Reply to Referee #2:

We deeply appreciate your helpful comments and suggestions, which enabled us to improve the quality of our present study. In our response, we use italicization in blue to indicate the reviewer's comments, and normal type in black for our response. Besides, we use boldface type to indicate changes in the manuscript.

Specific Comments:

1. Page 14, Line 267: Can you provide possible reasons why the model significantly underestimated PMcoarse? Missing emissions?

Response:

Yes, missing emissions could be the major reason for the underestimated PMcoarse simulation in the model. Different from the United states or European countries that national emission inventories are provided and updated frequently by the government (e.g. US National Emission Inventory NEI 05-08-11-14-17), the publicly available emission inventories for China are mainly established by several scientific research groups. In the US, NEI are established based upon data provided by state, local, and tribal air agencies for sources in their jurisdictions and are supplemented by data that developed by the US Environmental Protection Agency; thus the statistics are comprehensive and detailed. In China, the scientific research groups established EIs only by public released statistics of energy, activity, emission factor etc., which are usually limited and incomplete; thus the uncertainties of the publicly available emission inventories in China are relatively larger compared with others (US, European countries). It's a known problem that the fugitive dust emissions over the whole of China is still lack, which might cause the underestimated PMcoarse simulation in the model.

Some related statements have also been added in the discussion section of the manuscript as below.

"As for the significantly underestimated PMcoarse in the model, the results might relate to the missing emissions under current situations. Different from the United states or European countries that national emission inventories are provided and updated frequently by the government (e.g. US National Emission Inventory NEI 05-08-11-14-17), the publicly available emission inventories for China are mainly established by several scientific research groups. In result, the uncertainties of the publicly available emission inventories in China are relatively larger compared with others (US, European countries), and it's a known problem that the fugitive dust emissions over the whole of China is still lack, which might cause the underestimated PMcoarse simulation in the model."

2. Page 16, Line 305: Does Figure 4 show the vertical profiles of pollutant concentrations in the model simulations? It is not clear why the ozone levels are so low in the upper troposphere (9 km or above 15 km). Does the model account for stratospheric ozone boundary conditions? Please clarify.

Response:

Yes, Figure 4 (updated as Figure 5 in the manuscript) shows the vertical profiles of pollutants concentrations in the model simulations. Thanks for the kind reminder! The stratospheric ozone boundary conditions are not taken into account, which might be the reason for the low ozone level in upper tropopause.

3. Page 18, Line 356-360: I suggest move the definition of the threat score (TS) from the Supplement to the main text here. Also here in the text I suggest explain what the values of TS represent.

Response:

Accepted. The definition and explanation of the threat score (TS) have been moved from the supplement to the manuscript as below.

"To reflect the integrated DA effect of aerosols and gas-phase pollutants, the threat score (TS), one of the most commonly used criterions in the verifications of meteorological forecasts, is used for Air Quality Index (AQI) for six AQI levels. The threat score (TS) for AQI is calculated by

$$\mathbf{TS_i = \frac{H_i}{H_i+M_i+F_i}} \qquad \textbf{(6)}$$

where H, M, and F denote the counts of the hits, the misses, and the false alarms in the forecast of AQI, and i denotes the AQI levels from 1 to 6. In result, the TS is acquired at each AQI level ranging from 0 to 1, and the higher (lower) TS represents the better (worse) forecast performance."

4. Page 19, Line 379: Need to explain here "ALL_6h" is the "ALL" simulation in Table 2, right?

Response:

Accepted. The "ALL_6h" is the "ALL" experiment in Tab.2, and the explanation has been added in the manuscript as below.

"Figure 9 shows the domain-averaged bias and RMSE of the analysis as in Fig. 3, but for experiments with different DA frequencies (ALL_6h, ALL_3h, and ALL_1h; the ALL_6h is the ALL experiment in Tab.2)."

5. Page 20, Line 397-406:I do not think the discussion on ozone performance here and other places (e.g., abstract, conclusions) is convincing. The study also simulated a winter month (January 2017) when ozone photochemistry is very weak. Therefore, I do not think that the ozone photochemistry or NOxVOC ratios would explain the decreased forecast skill for ozone when increasing assimilation frequency. Since January is not an ozone pollution season, the conclusion that "assimilate O3 and NO2 every 6 h" would not be robust only based on results of this month. Please clarify and discuss the limits.

Response:

Accepted. Thanks for the great suggestion! It could possibly also be related with the NO_x titration due to the changed NO_2 concentrations. The discussion on the ozone performance in association with NO_2/VOC ratios has been weakened in the manuscript, the abstract, and the conclusion as below. The discussion on NO_x titration is added, and the limits of the current findings are also mentioned in the discussion section.

Statements in section 3.3 of the manuscript:

"However, the analysis and 24-hr forecast of O_3 become worse under higher cycling frequencies for this winter season (Fig. 9e and 11c). Given the

analysis is at 00 UTC, the worsen analysis in the experiments with higher DA frequencies (1-h, 3-h) could be mainly due to the unfavorable changes in the 1-h/3-h forecasts period (starting from 23 UTC, 21 UTC), which is different from the situation in the 6-h cycling experiment. As for the forecasts, the 24-hr performances starting from 00 UTC show complex changes along with the forecast range: compared to the 6-h cycling experiment, the biases in the experiments with higher DA frequencies decrease at 09-14 UTC but increase for other hours; the RMSE and correlations in the experiments with higher DA frequencies become worse in most of the hours (Fig. 11c). It should be mentioned that O_3 is a relatively short-lived chemical reactive species, and takes part in highly complex and photochemical reactions in association with NO_x and VOC (Peng et al. 2018, Lu et al., 2019). From this perspective, the performances of O_3 could also rely on the photochemistry and the NO_x titration, in addition to the IC. Although the winter month (January 2017) is investigated here when ozone photochemistry is relatively weaker compared to other seasons, the photochemistry and the NO_x titration still play their roles. Accordingly, when the assimilation of NO_2 changes the NO_2 concentration and leave the NO and VOC unadjusted due to the absence of NO and VOC measurements, two results might occur: firstly, the NO_2/VOC ratio which determine the photochemical reactions and even the regime might be changed (O_3 production/loss direction might change); secondly, the NO_x titration process might be changed due to the NO_2 concentration updates (but no change on

NO). Considering the relevant NO_x-VOC-O_3 reactions take place quickly, changing the O_3 concentration in a short period, the advantage of IC DA could compete with the disadvantages of the disordered photochemistry (inaccurate NO_2/VOC ratios) or the changed titration (adjusted NO_2 concentrations but not NO) resulting from the DA. Under this circumstance, the more frequent the O_3 and NO_2 were assimilated, the more incompatibilities could be brought into the related photochemical/titration reactions, resulting the model performs worse in the O_3 forecasts under higher cycling frequencies. It is noted that these statistics were only for the analysis at 00UTC and the 24-hr forecast starting from 00UTC for winter season. Since O_3 has strong diurnal and seasonal variations, more experiments and statistics at different time of the day and different season of the year should be conducted in the future."

Statements in the abstract:

"For O_3, although improvements are acquired at the 6-h cycling frequency, the advantage of more frequent DA could be consumed by the disadvantages of the unbalanced photochemistry (due to inaccurate precursor NO_x/VOC ratios) or the changed titration process (due to changed NO_2 concentrations but not NO) from assimilating the existing observations (only O_3 and NO_2, but no VOC and NO); yet the finding is based on the 00 UTC forecast for this winter season only and O_3 has strong diurnal and seasonal variations, more experiments should be conducted to draw further conclusions."

Statement in the conclusion:

"for O_3, compared to a better performance at the 6-h cycling frequency, its analysis at 00 UTC and the following 24-hr forecast become generally worse under higher cycling frequencies for this winter season, although the biases did decrease at 09-14 UTC in the 24-hr forecast. Considering the relevant NO_x-VOC-O_3 reaction system changes the NO_2/O_3 concentration in a short period, the advantage of IC DA could compete with the disadvantages of the disordered photochemistry (inaccurate NO_2/VOC ratios) or the changed titration (adjusted NO_2 concentrations but not NO) resulting from the DA. In future applications, it is better to assimilate $PM_{2.5}$, PM_{10}, SO_2, and CO every 1 h. For the frequency of O_3 and NO_2 assimilation, every 6 h is the best in this winter season in our study. Since O_3 has strong diurnal and seasonal variations, more experiments and statistics at different time of the day and different season of the year should be conducted in the future. Also, it might be helpful to assimilate NO/VOC simultaneously with O_3 and NO_2 after there are corresponding measurements."

6. Page 23, Line 460-462: As a future development, is it possible to directly constrain the coefficients of heterogeneous reactions using the data assimilation system?

Response:

Yes. The variables in the heterogeneous reactions include precursor concentrations, meteorology (RH), and uptake coefficients. Two DA approaches would be tried in the future to fully utilize the observations and constrain uptake coefficients.

The first approach is by the 3DVAR technique: (1) simultaneously assimilate precursor concentrations and meteorology (RH) in the model to generate the best initial condition (IC); (2) start from this IC, conduct sensitivity simulations with a series of adjusted uptake coefficients to best match the SNA species observations.

The second approach is by the Ensemble Kalman Filter (EnKF) technique which might be computing expensive: (1) perturb the uptake coefficients in the model and generate ensemble members through model forecasts; (2) use all observations (precursor concentrations, RH, species) in the EnKF system as constraints to optimize the uptake coefficients.

7. Page 50, Figure 12: Need to add the unit in the figure or in the caption.

Response:

Accepted. The units have been added in the caption of the figure, and the figure number has been updated as 13.

8. Page 51, Figure 13: The titles say "Used 79" and "Used 80". What do they mean?

Response:

They are the accumulated numbers of the used observations around Beijing area during January 16 (1600 UTC, 1606 UTC, 1612 UTC, and 1618 UTC). The corresponding description has been added in the caption in the manuscript as below, and the figure number has been updated as 14.

"Figure 14. Averaged scatter plot of (a, c) observation versus background and (b, d) observation versus analysis for (a, b) SO_2 and (c, d) NO_2 around Beijing area (red dots in Fig. 1) on January 16. The numbers on the title denote the accumulated numbers of the used observations around Beijing area during January 16 (1600 UTC, 1606 UTC, 1612 UTC, and 1618 UTC)"