Referee #1:

This study estimates daily NO_x emissions at a 3-km resolution in Beijing and its surrounding areas using an emission inversion framework. The framework assimilates TROPOMI NO_2 column concentrations with an Ensemble Kalman Filter coupled with CMAQ. The results reveal that proxy-based bottom-up emission datasets tend to overestimate NO_x emissions in densely populated areas, providing crucial insights for urban air quality regulations. Robust sensitivity analyses further strengthen the study by evaluating the effects of satellite retrieval parameters (e.g., a priori profiles and averaging kernels) and an observation localization radius parameter on the inversion results. Specific comments on the manuscript are outlined below.

Response:

We thank the referee for the constructive and positive comments on our paper. We have provided our point-by-point responses as follows and revised the manuscript accordingly.

Specific comments

1. Figure 3 and Figure S4: What ground air quality monitoring station data is used for this comparison? Is it based on a single station or a multi-station average? Additionally, how does this comparison vary across different ground stations, such as those in densely populated areas versus suburban or rural areas?

Response:

Thanks for your valuable comments. We use the national control stations maintained by China National Environmental Monitoring Center (CNEMC) to evaluate the model simulations. Fig. S1 shows the spatial location of the stations. These stations are primarily located in densely populated areas of each city.

According to your suggestion, we have classified the observation stations into two categories based on NO_2 concentration characteristics, population density, and emission patterns: low-emission areas and high-emission areas. The low-pollution areas refer to the two northern cities in the D03 domain, Zhangjiakou and Chengde, where NO_2 concentrations are relatively low. The observation stations in all other urban areas are classified as high-pollution stations.

Fig. S6 presents a comparison between CMAQ simulations and ground-based observations in different regions. Compared to the highly polluted urban areas, the posterior NO_2 simulations in Zhangjiakou and Chengde show much better consistency with observations during summer. This indicates that in regions with low surface emissions, the accuracy of posterior simulations in summer is relatively high. Furthermore, it reinforces the finding that in high-emission urban areas, the constraint of satellite NO_2 column measurements on surface emissions in summer is weaker, leading to an overestimation in posterior simulations.

We have added a description of the air quality monitoring stations in Sect. 2.6 (Lines 262-263) as follows:

"Fig. S1 shows the spatial location of the ground-based stations used to evaluate the CMAQ simulations. These stations are primarily located in densely populated areas of each city."

We have added the analysis of the evaluation results for different regions in Sect. 3.1.2 (Lines 328-336) as follows:

"Furthermore, we have classified the observation stations into two categories based on NO_2 concentration characteristics, population density, and emission patterns: low-pollution areas and high-pollution areas. The low-pollution areas refer to the two northern cities in the D03 domain, Zhangjiakou and Chengde, where NO_2 concentrations are relatively low. The observation stations in all other urban areas are classified as high-pollution stations. Fig. S6 presents a comparison between CMAQ simulations and ground-based observations in different pollution regions. Compared to other highly polluted urban areas, the posterior NO_2 simulations in Zhangjiakou and Chengde show much better consistency with observations during summer. This indicates that in regions with low surface emissions, the accuracy of posterior simulations in summer is relatively high. Furthermore, it reinforces the finding that in highly polluted urban areas, the constraint of satellite NO_2 column measurements on surface emissions in summer is weaker, leading to an overestimation in posterior simulations."

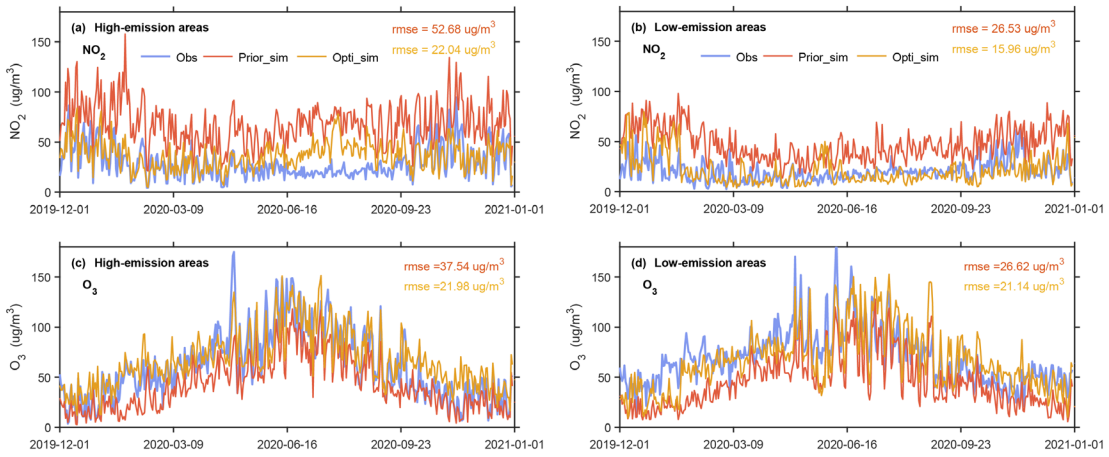

Figure S6. Comparison of the ground-based daily (a) NO_2 and (b) O_3 concentrations with the CMAQ 3-km simulations utilizing prior and posterior NO_x emissions as the model input. High-emission areas refer to the average of station measurements in the urban areas of Beijing, Tianjin, Baoding, Langfang, and Tangshan, while Low-emission areas refer to the average of station measurements in Zhangjiakou and Chengde.

2. Page 4, Lines 118-119: "The MEIC emission inventory is spatially and temporally allocated to match the CMAQ model domain using spatial proxies and empirical temporal profiles." What are the temporal and spatial resolutions of the MEIC inventory? What types of spatial proxies and temporal profiles are used to allocate emissions to the CMAQ model domain? Please elaborate further on these details in the paragraph.

Thanks for your good suggestion. We have added the details on the spatiotemporal allocation method for the MEIC emission inventory using spatial proxies and temporal profiles in Lines 126-132 (Sec. 2.1), as follows:

"The spatial proxies include total population density, urban population density, rural population density, and road length. These spatial proxies are updated annually to reflect interannual changes. The temporal profiles are unique for each major emission source. The

monthly profiles capture both seasonal variations and interannual trends in emissions, reflecting real activity levels. The allocation from monthly to daily values is achieved using sector-specific profiles that incorporate weekly and workday variations. In the CMAQ model, the MEIC inventory is mapped to the CMAQ model grids. Emissions from point sources are directly assigned to the grid cells where they are located, while emissions from area sources are first allocated to 1 km × 1 km grid cells based on the spatial proxies and then aggregated to the model grids based on WRF-CMAQ grid parameters."

3. Page 5, Lines 151-152: Does the inversion system presented in this study scale prior emissions on a daily scale? If so, how does the inversion system address hourly variations in NO_x emissions? Does the inversion system adjust the hourly profiles of the bottom-up emission inventory? Please provide additional details on the time steps used for assimilating TROPOMI NO_2 data to scale prior emission inventories.

Thank you for your valuable comments. Yes, the inversion system presented in this study adjusts prior emissions on a daily scale. Due to the once-daily overpass of the TROPOMI satellite, we are unable to utilize TROPOMI observations to resolve hourly variations in NO_x emissions. For hourly variations in NO_x emissions, we still follow the daily variation pattern in the prior MEIC inventory, using hourly profiles for temporal emission allocation. The assimilation time step for scaling prior emissions using TROPOMI NO_2 data is set to one day. Specifically, the differences between satellite observations and model simulations (spatiotemporally collocated for afternoon overpasses) are used to update the daily scaling factors of emissions.

We have clarified this issue in Sec. 4.3 and discussed the future outlook of using GEMS satellite hourly observation data for hourly emission estimation (Lines 494-496 and Lines 499-502).

4. Page 9, Lines 249-250: I recommend including additional error metrics, such as mean percentage error, to further illustrate the improvement in posterior emissions simulations. This would help address the question, "In which season is the most significant improvement observed after inversion?"

Thanks for your good suggestion. We chose to use the percentage change in the root mean square error (RMSE) to quantify the improvement rate of posterior emissions relative to prior emissions, which provides a more intuitive understanding that the improvements are greater in the summer and autumn seasons compared to winter and spring. We made the following modifications in Sect. 3.1.1, Lines 284-286:

"Compared to the prior simulations, the RMSE differences between the posterior simulated NO_2 concentrations and the observations were reduced by 58.32%, 59.10%, 69.73%, and 70.03%, respectively."

5. Page 9, Lines 251-252: What spatial proxies are used in MEIC? For example, does it utilize road network shapefiles? Providing specific examples would make this argument more compelling and relevant.

The spatial proxies include total population density, urban population density, rural population density, and road length. We have clarified this in Lines 287 and 125-126.

6. Page 9, Lines 253-254: "However, the NO_2 TVCDs from prior simulations indicate substantial overestimations in urban environments across various seasons…" Please specify the seasons or months to provide clarification.

Done. We have clarified this in Line 296 (Sect. 3.1.1).

7. Page 10, Lines 282-286: Why doesn't the simulated O_3 concentration exhibit the "summer bias" that is clearly evident in the comparison between simulated NO_2 and observations? Please provide a more detailed discussion of the factors that could explain the differences between the simulations of NO_2 and O_3.

Thanks for your good suggestion. We have added the discussion of the factors that could explain the differences between the simulations of NO_2 and O_3 in Lines 340-347 (Sect. 3.1.2), as follows:

"The overestimation of bottom-up NO_x emissions leads to negative biases in O_3 simulations throughout the year. However, during the summer, the deviations between simulated O_3 concentrations and observations are less pronounced compared to those observed in NO_2 simulations. This is because the simulation of NO_2 is significantly influenced by surface NO_x emissions. In contrast, simulated O_3 concentrations are affected by multiple factors, with NO_x emissions being only one of them. For instance, during summer, increased biogenic emissions and enhanced photochemical activity play critical roles. As a result, even in the prior O_3 simulations, the discrepancies between modeled and observed O_3 values are less pronounced compared to those in NO_2 simulations. Although the posterior O_3 simulations align well with observations, they remain slightly below the observed values, indicating that the posterior NO_x emissions may still be overestimated during the summer."

8. Page 10, Lines 295-296: "However, the posterior emission maps substantially reduce emissions from city centers and reallocate these emissions to other areas, such as increasing emissions from inter-city transportation, among other changes." This is an important finding, but it requires more supporting evidence. From Figure 4 alone, it is challenging to identify the locations of inter-city transportation, making it difficult to confirm whether the reduced emissions from city centers are reallocated to road networks. Consider including an additional figure that overlays the locations of major inter-city roadways with the areas of increased emissions in the posterior estimates.

Thanks for your good suggestion. We have added a spatial distribution map of emissions for 2020 in the supplementary materials, overlaying road information on the third column of the figure (see Fig. S7). The roads include national highways, provincial roads, and expressways. Since these three types of roads cover the main arterial network, we did not display county and rural roads to ensure the clarity of the figure.

Combined with the difference map between posterior and prior emissions in the third column of Fig. 4, the results show reduced emissions in urban areas, while certain roads exhibit

significant emission increases. Notably, the emission increases along roads are most pronounced during spring and summer.

We have made revisions in Sect. 3.2.1 (Lines 356-359):

"However, the comparison between posterior and prior emissions (third column of Fig. 4 and Fig. S7) reveals that the posterior emission maps substantially reduce emissions in urban centers and redistribute these emissions to other regions. For instance, emissions from inter-city transportation (see the third column of Fig. S7 with the overlay of the road map) are notably increased, particularly during spring and summer."

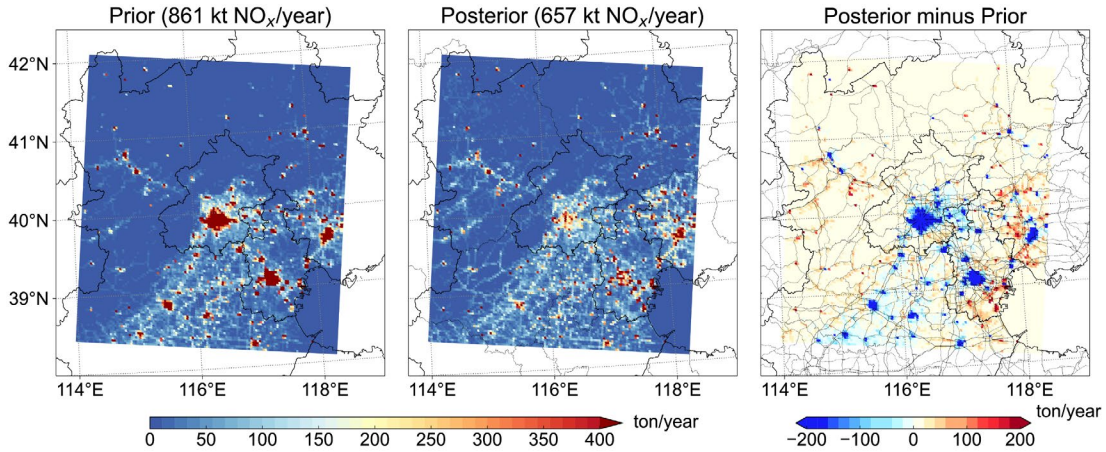

Figure S7. The spatial map of the prior and posterior emissions in 2020. The last column displays the differences between the two emissions, overlaid with a road map that includes national highways, provincial roads, and expressways.

9. Page 10, Lines 301-302: "The posterior NO_x emissions for the year 2020 (657 kt NO_x) decreased by 23.7% compared to the prior inventory (861 kt NO_x). The largest reductions occurred in winter and autumn, with declines of 44.5% and 36.4%, respectively." Does the bottom-up emission inventory (prior) account for the impact of COVID-19? If so, please provide an explanation in the paragraph.

Thanks for your good suggestion. We have clarified this in Lines 373-375 (Sect. 3.2.1)

"The bottom-up prior inventory is based on actual activity level data and emission factors, which can somewhat reflect the emission reductions during the COVID-19 lockdown period. However, due to statistical errors and the misrepresentation of emission factors, the prior emission inventory still fails to provide an accurate estimate of the regional total NO_x emissions."

10. Page 11, Lines 319-321: "The seasonal variation of the posterior NO_x emission estimate in our research is similar to the results obtained by previous studies (Wang et al., 2007; Qu et al., 2017; Miyazaki et al., 2017). Qu et al. (2017) utilized OMI measurements to infer the NO_x emissions in China, and the seasonal pattern of NO_x emissions for China and Beijing City is consistent with our study." How similar are these findings? Consider adding quantitative metrics to describe the seasonality of NO_x emissions as observed in this study and in previous studies.

Sorry for not explaining it clearly. We found that in the studies listed (Wang et al., 2007; Qu et al., 2017; Miyazaki et al., 2017), the seasonal variation pattern of NO_x emissions derived from satellite observations is similar to our study. Specifically, the posterior emissions based on satellite observations show significant seasonal variation, in contrast to the relatively insignificant seasonal changes in the prior emission inventories. Additionally, while the prior emissions tend to have the lowest levels in summer, the posterior emissions are higher in summer, comparable to or even higher than winter emissions (e.g., see Figures 12 and 14 in Qu et al., 2017). We compared the seasonal variation plots from these studies, but since the study areas differ in size, we did not perform a quantitative comparison of the changes in total regional emissions.

We have made modifications in Lines 412-414 (Sect. 3.2.2), as follows:

"These studies, along with the results of our study, indicate that while the prior emissions tend to have the lowest levels in summer, the posterior emissions are higher in summer, comparable to or even higher than winter emissions."

11. Page 12, Lines 364-365: "To evaluate the impact of different L values on the NO_x emission inversions, we perform two additional experiments with L = 3 km (Exp_L3km) and L = 81 km (Exp_L81km), respectively." What are the reasons for choosing these specific L values, 3 km and 81 km, for the sensitivity analysis? Please provide an explanation.

Thanks for your good suggestion. Firstly, based on theoretical analysis, we initially determined that a localization radius of 36 km is appropriate. As elaborated in Section 2.4, the selection of this radius (36 km) was guided by the typical lifetime of NO_2 (~4 hours) and wind speed (~3 m/s) in the Beijing region (Wu et al., 2021) (see Lines 205-208 in our manuscript). However, this value serves only as a preliminary estimate, as the lifetime of NO_2 varies with environmental conditions and wind speed is not constant. To assess the sensitivity of emission inversion results to localization radius, we tested values smaller and larger than 36 km. We chose 3 km because the resolution of our model is 3 km, and this choice can be roughly understood as ignoring the effects of emission transport between grids. The choice of 81 km is based on the typical lifetime of NO_2 and wind speed, providing a relatively larger transport distance. Additionally, following the suggestion of Referee #2, we added another experiment with L = 10 km to test whether the results would be better than those from the L = 36 km experiment. We have made modifications in Sect. 4.2 (see Lines 466-481).

12. Figure 5: Consider adding a visual marker to highlight the implementation and relaxation of COVID-19 containment measures, as well as notable events such as the Chinese Lunar New Year holiday.

Thanks for your good suggestion. We have made revisions according to your comments as below (see Lines 380-386 and Fig. 5).

"The reduction of NO_x emissions due to the pandemic lockdown measures lasted from early 2020 to mid-March 2020, with emission levels gradually returning to normal by late April. However, although the prior emission inventory partially reflects real monthly production activity levels, it fails to accurately capture such dynamic changes in emissions. The prior emission inventory only distinctly captures the emission reduction that occurred from early

February to mid-February 2020, mainly caused by the Chinese Lunar New Year. In addition, the posterior emission estimates also indicate a period of emission reduction and rebound from mid-June to mid-July 2020, coinciding with the sudden outbreak of the epidemic and the subsequent lockdown measures implemented at the Xinfadi market in Beijing, China."

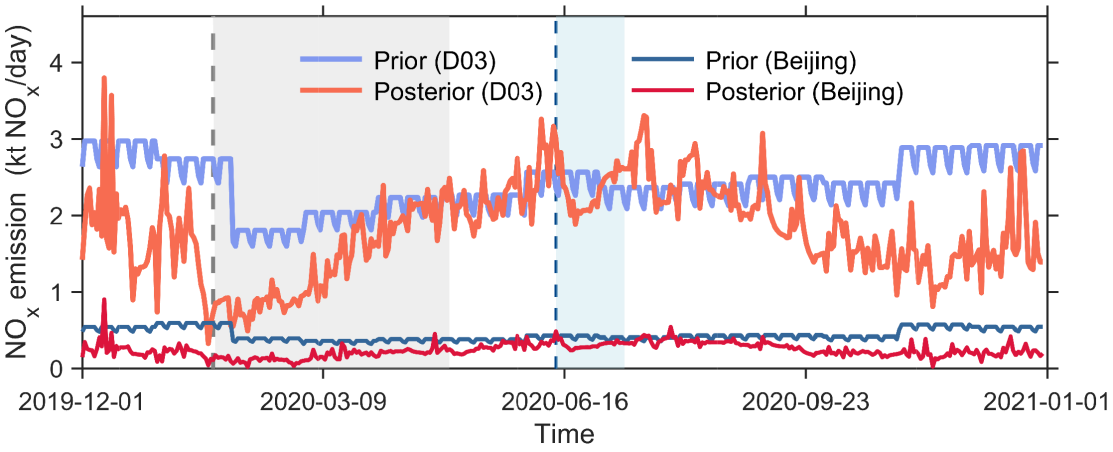

Figure 5. Time series of the bottom-up and top-down daily NO_x emissions for domain D03 and Beijing City. The gray dashed line indicates the Chinese Lunar New Year, which also marks the date when Beijing began implementing COVID-19 control measures. The blue dashed line represents the start date of control measures following the sudden outbreak at the Xinfadi market in Beijing. The gray shaded area represents the period affected by COVID-19 measures in 2020, and the light blue shaded area highlights the time frame impacted by the Xinfadi market outbreak.

13. Figure 5: It appears that the prior emission inventory exhibits a consistent diurnal cycle of hourly NO_x emissions. How does this compare to the diurnal cycle in the posterior NO_x emissions? Does the inversion system reveal a similar pattern? Please consider adding a figure and/or a paragraph to discuss this comparison.

Thank you for your valuable feedback and comments. Actually, Fig. 5 shows the time series of daily emissions, not hourly NO_x emissions. The daily time series of the prior inventory shows periodic variations between weekdays and weekends, i.e., the MEIC inventory shows a clear weekly pattern. This is because, for the prior inventory, the allocation from monthly emissions to daily values is achieved using sector-specific profiles that incorporate weekly and workday variations. However, our satellite-derived emissions do not display a distinct weekly pattern. Previous studies on satellite observations and ground concentration variations have also indicated that such a weekly pattern is not prominent in China (Wei et al., 2022).

As for the MEIC inventory, it does include a diurnal cycle of hourly NO_x emissions, as the MEIC inventory uses hourly profiles to allocate hourly emissions. However, since the TROPOMI satellite observations used in this study only provide afternoon overpasses per day, they do not support the derivation of hourly emission variation patterns. We have clarified this issure and added a discussion on the future aim to utilize GEMS hourly data to reveal the hourly variation patterns of emissions in Sect. 4.3 (Lines 494-496 and Lines 499-502).

References

Wang, Y., McElroy, M. B., Martin, R. V., Streets, D. G., Zhang, Q., and Fu, T.-M.: Seasonal variability of NO_x emissions over east China constrained by satellite observations: Implications for combustion and microbial sources, Journal of Geophysical Research: Atmospheres, 112, https://doi.org/10.1029/2006JD007538, 2007.

Miyazaki, K., Eskes, H., Sudo, K., Boersma, K. F., Bowman, K., and Kanaya, Y.: Decadal changes in global surface NO_x emissions from multi-constituent satellite data assimilation, Atmos. Chem. Phys., 17, 807-837, 10.5194/acp-17-807-2017, 2017.

Qu, Z., Henze, D. K., Capps, S. L., Wang, Y., Xu, X., Wang, J., and Keller, M.: Monthly top-down NO_x emissions for China (2005–2012): A hybrid inversion method and trend analysis, Journal of Geophysical Research: Atmospheres, 122, 4600-4625, https://doi.org/10.1002/2016JD025852, 2017.

Wu, N., Geng, G. N., Yan, L., Bi, J. Z., Li, Y. S., Tong, D., Zheng, B., and Zhang, Q.: Improved spatial representation of a highly resolved emission inventory in China: evidence from TROPOMI measurements, Environ Res Lett, 16, ARTN 084056
10.1088/1748-9326/ac175f, 2021.

Wei, J., Liu, S., Li, Z., Liu, C., Qin, K., Liu, X., Pinker, R. T., Dickerson, R. R., Lin, J., Boersma, K. F., Sun, L., Li, R., Xue, W., Cui, Y., Zhang, C., and Wang, J.: Ground-Level NO_2 Surveillance from Space Across China for High Resolution Using Interpretable Spatiotemporally Weighted Artificial Intelligence, Environ. Sci. Technol., 56, 9988–9998, https://doi.org/10.1021/acs.est.2c03834, 2022