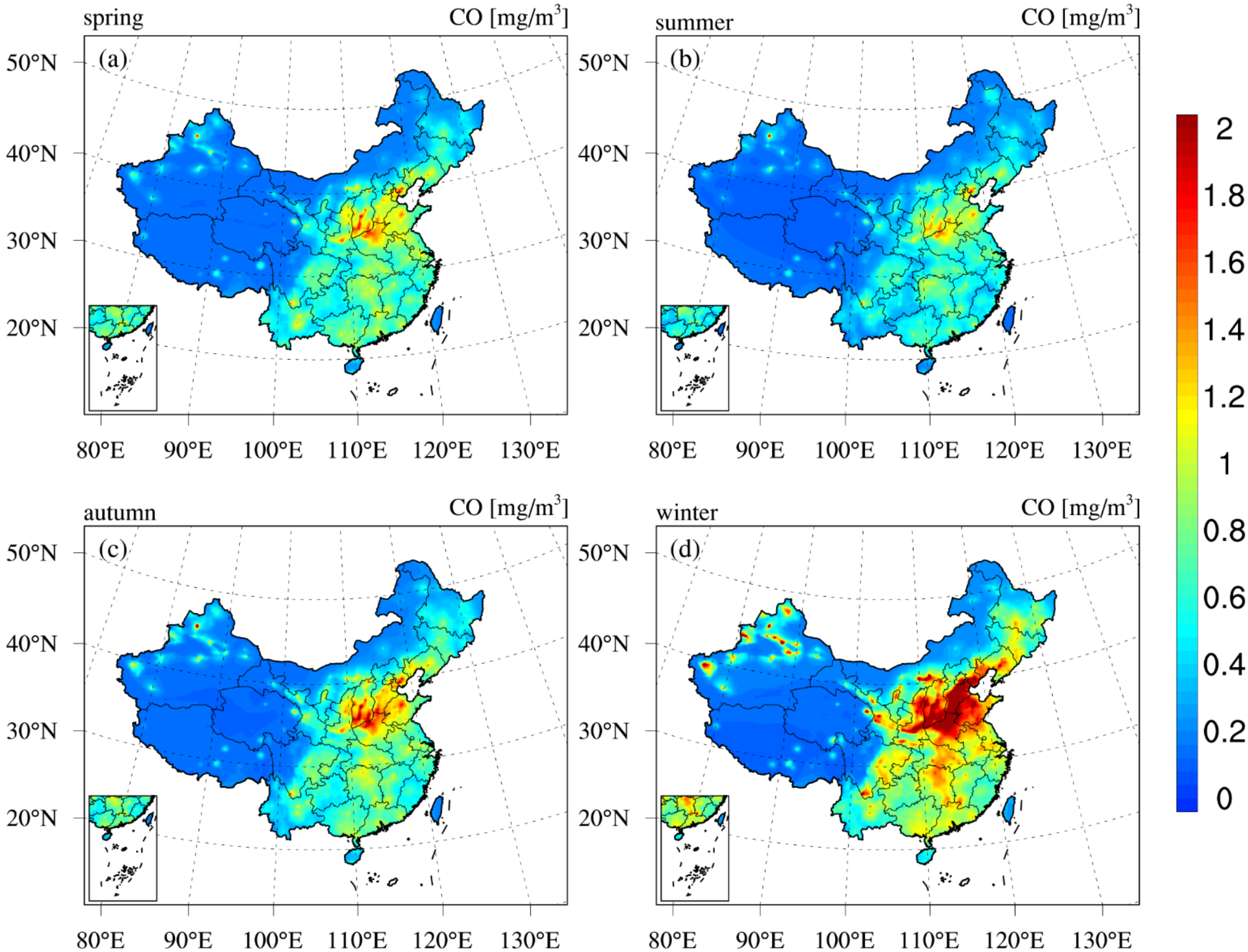


Figure S10: Same as Fig. S7 but for CO concentrations.

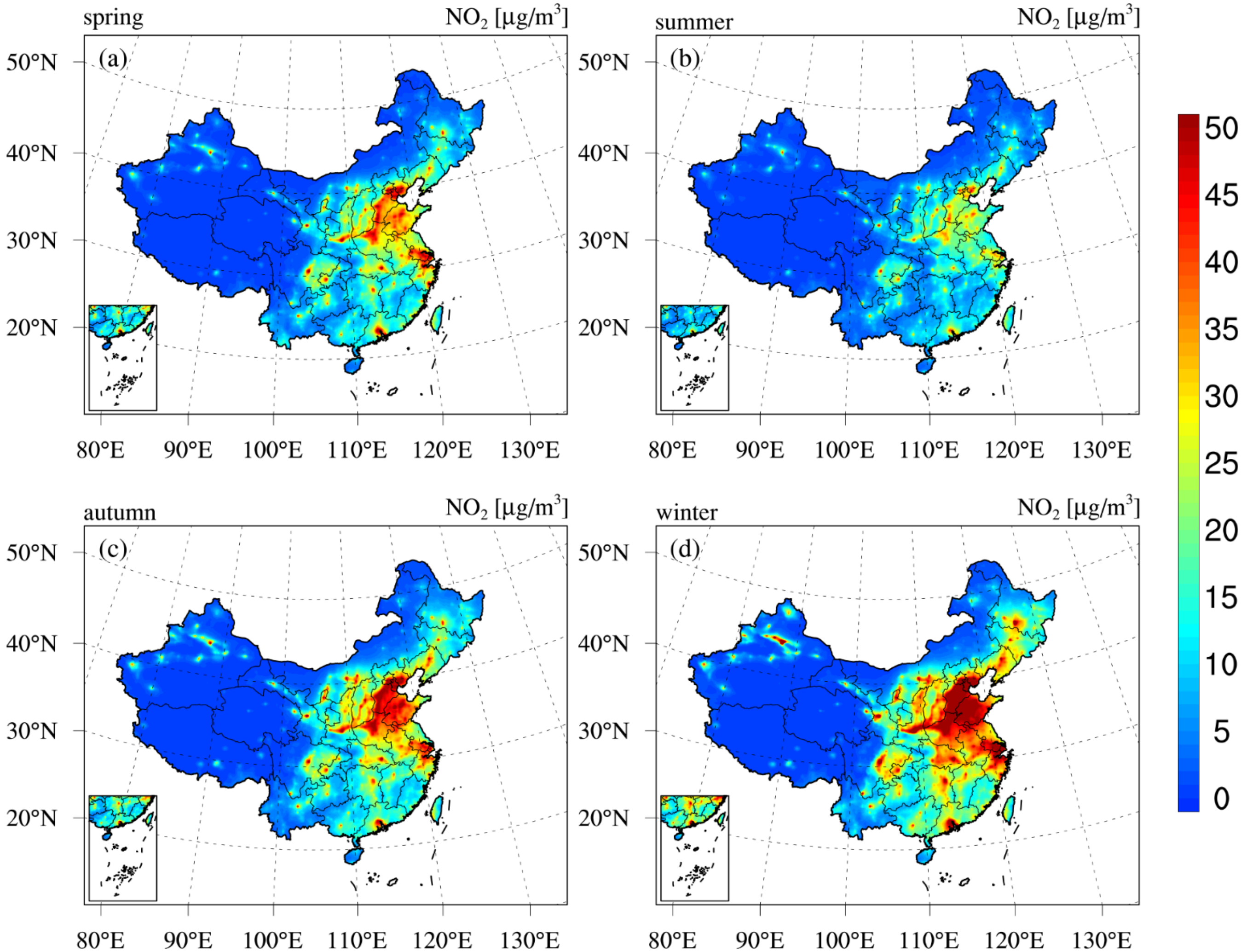


Figure S11: Same as Fig. S7 but for NO_2 concentrations.

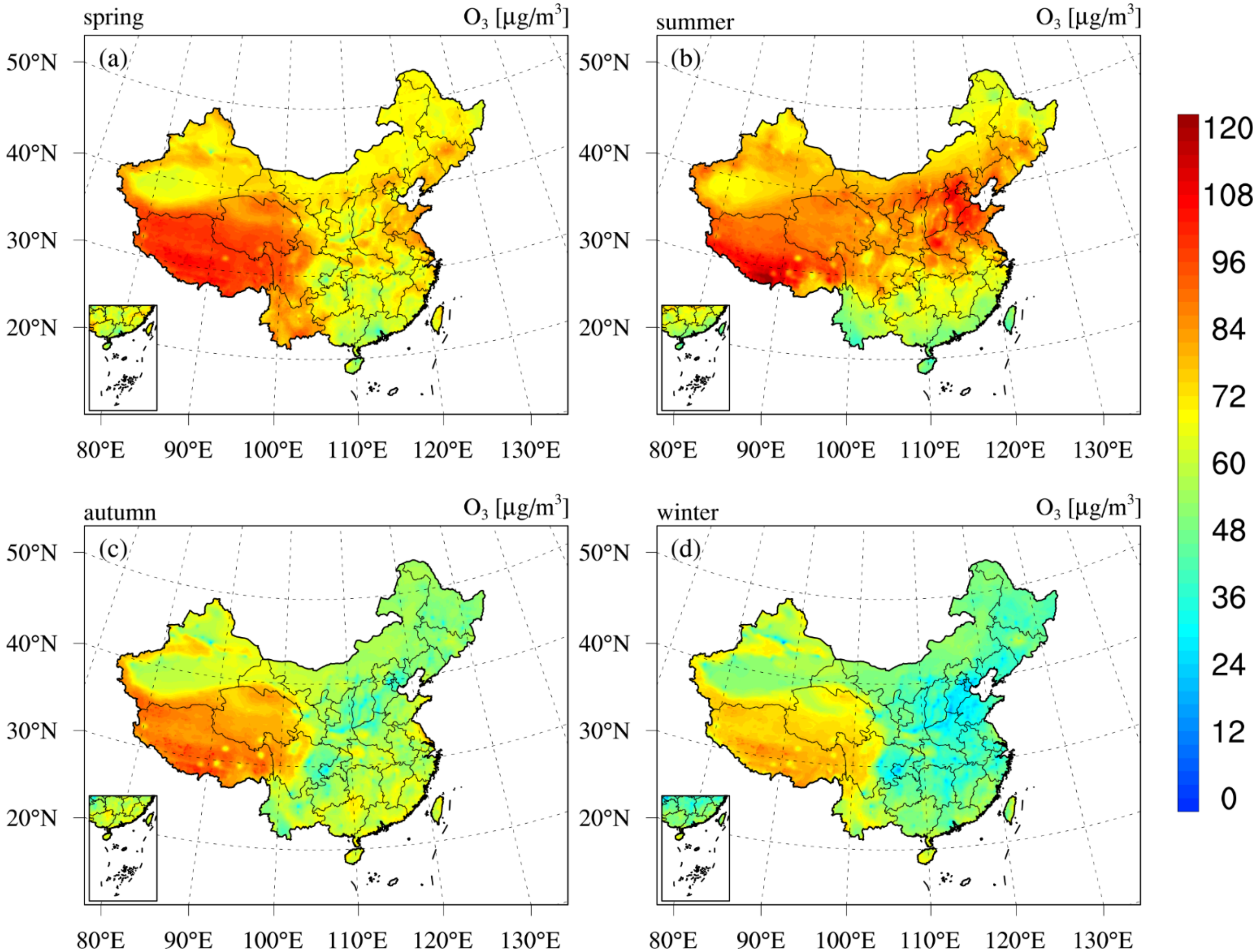


Figure S12: Same as Fig. S7 but for O_3 concentrations.