

Dear Editor and Reviewers,

We are grateful to the reviewers for their helpful comments. We have made the modifications in response to their comments. Attached is a point-by-point response to the comments. We hope that you and the referees will find the changes satisfactory, and we look forward to hearing from you soon.

### **RC1: 'Comment on egosphere-2024-558', Referee #1**

This study leverages advanced satellite measurements and machine learning techniques to estimate ground-level NO<sub>2</sub> concentrations in China. The use of the GEMS measurements combined with a nested machine learning model marks an advanced approach to addressing the challenge of translating satellite-derived VCDs of NO<sub>2</sub> into actionable ground-level concentration data. Incorporating the NMH into the prediction model not only demonstrates a methodological advancement but also highlights the crucial role of meteorological conditions in the dispersion of atmospheric pollutants. As the study achieves remarkable accuracy and provides comprehensive analyses of NO<sub>2</sub> distribution patterns, I recommend the publication of this paper for Atmospheric Chemistry and Physics after minor revisions.

Specific Comments:

#### **Comment 1:**

The planetary boundary layer (PBL), represented as NMH in this study, is identified as a significant factor influencing the conversion of VCDs of NO<sub>2</sub> to ground-level concentrations. Due to its importance as illustrated in Figure 5, there should be more discussions on the relationship between PBL and surface air pollution. I also suggest the authors acknowledge the previous study investigating the Relationships between the PBL and surface pollutants over China, as well as the influencing factors.

Response: Thank you for your valuable comments. We have added a new paragraph in the Introduction section (lines 89-103) and a new paragraph in the Discussion section (lines 485-505) to provide more elaboration on the impacts of the PBL on air pollution.

“Numerous past studies have highlighted the importance of the boundary layer structure in governing the occurrence and evolution of extreme air pollution episodes (Shi et al., 2020). A significant relationship between a surge in surface air pollutant concentrations and a shallow PBLH has been extensively reported (Miao et al., 2019; Su et al., 2020). It has also been recognized that air pollutants aloft can play a core role in the evolution of surface extreme pollution episodes via vertical mixing (Zhang and Rao, 1999). When the top of the mixing layer reaches the aloft pollutant-rich layer during the daytime, air pollutants can be entrained downwards, which rapidly increases surface air pollutant concentrations (Zhang et al., 2016). In addition to the vertical exchange, radiative absorption and scattering by pollutants can modify the boundary layer structure and consequently affect ground-level pollutant concentrations. For instance, high loadings of scattering pollutants can cool the air near the ground and result in a more stable boundary layer, which further worsens air quality (Li et al., 2017). As a result, the PBLH has been used as a proxy of the NMH because of its ability to regulate near-surface pollution levels. However, as NO<sub>2</sub> may not be uniformly distributed within the planetary boundary layer, a significant difference may exist between the PBLH and NMH. It is important to develop a conversion model that directly considers the impacts of the NMH, which paves the way to refine the processes of converting satellite-derived columnar measurements into ground-level NO<sub>2</sub> concentrations (Ahmad et al., 2024).”

“PBL characteristics are pivotal in regulating the vertical dispersion and horizontal transport of atmospheric pollutants, subsequently determining the vertical variations of NO<sub>2</sub> and its concentration at the Earth's surface (Akther et al., 2023; Xiang et al., 2019). Results in this study highlight the key role of the mixing height of NO<sub>2</sub> in linking satellite-derived VCDs of NO<sub>2</sub> with ground-level concentrations. To convert the VCDs of NO<sub>2</sub> into ground-level NO<sub>2</sub> concentrations, previous conversion models have used PBLH as a proxy of the NMH, because of its ability to regulate ground-level pollution levels. For example, within a stable PBL, pollutants like NO<sub>2</sub> from ground sources mainly accumulate near the

ground surface (Levi et al., 2020). Intense solar heating can induce elevated temperatures, fostering an unstable PBL that is conducive to the upward dispersion of air pollutants including NO<sub>2</sub> (Kalmus et al., 2022; Su et al., 2020). The wind pattern is connected to atmospheric stability and can impact NO<sub>2</sub> levels by modifying pollutants' dispersion and horizontal transport (Yin et al., 2019). High surface air pressure often leads to large-scale sinking air motion, resulting in the limited vertical diffusion of NO<sub>2</sub> (Chow et al., 2018). Elevated relative humidity levels act as a suppressive factor, constraining the PBLH and exacerbating the accumulation of pollutants near the ground (Xiang et al., 2019). Therefore, different meteorological factors significantly impact the vertical distribution of NO<sub>2</sub> in the atmosphere (Huang et al., 2021). This study developed a conversion model that directly considers the impacts of the NMH. The predictions of NMH from the inner model directly incorporated the impacts of meteorological parameters (T, P, WS, RH, DP, VIS, and PRECIP). It was found that temperature, wind speed, dew point, and visibility were positively correlated with NMH, while relative humidity and air pressure mainly demonstrated an inverse relationship (Ahmad et al., 2024). The atmosphere's dynamic and thermodynamic aspects played crucial roles in developing the vertical structure of NO<sub>2</sub>. The incorporation of the NMH in the model paved the way to refine the processes of converting satellite-derived columnar measurements into ground-level NO<sub>2</sub> concentrations.”

**Comment 2:**

Section 2.6 is the key section for this paper, since it present the details of machine learning model for this study. While the nested machine learning model demonstrates superior performance in estimating ground-level NO<sub>2</sub> concentrations, the methodology section could benefit from a more clear discussion of the advantage of XGBoost regression model, as well as feature selection process, and the rationale behind choosing specific meteorological parameters as predictors.

Response: Thanks for your comments. The XGBoost algorithm has proven to be useful in various air quality studies, including those focusing on the conversion between satellite-based column measurements and ground-level concentrations (Shao et al., 2023; Zhao et al., 2023). We have added a new paragraph to discuss the advantages of the XGBoost regression model in lines 197-214.

“XGBoost stands out as a notably efficient end-to-end gradient boosting tree framework, adept at transforming numerous weak learners into robust ones through boosting. This framework frequently demonstrates reduced computational overhead and enhanced predictive accuracy when compared with alternative ensemble tree models (Chen and Guestrin, 2016). Moreover, XGBoost exhibits a lower susceptibility to overfitting by mitigating the bias within the context of bias-variance decomposition. XGBoost has been empirically demonstrated to adeptly capture nonlinear relationships between predictions and predictors, yielding precise estimations through its regularized boosting methodology. This approach constructs the ultimate model by iteratively refining simpler and weaker models, each subsequent tree learning from its predecessors and updating residual errors via gradient descent to optimize the loss function. Within the XGBoost framework, an augmented penalty term is incorporated into the error function to fine-tune the objective function, thereby smoothing the final learned weights and mitigating overfitting tendencies. Additionally, to further mitigate overfitting, feature sub-sampling and shrinkage techniques are integrated (Liu 2021). The study by Van et al. (2022) also demonstrated the XGBoost algorithm as the most suitable lightweight algorithm based on the comparative analysis of three machine learning models, i.e., XGBoost, Decision Tree, and Random Forest. The XGBoost algorithm has proven to be useful in various air quality studies, including those focusing on the conversion between satellite-based column measurements and ground-level concentrations (Shao et al., 2023; Zhao et al., 2023). More details on the XGBoost regression model can be found in Chi et al. (2022). The XGBoost model was implemented in this study to convert columnar measurements into ground-level NO<sub>2</sub> concentrations.”

All common meteorological variables that are available from the ground monitoring network were used in this study. Therefore, we did not choose specific meteorological factors. The abilities of these meteorological variables

to regulate near-surface NO<sub>2</sub> levels are ranked as the feature importance in the XGBoost regression model. We clarified the use of meteorological variables in lines 225-232:

“All common meteorological variables available from the ground monitoring network were used in this study. The ability of these meteorological variables to regulate near-surface NO<sub>2</sub> levels is ranked by feature importance in the XGBoost regression model. In our previous study, these meteorological parameters were shown to impact the vertical mixing of NO<sub>2</sub> to varying extents (Ahmad et al., 2024). For instance, elevated temperatures are conducive to the upward mixing of air pollutants. Increased wind speed is associated with an unstable atmosphere and can impact NO<sub>2</sub> levels by modifying the vertical dispersion and horizontal transport of air pollutants. Increased surface air pressure often leads to large-scale sinking air motion, which suppresses the vertical dispersion of NO<sub>2</sub>.”

**Comment 3:**

The study mentions the challenges posed by cloudy conditions and the lack of nighttime data in interpreting GEMS measurements. While correction factors were applied to mitigate these issues, a more detailed discussion on the limitations and potential biases introduced by these factors would be beneficial. This discussion of limitations can be also included or mentioned in the conclusion section.

Response: Thank you very much for your comments. The correction factors were applied to mitigate the issues of missing data. We created a new section (Section 2.7) to summarize the calculation of correction factors. The correction factors are only based on the ground NO<sub>2</sub> measurements, which results in reduced and minimized biases associated with them. However, some limitations still exist, as these correction factors rely on an ancillary data source with a low spatial resolution. Spatially, the spatial distributions of the correction factors were obtained from interpolation of the ground monitoring data. We made the assumption that the correction factors vary smoothly in the areas between different stations. However, the atmospheric conditions and NO<sub>2</sub> emissions can vary significantly across different regions at different times of the day. Additionally, we applied a constant correction factor for seasonal and annual averages, which may not be able to correct the detailed bias from hour to hour. We add the limitations for the correction factors in lines 602-611:

“The lack of nighttime data and cloudy conditions leads to skewness in the GEMS measurements, especially for phenomena that exhibit diurnal variations. To align the satellite-estimated NO<sub>2</sub> with ground-measured NO<sub>2</sub>, correction factors were applied for hourly, seasonal, and annual averages (see Sec. 2.7). These correction factors are based solely on the ground NO<sub>2</sub> measurements, which results in reduced and minimized biases associated with them. However, some limitations still exist, as these correction factors rely on an ancillary data source with low spatial resolution. Spatially, the spatial distributions of the correction factors were obtained by interpolating the ground monitoring data. We made the assumption that the correction factors vary smoothly in the areas between different stations. However, atmospheric conditions and NO<sub>2</sub> emissions can vary significantly across different regions at different times of the day. Additionally, we applied a constant correction factor for seasonal and annual averages, which may not be able to correct the detailed bias from hour to hour.”

We also included this limitation in the conclusion (lines 645-647):

“Some limitations still exist, as these correction factors rely on an ancillary data source with low spatial resolution. Additionally, we applied a constant correction factor for seasonal and annual averages, which may not be able to correct the detailed bias that occurs from hour to hour.”

**RC2: 'Comment on egusphere-2024-558', Referee #2**

Overview:

This paper introduced a machine learning model to estimate ground-level NO<sub>2</sub> concentrations from geostationary satellite-derived NO<sub>2</sub> vertical column densities (VCDs). The overall conclusions are that utilizing NO<sub>2</sub> mixing

height (NMH) can improve the accuracy of ground-level NO<sub>2</sub> concentration estimates, and that satellite-derived ground-level NO<sub>2</sub> concentration presents a population-based gradient.

Although this manuscript provides a few pieces of information that I believe are suitable for publication, it is riddled with grammar and technical issues and requires major revisions. Extensive simple grammar corrections should not be on the peer reviewers to fix at this stage, and such issues did make it difficult to understand the authors' justification behind their conclusions. I also found the present document more like a technical report rather than a research paper, as plenty of scientific discussions are missing.

Major Comments:

**Comment 1:**

The weakest point in the manuscript is the discussion of the results. More than two-thirds of the 'Discussions' section repeats what have already been presented in the 'Results' section. The authors should expand more on the scientific principles underlying the results in the 'Discussions' section.

Response: Thanks for your valuable comments. We have made efforts to improve the grammar throughout the manuscript. Additionally, we have thoroughly revised the Discussion section. Specifically, we have added a paragraph at the beginning of the Discussion section to summarize the scientific contributions of this study (lines 472-484).

"The scientific contributions of this study are summarized as follows. First, the results of this study have contributed to enriching our scientific understanding of the relationship between columnar NO<sub>2</sub> and ground-level NO<sub>2</sub>. We have proven that the mixing height of NO<sub>2</sub> plays a key role in linking satellite-derived VCDs of NO<sub>2</sub> with ground-level concentrations, though the impacts of NMH were rarely considered in a direct manner in previous studies. Secondly, the analyses in this study have improved our understanding of the spatiotemporal variations of NO<sub>2</sub>, particularly the diurnal variations that cannot be obtained from common polar-orbiting satellite measurements. The diurnal variations in NO<sub>2</sub> concentration differ between urban and rural areas, resulting from the different emission sources and pollutant dispersion characteristics. Thirdly, the analyses of NO<sub>2</sub> variation have policy implications for air pollution control. It was found that the spatial coincidence between NO<sub>2</sub> concentrations and population density increased overall population exposure and the associated health impacts. This suggests that for more effective reduction of overall population exposure and better protection of public health, control efforts should be further targeted at highly populated and highly polluted areas. Additionally, land-use and city planning should encourage population redistribution away from the most heavily polluted regions."

Following the comments from the reviewers, we have expanded the discussions on the scientific principles underlying the results, focusing on the three key aspects mentioned above. First, we discussed the impacts of the mixing height of NO<sub>2</sub> (lines 485-537):

"PBL characteristics are pivotal in regulating the vertical dispersion and horizontal transport of atmospheric pollutants, subsequently determining the vertical variations of NO<sub>2</sub> and its concentration at the Earth's surface (Akther et al., 2023; Xiang et al., 2019). Results in this study highlight the key role of the mixing height of NO<sub>2</sub> in linking satellite-derived VCDs of NO<sub>2</sub> with ground-level concentrations. To convert the VCDs of NO<sub>2</sub> into ground-level NO<sub>2</sub> concentrations, previous conversion models have used PBLH as a proxy of the NMH, because of its ability to regulate ground-level pollution levels. For example, within a stable PBL, pollutants like NO<sub>2</sub> from ground sources mainly accumulate near the ground surface (Levi et al., 2020). Intense solar heating can induce elevated temperatures, fostering an unstable PBL that is conducive to the upward dispersion of air pollutants including NO<sub>2</sub> (Kalmus et al., 2022; Su et al., 2020). The wind pattern is connected to atmospheric stability and can impact NO<sub>2</sub> levels by modifying pollutants' dispersion and horizontal transport (Yin et al., 2019). High surface air pressure often leads to large-scale sinking air motion, resulting in the limited vertical diffusion of NO<sub>2</sub> (Chow et al., 2018). Elevated relative humidity levels act as a suppressive factor, constraining the PBLH and exacerbating the accumulation of pollutants near the ground (Xiang et al., 2019). Therefore, different

meteorological factors significantly impact the vertical distribution of NO<sub>2</sub> in the atmosphere (Huang et al., 2021). This study developed a conversion model that directly considers the impacts of the NMH. The predictions of NMH from the inner model directly incorporated the impacts of meteorological parameters (T, P, WS, RH, DP, VIS, and PRECIP). It was found that temperature, wind speed, dew point, and visibility were positively correlated with NMH, while relative humidity and air pressure mainly demonstrated an inverse relationship (Ahmad et al., 2024). The atmosphere's dynamic and thermodynamic aspects played crucial roles in developing the vertical structure of NO<sub>2</sub>. The incorporation of the NMH in the model paved the way to refine the processes of converting satellite-derived columnar measurements into ground-level NO<sub>2</sub> concentrations.

Two models were tested and trained: Model I, which did not consider NMH, and a nested Model II, which incorporated NMH. The validation results demonstrated that nested Model II exhibited more promising outcomes than Model I, suggesting that including NMH significantly influenced the model's performance. Including NMH as an input parameter in the machine learning model could better capture the vertical distributions of NO<sub>2</sub> and thus predict ground-level NO<sub>2</sub> concentrations with improved accuracy and performance. Additionally, the hour-by-hour 10-fold cross-validation depicted a distinct improvement in the ground-level NO<sub>2</sub> estimations for nested Model II considering NMH as an input parameter (Fig. S5 for Model I without NMH and Fig. S6 for nested Model II with NMH). The R<sup>2</sup> values for Model I without NMH were 0.63 for 08:00 AM, 0.70 for 09:00 AM, 0.69 for 10:00 AM to 01:00 PM, 0.55 for 02:00 PM, and 0.39 for 03:00 PM. The improved R<sup>2</sup> values for nested Model II, which includes NMH, were 0.85 for 08:00 AM, 0.90 for 09:00 to 11:00 AM, 0.91 for 12:00 PM, 0.93 for 01:00 PM, 0.89 for 02:00 PM, and 0.85 for 03:00 PM. Similarly, nested Model II, considering the NMH, depicted significantly reduced biases compared to Model I without NMH. The ground-level NO<sub>2</sub> estimations for all hours were significantly improved when considering NMH, as it directly incorporates the vertical distributions of NO<sub>2</sub>. During the early morning hours, most of the NO<sub>2</sub> is distributed near the ground. However, as the day progresses, NMH increases, and the ground-level NO<sub>2</sub> tends to be mixed vertically. Further, the improvements in ground-level NO<sub>2</sub> estimations were assessed using 10-fold cross-validation for different population categories, i.e., lightly populated, moderately populated, highly populated, and supremely highly populated. The nested Model II, considering NMH, depicted notable improvements compared to Model I without NMH (Fig. S7). The improved R<sup>2</sup> values for nested Model II considering NMH were 0.91 for lightly populated areas and 0.92 for the other three population categories compared to Model I without NMH, which depicted an R<sup>2</sup> value of 0.63 for lightly populated, 0.73 for moderately populated, 0.77 for highly populated, and 0.74 for supremely highly populated areas. The RMSE for nested Model II considering NMH was improved and observed below 5 µg/m<sup>3</sup> for all population categories compared to Model I without NMH, which depicted RMSE values around 8-9 µg/m<sup>3</sup> for different population categories. The MAPE for nested Model II considering NMH was also improved for all population categories, and around 15 % and lower values were observed. These improvements depict that nested Model II considering NMH effectively captures the spatial distributions of vertical mixing of ground-level NO<sub>2</sub> across all population categories. The spatiotemporal distributions and diurnal patterns of NMH are previously described by Ahmad et al. (2024). Compared to Model I without NMH, the performance of the ground-level NO<sub>2</sub> estimations through nested Model II considering NMH showed significant improvement at the grid points where ground-based observations were available (Fig. S8). The correlation coefficients for grid-based 10-fold cross-validation were improved to 0.8-1.0 for nested Model II considering NMH compared to Model I without NMH, which depicted lower correlation coefficients. Furthermore, nested Model II considering NMH also depicted lower RMSE values for grid-based estimations.”

Subsequently, we discussed the contribution of the new-generation geostationary satellite in improving our understanding of the spatiotemporal variations of air pollution, particularly the diurnal variations that cannot be obtained from common polar-orbiting satellite measurements (lines 538-571).

“GEMS, the world's first GEO-based environmental satellite instrument, offers a new opportunity for monitoring air quality across extensive regions, providing unprecedented spatial and temporal resolution. The quality of GEMS NO<sub>2</sub> VCDs, obtained from the level 2 product, has been evaluated using ground-

based instruments in various regions. Encouragingly, a good agreement has been observed between the GEMS NO<sub>2</sub> VCDs and measurements from various ground-based instruments (Ahmad et al., 2024; Kim et al., 2023; Li et al., 2023). The results presented in this study emphasize the significant advantage of geostationary satellites in providing air pollution information at an hourly resolution. They enable the assessment of diurnal variations in air pollution across different areas, ranging from lightly populated to supremely highly populated regions. This represents a substantial improvement over traditional LEO-based satellite instruments. Furthermore, these GEO-based measurements are valuable supplements to traditional measurements from ground-based air quality monitoring networks, primarily concentrated in urban areas, leaving vast rural regions without observations.

The diurnal variations of ground-level NO<sub>2</sub> concentrations across China depicted distinct gradients across all subregions and population categories. This gradient reflects regional disparities in industrialization, urbanization, and transportation infrastructure of Chinese megacities and rural areas. Highly populated areas depicted the highest concentrations of ground-level NO<sub>2</sub> during the early morning hours, attributed to intensified vehicular traffic in the early morning hours and higher industrial emissions. In contrast, lightly populated areas exhibited lower ground-level NO<sub>2</sub> concentrations and a delayed peak of around one to two hours, indicating lesser anthropogenic influence and more contribution from regional transport contributed by the NO<sub>2</sub> emissions from highly populated areas. Various driving factors influence these diurnal variations in ground-level NO<sub>2</sub> concentrations, each contributing differently across different regions. For instance, anthropogenic emissions dominate in highly populated urban and suburban areas, characterized by traffic emissions peaking in the morning and late afternoon (Liu et al., 2018; Naiudomthum et al., 2022). This phenomenon is particularly pronounced in highly populated areas with high traffic density. As morning rush hours subside, reduced vehicular traffic activities in highly populated areas lead to a gradual decline in NO<sub>2</sub> emissions. However, atmospheric processes such as higher mixing height of NO<sub>2</sub>, more dispersion, and dilution also come into play, resulting in reduced ground-level NO<sub>2</sub> concentrations. Increased turbulent mixing in the lower atmosphere helps disperse pollutants from their sources in highly populated areas, gradually decreasing ground-level NO<sub>2</sub> concentrations. Additionally, photochemistry also influences the diurnal variations of NO<sub>2</sub> concentrations. The ratio of NO<sub>2</sub> to NO is influenced by radiation, ozone, and peroxy radicals. During the daytime, NO<sub>x</sub> undergoes oxidation through radical-mediated reactions, forming nitric acid and organic nitrates, with their levels depending on radiation, ozone, and volatile organic compounds. As a result, the lifetime of NO<sub>2</sub> reaches its lowest point around noon, typically lasting a few hours during summer. Furthermore, atmospheric transport contributes to the diurnal variation of NO<sub>2</sub>, particularly in highly populated areas and their surrounding regions (Zhang et al., 2023). The hourly ground-level NO<sub>2</sub> concentration results presented in this study provide high-resolution information on the diurnal variations in ground-level NO<sub>2</sub> pollution levels across different regions and demographic patterns.”

Then, policy implications for air pollution control are discussed (lines 572-594).

“The spatial distribution of ground-level NO<sub>2</sub> concentrations in the study region revealed significant regional disparities, with higher levels observed in urban agglomerations with high population densities (e.g., BTH, YRD, and PRD regions) than in lightly populated areas (e.g., western China). Even within the NC region, the highly populated urban areas had NO<sub>2</sub> concentrations nearly double those of lightly populated rural areas. These spatial disparities are due to distributions of NO<sub>2</sub> emission sources that vary with population densities, decreasing from highly populated to lightly populated areas. In highly populated urban areas in regions like BTH, YRD, and PRD, mobile NO<sub>x</sub> emissions from dense road networks contribute to pronounced increase in NO<sub>2</sub> levels. Moreover, the short lifespan of NO<sub>2</sub> due to atmospheric chemical reactions results in elevated concentrations near emission sources in highly populated areas, such as roadways, accompanied by rapid declines in NO<sub>2</sub> concentrations with increasing distance from highly populated areas (Lee et al., 2018). Furthermore, the diverse terrains, land cover, and climates observed in subregions with different population categories collectively influence vertical and horizontal airflows, rates of NO<sub>2</sub> formation and deposition, and contribute to spatiotemporal variations in ground-level NO<sub>2</sub> concentrations between the highly populated and lightly populated areas across China. Additionally, the population-weighted mean NO<sub>2</sub> concentrations were consistently higher than

the spatial mean NO<sub>2</sub> concentrations in most provinces across China. This is due to the spatial coincidence between NO<sub>2</sub> concentrations and population density. These results indicate that the use of simple spatial average concentrations can lead to a systematic underestimation of overall population exposure and the associated health impacts. It is important to use high-resolution NO<sub>2</sub> data to accurately quantify true population exposure. Furthermore, the adverse impacts of high NO<sub>2</sub> concentrations in highly populated urban areas suggest that for more effective reduction of overall population exposure and better protection of public health, control efforts should be further targeted at highly populated and highly polluted areas. Targeted control programs to reduce pollutant levels at population hotspots should be more cost-effective than trying to reduce pollutant concentrations everywhere. Additionally, control policies can be implemented by encouraging the public to relocate to less polluted areas through land-use development and urban planning.”

#### Comment 2:

The title and abstract indicate that this paper aims at improving ground-level NO<sub>2</sub> estimation. However, the only figures that present such improvements are Figures 4 and 12. The manuscript also keeps talking about different patterns of ground-level NO<sub>2</sub> concentration between highly and lightly populated areas. But how the improvements differ between these regions (and at different hours of the day)? How the estimates perform at the grid points where ground-based observations are available?

Response: Thank you very much for your comments. In this study, we aimed to develop a nested model to improve the estimation of ground-level NO<sub>2</sub> and enrich our understanding of the spatial and temporal variations in NO<sub>2</sub> concentration using measurements from new geostationary satellite. The title of the manuscript has been revised as:

“Estimation of Ground-Level NO<sub>2</sub> and its Spatiotemporal Variations in China Using GEMS Measurements and a Nested Machine Learning Model”.

Additionally, the following part has been added to assess the improvements at different hours of the day, improvements between different regions and performance of estimates at grid points where ground-based observations are available (lines 506-537).

“Two models were tested and trained: Model I, which did not consider NMH, and a nested Model II, which incorporated NMH. The validation results demonstrated that nested Model II exhibited more promising outcomes than Model I, suggesting that including NMH significantly influenced the model's performance. Including NMH as an input parameter in the machine learning model could better capture the vertical distributions of NO<sub>2</sub> and thus predict ground-level NO<sub>2</sub> concentrations with improved accuracy and performance. Additionally, the hour-by-hour 10-fold cross-validation depicted a distinct improvement in the ground-level NO<sub>2</sub> estimations for nested Model II considering NMH as an input parameter (Fig. S5 for Model I without NMH and Fig. S6 for nested Model II with NMH). The R<sup>2</sup> values for Model I without NMH were 0.63 for 8:00 AM, 0.70 for 9:00 AM, 0.69 for 10:00 AM to 1:00 PM, 0.55 for 2:00 PM, and 0.39 for 3:00 PM. The improved R<sup>2</sup> values for nested Model II, which includes NMH, were 0.85 for 8:00 AM, 0.90 for 9:00 to 11:00 AM, 0.91 for 12:00 PM, 0.93 for 1:00 PM, 0.89 for 2:00 PM, and 0.85 for 3:00 PM. Similarly, nested Model II, considering the NMH, depicted significantly reduced biases compared to Model I without NMH. The ground-level NO<sub>2</sub> estimations for all hours were significantly improved when considering NMH, as it directly incorporates the vertical distributions of NO<sub>2</sub>. During the early morning hours, most of the NO<sub>2</sub> is distributed near the ground. However, as the day progresses, NMH increases, and the ground-level NO<sub>2</sub> tends to be mixed vertically. Further, the improvements in ground-level NO<sub>2</sub> estimations were assessed using 10-fold cross-validation for different population categories, i.e., lightly populated, moderately populated, highly populated, and supremely highly populated. The nested Model II, considering NMH, depicted notable improvements compared to Model I without NMH (Fig. S7). The improved R<sup>2</sup> values for nested Model II considering NMH were 0.91 for lightly populated areas and 0.92 for the other three population categories compared to Model I without NMH, which depicted an R<sup>2</sup> value of 0.63 for lightly populated, 0.73 for moderately populated, 0.77 for highly populated, and 0.74 for supremely highly populated areas. The RMSE for nested Model II considering NMH was improved and observed below 5 μg/m<sup>3</sup> for all population categories

compared to Model I without NMH, which depicted RMSE values around 8-9  $\mu\text{g}/\text{m}^3$  for different population categories. The MAPE for nested Model II considering NMH was also improved for all population categories, and around 15 % and lower values were observed. These improvements depict that nested Model II considering NMH effectively captures the spatial distributions of vertical mixing of ground-level  $\text{NO}_2$  across all population categories. The spatiotemporal distributions and diurnal patterns of NMH are previously described by Ahmad et al. (2024). Compared to Model I without NMH, the performance of the ground-level  $\text{NO}_2$  estimations through nested Model II considering NMH showed significant improvement at the grid points where ground-based observations were available (Fig. S8). The correlation coefficients for grid-based 10-fold cross-validation were improved to 0.8-1.0 for nested Model II considering NMH compared to Model I without NMH, which depicted lower correlation coefficients. Furthermore, nested Model II considering NMH also depicted lower RMSE values for grid-based estimations.”

#### Minor Comments:

**Comment:** Line 122: What is the nominal spatial resolution of GEMS  $\text{NO}_2$  product used in this study?

**Response:** Thanks for your comments. We clarified the spatial resolution of the GEMS data in lines 136-139:

“The nominal spatial resolution of the GEMS  $\text{NO}_2$  product used in this study was  $7 \text{ km} \times 7.7 \text{ km}$ . Despite the irregular shape of satellite measurement pixels due to east-to-west scans, this study performed re-gridding, which standardized the VCDs of  $\text{NO}_2$  onto a regular grid of  $0.2^\circ \times 0.4^\circ$  by calculating the average of all the  $\text{NO}_2$  VCDs within the  $0.2^\circ \times 0.4^\circ$  grid from 8:00 AM to 3:00 PM local time in China.”

**Comment:** Line 124: Please provide some information on how  $\text{NO}_2$  VCDs are standardized. Line 160 mentioned bi-linear interpolation, but it is for meteorological variables.

**Response:** Thank you very much for your comments. In this study, we standardized the VCDs of  $\text{NO}_2$  onto a regular grid of  $0.2^\circ \times 0.4^\circ$  by calculating the average of all the  $\text{NO}_2$  VCDs within the  $0.2^\circ \times 0.4^\circ$  grid. We clarified it in lines 136-139:

“Despite the irregular shape of satellite measurement pixels due to east-to-west scans, this study performed re-gridding, which standardized the VCDs of  $\text{NO}_2$  onto a regular grid of  $0.2^\circ \times 0.4^\circ$  by calculating the average of all the  $\text{NO}_2$  VCDs within the  $0.2^\circ \times 0.4^\circ$  grid from 8:00 AM to 3:00 PM local time in China.”

**Comment:** Line 135: ... divided the study region into four areas ... -> ... divided the study area into four categories ...

**Response:** Thanks for your comments. We have revised it accordingly (lines 149-150).

“Based on population density, we divided the study region into four categories.”

**Comment:** Line 253: How is the month of the year numbered exactly? If 1 to 12 is used for January to December, then cold months would be around 12 to 2, which may affect SHAP values shown in Figure 6.

**Response:** Thanks for your valuable comments. We used a common method to number the months. The months are numbered from 1 to 12, corresponding to January through December, exactly as per the real months of the observations. Using alternative numbering methods may increase the complexity. Additionally, the month variable has a relatively small contribution of only 3.23% to the model's performance. The month variable served mainly as an auxiliary factor, and the SHAP values were mostly clustered around zero. The major variables are GEMS  $\text{NO}_2$  VCDs and NMH. We clarified the numbering of the month variable in lines 225-226:

“The months are numbered from 1 to 12, corresponding to January through December, exactly as per the real months of the observations.”



**Comment:** Line 259: Figure 6 indicates that lower T corresponds to lower NO<sub>2</sub>. How does it relate to ‘worsened’ ground-level NO<sub>2</sub> pollution? And your reasoning ‘air stagnation’ may be wrong here.

**Response:** Thanks for your comments. When the feature values of temperature are large, the SHAP value is positive and may have a positive impact on the ground-level NO<sub>2</sub> predictions, but the impact value is not large. Some values with smaller feature values also have a positive impact on the model. It is noted that the SHAP values for the meteorological variables, including temperature, are all small, clustered around zero, and have limited influence on the prediction results. The major and distinct impact on the model’s performance for predicting ground-level NO<sub>2</sub> concentrations is observed for GEMS NO<sub>2</sub> VCDs and NMH. We have rewritten the paragraph to highlight our focus (lines 330-333):

“However, it is noted that the SHAP values for the meteorological variables, including temperature, are all small, clustered around zero, and have limited influence on the prediction results. The major and distinct impact on the model’s performance for predicting ground-level NO<sub>2</sub> concentrations is observed for GEMS NO<sub>2</sub> VCDs and NMH.”

**Comment:** Line 260: Figure 6 does not indicate this pattern. Please either quantify the impact of RH and dew point explicitly or remove this sentence.

**Response:** Thanks for your comments. We have removed the sentence.

**Comment:** Line 265: In this and the following sections, are ground-level NO<sub>2</sub> concentration from ground-based observations or satellite-based estimates? Please clarify.

**Response:** Thanks for your comment. In this and the following sections, ground-level NO<sub>2</sub> concentrations are from satellite-based estimates. We have clarified it in the manuscript (lines 338-339):

“Based on the satellite-derived ground-level NO<sub>2</sub> concentrations (mentioned as ground-level NO<sub>2</sub> concentrations from hereon)”.

**Comment:** Line 266: Since this paragraph is talking about Fig. S1, I would suggest presenting the figure in the main text. Also, as the correction factor is important to the results of this study, how it is calculated should be presented in the main text or as an appendix. Related to the computation of correction factor, what is the possible maxima of  $m$ ? Is it up to 24 (hours of a day)?

**Response:** Thanks for your comments.

(1) Fig. S1 has been moved to the main text as Fig. 7.

(2) The calculation of the correction factor is moved to the main text as Section 2.7.

(3) For a specific hour, the maximum value of  $m$  in Eq. 1 is 365 for one year. We clarified it in line 263-264.

“For a specific hour, the maximum value of  $m$  index in Eq. 1 is 365 for one year.”

For the annual correction factor, the maximum value of  $m$  in Eq. 1 is 8760 for one year. We clarified it in line 280-281.

“For the annual correction factor, the maximum value of  $m$  index in Eq. 1 is 8760 for one year.”

**Comment:** Line 350: Since Fig. S6 is discussed here, considering presenting the figure in the main text.

**Response:** Thanks for your comments. The figure has been moved to the main text as Fig. 12.

**Comment:** Line 425: Are NO<sub>2</sub> and NO really in chemical equilibrium in the real atmosphere?

**Response:** Thanks for your comments. We revised the sentence and removed this particular description.

**Comment:** Line 444: The reasoning given here is too general. Consider adding some details/analysis specific to your results.

Response: Thanks for your comments. We had added more discussions on the spatial disparities of NO<sub>2</sub> concentration and its implication for air pollution management in lines 572-594:

“The spatial distribution of ground-level NO<sub>2</sub> concentrations in the study region revealed significant regional disparities, with higher levels observed in urban agglomerations with high population densities (e.g., BTH, YRD, and PRD regions) than in lightly populated areas (e.g., western China). Even within the NC region, the highly populated urban areas had NO<sub>2</sub> concentrations nearly double those of lightly populated rural areas. These spatial disparities are due to distributions of NO<sub>2</sub> emission sources that vary with population densities, decreasing from highly populated to lightly populated areas. In highly populated urban areas in regions like BTH, YRD, and PRD, mobile NO<sub>x</sub> emissions from dense road networks contribute to pronounced increase in NO<sub>2</sub> levels. Moreover, the short lifespan of NO<sub>2</sub> due to atmospheric chemical reactions results in elevated concentrations near emission sources in highly populated areas, such as roadways, accompanied by rapid declines in NO<sub>2</sub> concentrations with increasing distance from highly populated areas (Lee et al., 2018). Furthermore, the diverse terrains, land cover, and climates observed in subregions with different population categories collectively influence vertical and horizontal airflows, rates of NO<sub>2</sub> formation and deposition, and contribute to spatiotemporal variations in ground-level NO<sub>2</sub> concentrations between the highly populated and lightly populated areas across China. Additionally, the population-weighted mean NO<sub>2</sub> concentrations were consistently higher than the spatial mean NO<sub>2</sub> concentrations in most provinces across China. This is due to the spatial coincidence between NO<sub>2</sub> concentrations and population density. These results indicate that the use of simple spatial average concentrations can lead to a systematic underestimation of overall population exposure and the associated health impacts. It is important to use high-resolution NO<sub>2</sub> data to accurately quantify true population exposure. Furthermore, the adverse impacts of high NO<sub>2</sub> concentrations in highly populated urban areas suggest that for more effective reduction of overall population exposure and better protection of public health, control efforts should be further targeted at highly populated and highly polluted areas. Targeted control programs to reduce pollutant levels at population hotspots should be more cost-effective than trying to reduce pollutant concentrations everywhere. Additionally, control policies can be implemented by encouraging the public to relocate to less polluted areas through land-use development and urban planning.”

**Comment:** Line 470: The wording and the order of the sentence starting with ‘The average ground-measured NO<sub>2</sub> concentrations’ is confusing, please revise.

Response: Thanks for your comments. We revised the sentence from “The average ground-measured NO<sub>2</sub> concentrations, when satellite data was available, consistently underestimated the average NO<sub>2</sub> concentrations from all ground measurements for each hour.” to (line 618):

“The issue of missing data consistently underestimated the average NO<sub>2</sub> concentrations for each hour.”

**Comment:** Figure 3: How model 1 (i.e., without NMH) differs from model 2 (with NMH) is not clearly shown in the diagram. Please either split the flowcharts or add some description in the caption.

Response: Thanks for your valuable comments. We re-plotted the flowchart to highlight the role of inner model. Additionally, we added description on the difference between basic model and nested model in the caption of Figure 3.

“The basic model (Model I) does not consider NMH from the inner model and utilizes only ten input variables for testing and training, namely: satellite NO<sub>2</sub>, two temporal variables, and seven meteorological variables. The nested model (Model II) considers the NMH from the inner model as an additional input variable, along with the other ten input variables used for the basic model. Therefore, the nested model utilizes eleven input variables for testing and training: satellite NO<sub>2</sub>, two temporal variables, seven meteorological variables, and the NMH predictions from the inner model.”

**Comment:** Figure 4: Please clarify the meaning of each figure element (dots with colors, lines, etc.).

Response: Thank you very much for your comments. We clarified the meaning of the figure elements in lines 303-306:

“The red dotted line represents a 1:1 relationship. The solid black line is the line of best fit between the ground-measured NO<sub>2</sub> and the satellite-estimated NO<sub>2</sub>. The scattered dots represent the individual NO<sub>2</sub> values for each ground measurement and satellite-based estimation. The color scale ranging from red to blue represents the density of the NO<sub>2</sub> values, with red indicating high density and blue representing low density.”

**Comment:** Figure 7: Is this figure corresponding to ground-based observations or satellite-based estimates? Is it an average of 8 AM to 3 PM local time or daily average? Please clarify. Also, mark the province if possible so that readers unfamiliar with China can have a better sense of the regions you are referring to.

Response: Thanks very much for your comment. This figure presents satellite-based estimates of the annual average ground-level NO<sub>2</sub> concentration, which represents the 24-hour average throughout the year 2021, after bias correction for the missing data issue. We have clarified this information in the caption of the figure (lines 364-368). Additionally, provinces are marked in the figure.

“Spatial distributions of annual average ground-level NO<sub>2</sub> concentrations for 2021 derived from satellite measurements in the study region (left panel) and in the four major urban agglomerations in China (right panel): Beijing-Tianjin-Hebei (BTH), Yangtze River Delta (YRD), Pearl River Delta (PRD), and Sichuan Basin (SCB). This annual average concentration represents the 24-hour average throughout the year of 2021 after the bias correction for the missing data issue.”

**Comment:** Figures 9 through 12: What are the vertical bars in each plot? Please clarify.

Response: Thanks for your comment.

(1) The vertical bars in figure 9 (now figure 10), 10 (now figure 11) and 11 (now figure 13) represent one standard deviation. The description is added in the caption of the figures.

(2) The vertical bars in figure 12 (now figure 14) represent whiskers that extend to the most extreme data points within 1.5 times the interquartile range from quartile 1 (25th percentile of data) and quartile 3 (75th percentile of the data). The description is also added in the caption of the figure.