Response to the Referee #2 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 #2

General Comments:

This manuscript investigates daytime hourly Nitrogen Oxides (NOx) emissions over the Bangkok Metropolitan Region (BMR) using hourly Geostationary Environment Monitoring Spectrometer (GEMS) Nitrogen Dioxide (NO2) observations. The authors apply the Cross-Sectional Flux (CSF) method via the ddeq Python library to derive satellite-based emission estimates, which are then used to update the diurnal NOx profile in nested WRF-Chem simulations at 20 km and 4 km resolutions. The updated model is evaluated against independent multi-platform measurements from the 2024 ASIA-AQ campaign, including the Thailand Pollution Control Department (PCD) surface monitors, Pandora, the GEOstationary Coastal and Air Pollution Events (GEO-CAPE) Airborne Simulator (GCAS), DC-8, and the High Spectral Resolution Lidar (HSRL)-2 lidar observations.

This work is well-motivated given the growing availability of hourly Geostationary Earth Orbit (GEO)-based composition retrievals and the persistent uncertainty in urban NOx inventories across Southeast Asia. The effort to constrain the daytime structure of emissions, rather than applying a single scaling factor, is a meaningful methodological advance, and the multi-platform evaluation substantially strengthens confidence in the results.

However, several methodological assumptions, particularly the inconsistency in meteorological fields used across different stages of the analysis, require further clarification. Additionally, while the authors correctly acknowledge a systematic low bias in the GEMS v3 NO2 product, the manuscript would benefit from a clearer discussion of how this bias propagates into the derived emission estimates.

Response: Thank you for the helpful comments. We address these assumptions in detail in the responses below. In the revised manuscript, we now explicitly account for the systematic low bias in the GEMS product and introduce an additional simulation that incorporates this correction ($WRF_{Updated+BC}$). Details of these updates are provided in the responses that follow.

Specific Comments:

Comment #1: *The use of different meteorological fields at different stages of the analysis needs clarification. ERA5 winds are used for the CSF inversion because WRF-Chem winds are biased high, yet the NOx lifetimes fed into the optimization operator H appear to come from WRF-Chem. The remaining negative biases in WRF-Updated are subsequently attributed to WRF-Chem wind overprediction. The authors should add a short paragraph that clearly separates (a) how the wind bias affects the initial inversion step and (b) how it affects the final model evaluation. It should also be clarified whether the lifetimes used in H are physically consistent with ERA5- or WRF-Chem-based transport, and whether using ERA5-derived lifetimes would materially change the optimized emission profile.*

Response #1: We thank the reviewer for this comment and have clarified this point in the manuscript. The NO_x lifetimes used in the forward operator (H) are not derived-from WRF-Chem. Instead, they are obtained from the GEMS-based CSF analysis, where lifetime is fitted from the observed along-plume decay using ERA5 wind fields. As such, the lifetimes are internally consistent with ERA5-based transport. This makes the satellite emission estimates completely independent of the WRF-Chem simulations.

WRF-Chem wind biases affect primarily the model evaluation stage. When the updated emissions are simulated within WRF-Chem, the model's tendency to overestimate wind speed contributes to the negative remaining biases in $WRF_{Updated}$. Because the lifetime is derived from the GEMS + ERA5 framework, we are already including it within our optimization approach.

Revised Manuscript (Lines 429 – 434): The three-hour window provided the best agreement with the model-derived CSF emissions as shown in Fig 3b, 3c. Within this window, the contribution from each prior hour depends on the atmospheric lifetime of NO_2 for that hour. We use hourly lifetimes, τ, derived from the GEMS-based CSF inversion in Eq. (7), where τ is fitted from observed along-plume decay using ERA5 winds, and is therefore independent of WRF-Chem meteorology. This ensures the satellite-derived emission estimates are completely independent of the WRF-Chem simulations. For reference, WRF-Chem-derived decay times are generally longer than those inferred from GEMS, reflecting differences in modeled vs. observed loss and dilution processes.

Comment #2: *As the authors acknowledge, the remaining negative biases in WRF-Updated are consistent with the known low bias in GEMS v3 NO2. However, this bias is compounded under high-NO2 conditions, precisely the conditions characteristic of Bangkok, meaning that emissions derived directly from the uncorrected satellite data are likely biased low. Consequently, the conclusion that EDGAR v5 overestimates NOx by ~75% should be stated with appropriate caveats, as the true magnitude of overestimation may be smaller once the satellite bias is accounted for.*

Response #2: Thank you for this helpful suggestion. 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 $\approx -9.2 \times 10^{15}$ 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 $\mathrm{WRF_{Base}}$ (B) $\mathrm{WRF_{Updated}}$ (U), and $\mathrm{WRF_{Updated+BC}}$ (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 (ppbV)	**B**	1.4	1.5	87	94	2.9	0.97
		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$_{\mathrm{Updated}}$ 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$_{\mathrm{Updated+BC}}$ 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$_{\mathrm{Updated+BC}}$. 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-GCAS$_{\mathrm{Updated+BC}}$/WRF-GEMS$_{\mathrm{Updated+BC}}$ 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$_{\mathrm{BC}}$ 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$_{\mathrm{Updated+BC}}$ 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$_{\mathrm{Updated+BC}}$ 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$_{\mathrm{Updated}}$ clearly outperforms the baseline model, and WRF$_{\mathrm{Updated+BC}}$ 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 #3: *The 24-hour emission scaling approach warrants further justification. Since GEMS constrains emissions only between 08:00 and 14:00 LT, applying a single average scaling factor to unconstrained evening and nighttime hours is a strong assumption, particularly given Bangkok's well-documented bimodal traffic patterns. It is unclear whether this extrapolation is physically reasonable, especially during the evening rush hour. The authors should provide a brief justification or, where possible, compare the implied evening scaling against an independent diurnal reference such as the THAI-KMUTT local inventory profile.*

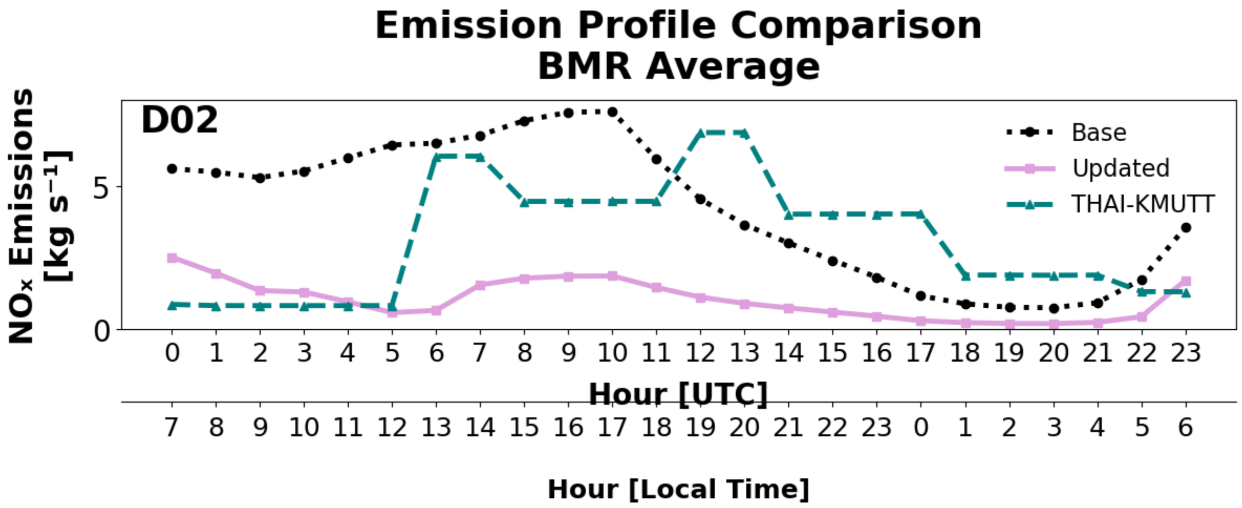
Response #3: Thank you for the helpful comment. We agree that applying a single mean scaling factor derived from the GEMS-constrained window (08:00-14:00 LT) to the full 24-hour profile is an assumption and warrants justification. In our framework, the satellite observations primarily constrain the emissions during the daytime window. Applying the scaling uniformly during the nighttime is necessary to maintain physical consistency. If the scaling was applied only during the GEMS-constrained hours, it would introduce an artificial discontinuity between daytime and nighttime emissions as the GEMS constrain substantially reduces daytime NO_x (leaving nighttime emissions unscaled would result in an unrealistic step change in the diurnal profile).

To evaluate the physical plausibility of the resulting 24-hour profile, we compared it with the THAI-KMUTT local inventory (see below). The prior inventory (EDGAR) already includes a bimodal structure, although the second (evening peak) is partially obscured by an overall high emission bias. The updated profile preserves this temporal structure during the evening hours while reducing the overall magnitude. The amplitude/timing of the peak is weaker than in THAI-KMUTT, suggesting greater uncertainty outside the GEMS-constrained window. It is also important to note that NO_x emissions in the model represent a combination of sources (e.g., traffic, area sources, and other anthropogenic activities), not traffic alone. As such, perfect agreement with a traffic-based or locally derived diurnal profile is not expected.

We also emphasize that the prior inventory provides a regionally representative temporal profile (developed for broader Asian emissions rather than specifically for Bangkok) which was available at the time and thus may not fully capture the local diurnal variability. Future work should incorporate the best available temporal profiles where available.

Lastly, this approach is consistent with prior satellite-constrained emission studies. For example, a similar methodology was applied using a single daytime satellite constraint to scale emissions without resolving the full diurnal profile in a KORUS-AQ field study over Seoul (Goldberg et al. 2019). Our use of GEMS extends this framework by leveraging multiple daytime observations but relies on the same assumption for extrapolation beyond the observational window.

We have added some sentences to the main manuscript to further clarify this.



Revised Manuscript (Lines 482 – 489): Because GEMS provides constraints only during the daytime overpass window, the derived scaling below primarily reflects daytime emission adjustments. We therefore apply the mean scaling factor uniformly during the nighttime hours to preserve the prior temporal structure and avoid introducing artificial discontinuities between the constrained and unconstrained hours, particularly given the large daytime emission reductions. This approach is consistent with previous satellite-constrained emission studies (e.g., Goldberg et al. (2019)) which constrain emission magnitude based on a single daytime value derived from OMI. Where available, locally derived temporal profiles should be used to better represent regional emission patterns; in this study, our prior profile reflects the best available regional representation for Asia at the time of study.

Technical Corrections:

Page 1, Line 24: *“College Park, postal code, USA” → “College Park, USA”*
Page 4, Line 103*: “e.g..,” → “e.g.,”*
Page 4, Line 106*: “(Kim et al., 2020; Park et al., 2025).Recent” → “(Kim et al., 2020; Park et al., 2025). Recent”*
Page 4, Line 122*: “1; Fig 1)(Agarwal et..” → “1; Fig 1) (Agarwal et …”*
Page 5, Line 134: *Does “1 km2 resolution” mean the 1 km × 1 km resolution?*
Page 8, Line 220: *I don’t know if it is better to write from “Sentinel5P-TROPOMI” to “Sentinel-5P TROPOMI”.*
Page 8, Line 225: *“LEO observations For our” → “LEO observations. For our”.*
Page 8, Line 227: *“Kuhlmann et al. (2014)” → “Kuhlmann et al. (2024)”.*
Page 14, Line 376: *“where larger correction (e.g., 1.33) where present” → “where larger*

correction (e.g., 1.33) were present”.
Page 25, Line 610: *“from that fact that” → “from the fact that”.*
Page 32, Line 803: *“(St. Clair et al., 2019) .We” → “(St. Clair et al., 2019). We”.*
There is no definition of the HSRL-2.
This is not significant, but authors can use the consistent word:
1. modelled or modeled
2. re-grided or re-gridded

Response: Thank you for the clarifications. In the revised manuscript we have addressed the above technical corrections through tracked changes.