# Response to the Referee #1 for $NO_x$ Emissions Constraints from GEMS $NO_2$ Retrievals: Inversion Methodology and Air Quality Model Evaluation in Bangkok using ASIA-AQ Multi-Platform Observations

Christopoulos et al.

We thank the reviewer for their careful reading of the manuscript and for the constructive comments that have helped improve its quality. Below, we respond to each comment (italics). Revisions made in the manuscript are indicated in green text.

**Anonymous Referee #1**

**General Comments:**

**Comment #1:** *Section 3: The study relies heavily on the CSF method for deriving top-down NOx emissions. While this approach is computationally inexpensive, it has known limitations. For instance, how does the algorithm determine the optimal length for line density integration? Additionally, what is the sensitivity of the inferred emissions to transport wind speed errors and the fitting process for line density and chemical lifetime? I suggest the authors include a brief discussion on the structural uncertainties inherent to the CSF method.*

**Response #1:** Thank you for this helpful comment. We agree that the CSF method introduces several structural uncertainties that should be clearly described and so have revised Section 3 to explicitly address these points. Specifically, we now clarify that the integration bounds used to compute line densities are defined by the plume subregion identified using the *ddeq* framework, based on GEMS-observed enhancements and ERA5 wind direction. We also added a discussion of the uncertainty of inferred emissions to transport wind speed, noting that errors in wind magnitude and direction directly propagate into emission estimates through the flux calculation (referring to Figure S5). In addition, we now describe uncertainties associated with the Gaussian fitting of line density and the estimation of the effective lifetime, both of which influence the derived emissions. These additions are included in a new paragraph (Section 3.3.5).

Revised Manuscript (Lines 304 – 307): The integration bounds $y_1$ and $y_1$ are defined by the plume subregion (polygon) identified using the ddeq framework, which delineates the area of enhanced $NO_2$ associated with the source based on plume detection and wind direction (Kuhlmann et al., 2024). As such, the effective integration length across the plume is determined by the spatial extend of this polygon and the fitted Gaussian representation.

Revised Manuscript (Lines 3336 – 348)

### 3.1.5. Structural uncertainties in the CSF framework

While the CSF framework provides a computationally efficient approach for deriving satellite-derived emissions, we acknowledge the method contains some structural uncertainties. First, the integration bounds used to compute the line densities (Eq. 4) are defined by the plume polygon identified from the satellite observations and wind field, such that the inferred top-down emissions depend on how the plume extent is delineated. As a result, inferred satellite-derived emissions can be sensitive to the selected plume extend, particularly under conditions of weak enhancements or overlapping sources. Second, the satellite-derived emission estimate scales directly with the effective wind speed (Eq. 1), such that errors in wind magnitude or direction propagate into the derived fluxes and satellite-derived emissions (see Fig. S4). Although ERA5 winds averaged over the lowest model layers are used to approximate plume transport, they may not fully represent local meteorological variability. Third, the inferred satellite-derived emission rate depends on the assumptions within the fitting framework, including the Gaussian representation of plume structure (Eq. 5), the parametrization of plume spreading (Eq. 6), and the effective lifetime $\tau$ (Eq. 7), which represent a simplified treatment of chemical decay and dispersion. Together, these factors represent inherent uncertainties in the CSF approach that should be considered when interpreting the satellite-derived emissions.

**Comment #2:** *The authors develop a novel approach for updating diurnal emission profiles; however, the inversion is essentially simplified to a 1-D problem (magnitude over time for a single urban box). Inaccuracy in the spatial distribution of NOx sources also contributes significantly to model error. Do the authors have thoughts on how to enhance this method to quantify emissions on a finer spatial scale?*

**Response #2:** Thank you for the comment. We would like to clarify that while the inversion is formulated as a temporal optimization (i.e., solving for hourly scaling factors), these scaling factors are applied to each grid cell within the Bangkok box of the domain. As a result, the emissions retain their spatial structure from the prior inventory but are adjusted uniformly in time across all grid cells.

We agree, however, that this approach does not explicitly optimize the spatial distribution of emissions, and therefore any spatial inaccuracies in the prior (e.g., EDGAR downscaled with ODIAC) spatial pattern are not corrected by the inversion. Given the current framework, which treats emissions as area sources, improving spatial patterns is inherently limited. Addressing this would likely require a more detailed treatment, such as performing separate inversions for distinct source types (e.g., point and area sources, including features like the economic corridor southeast of Bangkok). In addition, the use of a higher-resolution local inventory could further improve the representation of regional variability. However, this study relied on the best available data at the time (EDGAR).

We have clarified this distinction and added a brief discussion in the revised manuscript.

Revised Manuscript (Lines 1051 – 1060):

### 7.2 Limitations and potential extensions of the inversion framework

The current model emission update framework optimizes the temporal evolution of emissions but does not explicitly resolve regional variability, as a single set of hourly scaling factors is applied uniformly across all grid cells within the BMR. As a result, the inversion preserves the spatial structure of the prior inventory,

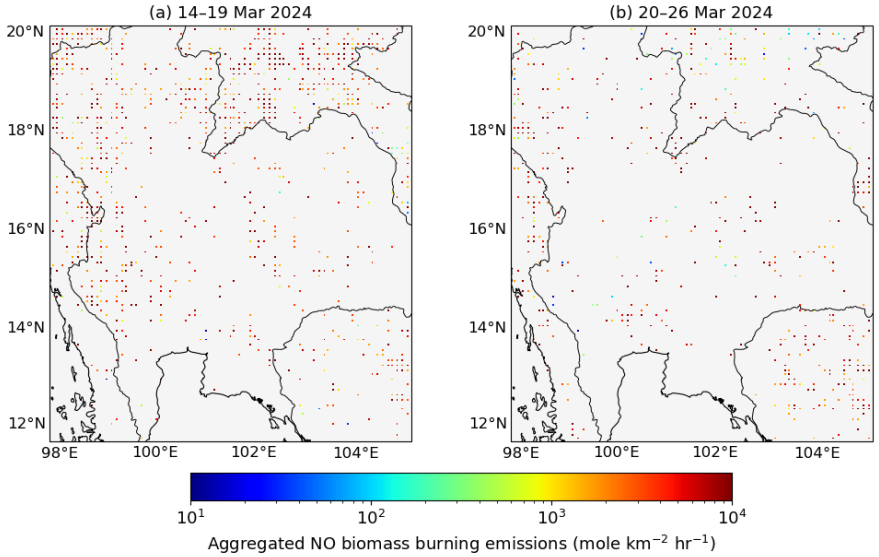
and any inaccuracies in the spatial distribution of emissions are not corrected.

Extending this approach to resolve emissions at finer scales would require allowing emissions to vary across grid cells or subregions, supported by additional constraints. These constraints could include regional higher-resolution regional prior inventories when available, as well as higher-resolution observations (e.g., GCAS) that can better resolve the urban variability in $NO_2$ as shown in this analysis.

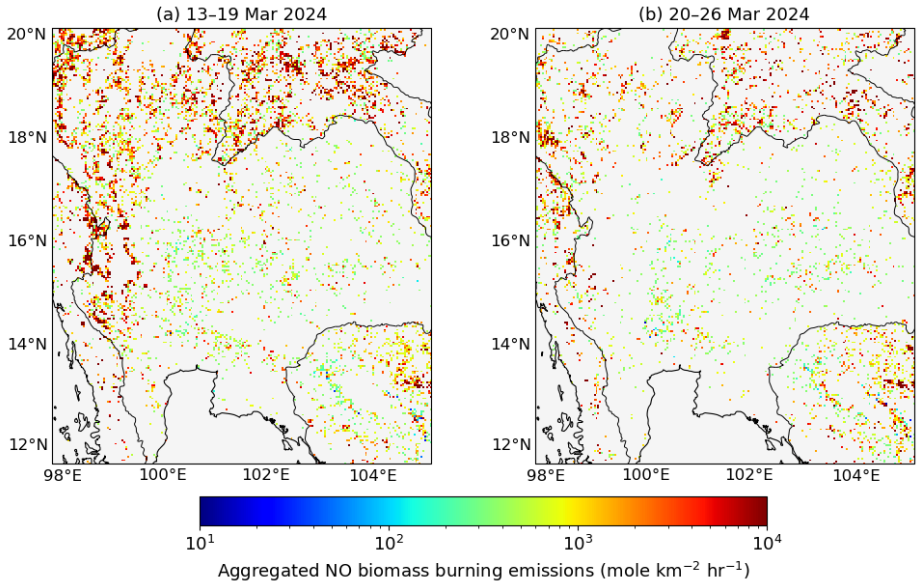
**Comment #3:** *Biomass burning in Southeast Asia significantly influences regional air quality. Was the study period affected by any specific biomass burning events? I would appreciate a brief discussion on how such events might impact the NOx and O3 simulations, as well as the impact on the NOx emission inversion.*

**Response #3:** Thank you for your important comment. To address this, we include an additional plot of biomass burning emissions (aggregated NO emissions) below derived from the QFED (used in this study) and FINN v2.5 inventories. These inventories provide a reasonable range of fire activity, with QFED generally representing a lower-bound estimate (particularly with agricultural fires in SE Asia) and FINN representing an upper-bound estimate. We separate the plots to illustrate two different regimes during our study period (13 – 19 March and 20 – 27 March 2024). During the study period, biomass burning activity was present primarily outside of the Bangkok region with burning focused on regions in northern Thailand and Myanmar. Peak fire activity was seen from 10 – 15 March. During this period, weak synoptic flow likely reduced efficient long-range transport of biomass burning plumes, likely contributing to buildup in the source regions and minimal influence in the Bangkok area given chemical lifetime of $NO_x$. Following 20 March, precipitation in the region substantially reduced fire activity as depicted in the below plots (b), indicating substantially cleaner conditions. While burning can contribute to regional $NO_x$ and $O_3$ enhancements, its impact on the inversion results presented here is expected to be somewhat limited given the burning locations, which substantial impacts likelier seen in cities in the north (e.g., Chiang Mai). We have added clarifying information on this in the independent evaluation to frame interpretation of the results.

**QFED NO Emission:**



**FINN NO Emission:**




Revised Manuscript (Lines 675 – 680): Biomass burning was active in the region during the study period, particularly over northern Thailand and Myanmar, with peak activity occurring in mid-March. However, transport to Bangkok was likely limited due to weak synoptic flow and short chemical lifetime of $NO_x$, reducing its impact on urban concentrations of $NO_x$ in the BMR. Following March 20, precipitation suppressed fire activity, further minimizing its impact. While biomass burning may contribute to broader regional enhancements, its impact on the inversion results and coincident model evaluation over Bangkok is expected to be minor.

**Comment #4:** *Section 3&4: The current terminology used to describe emissions is inconsistent and may be confusing to the reader. Terms such as 'prior NOx emissions,' 'CSF-derived emissions,' and 'emissions after applying the forward operator' are used interchangeably or in ways that overlap. I suggest the author review these sections to consolidate the terminology and ensure that uniform terms are used consistently throughout the manuscript to distinguish between the different emissions datasets.*

**Response #4:** Thank you for this helpful comment. We agree that the terminology describing the different emission datasets was not sufficiently consistent and may lead to confusion. As such, we have revised Section 3 and 4 to distinguish terms throughout the manuscript. Specifically, we now consistently refer to the input inventory as the *prior emissions*, the GEMS emissions derived from the CSF method as the *satellite-derived emissions*, the final daytime profile after applying the optimization as the *posterior emissions,* and the optimized input emissions as the *updated emissions*. We hope these revisions improve clarify and ensure a clear distinction between the different emissions referred to in this study.

Revised Manuscript (Lines 276 – 279): In the following subsections and in Section 4, we refer to the WRF-Chem input inventory based on EDGAR as the *prior emissions*, the CSF-based estimates from GEMS as the *satellite-derived emissions*, the CSF-based estimates from the model-simulated columns as the model-

derived emissions, and the final emissions after applying the optimization framework, as the *posterior emissions*.

**Comment #5:** *The citations for figures in the Supplement appear disorganized. Please verify the numbering and ensure they are cited sequentially.*

**Response #5:** We have carefully reviewed all Supplementary figure citation and ensured they are numbered and referenced sequentially throughout the manuscript. The figures in the Supplement have been reordered to flow with order of reference in the manuscript to improve clarity and consistency.

**Specific comments:**

**Comment #6:** *Line 49-51: The statement that GEMS NO2 product bias "cannot be diagnosed using satellite data alone" seems somewhat redundant, as independent observations are inherently required to assess retrieval quality.*

**Response #6:** We have corrected the manuscript text as follows:

Revised Manuscript (Lines 53 – 54): Remaining negative biases are consistent with a systematic low bias in the GEMS v3 $NO_2$, highlighting the importance of multi-platform evaluation using independent observations.

**Comment #7:** *Line 125: The reanalysis dataset should be formally cited as NCEP FNL.*

**Response #7:** The manuscript has been updated to reflect this citation.

Revised Manuscript (Lines 128 – 130): The model was driven by reanalysis meteorology, NCEP FNL (Final), and Copernicus Modeling Service (CAMS) chemical boundary conditions (Commerce, 2000; Inness et al., 2019).

**Comment #8:** *Figure 1: Since NOx is primarily emitted as NO, it would be more appropriate to plot either NO or total NOx emissions rather than just NO2.*

**Response #8:** We have replaced Figure 1 with a plot of total $NO_x$ emissions obtained from the EDGAR inventory.

Revised Manuscript (Lines 155 – 159):

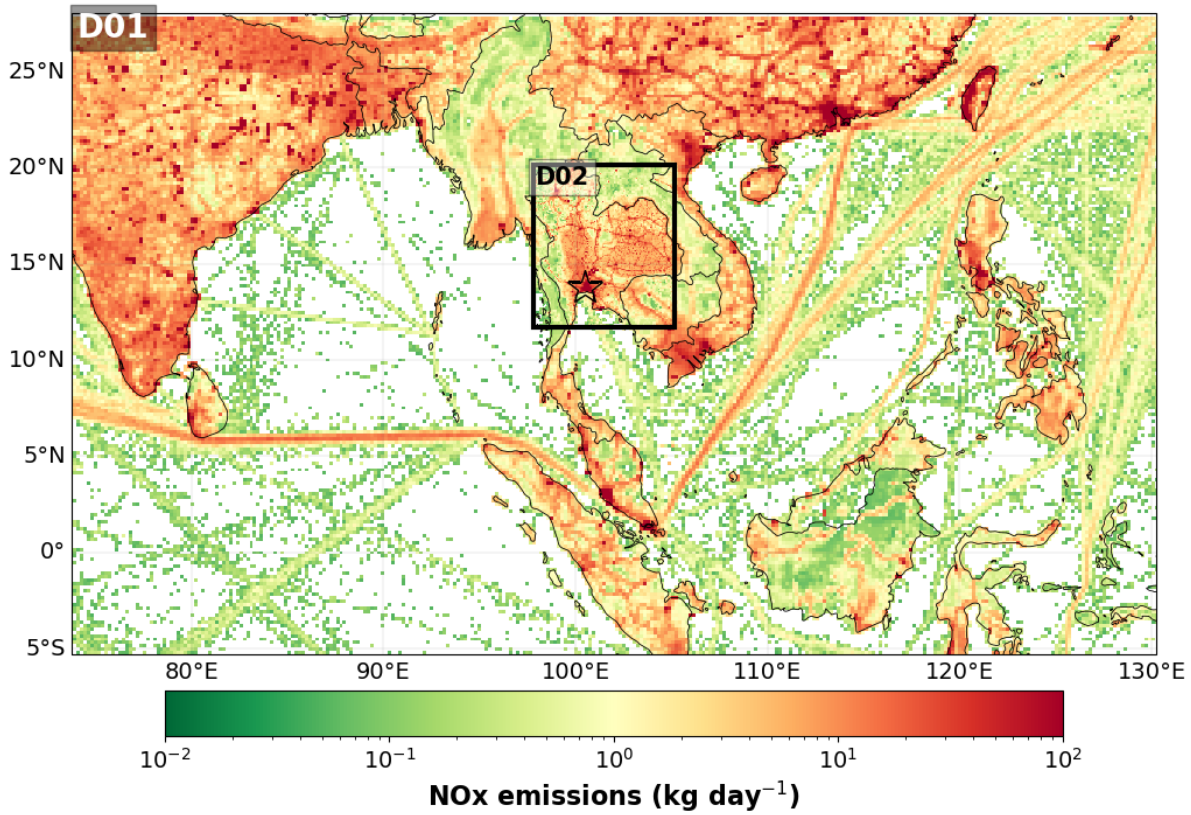



Figure 1. Spatial illustration of the WRF-Chem model domain configuration, including D01 (20 km) and D02 (4 km), and average base-model $NO_x$ input emissions from EDGAR v5. D02 is centered over Thailand, and the urban signal associated with the Bangkok Metropolitan Region (BMR) is highlighted by the star.

**Comment #9:** *Table 1 and Section 6.1.2: Since model NOx is sensitive to the accuracy of local wind field simulations, were any measures like grid or observational nudging applied to the model meteorology? Furthermore, why was 1-degree NCEP FNL data utilized instead of higher- resolution 0.25-degree data?*

**Response #9:** Yes, meteorology was constrained using four-dimensional data assimilation (FDDA) via grid nudging. Nudging was applied to d01 and d02. We also clarify that the simulations here used the 0.25° NCEP FNL data; the former reference was a typo. We have updated Table 1 and added clarifying information to the manuscript.

Revised Manuscript (Lines 130 – 131): Meteorology was constrained using four-dimension data assimilation (FDDA) via grid nudging applied to model domains using the 0.25° NCEP FNL reanalysis fields to improve large-scale wind representation.

Revised Manuscript (Lines 153 – 155): **Table 1** WRF-Chem base ($WRF_{Base}$) model configuration and input datasets.

| **Component** | **Configuration/Source** |
|---|---|

| | |
|---|---|
| **Model Version** | WRF-Chem v4.2.2 |
| **Domain** | D01: 20 km (regional); D02: 4 km (nested) |
| **Meteorological IC/BC** | NCEP FNL reanalysis (0.25v x 0.25°; 6-hourly) |
| **Chemical IC/BC** | CAMS global model output |
| **Chemistry Mechanism** | RACM-MADE-VBS (with aqueous reactions) |
| **Anthropogenic Emissions** | EDGAR v5; downscaled with ODIAC $CO_2$ |
| **Biogenic Emissions** | MEGAN v2.1 (Guenther et al., 2006) |
| **Biomass Burning Emissions** | QFED v2.6 (Grell et al., 2011; Koster et al., 2015) |
| **Dust Emissions** | GOCART scheme (Zhao et al., 2010) |
| **Sea Salt Emissions** | Gong et al. (1997) parameterization (Gong et al., 1997) |
| **Radiation Scheme** | RRTMG (longwave); Goddard (shortwave) |
| **Cumulus Parameterization** | Grell-Freitas (D01 only) |
| **Microphysics** | Morrison (double-moment) (Ye et al., 2021) |
| **PBL Scheme** | Mellor-Yamada-Janjic TKE scheme |
| **Simulation Period** | 14 – 27 March 2024 |

**Comment #10:** *Line 227: Correct the citation for Kuhlmann et al. to (2024)*

**Response #10:** The citation has been corrected in the manuscript text.

Revised Manuscript (Lines 254 – 255): A brief description of inversion methods and specifics is provided here. A complete description of the inversion methods and algorithm can be found in Kuhlmann et al. (2024).

**Comment #11:** *Line 298: Please provide more justification for the fixed conversion factor fq = 1.32, as the NO2-to-NOx ratio is highly dependent on environmental conditions and emissions sources. Could WRF-Chem provide a localized factor for Bangkok?*

**Response #11:** Thank you for the comment. We agree that the NO2-to-NOx ratio can vary with chemical regime, emissions source, and meteorological conditions. However, in our study case, we did not use a WRF-Chem-derived localized conversion factor because the prior simulation at the time showed substantial biases in $NO_x$ and related chemistry before inversion, including unrealistically long $NO_x$ lifetimes. In that case a model-derived local NO2-to-NOx ratio would likely reflect model error rather than provide an independent or more accurate constraint. We therefore retained the fixed conversion factor, $fq$ = 1.32, as suggested in the *ddeq* framework for consistency and to avoid introducing additional uncertainty from biases in the prior chemistry.

Revised Manuscript (Lines 331 – 334): Following the *ddeq* implementation, we use:

$$Q_{NO_x} = f_q \cdot Q_{NO_2} \quad (8)$$

to obtain $NO_x$ estimates, where $f_q$= 1.32, as implemented in the *ddeq* algorithm; a localized factor was not derived from WRF-Chem due to known $NO_2$ biases in the prior simulation that would propagate into the conversion.

**Comment #12:** *Line 343-347: Do the author know why does the CSF method perform better than the GP or IME approaches in recovering the model's prior emissions?*

**Response #12:** We suspect the performance of the CSF method is likely related to how the framework separates and constrains physical processes. In the CSF method, emissions are derived from the downwind $NO_2$ flux, and multiple cross-sections along the plume are used to independently estimate fluxes at different downwind distances. This provides multiple constraints on plume evolution and might reduce the sensitivity to local errors in ERA5 wind fields, plume definition, or missing satellite data.

In contrast, the IME method relies on a single integrated plume mass, making it more sensitive to plume boundaries, wind errors, and assumed decay corrections, while the GP methods depend on fitting the full 2D plume structure, which introduces sensitivity to assigned plume shape and spatial variability. Overall, the CSF method might provide a more robust mapping between observed $NO_2$ and emissions by leveraging multiple cross-sectional constraints. We have clarified this point in the revised manuscript.

Revised Manuscript (Lines 404 – 407): Overlaying the prior emissions shows that the model-derived CSF method most accurately recovers them, likely due to its use of downwind fluxes and multiple cross-sections across the plume, which provide additional constraints on plume evolution and reduce sensitivity to local errors in wind fields, plume definition, and missing data compared to the IME and GP methods.

**Comment #13:** *Figure 3b: Is the black dashed line (WRF Prior) representing the three-hour backward average of EDGAR v5 emissions? It appears visually inconsistent with the prior curve shown in Figure 4c.*

**Response #13:** Thank you for the comment. Yes, the black dashed line in Figure 3b represents the three-hour backward average of EDGAR v5 emissions used as an input to the optimization. This is consistent with the prior emissions shown in Figure 4a.

The apparent discrepancy with Figure 4c arises from differences in the temporal window: Figures 3b and 4a include the 23 UTC hour from the previous day in the backward averaging, whereas Figure 4c does not. Figure 4c represents the full diurnal profile of the input emissions. Values from 00 – 07 UTC are identical across figures; differences at earlier times are solely due to inclusion of the 23 UTC hour in the backward averaging. We have edited the figure caption to clarify this distinction.

Revised Manuscript (Lines 550 – 555): **Figure 4.** Results of the emissions optimization, shown as averages over the BMR. (a) Posterior optimized (solid) and prior (dashed) daytime $NO_x$ emissions in model space, where the prior represents the three-hour backward average of EDGAR v5 emissions used as input to the optimization. (b) Same emissions as in (a) but transformed into observational space using the observational operator. The GEMS profile is shown in blue. (c) and (d) are representative of optimization results using the GEMS bias-corrected CSF results as the constraint. (e) Full diurnal cycle of updated (solid),

updated+BC (solid pink), and prior (dashed) $NO_x$ emissions. Note that, unlike (a), the prior in (e) does not include the 23 UTC hour from the previous day in the backward averaging.

**Comment #14:** *Line 370-378: I have concerns regarding the GEMS NO2 column scaling factor in Equation 11. Early morning model errors in boundary layer height and NO2 profiles can introduce artifacts into this scaling factor. Would it be more robust to correct GEMS NO2 columns using independent observations (e.g., Pandora, MAX-DOAS) prior to the inversion?*

**Response #14:** Thank you for this comment. We agree that early morning model errors in boundary layer height and $NO_2$ vertical structure can influence the scaling factor in Eq. (11). Our intent in applying this correction is not to adjust absolute GEMS column biases, but to account for the reduced surface sensitivity inherent to the GEMS averaging kernel in a way that is internally consistent with the model vertical distribution used in the inversion framework.

At the recommendation of both reviewers, we conducted a sensitivity analysis in which GEMS $NO_2$ columns within the BMR are adjusted using an independent, Pandora-derived bias prior to inversion. Based on the Pandora Bangkok evaluation (new Fig. S1), we apply an approximate scaling factor of two to the GEMS columns within the BMR. To assess the impact of this retrieval bias, we perform the inversion and subsequent model optimization framework both with and without the bias corrected emission results. The results from both configurations are now included in the main manuscript figures to explicitly illustrate the sensitivity of inferred $NO_x$ emissions to the assumed satellite bias. In addition to $WRF_{Updated}$, we introduce a bias-corrected simulation, $WRF_{Updated+BC}$. In most cases, $WRF_{Updated+BC}$ shows improved agreement relative to $WRF_{Updated}$; however, it can lead to overestimation at times as indicated by comparisons with DC-8 and GCAS observations. We therefore retain both simulations in the manuscript to highlight the influence of the bias correction and the associated trade-offs in model performance. The manuscript has been revised throughout the reflect these additions in addition to updated statistics in Table 2.

Revised Manuscript:

(Lines 48 – 49): GEMS-constrained $NO_x$ emissions for March 2024 are estimated to range from 2.7 to 4.3 kT $month^{-1}$ after accounting for known low biases in the GEMS retrievals.

(Lines 149 – 151): Model cases with updated anthropogenic $NO_x$ emissions will be referred to as $WRF_{Updated}$ and $WRF_{Updated+BC}$ (later introduced).

(Lines 213 – 232): **3.1 Correcting for GEMS bias prior to inversion**

Recent independent validation studies of the operational GEMS v3 product over Bangkok and South Korea report low biases in $NO_2$ columns relative to ground-based sun-photometer and DOAS measurements (Bae et al., 2025; Jung et al., 2025). Bae et al. (2025) shows that GEMS v3 increasingly underestimates $NO_2$ relative to Pandora under high-$NO_2$ conditions (>$1 \times 10^{16}$ molecules $cm^{-2}$) as is the case for Bangkok pollution levels. In Jung et al. (2025), validation results over Bangkok indicate a pronounced low bias in GEMS tropospheric $NO_2$ columns relative to Pandora, with regression slopes of ~0.35 for v2.0 and ~0.28 for v3.0, indicating increasing underestimation at higher $NO_2$ levels. While moderate correlations ($r \approx 0.6$–

0.7) suggest that GEMS captures temporal variability, column magnitudes are substantially underestimated, particularly under polluted conditions. The persistence of this behavior in the v3.0 product indicates that the low bias is not fully corrected by recent algorithm updates and is consistent with retrieval sensitivity limitations in highly polluted urban environments (Jung et al., 2025).

To address the low bias in GEMS prior to the $NO_x$ satellite emission inversion, we first quantify the GEMS bias relative to Pandora measurements over our period of interest (14 – 27 March 2024). We compare total column GEMS $NO_2$ to Pandora Level 2 direct-sun total column retrieval, filtering for high quality measurements (quality flag = 10) and averaging to hourly means. GEMS columns are sampled at the nearest grid cell to each Pandora site and temporally collocated. Results are shown in Fig. S1. GEMS $NO_2$ columns are generally about a factor of two lower than Pandora measurements (mean bias ≈ $-9.2 \times 10^{15}$ molecules $cm^{-2}$). This factor-of-two difference persists throughout March 2024 (Fig. S1).

To assess the sensitivity of the top-down $NO_x$ inversion framework to this bias, we apply a simple correction factor to GEMS prior to inversion. Specifically, GEMS $NO_2$ columns over the BMR are scaled by a factor of 1.67 derived from the Pandora comparison. The inversion is then performed with and without the bias-corrected columns.

(Lines 428 – 429): Two versions of the optimization are performed: one using the constraint derived from the bias-corrected retrievals, and one without it.

(Lines 657 – 672):

**Table 2.** Summary of WRF-Chem D02 validation statistics for $WRF_{Base}$ (B) $WRF_{Updated}$ (U), and $WRF_{Updated+BC}$ (U+BC) simulations evaluated against independent observational datasets. Metrics include mean bias (MB), mean error (ME), normalized mean bias (NMB), normalized mean error (NME), root-mean-square error (RMSE), and Pearson correlation coefficient (CORR).

| Dataset | Species | Model Run | MB | ME | NMB | NME | RMSE | CORR |
|---|---|---|---|---|---|---|---|---|
| GCAS | Tropospheric $NO_2$ column (molecules $cm^{-2}$) | **B** | 4.0E+15 | 5.1E+15 | 93 | 1.2E+2 | 9.3E+15 | 0.64 |
| | | **U** | -2.4E+14 | 1.9E+15 | -5.6 | 45 | 3.4E+15 | 0.54 |
| | | **U+BC** | 6.8E+14 | 2.3E+15 | 16 | 53 | 4.1E+15 | 0.4 |
| DC-8 | $NO_xO_3$ $NO_2$ (ppbV) | **B** | 1.3 | 1.4 | 82 | 87 | 2.8 | 0.97 |
| | | **U** | -0.81 | 0.8 | -49 | 49 | 1.4 | 0.93 |
| | | **U+BC** | -0.38 | 0.49 | -23 | 30.3 | 0.89 | 0.94 |
| | CANOE $NO_2$ | **B** | 1.4 | 1.5 | 87 | 94 | 2.9 | 0.97 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | (ppbV) | **U** | -0.76 | 0.76 | -48 | 48 | 1.3 | 0.93 |
| | | **U+BC** | -0.34 | 0.50 | -21 | 31 | 0.88 | 0.94 |
| Ground Monitors | $NO_2$ (ppbV) | **B** | 12 | 12 | 1.3E+2 | 1.3E+2 | 15 | 0.61 |
| | | **U** | -3.0 | 4.5 | -32 | 47 | 5.6 | 0.53 |
| | | **U+BC** | 0.1 | 4.0 | 0.6 | 43 | 5.5 | 0.55 |
| | $NO_x$ (ppbV) | **B** | 14 | 15 | 1.4E+2 | 1.4E+2 | 18 | 0.59 |
| | | **U** | -3.6 | 4.9 | -34 | 47 | 6.7 | 0.61 |
| | | **U+BC** | 0.02 | 4.8 | 0.2 | 46 | 7.1 | 0.57 |
| | $O_3$ (ppbV) | **B** | -5.6 | 8.2 | -18 | 27 | 10 | 0.87 |
| | | **U** | 4.9 | 8.5 | 16 | 27 | 11 | 0.81 |
| | | **U+BC** | 3.1 | 6.6 | 10 | 22 | 9.2 | 0.86 |
| Pandora | Tropospheric $NO_2$ column (molecules $cm^{-2}$) | **B** | 7.9E+15 | 8.6E+15 | 55 | 59 | 9.6E+15 | 0.71 |
| | | **U** | -6.6E+15 | 6.8E+15 | -46 | 47 | 8.2E+15 | 0.67 |
| | | **U+BC** | -3.8E+15 | 4.9 E+15 | -26 | 33 | 6.4 E+15 | 0.70 |

(Lines 991 – 1042): **7.1 Implications of GEMS $NO_2$ retrieval biases**

In our study, the comparisons with independent aircraft and ground-based observations indicate the satellite-constrained emissions without bias-corrected presented here may represent a low estimate. In particular, the systematic low bias in GEMS $NO_2$ is consistent with the negative biases often seen in the $WRF_{Updated}$ simulations. Importantly, this bias could not be diagnosed using the satellite data alone.

To further evaluate the relative behavior of GEMS and airborne GCAS $NO_2$ columns, we compare their tropospheric columns using $WRF_{Updated+BC}$ as a common transfer framework (Fig. S21). Figure S21 compares the observed GCAS/GEMS $NO_2$ column ratios with ratios calculated after both datasets are mapped through $WRF_{Updated+BC}$. In the observations, GCAS columns are consistently higher than the raw GEMS columns as illustrated by the orange line, with GCAS/GEMS ratios of ~2–3 on most days and values reaching ~6–7 on 21 March. In contrast, the corresponding $WRF\text{-}GCAS_{Updated+BC}/WRF\text{-}GEMS_{Updated+BC}$

ratios are closer to 1:1 (~1.1–1.7), even on 21 March. This behavior is consistent with a low bias in GEMS $NO_2$ columns relative to GCAS. After applying the bias correction, the GCAS/$GEMS_{BC}$ ratios approach 1:1 on most days, indicating improved consistency between the datasets, and supporting the effectiveness of the applied bias correction.

When incorporating bias-corrected retrievals into the top-down inversions and $NO_x$ emission optimization process, the resulting $WRF_{Updated\ +BC}$ run reduces the negative biases in the independent validation, in some cases bringing the model results even closer to observations as seen in the surface air quality analysis. However, this improvement is not uniform. In certain comparisons (e.g., GCAS and DC-8), $WRF_{Updated\ +BC}$ introduces a tendency toward overprediction in morning hours, highlighting tradeoffs associated with the bias correction. These results suggest that while accounting for the retrieval bias can improve mean model performance, additional uncertainties such as overestimated wind speeds and associated transport errors continue to influence the representation of modeled $NO_2$.

Here, the integration of ground-based, airborne, satellite, and model data provides a powerful framework not only for improving emissions but also for identifying limitations within individual observing systems. While $WRF_{Updated}$ clearly outperforms the baseline model, and $WRF_{Updated\ +BC}$ offers targeted improvements in reducing systematic bias, the combined observational evidence highlights the necessity of a multi-platform validation to fully interpret the satellite-based emission estimates.

Despite this low bias in the GEMS retrieval, the high-frequency daytime sampling provided by geostationary observations offers critical constraints on daytime variability and plume evolution that are particularly valuable for emission inversion and air quality modelling. For example, the GEMS-constrained emission adjustments presented here were critical for improving the temporal evolution of $NO_x$ in WRF-Chem, resulting in substantial and robust improvements in model performance across independent evaluations. Future work may further benefit from continued refinement of bias-corrected GEMS products and upcoming algorithm improvements (e.g., v4), alongside improved representation of spatial representation of emissions within the model.

**Comment #15:** *Section 4.1: Since the authors did not directly assimilate NO2 columns, the term "satellite observations" in this context is imprecise (e.g., Line 422-424, Line 434-435). I suggest using "GEMS-informed emissions" or a similar term.*

**Response #15:** Yes, following the revised wording in this section, we have changed the term to reflect "satellite-derived emissions."

Revised Manuscript (Lines 490 – 492): The result is a corrected model hourly emission vector, $\hat{x}$, that balances the satellite-derived emissions with the model's prior estimate while accounting for chemistry and transport behavior (Fig. 4). Figure 4b shows the posterior profile closely reproduces the GEMS-derived daytime pattern, indicating the optimization effectively aligns model emissions with satellite-derived emissions while preserving components of the prior model behavior.

**Comment #16:** *Figure 4: The caption for Figure 4a and 4b requires more detail to clearly distinguish between the optimized profiles and those viewed through the observational operator.*

**Response #16:** We have edited the caption for Figure 4 accordingly.

Revised Manuscript (Lines 550 – 555): **Figure 4.** Results of the emissions optimization, shown as averages over the BMR. (a) Posterior optimized (solid) and prior (dashed) daytime $NO_x$ emissions in model space, where the prior represents the three-hour backward average of EDGAR v5 emissions used as input to the optimization. (b) Same emissions as in (a) but transformed into observational (GEMS) space using the observational operator. The GEMS profile is shown in blue. (c) Full diurnal cycle of updated (solid purple), updated+BC (solid pink), and prior (dashed) $NO_x$ emissions. Note that, unlike (a), the prior in (c) does not include the 23 UTC hour from the previous day in the backward averaging.

**Comment #17:** *Line 527-528: Beyond meteorology, how do inaccuracies in the spatial pattern of emissions affect the inversion framework? Do the authors have any thoughts on addressing this issue under this study's emissions inversion framework?*

**Response #17:** Thank you for the comment. We agree that this approach does not explicitly optimize the spatial distribution of emissions, and therefore any spatial inaccuracies in the prior are not inherently corrected by the inversion. Such errors can propagate into the forward model through biases in plume structure and transport, potentially affecting the agreement with observations as shown in this section.

As noted above, extending this framework to resolve emissions at finer scales would require additional steps. This could be achieves by performing separate inversions for different source types. A first step could be to incorporate high-resolution local priors when available, together with higher-resolution observations (e.g., GCAS) that can better resolve variability in $NO_2$.

Revised Manuscript (Lines 621 – 623): Additionally, while the inversion framework adjusts emission magnitudes temporally, it does not correct spatial errors in plume transport or source distribution (see Sect. 7.2 for further discussion).

**Comment #18:** *Line 643-644: Note that the highest observed O3 peak actually occurred on March 22, rather than March 20–21.*

**Response #18:** We have updated the phrasing here to address this comment.

Revised Manuscript (Lines 747 – 750): The 20 – 21 March is shaded in grey to indicate the period where the model mis-represented synoptic conditions resulting in enhancements in the observed concentrations due to stagnation. During this time, $NO_2$ and $NO_x$ reached as high as 49, and 59 ppbV, respectively. $O_3$ peaked at 103 ppbV the following day, on 22 March.

**Comment #19:** *Line 655-661: The shift in O3 peak time to the early morning in the WRFupdated rum (Figure S12a) does not align with observations. Does this discrepancy suggest the emission inversion timing might be incorrect?*

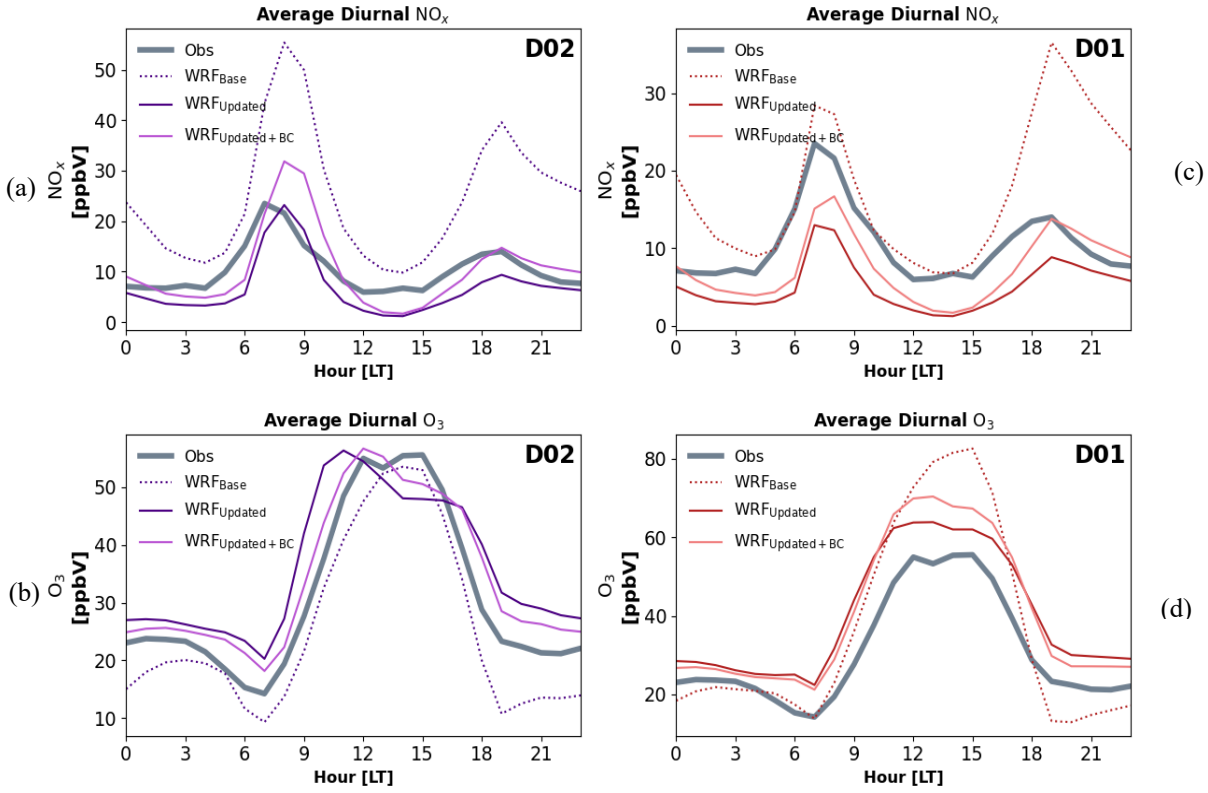
**Response #19:** Thank you for the comment. We agree that the earlier $O_3$ peak in $WRF_{Updated}$ D02 indicates the model does not fully reproduce the observed diurnal evolution of ozone. However, we do not interpret this as evidence that the inferred $NO_x$ emission timing is incorrect.

$O_3$ production in Bangkok occurs under a mixed $NO_x$-VOC sensitivity regime, as supported by recent ASIA-AQ analyses (Cho et al., 2026). In such regimes, the timing and magnitude of $O_3$ are not only controlled by $NO_x$ emissions alone, but also depend strongly on VOC availability, boundary layer evolution, and transport processes. Importantly, the same emission inversion is applied in both D01 and D02, yet the early $O_3$ peak is only present in D02 (Figure S12a-b). This suggests that the discrepancy is more likely driven by domain-dependent differences in meteorology and chemical processing, rather than an error in emission timing. We therefore interpret the early $O_3$ peak in D02 as reflecting remaining model chemistry (as VOC emission was not altered in this study) rather than a direct bias in the inferred $NO_x$ emission timing.

We have updated the supplemental figure (now Figure S12) to illustrate the $NO_x$ diurnal cycles as well to support this comparison.

Revised Manuscript (Lines 762 – 768): Average diurnal cycles of $NO_x$ for D02, D01 are further illustrated in Fig. S12a, c. Observed $NO_x$ shows a clear morning peak and lower midday concentrations driven by boundary layer evolution. $WRF_{Base}$ overestimates $NO_x$ and exaggerates the morning peak while $WRF_{Updated}$ (dark purple) substantially improves both magnitude and timing but underestimates the evening peak values. $WRF_{Updated+BC}$ shows mixed performance with an overprediction of the morning peak while the evening peak aligns more closely with observations compared to $WRF_{Updated}$, indicating a trade-off in how the bias correction might redistribute $NO_x$ across the diurnal cycle in the updated models.

Revised Manuscript (Lines 776 – 780): While $WRF_{Updated}$ exhibits an earlier $O_3$ peak in D02 compared to observations, this shift is not present in D01 despite the same emission inversion being applied. This likely suggests the discrepancy is driven by domain-dependent differences in meteorological or chemical processes, rather than an error in the inferred $NO_x$ emission timing.




Revised Supplement (Lines245 - 250): **Figure S12.** Average diurnal cycles of $NO_x$ and $O_3$ for D02 (a,b) and D01 (c,d). Comparison of $WRF_{Base}$ (dotted), $WRF_{Updated}$, and $WRF_{Updated+BC}$ simulations against Thailand Pollution Control Department (PCD) ground monitor network observations during the ASIA-AQ deployment period (14-27 March 2024).

**Comment #20:** *Line 677-679: Please clarify which Pandora data (e.g., direct sun or sky scan), data screening, and temporal collocation methods were used for the evaluation*

**Response #20:** Thank you for the comment. The Pandora evaluation was conducted using the Bangkok Level 2 Pandora $NO_2$ direct-sun total column product. Only high-quality measurements (quality flag = 10) were retained. The observations were aggregated to hourly means prior to comparison. To approximate the tropospheric columns, the Pandora total column was corrected by subtracting the closest GEMS stratospheric $NO_2$ column at the nearest satellite pixel. WRF-Chem $NO_2$ columns were computed at the nearest model grid cell to the Pandora site and were temporally collocated with the hourly Pandora data. We have revised the manuscript to provide this clarifying information.

Revised Manuscript (Lines 796 – 800): Pandora $NO_2$ evaluation used Level 2 direct-sun total column retrievals, filtered for high-quality measurements (quality flag = 10), and averaged to hourly means. Tropospheric columns from Pandora were estimated by subtracting coincident or closest GEMS stratospheric $NO_2$ at the nearest satellite pixel. Model columns were sampled at the nearest grid cell to the Pandora site and temporally matched to observations.

**Comment #21:** *Equation 17: Please define SW (Scattering Weights).*

**Response #21:** We have revised the manuscript to include this definition.

Revised Manuscript (Lines 879 – 880): Here, $SW_i$ represents the scattering weight for layer $i$, representing the sensitivity of the measured radiance to $NO_2$ in that layer.

**Comment #22:** *Figures 8–9: I suggest converting the time from UTC to Local Standard Time (LST) for better interpretability.*

**Response #22:** In Figure 8 and corresponding Figure S16 we have changed the times to reflect local standard time in the figure. We have edited the GEMS image to reflect local time in Figure 9c.

**Comment #23:** *S19 is provided in supplement materials but not used in the manuscript.*

**Response #23:** As suggested in the earlier comment, we have revised the manuscript and Supplement figure references to ensure each figure is referenced accordingly.

**Comment #24:** *Line 804: This reference should likely be Figure S17.*

**Response #24:** We have adjusted the figure reference accordingly to match the updated Supplement.

**Comment #25:** *Lines 925–929: Any model output (e.g., WRFbase) can be used as a transfer standard for accessing GEMS vs. GCAS, though high-resolution models are preferred.*

**Response #25:**
Thanks for the comment. We agree that the model output can serve as a transfer standard to facilitate comparisons between different observing systems, including GEMS and GCAS and have revised the text to clarify this point.

Revised Manuscript (Lines 1003 – 1005): While model output (e.g., $WRF_{Base}$) can serve as a transfer standard for comparing different observing systems, these results highlight the importance of integrating ground-based, airborne, satellite, and model information to robustly identify biases and improve emission estimates.