Response to Reviewers

This document contains our responses to the reviewers' comments. We have classified the responses using three colors for clarity: black, blue, and red. *Italics black* text is the original comments from the reviewers. Normal black text is the original text from the manuscript. Blue text is our response to the reviewers' comments. Red text is the revised text in the manuscript.

Reviewer #1

This study uses NO2 observations from the GEMS satellite to re-estimate NO_x emissions over Thailand. It updates emissions in two ways: where they are higher or lower across regions (spatial pattern), and how they change through the day (diurnal variation). Using a geostationary satellite like GEMS to capture the daily cycle of emissions is a worthwhile contribution, since this is something the low orbit satellites cannot do. The method itself (IFDMB) is taken from earlier work, but applying it to Thailand and using GEMS to refine the diurnal cycle is a genuinely useful effort. The main issues to address are (1) how trustworthy the resulting emissions are, and (2) whether the diurnal results are physically reasonable. I recommend major revisions.

Major comments

1. The North Thailand emissions are likely contaminated by a spurious satellite signal. : The paper concludes that emissions increase in the North, but this comes from an abnormally high GEMS signal over Lampang. As the authors themselves show (Section 3.5), this signal does not appear in TROPOMI or in the alternative GEMS retrieval (DVCF). It originates from an error in the prior inventory (ASIA-AQv3) used inside the GEMS retrieval, which then propagates through the inversion and shows up as a false "emission increase". Presenting this as a real finding is misleading. Please state clearly in the abstract and conclusions that the North increase is most likely a retrieval artifact, and report the Thailand-total posterior emission both with and without the North so readers can judge its influence.

Response: We agree with the reviewer that the increase in emissions in North Thailand is largely driven by an abnormally high NO_2 from the GEMS retrieval. They are correct that presenting this as a physical increase could mislead readers. In response to the reviewer's comment, we have added clear statements explicitly noting that the observed emission increase in the North is most likely a retrieval uncertainty driven by the a priori used in GEMS. Moreover, we presented the posterior emissions for Thailand both with and without the North, ensuring readers can accurately assess how this influences the national total.

Modified text:

Abstract

Nitrogen oxides ($NOx = NO + NO_2$) are key pollutants that contribute to ozone and secondary aerosol formation, posing environmental and health risks. Accurate simulation and forecasting of NO_x pollution is essential for developing mitigation strategies. Local inventories in Thailand are infrequently updated, leading researchers to use global inventories such as CAMS-GLOB-ANT for simulation. Global inventories carry uncertainties due to assumptions in emission factors, outdated activity data, and coarse temporal resolution. To address these limitations, this study applies a top-down approach to update NO_x emissions in Thailand using the iterative finite difference mass balance (IFDMB) method. Tropospheric NO_2 vertical column densities (VCDs) from the GEMS version 3 are integrated with the WRF-Chem to refine CAMS-GLOB-ANT emissions for September 2023. The simulations with posterior emissions are evaluated against TROPOMI NO_2 VCDs and surface NOx concentration. Results show that the baseline simulation overestimates NO_2 VCDs across Thailand compared with GEMS, except in North Thailand. GEMS reports substantially higher NO_2 than TROPOMI—particularly over Lampang—reflecting retrieval uncertainty associated with the a priori profiles used in version 3. Consequently, IFDMB reduces NO_x emissions across most regions but increases in the North. The increase is attributed primarily to a retrieval artifact rather than to a genuine emission underestimate. Baseline NO_x emissions in Thailand are 35,852 Mg month^{-1} with North Thailand and 30,699 Mg month^{-1} without North. Our posterior emissions over three IFDMB schemes are 25,994, 34,808, and 36,870 Mg month^{-1} with North, and 15,376, 26,523, and 27,087 Mg month^{-1} without North. These adjustments improve model bias and error relative to GEMS. However, when evaluated against TROPOMI, we find an increase in the bias for North Thailand, likely due to discrepancies between GEMS and TROPOMI retrievals. Discrepancies between GEMS and TROPOMI These discrepancies highlight the importance of future retrieval algorithm improvement and calibration across satellite products. Comparisons to surface observations indicate that IFDMB shifts the NOx peak to later than observations. This is because observations are strongly influenced by local transportation sources, which are hard to observe and simulate retrieved by GEMS and the model, respectively.

Conclusion

This study applies high-temporal-resolution GEMS satellite data to update anthropogenic NO_x emissions over Thailand during September 2023 to improve the air quality modeling and access diurnal variability of emissions. We compare WRF-Chem simulations using posterior emissions with observations from both the satellite and surface PCD stations to highlight the strengths and limitations of the IFDMB method using GEMS. We find that the baseline simulation from CAMS-GLOB-ANT v5.3 significantly overestimates tropospheric NO_2 VCDs across Thailand in comparison to both GEMS and TROPOMI, except in North Thailand where GEMS indicates this inventory is an underestimation. We attribute this underestimation primarily to retrieval uncertainty associated with the a priori in GEMS version 3 rather than to a genuine emission deficit. By applying the IFDMB inversion technique, the fixed diurnal update increases NO_x emission by about 20 mol km^{-2} hr^{-1} in North and West Thailand while decreasing them by about 30 mol km^{-2} hr^{-1} across most other regions. In contrast, the daylight diurnal and temporal average updates produce generally smaller adjustments. Including North Thailand, total NO_x emissions are 35,852 Mg month^{-1} in baseline, compared with posterior totals of 25,994 Mg month^{-1} (fixed diurnal), 34,808 Mg month^{-1} (daylight diurnal), and 36,870 Mg month^{-1} (temporal average). Excluding North Thailand, the corresponding totals are 30,699 Mg month^{-1} (baseline), 15,376 Mg month^{-1} (fixed diurnal), 26,523 Mg month^{-1} (daylight diurnal), and 27,087 Mg month^{-1} (temporal average). Because North Thailand accounts for about 14% of national NO_x emissions, the GEMS retrieval uncertainty in this region propagates to a 10%–20% effect on the national posterior totals. WRF-Chem simulations using posterior emissions show improved agreement with TROPOMI in most regions, especially the Bangkok Metropolitan Region. In this area, the fixed diurnal and temporal average updates achieve NMB reductions of 83% and 81%, and NRMSE reductions of 76% and 61%, respectively. The daylight diurnal update reduces NMB by 54% and NRMSE by 39%. However, the improvement in North Thailand is limited due to the discrepancy between GEMS and TROPOMI data in this area. Further refinement of retrieval algorithms and cross-satellite calibration would be expected to improve IFDMB performance there. Evaluation against PCD stations shows that the baseline simulation underestimates NO_x mixing ratios by 10–15 ppbv at roadside site. WRF-Chem using posterior emissions further exacerbates this underestimation at roadside PCD stations. This is attributed to strong differences in spatial resolution combined with the short lifetime of NO_x. The roadside PCD stations are mainly influenced by local transportation emissions, whereas GEMS captures broader urban and industrial emissions. Despite this limitation, GEMS demonstrates the potential to constrain the area-specific daytime diurnal variability of NO_x emissions, allowing diurnal emission profiles to be updated for each region rather than applying a uniform diurnal profile. This leads the model to be more accurate and reflects real-world emission patterns.

Our results are broadly consistent with previous study using GEMS data to constrain NO_x emissions over Southeast Asia during the haze season (March–April 2022). Both simulations overestimate NO_2 VCDs relative to GEMS across most regions which results in NO_x emission reductions across Thailand. However, our IFDMB approaches increase the NO_x emissions in North Thailand during ~~October~~September 2023 whereas previous study decreases the emissions. This key divergence is mainly due to the high biomass burning emission in haze season and the differences between GEMS versions. Moreover, our study highlights the limitations of mass balance method using GEMS observation for refining NO_x emissions in Thailand. The coarsening of GEMS and WRF-Chem data to 45 x 45 ~~81 x 81~~ km^2 for mass balance method reduces smearing errors, but introduces resolution errors that prevent accurate representation of local traffic emission. A further limitation is the discrepancy between GEMS and TROPOMI NO_2 in North Thailand, which is attributed to the a priori emission inventory used in the GEMS retrieval algorithm. This emphasizes the calibration and bias correction for satellite data in Thailand in the future.

In summary, this study demonstrates both the value and limitations of GEMS observations for refining NO_x emissions in Thailand. GEMS provides valuable insights into the spatial distribution and daylight diurnal variability of NO_x across the region. By integrating GEMS data with mass balance methods, we improve the performance of NO_x simulations using WRF-Chem, particularly in the Bangkok Metropolitan Region, which showed substantial improvement when compared to TROPOMI observations. These results underscore the benefits of geostationary satellites such as GEMS for near-real-time emission constraints. However, limitations related to satellite retrieval uncertainty, spatial resolution, and local emission representation remain challenges Thailand. For future work, we recommend combining GEMS observations with complementary datasets—such as PCD stations, LEO satellite, and data from the recent ASIA-

AQ field campaign—to enhance data assimilation and reduce potential biases in GEMS retrievals.

2. The regularization parameters (C_1, C_2) controlling the strength of the emission adjustment are insufficiently justified. : This study uses two values, C_1 and C_2, which essentially decide how much to trust the satellite and how aggressively to change the emissions. These values drive the over-adjustment in the North raised in comment 1. Yet the chosen values are justified only briefly, and the sensitivity test (Table S1) shows the results change substantially with them (e.g., NMB improvement ranging ~30–66%). Please move a short sensitivity analysis from the Supplement into the main text, and discuss letting C_1 vary by region. Also, the signs in the denominators of Eqs. (2) and (4) appear to contradict the text, which says C_1 is added to "inflate" the denominator. Please double-check.

Response: We thank the reviewer for this constructive feedback regarding the regularization parameters (C_1 and C_2). In our work, we aim to treat these parameters as known, fixed, statistical values rather than tuning variables. Their values were determined based on the satellite column detection limit (C_1 = 1x10^{14} molecules cm^{-2}) and typical background emission near the Bangkok area (C_2 = 10 mol km^{-2} hr^{-1}). We note that even fixed values can serve to regularize an inverse problem, even if they are not tuned or optimized parameters. However, we do recognize that our knowledge of these error terms is incomplete, and that there could indeed be regional variations. Thus, we do consider how sensitive our results are to these values. With our available computational resources, we explore two more regularization parameters: C_1 = 2x10^{15} molecules cm^{-2}, representing the background tropospheric column in urban areas near Bangkok, and C_2 = 1 mol km^{-2} hr^{-1}, representing emissions in rural areas. To bracket the model's sensitivity, we evaluated four representative combinations (C_1: 1x10^{14} and 2x10^{15}; C_2: 1 and 10).

We agree with the reviewer that the sensitivity analysis of these parameters would help readers better understand the impact of regularized parameters on the IFDMB. Following their suggestion, we have moved the sensitivity analysis and its associated discussion (formerly Table S1 in the Supplement) into the main text. Furthermore, we agree that allowing C_1 and C_2 to vary by region—such as applying different C_1 between North Thailand and Bangkok—would be an effective strategy to prevent over-adjustment in areas like the North.

While conducting new regionally varying parametric WRF-Chem simulations is beyond the computational and time resources of this study, we have added a dedicated discussion in the main text highlighting the trade-offs of static regularization and outlining the potential benefits of implementing spatially varying C_1 and C_2 parameters in future studies.

Finally, we double-checked the denominators in Equations (2) and (4). We confirm that there was a typographical error in our main text. The parameters C_1 and C_2 are indeed intended to be added to the denominator to prevent division by very small numbers (inflating the denominator and damping the adjustment). We have corrected Eq. (2) in line 218 and Eq. (4) in line 254 to ensure complete consistency.

Modified text:

…

$$J = J_{prediction} + J_{parameter} \quad (1)$$

$$= \frac{1}{2}\frac{(\Omega_o - \Omega_a)^2}{(\sigma_o^2 - + C_1^2)} + \frac{1}{2}\frac{(E_t - E_a)^2}{(\sigma_a^2 - + C_2^2)} \quad (2)$$

…

As σ_o is often quite low (on the order of 10% or less of Ω_o), the cost function in Eq. (2) can become overly sensitive to the prediction error term, potentially leading to overfitting. To prevent this, we introduce a regularization term C_1 to inflate the denominator of the first term. In contrast, σ_a is significantly small in background areas, where emissions are close to zero. This leads Eq. (2) to become sensitive to the parameter error term and results in underfitting. Consequently, we add another regularization term C_2 in the second term. These regularization parameters effectively limit the influence of the prediction/parameter error term, thereby balancing the relative weights of the model–observation mismatch and the prior deviation terms. The low value of C_1 leads to overfitting the solution to observations, while the high value of C_1 leads to overfitting the solution to the prior, and vice versa for C_2. As σ_o and σ_a only quantify relative errors, the C_1 and C_2 terms can also be interpreted as absolute contributions to the total observation and emissions errors, respectively, such as a detection limit for the former. In this study, we treat these regularization parameters as known (or estimated) fixed statistical values rather than tuning variables. For the C values, we use 1x10^{14} molecules cm^{-2} for C_1 and 10 mol km^{-2} hr^{-1} for C_2. These numbers are based on the GEMS detection limit and background NO_x emission near the Bangkok area. However, we also explore two more regularization parameters: C_1 = 2x10^{15} molecules cm^{-2}, representing the background tropospheric column in urban areas near Bangkok, and C_2 = 1 mol km^{-2} hr^{-1}, representing emissions in rural areas. The sensitivity test of four combinations is performed to understand the impact of regularized parameters on the IFDMB. The short information is presented in Sect. 3.2, and full information is in Sect. S1.

…

$$g = \frac{h\sigma_a^2}{h^2(\sigma_a^2 - + C_2^2) + (\sigma_o^2 - + C_1^2)} \quad (4)$$

…

3.2 Evolution and optimization of mass balance methods

…

For the fixed diurnal and daylight diurnal update schemes, iteration 2 provides the lowest J and is thus selected as optimal. For the temporal average update scheme, iteration 4 yields the optimal performance. ~~Furthermore, analyses of the sensitivity of the cost function in the fixed diurnal scheme the parameters C_1, C_2, demostrate consistent model performance patterns across different parameter values (Sect. S1).~~

Additionally, we performed a sensitivity analysis of the regularization parameters in the IFDMB framework, using a fixed diurnal update, to evaluate the impacts of C_1 and C_2. We found that varying these parameters alters the results significantly—yielding differences of ~30% to 60%—

with the model showing particular sensitivity to C_1 (Table S1). Increasing C_1 from 1x10^{14} to 2x10^{15} molecules cm^{-2} limits the influence of satellite constraints, resulting in higher NMB and NRMSE values, particularly in Central and East Thailand (Figs. S1 and S2). A low value for C_1 permits large emission adjustments across all of Thailand, including regions with high GEMS retrieval errors, such as North Thailand (Fig. S3). Conversely, a higher C_1 value restricts these adjustments. In North Thailand, where retrieval uncertainty is high, increasing C_1 benefits the IFDMB by preventing the model from overfitting to highly uncertain GEMS data. However, this same increase has a detrimental impact on the Bangkok Metropolitan Region. Because uncertainty is comparatively low here, a high C_1 limits emission improvements. Regarding C_2, increasing its value from 1 to 10 mol km^{-2} hr^{-1} has minimal impact when C_1 is low.

This is because a low C_1 permits aggressive emission adjustments nationwide; variations in C_2 become negligible. The impact of C_2 is much more pronounced when C_1 is high. C_2 primarily governs adjustments for smaller emission sources in rural areas. Consequently, a high C_2 allows for greater adjustments in these low-emission zones, which leads to overfitting to the GEMS data in the rural North. A low C_2 leads to underfitting in Central Thailand and the Bangkok Metropolitan Region. Overall, the adjustable parameters in the IFDMB offer valuable flexibility to manage the mismatch between GEMS and TROPOMI by moderating emission updates in high-error regions and balancing prior and posterior estimates. In this study, we applied fixed regularization parameters across the entirety of Thailand. However, our findings indicate that a large C_1 benefits high-uncertainty areas like North Thailand, whereas a low C_1 benefits the Bangkok Metropolitan Region and Central Thailand. This highlights the potential of using spatially varying parameters based on regional uncertainty rather than fixed domain-wide values. Nevertheless, because we aim to treat these parameters as fixed statistical values rather than tuned parameters, an exhaustive fine-tuning exploration is beyond the scope of this paper.

3. The diurnal emission update is a good direction, but the results need to be made more physically realistic. : Using GEMS to capture the daily cycle of emissions is a real strength of this paper. However, the current results need a few improvements before they are usable. Because the satellite only sees during daytime, the updated emission curve drops back to the prior at night, creating an unnatural, discontinuous shape. The authors mention applying a smoothing step — please actually apply it and show the result.

Response: We thank the reviewer for pointing this out. We agree that capturing the daylight diurnal cycle of emissions is a key strength of using GEMS, and that the abrupt drop to the prior at night requires smoothing to be physically realistic.

While our IFDMB method is computationally cheaper than 4D-Var or similar assimilation techniques, the iterative process remains computationally expensive. Rerunning our entire suite of simulations from scratch with the smoothing technique applied iteratively falls outside the feasible timeframe for this revision. However, to address your concern and demonstrate the effectiveness of this approach, we applied a smoothing technique to our posterior emissions and performed a new simulation to show the benefits of the smoothing technique. We have added this section in the Supplement. To conserve computational resources, we applied this smoothing technique exclusively to the posterior emissions of the IFDMB with the daylight diurnal update. Our comparison focuses primarily on PCD stations, as the emissions within the GEMS observation window remain identical; therefore, the smoothing has no impact on TROPOMI comparisons and only a minor impact on GEMS at 08:00.

Modified text:

3.6 Comparison with data from PCD stations

…

However, using the GEMS satellite to only update the diurnal variability introduces an unrealistic diurnal profile in the model – an abrupt increase and drop in the NO_x emissions profile (Fig. 9). We recommend implementing a smoothing technique in future work to mitigate such discontinuities. In this study, we applied a 5-point moving average to the posterior emissions from the IFDMB with the daylight diurnal update (Sect. S4). While this smoothing technique has a small impact on absolute NO_x concentrations, it yields a more physically realistic NO_x concentration profile by mitigating the abrupt changes in the transition zones bordering the GEMS observation window. In summary, while the fixed diurnal update does not improve model–observation agreement due to its unchanged emission profile, the daylight diurnal and temporal average updates offer some potential for spatially resolved diurnal refinements. However, the lack of satellite observations during critical rush-hour periods (early morning and evening) limits their effectiveness. Although GEMS-based updates alone do not substantially improve NO_x simulations at PCD stations due to resolution and timing mismatches, the satellite provides valuable information for refining regionally specific diurnal patterns, which could enhance WRF-Chem performance in future studies.

S4 Moving Average for IFDMB daylight diurnal update

Because our IFDMB method only updates emissions during the GEMS observation window and reverts to prior emissions outside this period, it generates an unnatural, discontinuous emission profile. In this section, we test the ability of a smoothing technique to alleviate discontinuities caused by the temporal limitations of GEMS data. We apply the smoothing technique exclusively to the posterior emissions of the IFDMB with the daylight diurnal update. Our comparison focuses primarily on PCD stations, as the emissions within the GEMS observation window remain identical; therefore, the smoothing has no impact on TROPOMI comparisons and only a minor impact on GEMS at 08:00.

Specifically, we applied a 5-point moving average to bridge the GEMS daylight observation window (08:00–15:00) and the nighttime prior. The data within the GEMS daylight window remains unsmoothed to fully preserve the observational updates. For the transition periods (02:00–07:00 and 16:00–21:00), we applied the moving average using a distance-decaying weight (equation below).

$$E_{t,mov} = \frac{E_{t-2}+E_{t-1}+E_t+E_{t+1}+E_{t+2}}{5} \qquad \text{(S2)}$$

$$E_{t,smooth} = (w-1)E_t + wE_{t,mov} \quad \text{(S3)}$$

where $E_{t,mov}$ is the moving average of posterior emission at time t (E_t). The smoothed posterior emissions at time t ($E_{t,smooth}$) are calculated based on a weighted average between E_t and $E_{t,mov}$. During the GEMS observation window, the weight (w) is manually set to zero to preserve information from observation. For the 3 hours adjacent to the GEMS window (e.g., 05:00, 06:00, 07:00), we set w to 1 for fully smooth as shown in Figure S8. This is because these points are the transition zone between the GEMS observation window and the no-observation zone. As the time moves further from the GEMS window, this weight is reduced by 25% per hour until it reaches 0 (Figure S8). There is no GEMS information in this zone, but the smoothing is applied for smoothness.

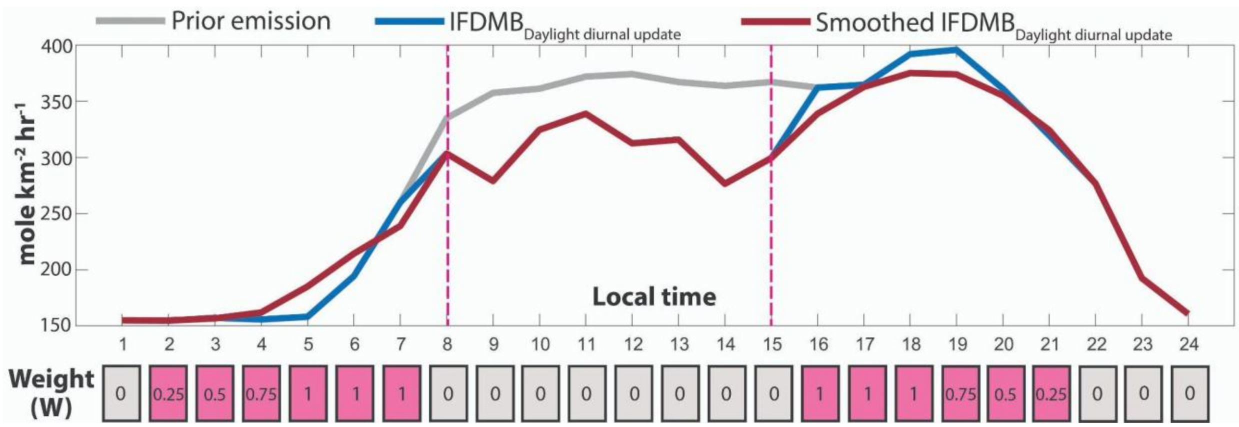

Figure S8: Hourly NO_x emissions in Bangkok comparing IFDMB with and without 5-point moving average using a distance-decaying weight: prior emission (gray), IFDMB with daylight diurnal update (blue), Smoothed IFDMB with daylight diurnal update (Red). The pink line represents the GEMS window for Thailand, and the boxes show the weight of each time for Eq. (S3).

For example, the data at 04:00 uses the moving average of 02:00 to 06:00 with a weight of 0.75. The data at 03:00 uses a weight of 0.50, and at 02:00 uses a weight of 0.25. The evening transition (16:00–21:00) follows the exact same tapering logic. This smoothly relaxes the posterior daytime emissions back to the prior at night, successfully removing the unnatural discontinuity. However, the emission at 05:00 will be slightly oversmoothed. As this point is in the twilight zone, the 5-point moving average is used to smooth the emissions of daytime and nighttime to calculate smoothed posterior emissions.

When comparing against PCD stations, we found that while the smoothing method has a negligible impact on absolute NO_x concentrations (Fig S9), it produces a much smoother and more physically realistic NO_x concentration profile (Fig S10). Specifically, the smoothed IFDMB with the daylight diurnal update prevents the abrupt spikes and drops in concentration during the transition periods (15:00–17:00 and 06:00–08:00) caused by the sudden reversion of posterior emissions to prior emissions. This demonstrates the benefit of the smoothing technique. Even though the absolute concentrations do not change significantly, the resulting profile is much more realistic than the non-smoothed version. Future work employing more advanced smoothing techniques and incorporating nighttime observations will further refine the accuracy of these emission profiles.

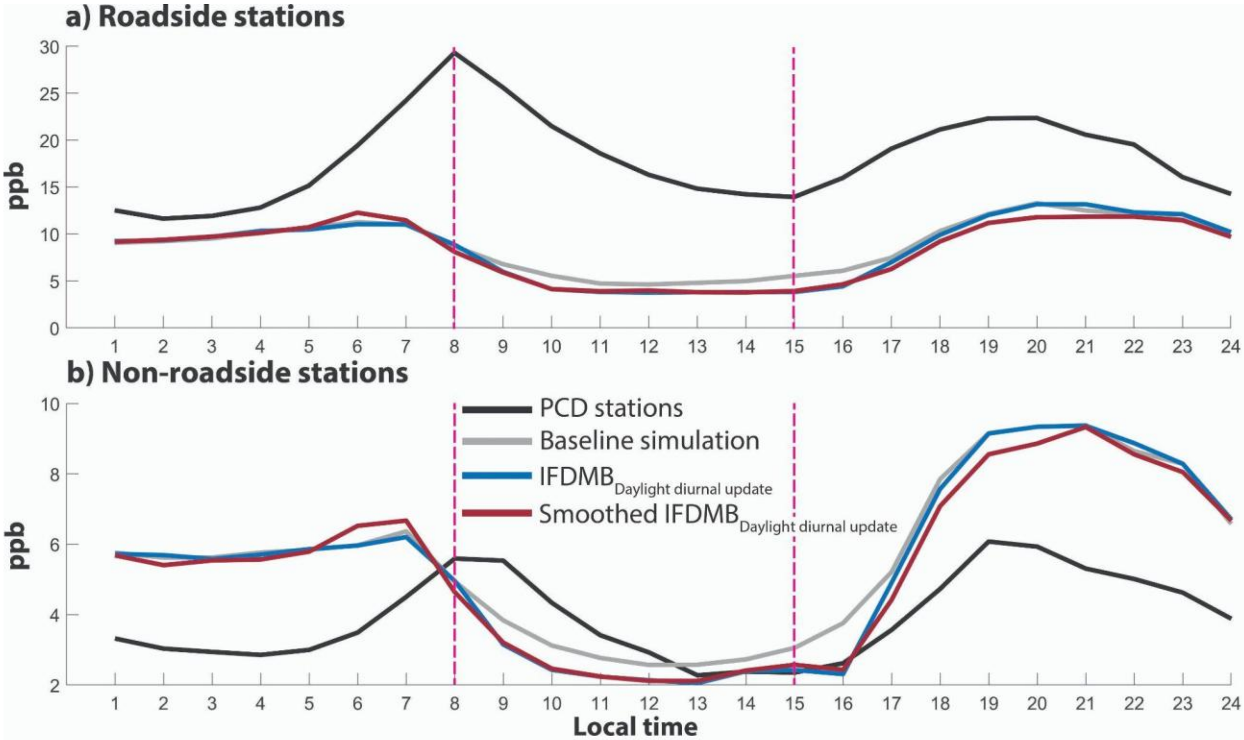

Figure S9: The monthly average diurnal profile of surface NO_x mixing ratios from PCD stations (black) and WRF-Chem: the baseline simulation (gray), IFDMB with daylight diurnal update (blue), and smoothed IFDMB with daylight diurnal update (red). PCD stations are classified into roadside stations (a) and non-roadside stations (b). The pink lines mark the GEMS overpass times in Thailand from 08:00 - 15:00 local time.

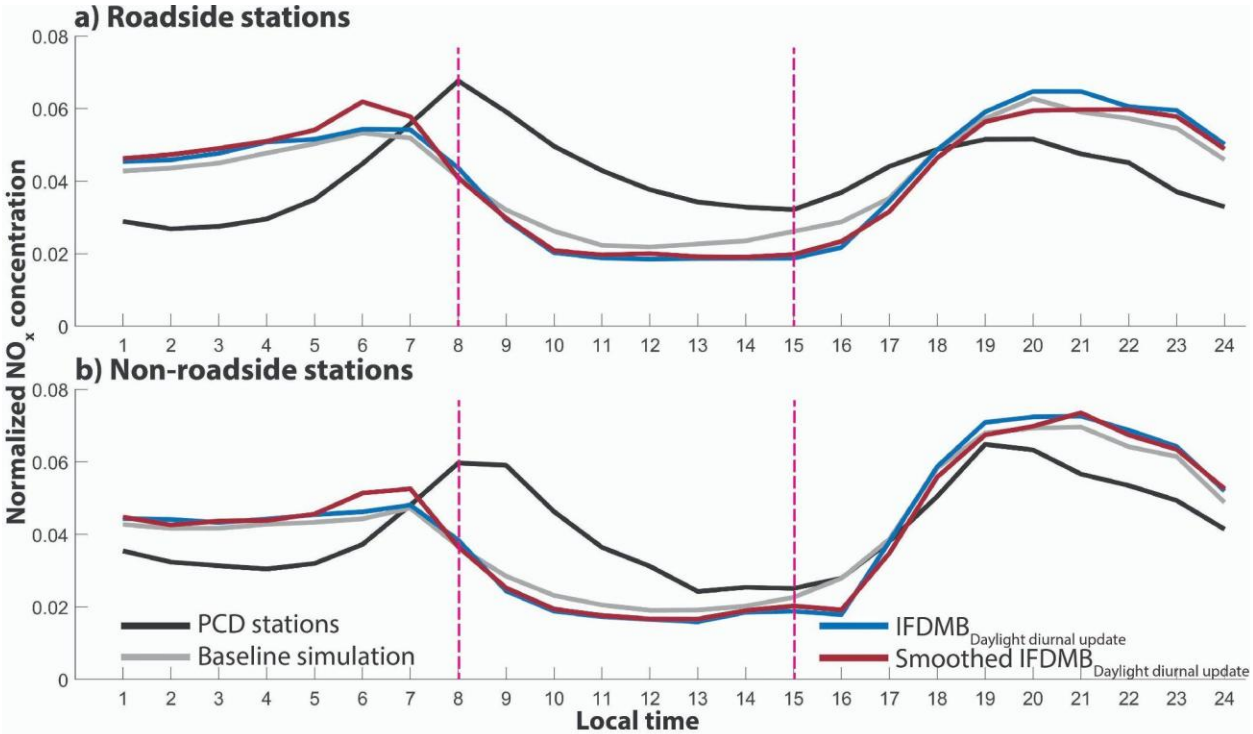

Figure S10: The diurnal variability of normalized NO_x concentrations from PCD stations (black) and WRF-Chem: the baseline simulation (gray), IFDMB with daylight diurnal update (blue), and smoothed IFDMB with daylight diurnal update (red). PCD stations are classified into roadside stations (a) and non-roadside stations (b). The pink line is the GEMS scanning time in Thailand from 08:00 - 15:00 local time.

4. GEMS misses the rush-hour periods (early morning and evening) when emissions are highest.: Please quantify how much this gap affects the updated diurnal profiles. The conclusion ends somewhat vaguely ("no single scheme is consistently best"). A clear, practical recommendation — e.g., which scheme to use when spatial accuracy matters versus when capturing the daily cycle matters — would be much more useful to future users.

Response: We thank the reviewer for pointing this out. We add the recommendation about each scheme in the main text.

Modified text:

3.7 Inversion approaches

…

For the update schemes, ~~while~~ all three schemes for IFDMB offer improvements over the prior simulation when evaluated against GEMS data.~~, the analysis reveals that no single scheme is consistently the best.~~ However, each scheme involves trade-offs in how it treats diurnal variability, and these trade-offs determine which scheme is best suited to a given application. For applications focused on monthly-average NO_2 improvement or on updating emissions at the national or regional scale where the diurnal cycle matters less than spatial accuracy, we recommend the fixed diurnal update. This scheme applies a uniform monthly-averaged correction factor to emissions at all hours and produces the most substantial improvements against GEMS observations. However, this method could not correct the diurnal variability of emissions. The ground-based PCD data clearly show that the baseline simulation misrepresents the diurnal cycle, with the model peaking at night instead of during morning and evening rush hours. The high NO_x concentration during the nighttime from the model is likely due to the low planetary boundary layer at that time (Thongsame et al., 2024). The fixed diurnal update simply scales down the incorrect diurnal profile, without improving the diurnal variability. We find that it lowers the correlation at 08:00 from 0.70 to 0.60 at roadside stations and from 0.70 to 0.66 at non-roadside stations. Even so, the fixed diurnal update remains a simple and effective method, as it is the only scheme able to update emissions during both daytime and nighttime using monthly-average data; because GEMS only provides information during daylight hours, the other two schemes must leave nighttime emissions unchanged from the prior.

The daylight diurnal and temporal average updates are designed to address diurnal issues by incorporating hourly GEMS data. These schemes only adjust emissions during daylight hours, resulting in more modest improvements in GEMS agreement. Their key weakness, however, is the introduction of an unrealistic diurnal profile. Because the algorithms have no observational constraints outside the GEMS timing window, they revert to the prior profile at night. This is due to the limitations of our update schemes. For studies focused on daytime hourly NO_x concentrations and diurnal variability, we recommend the daylight diurnal update scheme. Because the daylight diurnal and temporal average schemes yield no significant difference in performance, we favor the daylight diurnal update for its simpler implementation. It improves the correlation at 08:00 from 0.70 to 0.75 at non-roadside PCD stations, while the correlation at roadside stations remains unchanged at 0.70. By applying the smoothing technique, the impact of the unrealistic diurnal profile would be reduced. Unlike the fixed diurnal update, it can adjust the diurnal profile region by region rather than applying a single uniform profile across the domain, though this benefit remains limited by GEMS's spatial resolution. The challenge of this scheme is how to refine ~~This limitation reflects the challenges of refining~~ emissions outside the GEMS coverage period, as no UV/Vis satellite observations are available at night. Future research should consider integrating additional datasets—such as from the recent ASIA-AQ field campaign data, Pandora spectrometer measurements, and PCD stations—with satellite observations to develop accurate diurnal profiles. Nevertheless, GEMS provides valuable information for constraining daylight diurnal variability across different regions of Thailand.

5. Averaging the data to 81 km removes the ability to constrain local sources. : The analysis groups the data into 81 × 81 km boxes (9 times coarser than the model). Please justify quantitatively why 9× and not, say, 3× or 5×. If possible, test one or two intermediate resolutions to show the trade-off between reducing the spread of signal into neighboring cells and losing the ability to constrain local sources such as traffic. This directly relates to why the roadside surface comparison gets worse.

Response: We thank the reviewer for raising this important point. First, we must clarify and apologize for a typographical error in the manuscript: the spatial aggregation used in our primary analysis is actually 45 × 45 km (5 × 5 grid cells), not 81 × 81 km (9 × 9 grid cells) as incorrectly stated. We have corrected this typo throughout the revised main text.

We agree that the degree of grid cell aggregation significantly impacts the posterior emissions, particularly for localized sources such as traffic. The primary motivation for aggregating data to the 45 × 45 km resolution (5×5 grid cells) is to minimize smearing errors caused by horizontal NO_2 transport between neighboring cells. However, coarser aggregation introduces resolution error, which degrades the ability of the IFDMB method to resolve localized emissions and explains why the comparison against roadside surface observations degrades.

To quantitatively explore the trade-off between smearing error and resolution error, we performed new sensitivity simulations using both 3 × 3 grid cell aggregation (27 × 27 km) and no aggregation (1 × 1, or 9 × 9 km). To manage computational costs, these sensitivity simulations were conducted using the IFDMB method with fixed diurnal updates. Because our primary 5 × 5 grid cell configuration converged after two iterations, we capped the 3 × 3 and 1 × 1 configurations at two iterations as well to ensure a consistent and equitable comparison.

Modified text:

2.6 Experimental Design

Although IFDMB assumes negligible horizontal transport due to the short atmospheric lifetime of NO_2, our high-resolution simulations (9 x 9 km^2) violate this assumption, as there may be significant horizontal transport of NO_2 to surrounding grid cells. To mitigate the smearing error, we regrid both GEMS and WRF-Chem NO_2 VCDs to a coarser 45 x 45 ~~81 x 81~~ km^2 resolution by spatially averaging over surrounding grid cells. To update NO_x emissions, we apply IFDMB in three different configurations: Fixed diurnal, Daylight diurnal, and Temporal average (Table 1). Each approach is described below. For each method, we use some form of a monthly average of the GEMS data to update emissions. This helps reduce the impact of noise and missing data in satellite observations. Additionally, the smearing error is minimized when considering monthly average NO_2 VCDs. Additionally, we perform the sensitivity simulation in the Supplement using three degrees of aggregation: 5 × 5 grid cells (45 × 45 km^2), 3 × 3 grid cells (27 × 27 km^2), and no aggregation (1 × 1, or 9 × 9 km^2). This sensitivity test is to explore the impact of grid cell aggregation on the IFDMB with fixed diurnal update scheme.

3.6 Comparison with data from PCD stations

…

All inversion schemes lead to an increase in bias and error at roadside stations. These discrepancies arise primarily from differences in spatial resolution between the GEMS (~3.5 × 8 km^2), WRF-Chem (9 × 9 km^2), and PCD stations (point-scale), as well as strong local contributions of transportation emissions at the PCD sites. Moreover, both GEMS observations and WRF-Chem simulations are regridded to 45 x 45 ~~81 x 81~~ km^2 for the mass balance calculation in this study to reduce smearing errors associated with horizontal transport.

…

Our findings also underscore the tradeoffs in the application of mass balance methods. The IFDMB approach, when applied at the 45 x 45 ~~81 x 81~~ km^2 resolution, offers computational efficiency and is well-suited for adjusting regional-scale emissions over long temporal periods. However, its coarse spatial resolution limits its ability to capture fine-scale urban variability and accurately constrain emissions from local sources such as traffic. Moreover, reducing or removing the aggregation for IFDMB provides no significant difference in roadside stations (Sect. S5). For non-roadside stations, the 45 x 45 km^2 resolution performs better than 27 x 27 km^2 resolution and 9 x 9 km^2 resolution (Fig S11). In contrast, other inversion techniques—such as 4D-Var or ensemble Kalman filter (EnKF) methods—coupled with high-resolution models, would be more effective for resolving local emissions. However, these approaches are computationally expensive and may not be feasible for long-term simulation. Therefore, the IFDMB with GEMS data provides practical and efficient adjustments to regional NO_x emissions in Thailand.

S5 The analysis for the degree of grid cell aggregation

In this section, we quantify the impact of grid cell aggregation on the IFDMB with the fixed diurnal update scheme. We compare three degrees of aggregation: 5 × 5 grid cells (45 × 45 km, the original IFDMB configuration used in the main text), 3 × 3 grid cells (27 × 27 km), and no aggregation (1 × 1, or 9 × 9 km).

For the comparison with GEMS (Table S2), no aggregation yields the lowest NMB and NRMSE among the three degrees of aggregation. This is because no aggregation overfits the GEMS satellite data by minimizing resolution error compared to others. However, the correlation coefficient for no aggregation is the lowest of the three schemes over Thailand, which we attribute to smearing error that leads to an underestimated adjustment in source areas. The 5-grid and 3-grid aggregations perform comparably in NRMSE and correlation, although the 5-grid aggregation gives a higher NMB over Thailand. Over Bangkok Metropolitan Region, no aggregation underestimates NO_2 VCD, whereas both the 5-grid and 3-grid aggregations overestimate it. This suggests that in large NO_2 emission hotspots such as Bangkok or industrial estates, smearing error is significant. The grid cell aggregation helps prevent the underestimation of emission adjustments in these areas. The 5-grid shows slightly lower NMB and NRMSE in Bangkok Metropolitan Region compared to 3-grid. However, no aggregation still provides the lowest NMB and NRMSE in this area.

For the comparison with TROPOMI (Table S3), no aggregation again provides the lowest NMB and NRMSE across Thailand as a whole, though the correlation coefficients are comparable across all three aggregation schemes, at 0.61 (5-grid), 0.68 (3-grid), and 0.63 (no aggregation). Because GEMS retrievals carry greater uncertainty over North Thailand due to the a priori profile used in the retrieval, we focus our comparison on the Bangkok Metropolitan Region instead. In this area, no aggregation gives the lowest NMB and NRMSE but also the lowest correlation coefficient. While the 5-grid and 3-grid aggregations perform comparably in NMB and NRMSE, the 5-grid aggregation achieves the highest correlation in this area. We interpret this as follows: in dense NO_2 source areas such as Bangkok Metropolitan Region or industrial estates, where NO_2 is transported further from its sources, smearing error is more important for the satellite data, so grid cell aggregation is beneficial because it reduces this error by preventing under-adjustment. In rural or agricultural areas, where NO_2 does not travel far from its source, resolution error dominates instead, so aggregation increases the bias and error in these areas. The 5-grid aggregation shows the highest correlation over the Bangkok Metropolitan Region in particular, likely reflecting the higher wind speeds typical of this coastal area.

For the comparison with PCD stations, all three aggregation schemes yield nearly identical results at roadside stations, with no significant difference in absolute concentration (Figs S11 and S12). This is because GEMS has a resolution of approximately 5 km and our WRF-Chem simulation a resolution of 9 km, neither of which is fine enough to resolve near-road transport of emissions. At non-roadside stations, however, the 5-grid aggregation performs better than the 3-grid and no aggregation (Fig S11). Because most non-roadside stations are located near urban and industrial areas, they are more strongly affected by smearing error associated with the transport of NO_2 away from its sources, and therefore benefit most from the 5-grid aggregation.

Table S2: Monthly statistical evaluation between GEMS NO_2 VCD and WRF-Chem for different degrees of grid cell aggregation using IFDMB with fixed update schemes for Thailand and BMR.

Thailand	**NMB**	**NRMSE**	***R***
Baseline simulation	83.52	151.94	0.65
5 grids	31.90	56.25	0.84
3 grids	16.26	53.61	0.84
no aggregation	6.80	39.74	0.81
BMR	**NMB**	**NRMSE**	***R***
Baseline simulation	278.67	336.69	0.96
5 grids	29.46	51.93	0.95
3 grids	40.02	66.99	0.96
no aggregation	-12.77	43.01	0.96

Table S3: Monthly statistical evaluation between TROPOMI NO_2 VCD and WRF-Chem for different degrees of grid cell aggregation using IFDMB with fixed update schemes for Thailand and BMR.

Thailand	**NMB**	**NRMSE**	***R***
Baseline simulation	0.80	1.21	0.79
5 grids	0.60	0.86	0.61
3 grids	0.45	0.71	0.68
no aggregation	0.39	0.66	0.63
BMR	**NMB**	**NRMSE**	***R***
Baseline simulation	3.46	5.26	0.90
5 grids	0.59	1.25	0.93
3 grids	0.55	1.47	0.86
no aggregation	-0.07	0.93	0.80

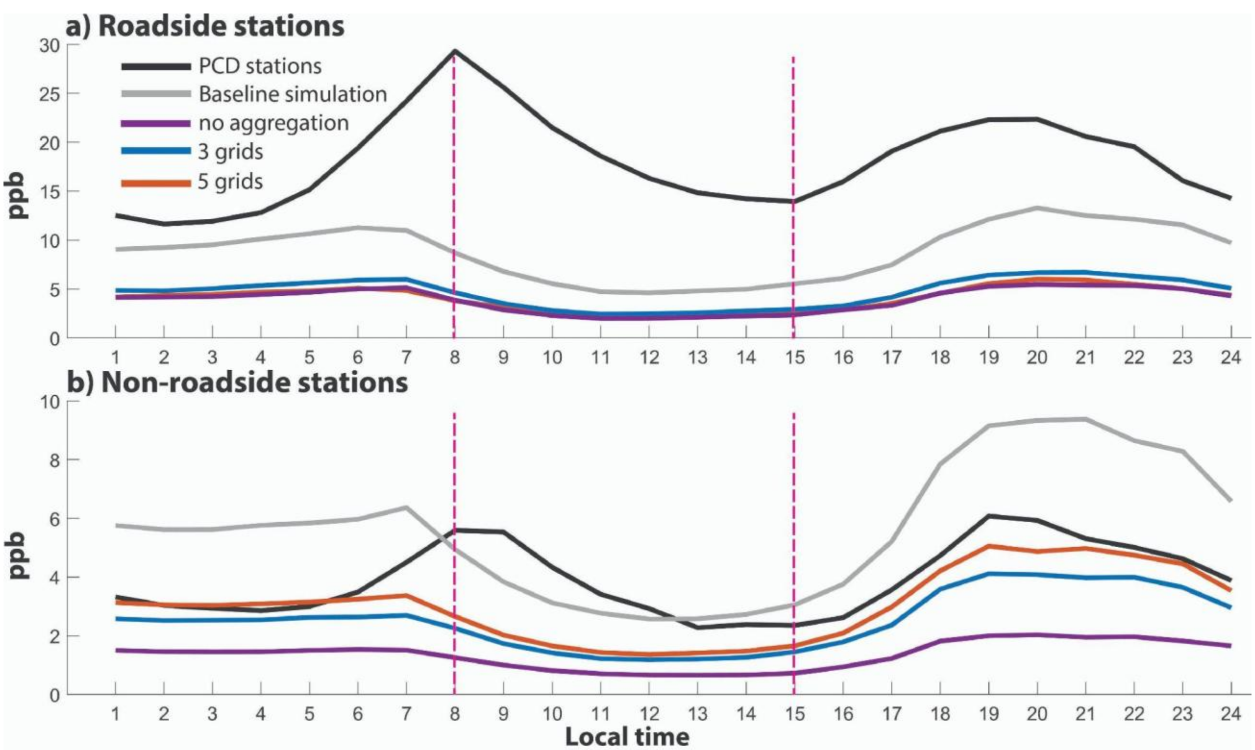

Figure S11: The monthly average diurnal profile of surface NO_x mixing ratios from PCD stations (black) and WRF-Chem: the baseline simulation (gray), no aggregation (purple), 3-grid aggregation (blue), and 5-grid aggregation (orange). PCD stations are classified

into roadside stations (a) and non-roadside stations (b). The pink lines mark the GEMS overpass times in Thailand from 08:00 - 15:00 local time.

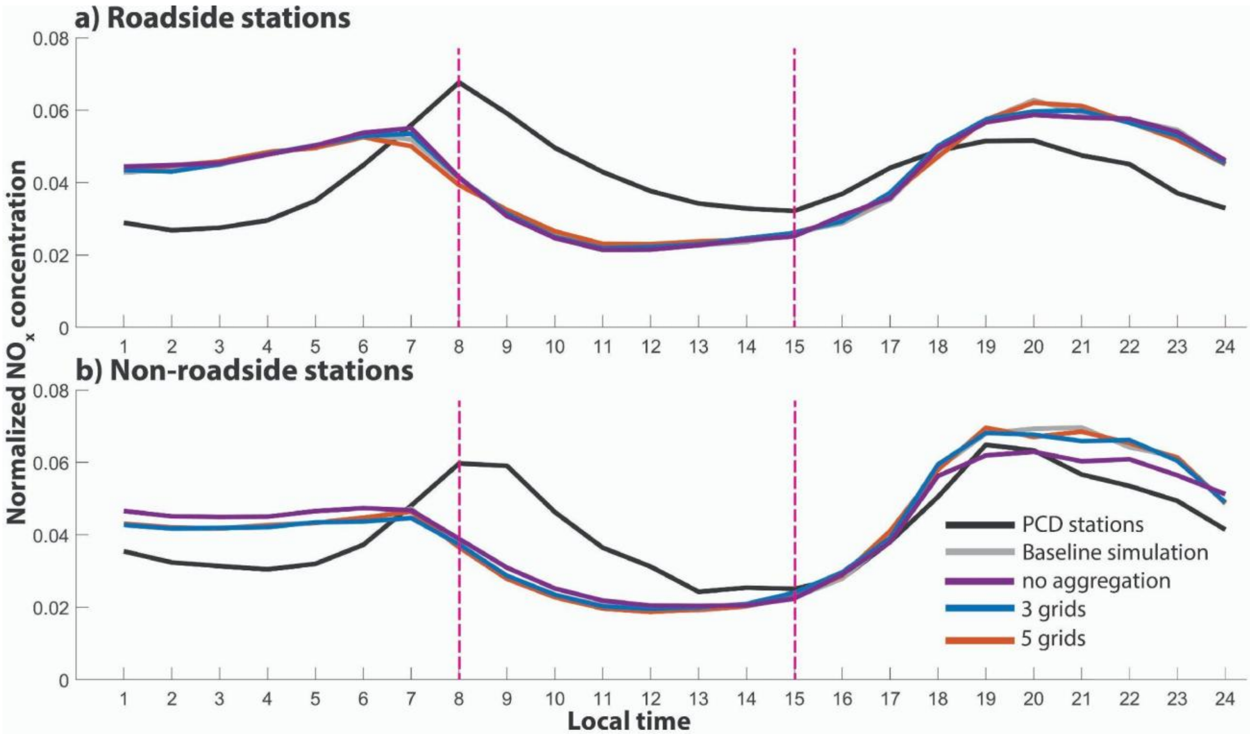

Figure S12: The diurnal variability of normalized NO_x concentrations from PCD stations (black) and WRF-Chem: the baseline simulation (gray), no aggregation (purple), 3-grid aggregation (blue), and 5-grid aggregation (orange). PCD stations are classified into roadside stations (a) and non-roadside stations (b). The pink line is the GEMS scanning time in Thailand from 08:00 - 15:00 local time.

Minor corrections

1. Some numbers are inconsistent between sections. For example, the NRMSE reduction for the temporal-average scheme is 57% in the text (line 581) but 61% in the conclusions. The daylight scheme is 36% (line 579) versus 39% (line 820).

Modified text: We thank the reviewer for catching this. The reviewer is correct that these values were inconsistent. We have re-checked the values against Table S4 and confirm that the figures reported in the Conclusions are the correct ones.

3.5 Comparison with TROPOMI

…

In this region, the daylight diurnal update reduces NMB and NRMSE by 54% and 39%~~36%~~, respectively, compared with the baseline simulation (Table ~~S2~~ S4). The temporal average update yields even larger improvements, reducing NMB by 81% and NRMSE by 61%~~57%~~ (Table ~~S2~~ S4).

2. Line 836 says "October 2023," but the study period is September 2023?

Modified text:
However, our IFDMB approaches increase the NO_x emissions in North Thailand during ~~October~~September 2023 whereas previous study decreases the emissions.

3. Figure references: line 410 "(Fig. 5a)" should be "(Fig. 6a)"; line 522 "10f" does not exist and is likely "Fig. 9f". Please check.
Modified text:
Our results show that *J* decreases substantially during iterations 1 and 2 across all update schemes, but it begins to increase beyond iteration 3 (Fig. ~~5~~ 6a).
…
The results indicate that TROPOMI retrieves higher NO_2 VCDs than GEMS across most regions, except in Lampang, where GEMS values remain substantially larger (Figs. 9d and ~~10~~ 9f).

6. The surface stations (PCD) measure NOx while the model is compared as NOy (lines 195198). Please quantify or at least bound how much this difference affects the results.

Response: We have found that for WRF-Chem on average NOy is higher than NOx around 5% at roadside stations and 6% at non-roadside stations.

Modified text:
2.4 Ground-based monitoring data
For ground-based monitoring data, we use the data from the Pollution Control Department (PCD) of Thailand air quality stations (http://air4thai.pcd.go.th/webV3/) to evaluate surface NO_x concentrations from WRF-Chem. The PCD stations' measurements are based on the chemiluminescence technique, which detects NO_y species (Pollution Control Department, 2022; Zhang et al., 2009). Hence, we compare WRF-Chem NO_y (which includes NO_X and its oxidation products) with NO_x from PCD. The PCD stations (20) are classified into roadside (14) and non-roadside stations (6) as shown in Fig. ~~S8~~ S14. On average, the simulated NO_y concentrations from WRF-Chem is higher than NO_x by 5% at the roadside stations and 6% at the non-roadside stations.

7. The reference list appears to contain a duplicate (Kim et al. 2023a and 2023b look like the same paper). Please also disambiguate the several "Li et al., 2023" citations in the text.

Response: The reviewer is correct. Kim et al. (2023a) and Kim et al. (2023b) are indeed the same publication. We have deleted the duplicate entry, and the single remaining reference is now cited as “Kim et al. (2023)” throughout the text. All in-text occurrences of “Kim et al., 2023a” and “Kim et al., 2023b” have been updated accordingly.

Modified text:

The Geostationary Environment Monitoring Spectrometer (GEMS) is a GEO satellite for air quality monitoring from space, providing unprecedented temporal resolution and continuous observational capabilities that are not feasible with LEO satellites (Kim et al., 2023~~a~~).

It is a geostationary orbit satellite, which was launched in February 2020, with a resolution of 3.5 × 8 km² (Kim et al., 2023~~a~~).

Tropospheric NO_2 VCDs are retrieved using a Differential Optical Absorption Spectroscopy (DOAS) algorithm, applied over the 432–450 nm spectral window (Kim et al., 2023~~a~~).

However, LEO satellites have limited temporal sampling, which can miss rapid changes or diurnal patterns of NO_2 (Kim et al., 2023~~b~~).

Kim, S., Kim, D., Hong, H., Chang, L.-S., Lee, H., Kim, D.-R., Kim, D., Yu, J.-A., Lee, D., Jeong, U., Song, C.-K., Kim, S.-W., Park, S. S., Kim, J., Hanisco, T. F., Park, J., Choi, W., and Lee, K.: First-time comparison between NO $_2$ vertical columns from geostationary environmental monitoring spectrometer (GEMS) and pandora measurements, Atmospheric Meas. Tech., 16, 3959–3972, https://doi.org/10.5194/amt-16-3959-2023, 2023~~a~~.

~~Kim, S., Kim, D., Hong, H., Chang, L.-S., Lee, H., Kim, D.-R., Kim, D., Yu, J.-A., Lee, D., Jeong, U., Song, C.-K., Kim, S.-W., Park, S. S., Kim, J., Hanisco, T. F., Park, J., Choi, W., and Lee, K.: First-time comparison between NO_2 vertical columns from Geostationary Environmental Monitoring Spectrometer (GEMS) and Pandora measurements, Atmospheric Meas. Tech., 16, 3959–3972, https://doi.org/10.5194/amt-16-3959-2023, 2023b.~~

Regarding the “Li et al., 2023” citations, we have modified text:

Moreover, NO_x highly depends on localized sources (Crippa et al., 2023; Miyazaki et al., 2017) and has a short atmospheric lifetime (Horner et al., 2024), leading to sharp spatial gradients and strong heterogeneity in ambient concentrations (Li et al., 2023a).

…

Li, C., Martin, R. V., Cohen, R. C., Bindle, L., Zhang, D., Chatterjee, D., Weng, H., and Lin, J.: Variable effects of spatial resolution on modeling of nitrogen oxides, Atmospheric Chem. Phys., 23, 3031–3049, https://doi.org/10.5194/acp-23-3031-2023, 2023a.

…

Several studies have demonstrated strong correlations between GEMS NO_2 and ground-based measurements in regions such as China and South Korea (Kim et al., 2023~~a~~; Li et al., 2023b; Seo et al., 2024).

…

Li, Y., Xing, C., Peng, H., Song, Y., Zhang, C., Xue, J., Niu, X., and Liu, C.: Long-term observations of NO2 using GEMS in china: validations and regional transport, Sci. Total Environ., 904, 166762, https://doi.org/10.1016/j.scitotenv.2023.166762, 2023b.

8. Typos: line 807 "access" → "assess"; line 478 "Thongsame et al. (2023)" → "Thongsame et al. (2024)."

Modified text:
This study applies high-temporal-resolution GEMS satellite data to update anthropogenic NO_x emissions over Thailand during September 2023 to improve the air quality modeling and ~~access~~ assess diurnal variability of emissions.

Our study reveals that the diurnal variability from Thongsame et al. (~~2023~~ 2024), which shows peaks aligned with morning and evening rush hours, is different from the GEMS diurnal variability.

Reviewer #2

The study by Thongsame uses a chemistry model and new geostationary satellite data to estimate nitrogen oxide emissions in Thailand. Overall, it is an interesting study using established methods to estimate these emissions. Secondary comparisons with TROPOMI and surface sites provide additional validation of the work. The discussion on the discrepancies between TROPOMI and GEMS are an interesting and valuable addition to the literature. The analysis is detailed, which suites the scope of ACP and suitable of publication subject to some minor corrections:

1. In the abstract, GEMS does not "simulate" NO2 (Line 29).

Response: Change from simulate to retrieve

Modified text:
Abstract
…is because observations are strongly influenced by local transportation sources, which are hard to observe and ~~simulate~~ retrieved by GEMS and the model, respectively.

2. The authors state that GEO has higher spatial resolution than LEO. If oversampling is used, the LEO datasets can be higher resolutions (i.e. < 1 km) (Line 90).

Response: We thank the reviewer for pointing this out. We agree that by applying spatial oversampling techniques—which leverage the overlapping swaths and varying pixel geometries of repeated satellite passes—LEO satellites can achieve effective spatial resolutions finer than GEO satellites.

In our original main text, our comparison referred strictly to standard native single-pixel resolutions without accounting for oversampling methodologies. We have revised the text to correct this statement and explicitly acknowledge that LEO satellites can achieve higher spatial resolution with oversampling techniques compared to GEO satellites.

Modified text:
For the past few decades, NO_2 monitoring has relied on sun-synchronous Low Earth Orbit (LEO) satellites (e.g., GOME, SCIAMACHY, OMI, and TROPOMI) that provide a global map of tropospheric NO_2 (Boersma et al., 2008; Martin et al., 2003; Tang et al., 2013). With oversampling techniques, these instruments can resolve spatial features finer than 1 km (Jin et al., 2025). However, LEO satellites have limited temporal sampling, which can miss rapid changes or diurnal patterns of NO_2 (Kim et al., 2023~~b~~). To address the challenges in data availability, Geostationary Earth Orbit (GEO) satellites, which have higher temporal ~~and spatial~~ resolution than the LEO satellite, have been launched (Choi, 2018; Seo et al., 2024; Yang et al., 2023). The Geostationary Environment Monitoring Spectrometer (GEMS) is a GEO satellite for air quality monitoring from space, providing unprecedented temporal resolution and continuous observational capabilities that are not feasible with LEO satellites (Kim et al., 2023a). This satellite is designed to measure air pollutants, including NO_2, sulfur dioxide (SO_2), formaldehyde (HCHO), and aerosols, over Asia (Kim et al., 2020). The hourly NO_2 data from GEMS can provide improvement for anthropogenic emissions (Park et al., 2023) and information about NO_2 diurnal variability (Park et al., 2024).

3. The discussion on the averaging kernels is unclear. Please add the equations used to apply the AKs and clear description. It would also be good to add a figure (e.g. in supporting information) comparing the GEMS and TropOMI AKs.

Response: We thank the reviewer for pointing out the need for greater clarity regarding the application of the Averaging Kernels (AKs). To ensure full transparency and reproducibility, we have revised the manuscript to explicitly define the mathematical formulation and provide a physical interpretation of the AK application, as well as a visual comparison between instruments.

Modified text:

2. Methodology

2.1 WRF-Chem

…

The simulation employs 42 vertical layers extending to approximately 63 hPa, with a horizontal resolution of 9 × 9 km^2. The model tropopause is located at approximately 100 hPa. To compare with GEMS, WRF-Chem NO_2 concentrations are interpolated to GEMS vertical pressure levels. Tropospheric NO_2 VCD from WRF-Chem is then calculated by integrating NO_2 from the surface to the GEMS-defined tropopause at 230 hPa. The GEMS averaging kernel is applied to the model-simulated NO_2 profiles to account for the satellite's retrieval sensitivity, following previous approaches (Cooper et al., 2020; Park et al., 2024). The averaging kernel accounts for the sensitivity of the satellite and is applied to the model to mitigate uncertainties in the retrieval algorithm. Further details regarding the averaging kernels and the applied equations are provided in Sect. S6.

S6 Averaging kernel

Retrieving NO_2 from ultraviolet and visible satellite observations relies on three stages (Cooper et al., 2020): calculating a slant column, removing the stratospheric contribution, and converting the slant column into a vertical column density (VCD) using an air mass factor (AMF). AMFs are calculated using a radiative transfer model and are dependent on viewing geometry, surface reflectance, and cloud cover (Eskes and Boersma, 2003). Furthermore, these calculations require an a priori estimate of the NO_2 vertical profile (Cooper et al., 2020; Eskes and Boersma, 2003). Uncertainties in the AMF and the a priori information used in the retrieval algorithm can degrade the quality of model–satellite comparisons beyond the inherent errors of either the model or the measurements alone (Boersma et al., 2007). To mitigate these errors, an averaging kernel (AK) is commonly applied to the model output. The AK accounts for the altitude-dependent sensitivity of the satellite, providing a consistent mathematical relationship between the retrieved quantities and the true tracer profile (Eskes and Boersma, 2003).

To account for the sensitivity of the satellite instrument, the AK is applied to the interpolated model profile using the following equation (Eskes and Boersma, 2003):

$$\Omega_{WRF_{satellite}} = \sum_{i=1}^{i=tropopause} AK_i \, \Omega_{WRF,i} \quad \text{(S4)}$$

where $\Omega_{WRF_{satellite}}$ is the tropospheric vertical column density (VCD) accounting for satellite sensitivity. AK_i is the averaging kernel from satellite data at layer i. $\Omega_{WRF,i}$ is the simulated WRF-Chem NO_2 vertical column density at layer i, after being mapped to the satellite's vertical pressure grid.

The AK is mathematically derived from the AMF and scattering weights. Because these values depend on the specific retrieval algorithm and the assumed a priori profiles, the resulting AKs differ inherently between GEMS and TROPOMI. To evaluate the relative sensitivities of these satellites, the monthly mean surface of the first layer (i = 1) AKs for both GEMS and TROPOMI over Thailand are compared in Fig. S13. The monthly mean AK_{GEMS} aggregates all daylight diurnal data points (00:45–04:45 UTC and 06:45–07:45 UTC), whereas the monthly mean $AK_{TROPOMI}$ is based solely on the single daily overpass at approximately 06:30 UTC. While the AK spatial distributions across Central Thailand and the Bangkok Metropolitan Region are similar between the two satellites, distinct disparities emerge in Northern Thailand. TROPOMI exhibits elevated surface AK over Chiang Mai and the Mae Moh power plant, whereas GEMS shows higher AK over the broader Lampang area. These regional discrepancies are largely attributable to uncertainties in the a priori profiles utilized by the GEMS retrieval algorithm. Additionally, significant AK variations are shown in Rayong, a major industrial estate containing extensive harbor and power plants. $AK_{TROPOMI}$ demonstrates high AK along the coastal areas, capturing emissions from shipping, coastal power plants, and near-shore industrial zones. Conversely, AK_{GEMS} exhibits its highest values in inland industries and power plants. Ultimately, these spatial discrepancies confirm that the retrieval algorithms of the two satellites possess distinctly different sensitivities to surface-level NO_2 concentrations.

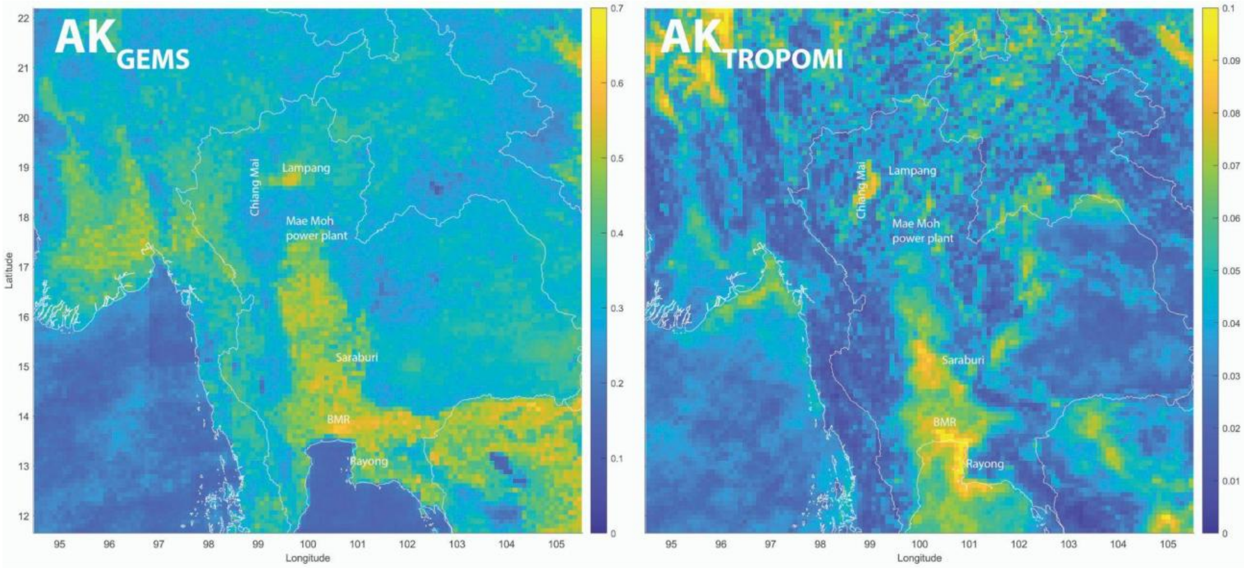

Figure S13: The monthly mean averaging kernel for GEMS and TROPMI at the surface (1st layer).

4. Why is 230 hPa used for the tropopause (Line 167)? It will also shift in the day, which can be important. Please justify the reasoning behind this.

Response: We thank the reviewer for raising this important point regarding the definition of the tropopause and its diurnal variability. Our use of a fixed 230 hPa pressure level for the tropopause is dictated by the operational GEMS Version 3 retrieval algorithm. The GEMS v3 product currently assumes a static 230 hPa tropopause across the domain. Because the operational product does not publicly release the underlying a priori model profiles required to change the tropopause in the algorithm, we are constrained to adopt the 230 hPa definition to ensure consistency between the satellite retrieval and our model comparisons.

However, we fully agree with the reviewer that diurnal shifts in tropopause height can be important depending on the atmospheric regime. We conducted a sensitivity check using our WRF-Chem simulations. Specifically, we compared the integrated tropospheric NO_2 columns calculated using a fixed 230 hPa boundary against those calculated using the model's dynamically varying tropopause. We found that the difference between the two NO_2 columns has a negligible impact during the daytime. This is because the NO_2 in Thailand during September is overwhelmingly dominated by near-surface emissions and the NO_2 beyond 230 hPa is almost zero everywhere in Thailand.

Modified text:

2.2 GEMS

…

To ensure data quality and improve the reliability of the emission adjustment process, we manually remove these from our analysis. The data coverage in North Thailand is limited, and retrieval errors are relatively high in both North Thailand and the Bangkok Metropolitan Region (Figs. 2b and 2c). ~~The tropopause pressure level used in GEMS retrievals is fixed at 230 hPa across all pixels.~~ In GEMS version 3, the tropopause pressure level is assumed to be fixed at 230 hPa across all pixels. As GEMS does not publicly release full a priori model profiles required to change the tropopause in the algorithm, we are constrained to adopt the 230 hPa definition to ensure consistency between the satellite retrieval and our WRF-Chem model. Our sensitivity test using WRF-Chem reveals that a fixed tropopause at 230 hPa and the model's varying tropopause have a negligible impact during the daytime. This is because the NO_2 in Thailand during September is overwhelmingly dominated by near-surface emissions, and the NO_2 beyond 230 hPa is almost zero everywhere in Thailand.

5. Please expand on how the RMSE is calculated as the error term for GEMS data (Line 226)? Would the RMSE be based on the comparison of GEMS with a secondary data set?

Response: In GEMS Level 2 output products, the Root Mean Square Error (RMSE) is defined as the spectral fitting residuals resulting from the DOAS (Differential Optical Absorption Spectroscopy) method between 432 and 450 nm. GEMS simultaneously fits NO_2, O_3, as well as a Ring spectrum, wavelength shift with respect to the reference spectrum, and 3rd-order closure polynomials.

Modified text:
In this study, we assume a 50% relative uncertainty for the prior emissions based on previous studies (Jung et al., 2022; Park et al., 2024; Souri et al., 2020). The observation error is derived from the root mean square error (RMSE) in GEMS data, which is calculated from the spectral fitting residuals resulting from the DOAS (Differential Optical Absorption Spectroscopy) method in the spectral window between 432 and 450 nm. All of the terms in the above equation pertain to values in a single grid cell. The approach is applied independently across all grid cells in Thailand.

6. In Section 2.7, how are the metrics (NMB) calculated? Is this done using the observational error? I ask as the MB is used for the surface sites and no error term is available, so I assume the satellite metrics are normalised via the error term. Please make this clear what is used.

Response: We thank the reviewer for pointing this out, and the reviewer is correct in the assumption. In our study, the Normalized Mean Bias (NMB) for the satellite data is calculated by normalizing the Mean Bias (MB) using the observation error (σ_o) provided by both the GEMS and TROPOMI satellite products. Because the surface measurements from the PCD stations do not include an observation error term, we evaluate the model against those surface sites using the non-normalized MB and Root Mean Square Error (RMSE). We have clarified this distinction in Section 2.7 of the revised manuscript.

Modified text:

2.7 Evaluation method

Three metrics, namely, normalized mean bias (NMB), normalized root mean square error (NRMSE), and correlation coefficient (*R*), are used to evaluate the performance of the IFDMB solution in comparison to our simulation using the prior emissions. The observation error (σ_o) provided by ~~from~~ GEMS and TROPOMI data is used to normalize mean bias (MB) and root mean square error (RMSE). We compare our results with tropospheric NO_2 VCD from GEMS and TROPOMI, and with in situ NO_x measurements from PCD stations. For GEMS and TROPOMI, the monthly average tropospheric NO_2 VCDs in the Thailand grid cells are compared with the monthly average WRF-Chem results. This is because only emissions in Thailand are updated, while emissions outside Thailand remain the same. For the satellite data, our metrics are normalized with the observation error. However, for PCD stations, we evaluate our results with non-normalized MB, non-normalized RMSE, and *R*. This is because ~~as~~ the PCD stations have not reported the observation error ~~of the instruments~~. Additionally, we compare the monthly average diurnal NO_x profile with PCD data.

7. In Section 3.1 and discussion of Figure 5, please add some more quantitative discussion as scientifically weaker as it is.

Response: We added more quantitative discussion in section 3.1 and discussion of Figure 5.

3.1 Comparison between baseline simulation and GEMS

This section shows the performance of the WRF-Chem simulation using the prior emissions (the CAMS-GLOB-ANT inventory), aka the baseline simulation, compared to GEMS NO_2 VCDs before applying any mass balance methods. Overall, the baseline simulation significantly overestimates the monthly average NO_2 VCDs in the Bangkok Metropolitan Region and East Thailand (Fig. 4). The baseline simulation shows that the NO_2 VCD is dominated by large point sources from many sectors. In North Thailand, the prior emissions identify the Mae Moh coal power plant (energy sector) as the main source of NO_2, with a signal that significantly overwhelms the urban emissions from Chiang Mai and Lampang (Fig. 1). In the Bangkok Metropolitan Region, the model suggests a mixed contribution from the dense residential and transportation sectors of the urban area. Additionally, industrial estates in Saraburi and Rayong produce high NO_2 – as strong as urban emissions in Bangkok. GEMS presents a contrasting picture that emphasizes the role of urban sources. In North Thailand, GEMS has the highest NO_2 VCDs centered over the urban areas of Chiang Mai and Lampang, which are dominated by residential and transportation emissions. The Mae Moh power plant appears as only a moderate signal in the GEMS observations. This pattern holds in the Bangkok Metropolitan Region, where GEMS again highlights the urban area, while the signals from major industrial estates in Rayong and Saraburi are notably weaker than the urban emissions. Additionally, the GEMS NO_2 VCDs in North Thailand are as large as in the Bangkok Metropolitan Region, while the baseline simulation shows significantly larger NO_2 VCDs in the Bangkok Metropolitan Region. This direct contradiction between the baseline simulation and GEMS highlights that the CAMS-GLOB-ANT is likely to overestimate energy and industrial emissions relative to the residential and transportation emissions in Thailand. The baseline simulation shows high NMB and NRMSE in most regions, with comparatively better performance in North and West Thailand (Figs. 5 and ~~S9~~ S15). However, GEMS data availability is high in Central Thailand and the Bangkok Metropolitan Region, moderate in the North and East, and very limited in West Thailand (Fig. 2). When averaged across all grid cells in Thailand, the baseline simulation yields an NMB of 83.52 and an NRMSE of 151.94 relative to GEMS. These metrics quantitatively support the widespread overestimation of modeled NO_2, though the correlation coefficient across the domain remains moderate (R = 0.65). ~~The correlation coefficient is moderate (R = 0.65).~~ In the Bangkok Metropolitan Region, the baseline simulation shows an even more pronounced overestimation, with an NMB of 278.67 and an NRMSE of 336.69. Despite these high biases, the correlation coefficient in this area is exceptionally strong (R = 0.96). This suggests that while the baseline NO_2 spatial distribution strongly agrees with the satellite, the absolute VCDs of the prior emissions in this region are severely overestimated.

3.3 Comparison with GEMS

This section presents the results from WRF-Chem simulations using the IFDMB method evaluated with GEMS. The IFDMB with all three update schemes reduces both NMB and NRMSE in GEMS compared to the baseline simulation (Fig. 5). The IFDMB with fixed diurnal update demonstrates the highest improvement among the three update schemes. The NMB and NRMSE reduce by about 60% for the fixed diurnal update averaged across all grid cells in Thailand. Reductions are most pronounced in Central Thailand and the Bangkok Metropolitan Region (Figs. 5 and S9). In Bangkok Metropolitan Region, the IFDMB with fixed diurnal update reduces NMB to 29.46 and NRMSE to 51.93. The correlation coefficient remains the same as 0.96. Overestimated NO_2 VCDs from transportation and residential sectors are substantially corrected. However, improvements in North and West Thailand are more modest. The improvement in those areas is limited due to the low data density and high errors in those areas. The correlation coefficient (R) improves from 0.65 (baseline simulation) to 0.84.

The IFDMB with daylight diurnal and temporal average updates produces similar results. Both methods improve the performance in urban areas, particularly in Bangkok and big cities in North Thailand, but are less effective in other areas (Figs. 5 and S9). The NMB is reduced by 14% for the IFDMB with daylight diurnal update and 9% for the temporal average update averaged across all grid cells in Thailand. However, within the Bangkok Metropolitan Region, these two schemes show lower performance. Specifically in the Bangkok Metropolitan Region, the NMB decreases to 160.26 under the daylight diurnal update and to 110.52 under the temporal average update. Similarly, the NRMSE in this region drops to 214.43 and 158.45 for the daylight diurnal and temporal average updates, respectively. Despite these remaining absolute biases, the correlation coefficient R remains exceptionally strong at approximately 0.96 for both methods. Notably, both updates exhibit small improvements during early morning hours (e.g., 8–9 AM) compared to the IFDMB with fixed diurnal update (Fig. ~~S10~~ S16). Since these methods only update emissions during the daylight, uncorrected nighttime emissions likely contribute to the error of NO_2 VCD in the early morning. In contrast, the fixed diurnal update, which adjusted emissions at all time steps, shows lower NMB and NRMSE throughout the day. The NRMSE reduction for these two methods is around 20% and correlation coefficients improve to 0.74 for the daylight diurnal update and 0.78 for the temporal average update.

8. Figure 6, would you not expect more iterations before the IFDMB converges. Iterations of 2-4 seems low or very quick to converge on a solution?

Response: We thank the reviewer for this observation. There are several reasons supporting the rapid convergence within 2–4 iterations. The iteration in IFDMB accounts for horizontal transport and the nonlinear chemistry of NO_2. Regarding transport, we reduce the smearing error by regridding both GEMS and WRF-Chem NO2 VCDs to a coarser resolution. We also use monthly averages of GEMS and WRF-Chem, which reduce the impact of noise and missing data in the satellite observations and further minimize the smearing error. Together, these steps largely confine NO2 transported out of a source grid cell to the same coarse grid cell. The transport effect that iteration would otherwise need to resolve is therefore absorbed before the inversion begins, which is why fewer iterations are required here.

Additionally, we have found that the relative magnitudes of the GEMS observation errors and our prior emission uncertainties help IFDMB converge within a few iterations. The GEMS observation error is on the order of 10^{13}, whereas the GEMS NO_2 VCD data is on the order of 10^{15}. This large difference in magnitude results in a dominant $J_{prediction}$ term in Eq. (1), which drives a substantial emission adjustment during the very first iteration, followed by much smaller adjustments in subsequent iterations. Furthermore, we add the regularization parameter (C_1 = $1x10^{14}$) to prevent over-adjustment by constraining $J_{prediction}$. This avoids oscillatory fluctuations during convergence. We also assume a 50% relative uncertainty of the prior emissions as emission error. However, the true uncertainty for specific sectors like power plants and industry may be lower due to direct emissions reporting. This large assumed emission uncertainty yields a relatively small $J_{parameter}$ penalty term, giving the IFDMB algorithm the flexibility to make large initial emission adjustments to reconcile the model with the GEMS observations. Finally, we tested running the model out to 6 iterations for the fixed diurnal update scheme. We found that $J_{prediction}$ actually began to increase after 4 iterations, indicating that 2–4 iterations represent the optimal convergence window before overfitting occurs.

Modified text:

3.2 Evolution and optimization of mass balance methods

To determine the optimal number of iterations for the IFDMB inversion, we calculate the total cost function J, $J_{prediction}$, and $J_{parameter}$ based on comparisons between NO_2 VCDs from WRF-Chem and GEMS at each iteration (Fig. 6). Our results show that J decreases substantially during iterations 1 and 2 across all update schemes, but it begins to increase beyond iteration 3 (Fig. ~~5~~ 6a). This increase in the total cost function is primarily driven by a rise in $J_{parameter}$ (Fig. 6c). The $J_{prediction}$ term slightly decreases after iteration 2. However, $J_{prediction}$ does not decrease monotonically due to overadjustment; it may increase in some iterations before decreasing again in subsequent steps (Fig. 6b). There are several reasons supporting the rapid convergence within 2–4 iterations. The iteration in IFDMB accounts for horizontal transport and the nonlinear chemistry of NO_2. Regarding transport, we reduce the smearing error by regridding both GEMS and WRF-Chem NO_2 VCDs to a coarser resolution. We also use monthly averages of GEMS and WRF-Chem, which reduce the impact of noise and missing data in the satellite observations and further minimize the smearing error. Together, these steps largely confine NO_2 transported out of a source grid cell to the same coarse grid cell. The transport effect that iterations would otherwise need to resolve is therefore absorbed before the inversion begins, which is why only few iterations are required. Additionally, the GEMS observation error is on the order of 10^{13} (Fig. 2c), whereas the GEMS NO_2 VCD data is on the order of 10^{15} (Fig. 2a). This large difference in magnitude results in a dominant $J_{prediction}$ term in Eq. (1), which drives a substantial emission adjustment during the very first iteration, followed by much smaller adjustments in subsequent iterations. Furthermore, our regularization parameters prevent over-adjustment and avoid oscillatory fluctuations during convergence. In this study, we assume a 50% relative uncertainty for the prior emissions error. However, the true uncertainty for specific sectors like power plants and industry may be lower due to direct emissions reporting. This large assumed emission uncertainty yields a relatively small $J_{parameter}$ penalty term, giving the IFDMB algorithm the flexibility to make large initial emission adjustments to reconcile the model with the GEMS observations. Our analysis shows that the daylight diurnal scheme exhibits significant overadjustment because it assumes an immediate response of NO_2 VCDs to emission changes. For the fixed diurnal and daylight diurnal update schemes, iteration 2 provides the lowest *J* and is thus selected as optimal. For the temporal average update scheme, iteration 4 yields the optimal performance. ~~Furthermore, analyses of the sensitivity of the cost function in the fixed diurnal scheme the parameters C_1, C_2, demostrate consistent model performance patterns across different parameter values (Sect. S1).~~

9. A more detailed discussion on the differences between TropOMI and GEMS. Why can you not directly compare the two products? Would that not provide a systematic bias between the two instruments which other studies could use.

Response: We thank the reviewer for raising this point, which we believe will help readers better understand cross-satellite comparison. TROPOMI and GEMS cannot be directly compared using absolute VCD values because the two products are retrieved with different algorithms. Converting slant column density (SCD) to VCD requires an air mass factor (AMF), and GEMS and TROPOMI each apply a different a priori profile to compute it: TROPOMI uses the TM5-MP global chemistry transport model, while GEMS uses WRF-Chem with the ASIA-AQv3 inventory. This difference in AMF produces differences in absolute tropospheric NO_2 VCD between the two products that stem from the retrieval algorithm rather than from the underlying satellite measurements. Our study therefore focuses on the spatial distribution and relative NO2 VCD values between the two satellites rather than on their absolute bias or concentration.

There are two ways to enable a more direct comparison. The first is to harmonize the a priori profiles used by GEMS and TROPOMI (Cooper et al., 2020). The second is to use a chemical transport model (CTM) as an intercomparison platform (Zhang et al., 2010). We adopt the CTM method here because our application uses GEMS to update emissions, and we are more interested in the relative sensitivity differences between GEMS and TROPOMI and how they affect the model than in absolute concentration. We also wanted to preserve the original GEMS NO2 VCDs to keep our comparison method consistent; harmonizing the a priori profiles would require recomputing the AMF every time emissions are updated using GEMS, which would work against that consistency. The two methods should yield consistent conclusions, but each suits a different objective: for applications comparing satellite values directly, or using satellite data alone, the a priori harmonization approach is more suitable, as the reviewer notes, since it would reveal the systematic bias between the instruments. For modeling applications like ours, the CTM method is more useful, since it applies the averaging kernel (AK) from each satellite to the same model output to assess how the satellites' differing sensitivities compare relative to the model.

Modified text:

3.5 Comparison with TROPOMI

In this section, we compare tropospheric NO_2 VCDs from the WRF-Chem simulations and GEMS with those from TROPOMI (Fig. 9). Because GEMS and TROPOMI differ in their averaging kernels and overpass times, a direct comparison is not feasible. Therefore, we focus on assessing the relative spatial distributions of NO_2. To compare the NO_2 VCD between GEMS and TROPOMI, WRF-Chem is used as the intercomparison platform, following the methodology of Zhang et al. (2010). This approach applies the AKs from each satellite to the same WRF-Chem outputs to assess how the distinct vertical sensitivities of the instruments differ relative to the model (Figs. 9b, 9d, and 9f). To ensure consistency, the GEMS and WRF-Chem VCDs closest to the TROPOMI overpass time are selected for comparison (Figs. 9c and 9d). The differences between the GEMS and TROPOMI overpass times are about 15 minutes. The monthly average GEMS NO_2 VCDs at the TROPOMI overpass times are slightly higher than the monthly average NO_2 derived from all GEMS observations (Figs. 9a and 9c). Moreover, they exhibit increased noise, particularly over rural areas. Overall, both satellites show similar spatial patterns of NO_2, except in North Thailand. High NO_2 is consistently observed in the Bangkok Metropolitan Region, Saraburi, and Rayong from both GEMS and TROPOMI. In North Thailand, both satellites detect high NO_2 associated with the Mae Moh power plant; however, GEMS shows anomalously high NO_2 over Lampang that is not observed by TROPOMI (Figs. 9c and 9e). The NO_2 levels in Lampang from GEMS exceed those in Bangkok, whereas TROPOMI indicates the highest NO_2 over the Bangkok Metropolitan Region. Additionally, TROPOMI generally has higher background NO_2 levels compared to source regions. To further examine the magnitude differences, we apply the respective averaging kernels of GEMS and TROPOMI to the baseline simulation.

10. For Section 3.5, could you not replace e.g. the GEMS apriori with that of TropOMI and recalculate the trop col NO2 around the TropOMI overpass time and GEMS hourly measurements. That would allow more direct and useful comparisons.

Response: We sincerely thank the reviewer for this insightful suggestion. In Figs. 9a, 9c, and 9e, we average GEMS and TROPOMI without replacing the a priori. Fig. 9a shows the monthly average NO2 VCD from GEMS using all available GEMS data, while Fig. 9e shows the monthly average from TROPOMI. Because GEMS observes from 00:45–04:45 UTC and 06:45–07:45

UTC, Fig. 9a is built from more data points than Fig. 9e. To compare the two products within the same time window, we selected the GEMS observations nearest the TROPOMI overpass time and averaged them to compute a GEMS NO2 VCD sampled at the TROPOMI overpass time (Fig. 9c), which lets us see how the GEMS and TROPOMI NO2 VCDs differ.

For a more direct comparison, we adopted the CTM method of Zhang et al. (2010), using WRF-Chem as the intercomparison platform between GEMS and TROPOMI (Figs. 9b, 9d, 9f). We chose the CTM method over updating the a priori because our application uses GEMS to update emissions for WRF-Chem, and we are more interested in the relative sensitivity differences between GEMS and TROPOMI and how they affect the model than in absolute concentration. The CTM method and the a priori-updating approach would be expected to yield the same overall trend.

Added References

Boersma, K. F., Eskes, H. J., Veefkind, J. P., Brinksma, E. J., Levelt, P. F., Stammes, P., Gleason, J. F., and Bucsela, E. J.: Near-real time retrieval of tropospheric NO_2 from OMI, Atmos Chem Phys, 2007.

Cooper, M. J., Martin, R. V., Henze, D. K., and Jones, D. B. A.: Effects of a priori profile shape assumptions on comparisons between satellite NO_2 columns and model simulations, Atmos. Chem. Phys., 20, 7231–7241, https://doi.org/10.5194/acp-20-7231-2020, 2020.

Eskes, H. J. and Boersma, K. F.: Averaging kernels for DOAS total-column satellite retrievals, Atmos Chem Phys, 2003.

Jin, X., Merchant, Z., and Sun, K.: Physics-Based Spatial Oversampling of TROPOMI NO_2 Observations to US Neighborhoods Reveals the Disparities of Air Pollution, GeoHealth, 9, e2025GH001423, https://doi.org/10.1029/2025GH001423, 2025.

Park, J., Choi, Y., Jung, J., Lee, K., and Yeganeh, A. K.: First top-down diurnal adjustment to NO_x emissions inventory in Asia informed by the Geostationary Environment Monitoring Spectrometer (GEMS) tropospheric NO2 columns, Sci. Rep., 14, 24338, https://doi.org/10.1038/s41598-024-76223-1, 2024.

Zhang, L., Jacob, D. J., Liu, X., Logan, J. A., Chance, K., Eldering, A., and Bojkov, B. R.: Intercomparison methods for satellite measurements of atmospheric composition: application to tropospheric ozone from TES and OMI, Atmos. Chem. Phys., 10, 4725–4739, https://doi.org/10.5194/acp-10-4725-2010, 2010.