**Response to Referee #1:** We would like to thank the referee for the careful review throughout the paper that help to improve our paper.

Our Reply follows (*the referee's comments are in italics and blue*)

*General Comments*

*Representation error has posed a challenge in achieving consistent comparisons between models and ground-based observations. This issue arises because model grids are relatively coarse, whereas site-specific observations are locally representative, especially in heterogeneous environments targeting short-lived pollutants. This manuscript addresses this problem for NO2 by introducing a land-use-based representative (LUBR) observational operator, enabling the processed NO2 observations to better represent the means of 0.5x0.625 grid cells. This algorithm is proved effective for short-lived NO2 and is well evaluated in the paper. This method is helpful for accurately interpreting the bias between models and ground-based observations and is applicable to data assimilation research. I recommend this manuscript for publication once the issues outlined below are addressed.*

*Major Comments*

*An assumption underlying this LUBR algorithm is that observations from urban/rural sites can represent the average conditions of the entire urban/rural areas within this grid cell, which is not necessarily accurate. In other words, this algorithm only partially corrects the representation error, a point that needs clarifications.*

**Reply:** Thanks for the comments and this was not clearly explained in our previous version. It is indeed true that various factors, including meteorology, climate, and land cover, can influence representation errors. We have now added remarks to illustrate that there are also weaknesses of the LUBR algorithm and discussed prospects for future research in the ***Conclusion*** section. To clarify this, remarks are now added in page 19, line 8-12 by saying: "***The LUBR algorithm, though effective, doesn't fully correct the representation error as urban/rural sites cannot fully represent the average conditions of the entire urban/rural areas within this grid cell. Future endeavors could explore employing deep learning models to reveal the intricate relationship between the average conditions of grid cells and various factors beyond urban/rural sites, such as meteorology, climate, and land cover.***"

*In section 2.3, the authors compare modeled surface NO2 with ground-based observations, and modeled NO2 column with OMI observations. They note an inconsistent performance of model in simulating surface and column NO2. In their interpretation throughout the paper, satellite* *observations are considered more representative, and the model-to-satellite bias is treated as the true bias for simulating NO2. However, this assumption may not be accurate, for reasons that are listed below. It is important to address these issues throughout the paper, although they do not compromise the paper's overall conclusion. (1) Satellite observations have their own representative issues and should be treated carefully. OMI provides observations only for the 1-2pm overpassing window and are most reliable under clear-sky conditions, when chemistry/meteorology might differ from monthly means. OMI retrievals require a prior NO2 profile shape, which can be a major source of retrieval error. A consistent comparison between OMI and GEOS-Chem requires the same sampling process for modeled NO2, and replacing the a prior NO2 profile shape in OMI retrieval with one simulated by GEOS-Chem. Only after these processes can the bias between the resampled model and reprocessed retrieval be considered the actual bias between the model and satellite observations. It appears that in this paper, the authors lack this preprocessing before determining the model-retrieval bias. This should be corrected. (2) Even with a correctly determined bias between model and satellite observations, it does not imply that this bias will align with the bias between model and ground-based observations. This is because satellites measure column density of NO2, capturing information not just from the surface but also from the troposphere and stratosphere (I assume they use total column density which includes stratospheric contribution – this needs to be clarified in the paper). Thus, it is entirely reasonable for column bias to differ from the surface bias. The authors should not regard column bias as the true bias for ground-level comparisons.*

**Reply:** Thanks for the in-depth comments, and we acknowledge that the model-to-satellite bias should not be considered as the true bias. Regarding your first point, we now recognize the significance of ensuring the same sampling time for modeled $NO_2$ and clear-sky conditions for OMI retrievals. Additionally, we now understand the importance of Aerosol Air Mass Factors in OMI retrievals and the necessity of substituting the priori $NO_2$ profile shape in OMI retrieval with one simulated by GEOS-Chem. It's also common in previous works to compare satellite data products with CTM directly. We would like to list some of these works directly using OMI standard products to compare with GEOS-Chem column concentrations below:

Wang, Yi, Jun Wang, Xiaoguang Xu, Daven K. Henze, Zhen Qu, and Kai Yang. "Inverse modeling of SO 2 and NO x emissions over China using multisensor satellite data–Part 1: Formulation and sensitivity analysis." Atmospheric chemistry and physics 20, no. 11 (2020): 6631-6650. https://acp.copernicus.org/articles/20/6631/2020/

Chen, Youfan, Lin Zhang, Daven K. Henze, Yuanhong Zhao, Xiao Lu, Wilfried Winiwarter, Yixin Guo et al. "Interannual variation of reactive nitrogen emissions and their impacts on PM2. 5 air pollution in China during 2005–2015." Environmental Research Letters 16, no. 12 (2021): 125004. https://iopscience.iop.org/article/10.1088/1748-9326/ac3695/meta

Wang, Zhe, Itsushi Uno, Keiya Yumimoto, Syuichi Itahashi, Xueshun Chen, Wenyi Yang, and Zifa Wang. "Impacts of COVID-19 lockdown, Spring Festival and meteorology on the NO2 variations in early 2020 over

China based on in-situ observations, satellite retrievals and model simulations." Atmospheric environment 244 (2021): 117972. https://www.sciencedirect.com/science/article/pii/S1352231020307068

Considering the importance of updating NO2 profiles, we utilize the OMI L2 product instead of L3 to realize them. Besides, we resampled the GEOS-Chem modeled $NO_2$ to maintain consistency with the OMI local overpassing window (13:00-14:00 pm). Further details regarding the OMI product will be provided in your next major comment. To clarify this, comments remarking this are adding in page 6, line 7-24 by saying: "***The following filters of pixels are applied, following Dang et al. (2023): (1) nearly clear-sky scenes, with effective cloud5 fraction < 0.3; (2)surface reflectivity < 0.3; (3)solar zenith angles < 75°; (4)viewing zenith angles < 65°. In addition, we also ensure that the 'vcdQualityFlag' possesses an even integer value to align with recommended data quality standards. The air mass factor (AMF) converts the satellite-observed slant column density (SCD) into the vertical column density (VCD) using the $NO_2$ vertical profile (n) as follows:***

$$VCD = \frac{SCD}{AMF(n)} \tag{1}$$

***AMF is mainly determined by atmospheric path geometry, $NO_2$ vertical profile, surface reflectance, and atmospheric radiative transfer properties. $NO_2$ exhibits optical thinness in the visible spectrum, facilitating the calculation of AMF (Lamsal et al., 2014). This calculation involves altitude-dependent scattering weights (sw) derived from a radiative transfer model and a priori profile shape of $NO_2$ as follows:***

$$AMF = \frac{\sum_l sw \cdot x_a}{\sum_l x_a} \tag{2}$$

***where xa is the partial $NO_2$ column, l denotes each layer, extending either from the ground to the tropopause or from the tropopause to the stratropopause. We updated the AMF of both tropopause and stratropopause separately using the $NO_2$ vertical profile simulated by GEOS-Chem in this study. The total column $NO_2$ concentration is calculated as the sum of the updated tropospheric vertical column density and stratospheric vertical column density. We regridded the total column amount of $NO_2$ to match the horizontal resolution of GEOS-Chem used in this study, which is 0.5 degrees latitude by 0.625 degrees longitude. Note that for comparison with OMI observations, we restrict our analysis to the time window between 13:00 and 14:00 local time, ensuring consistency with the OMI observation window.***" And in page 10, line 8-10 by saying: "***We averaged the model output between 13:00 and 14:00 local time for consistency with the timing of the Aura overpass for comparison with OMI observations.***" We also re-plotted the comparison figure to ensure that the model output is consistent with the sampling time of OMI, as shown in ***Fig. 4***.

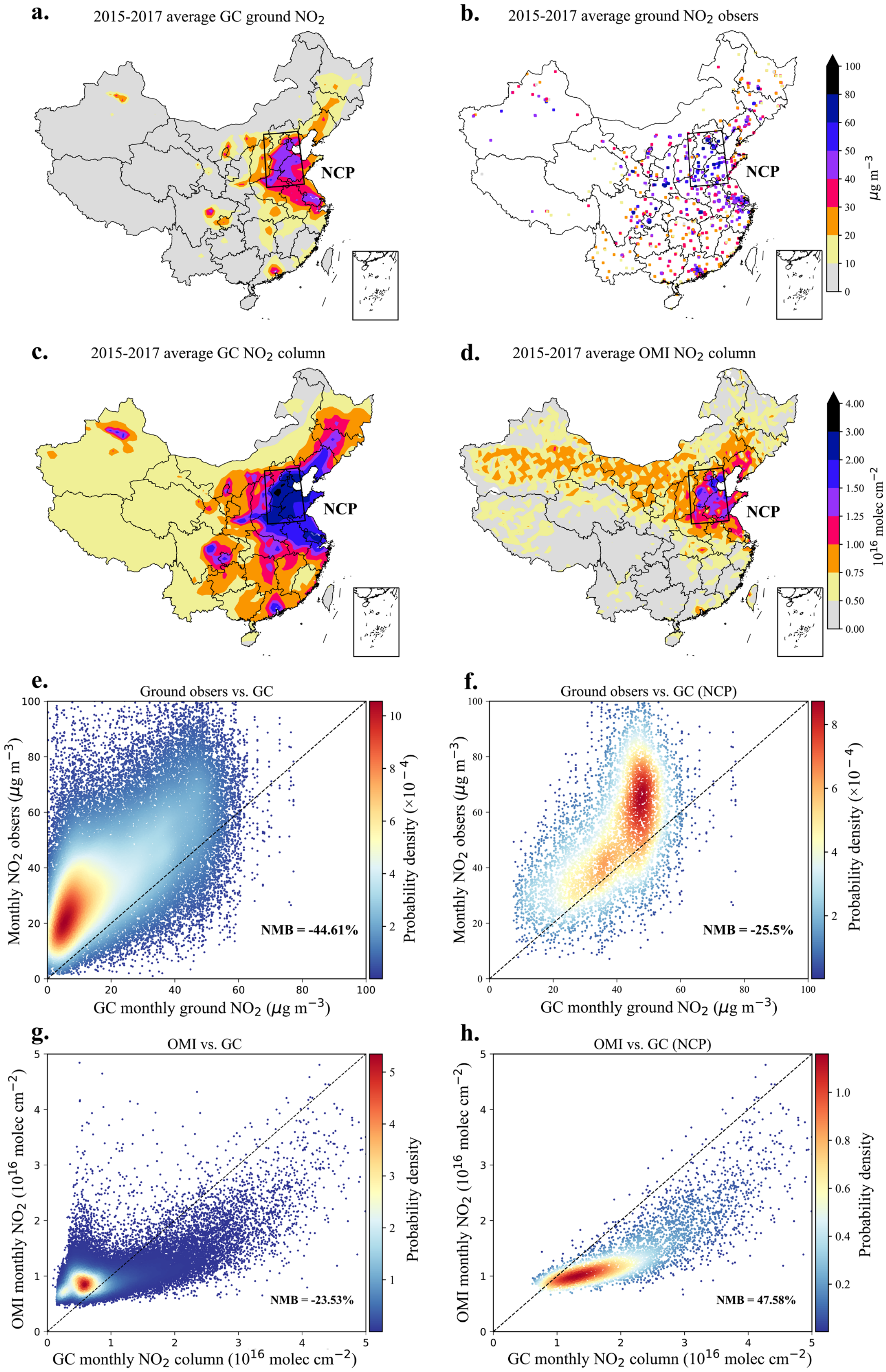


***Figure 4. The inconsistency between the observations and GEOS-Chem simulations is evident. Panels a and b depict the spatial distribution of ground-level $NO_2$ from GEOS-Chem and monitoring sites, while panels c and d***

***show the distribution of column-level $NO_2$ from GEOS-Chem and OMI. The NCP region, depicted by the black box, exhibits the most severe $NO_2$ pollution. The ground observations and model simulations represent the average conditions between 13:00 and 14:00 local time from 2015 to 2017. Panels e and g display scatter plots of the GEOS-Chem simulations and observations (monthly value), while panels f and h focus on the NCP region.***

Regarding your second point, we will not consider column bias as the true bias for ground-level comparisons but rather treat it as a point of comparison. Despite utilizing total column density of $NO_2$ data, which encompasses bias in both the troposphere and stratosphere, given the robust ability of $NO_2$ simulations in GEOS-Chem, we anticipate the overall tendency be similar. Remarks concerning this point are now added in page 11, line 3-9 by saying: "***The bias arises from uncertainties in both the retrieval algorithms of OMI products and the simulation of GEOS-Chem. For instance, Shah et al. (2020a) compared two OMI $NO_2$ retrievals, namely the European Quality Assurance for Essential Climate Variables (QA4ECV) project's $NO_2$ ECV precursor product (Boersma et al., 2018) and the Peking University POMINO product version 2 (Lin et al., 2015), with GEOS-Chem. They found that GEOS-Chem overestimates OMI $NO_2$ when using the QA4ECV retrieval, while underestimating it when using POMINO. In addition, MEIC tends to overestimate NOx emissions in cities with lower industrial emission intensities or fewer industrial facilities (Wu et al., 2021), which may contribute to the overestimation of GEOS-Chem in these areas.***" And in page 11, line 28-34 and page 13, line1-2 by saying: "***In panels (f) and (h) of Fig. 3, inconsistencies between observations and GEOS-Chem simulations in the NCP are evident: GEOS-Chem underestimates ground-level $NO_2$ while overestimating $NO_2$ column concentrations. Although the bias between model and satellite observations may not align with the bias between model and ground-based observations, as satellites measure the column density of $NO_2$, which captures information not only from the surface but also from the troposphere and stratosphere, it's worth noting that considering the model is the same and is popular and reliable, they should not diverge in opposite directions. The spatial disparity between model simulations and ground observations can indeed result in poor representation of grid cell observations, which is certainly one of the reasons for the differences. Therefore, our work primarily focuses on correcting the representativeness of ground observations and ensuring that the true correction direction closely aligns with the comparison results between model and satellite observations.***" And in page 13, line 6-7 by saying: "***This result aligns more closely with the trend of comparing GEOS-Chem and OMI observations.***"

*Additional information is needed regarding OMI product. Why was the OMAEROe product chosen? How does this product perform in comparison to ground based NO2 column observations and to other more popular OMI NO2 products, such as OMNO2 from NASA or the OMI product from KNMI?*

**Reply:** Thanks for point this, we apologize for misusing the OMI product, mistaking it for the one used in another study. In fact, the OMI product we used here is OMNO2 from NASA. Remarks concerning the detailed information are now added in supplement page 5, line 10-14 and page 6, line 1-6 by saying: "***Launched aboard the NASA EOS Aura satellite on July 15, 2004, OMI operates within a sun-synchronous ascending polar orbit. OMI conducts simultaneous measurements across a swath spanning 2600 km, partitioned into 60 Fields of View (FOVs). These FOVs range in dimension from approximately 13km x 24km near nadir to around 24km x 160km at the outermost FOVs. OMI provides observations only around 13:45(local time) overpassing window and is most reliable under clear-sky conditions. The NO2 total column concentrations utilized in this study were sourced from NASA Goddard Space Flight Center, specifically from the Goddard Earth Sciences Data and Information Services Center (GES DISC), through the OMI/Aura Nitrogen Dioxide Total and Tropospheric Column 1-orbit L2 Swath 13x24 km V003 (OMNO2) (Krotkov et al., 2019). The OMI $NO_2$ algorithm retrieves estimated columns (total, tropospheric, and stratospheric) of nitrogen dioxide from OMI Level-1B calibrated radiance and irradiance data. The current version, v4.0, improves on the retrievals in prior versions in several significant ways. The OMNO2 algorithm aims to infer as much information as possible about atmospheric $NO_2$ from OMI measurements, with minimal dependence on model simulations.***"

*I am curious if the urban/rural factor exhibits any seasonality, considering the longer lifetime of NO2 during winter compared to summer? Can soil NOx emissions during summer (dominant in rural areas?) influence the urban/rural factor? Please consider adding a discussion on this.*

**Reply:** Thanks for the suggestion. We consider it worthwhile to investigate the seasonality of urban/rural factors in future studies. Despite conducting searches for relevant papers, we could only find research focused on the city scale. Based on this study, we plotted the seasonal variation of $NO_2$ observations based on three years of ground observations from 2015 to 2017 in below ***Fig. 0***. We observe that the urban/rural factor tends to be larger in spring and summer compared to autumn and winter, which is contrary to its expected lifetime. However, the difference is not significant and does not totally consistent when research area changed. We think it could be due to the combined effects of various factors such as meteorological conditions, regional hotpots, human activities, biological sources, and topography. Therefore, it may be necessary to refine the research area and consider multiple factors rather than conclude solely from the ground observations.

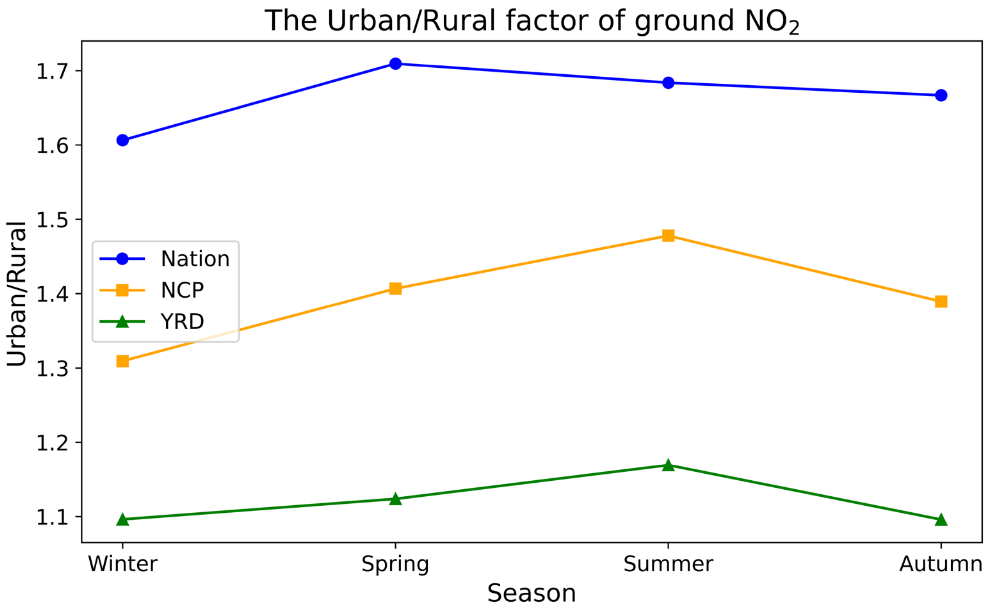


***Figure 0. The seasonal variation of the urban/rural factor of ground $NO_2$ concentrations. Each season is calculated from the average of 2015-2017 data, with blue, orange, and green representing the study areas of the Nation, NCP, and YRD respectively.***

For your second question, as reported by Lu et al., "The intensive nitrogen inputs to soil from fertilizer applications and nitrogen deposition lead to large soil NOx emissions via microbial processes, reaching 20% of the anthropogenic NOx emissions in summer over the NCP." These soil NOx emissions during summer can exert a significant influence, particularly as they constitute the main source for rural areas. However, it is challenging to provide concrete evidence based on the available data because we cannot distinguish the sources of NOx. Nevertheless, it does remind us of the importance of refining the Urban/Rural factor in the future.

Therefore, we would like to express our gratitude for the referee's insightful comments once again, as this direction appears promising. In the future, we plan to utilize more relevant data and employ advanced statistical models to conduct further in-depth research on the Urban/Rural factor.

*I don't see the point of figure 4, as it appears to convey ideas similar to those presented in figure 5 or 6. Please consider removing one of these figures to remain concise for evaluation section.*

**Reply:** Thanks for the comments. Figure 4 (now ***Fig. 6***) is intended to be directly compared with Figure 3(now ***Fig. 2***) as they convey a consistent message. On the other hand, Figure 5 and Figure 6 (now ***Fig. 7*** and ***Fig. 8***) present spatially averaged results, which may not contain as much information but provide a clearer view of the overall changes. Therefore, although they convey similar ideas, they serve different purposes, and we would like to retain them.

**Reference**

***Lu, X., Ye, X., Zhou, M. et al. The underappreciated role of agricultural soil nitrogen oxide emissions in ozone pollution regulation in North China. Nat Commun 12, 5021 (2021). https://doi.org/10.1038/s41467-021-25147-9***