Authors' Response
Preprint: https://doi.org/10.5194/egusphere-2026-3926)

We appreciate the editor and the anonymous referees for their time and constructive feedback. Our response to the referees' comments (RC1 and RC2) is given below. Thank you for your continued support during the revision process.

C: Referee Comment (in blue font); **AAR:** All Authors' Responses (in black regular font)
The modified text in the revised manuscript is in *black italics.*

Detailed Response to Referee #1 (RC1)

C1: "Monitoring Indian CO_2 Emissions: An Integrated Satellite-Based Dual-Species Approach Using TROPOMI and CO2M Retrievals" by Sukumaran et al. is a potentially useful manuscript that explores the possibility of monitoring CO_2 emissions from space based on real TROPOMI NO_2 observations and simulated XCO_2 observations corresponding to the future CO2M satellite constellation.

The manuscript makes a valuable qualitative conclusion about approaches to background definition, but the quantitative conclusions of the paper are questionable related to 2 key factors that the authors do acknowledge as caveats: namely, the neglect of the contribution of transport errors and the coarse resolution of the model in comparison to the satellite data, mentioned again later in this review.

AAR: We thank the referee for the careful review and positive remarks. We address this and discuss both factors more clearly and explicitly (see Sect. 4.5 in the revised manuscript). Accordingly, the quantification limit and related statements have been revised, and our main conclusions now emphasize the potential of using TROPOMI NO_2 observations for plume detection and for a relative comparison of background methods.

C2: Abstract: The reported "lower limit of quantification" of 20 $MtCO_2$/year is quite high and seems like a pessimistic/conservative lower limit for annual emission quantification from CO2M. Bovensmann et al. (2010) looked at the capability of quantifying emissions from a mid-sized power plant considered to be ~13 $MtCO_2$/year from CarbonSat, which eventually evolved into the 3-satellite CO2M constellation (with added aerosol and NO_2 observations further enhancing capabilities). Nassar et al. (2025) have demonstrated quantification of sources on the order of ~5 $MtCO_2$/year from a single overpass in the U.S. without even bringing in supporting NO_2 information, but they only conducted and instantaneous emission estimate rather than constrain annual emissions. Figure 12 of Bovensmann et al. (2010) shows the power plant emission

statistics for countries including India (reasonably current at the time of that publication). With a cut-off of 20 $MtCO_2$/yr, this would only cover about 20% of India's power plant emissions. With a cut-off of 5 $MtCO_2$/yr, about 75% of India's power plant CO_2 emissions would be included. Thus, the lower limit of quantification is quite an important point, thus the authors should refrain from reporting a number if there are major caveats related to it.

AAR: We agree with this suggestion. Given the current implementation of the method and the potential of satellite observations, the lower bound stated in the manuscript is overly conservative and may be misinterpreted. Hence, we remove such statements from the manuscript, as they seem ambiguous.

C3: Line 35: Introduction should mention Nassar et al. (2017) following Bovensmann et al. (2010), since Nassar et al. (2017) was the first publication to quantify power plant CO_2 emissions and in fact did so for Indian power plants, unlike Bovensmann et al. (2010).

AAR: We thank the reviewer for pointing out the missing reference. Added.

C4: Line 43: "Signal-to-noise ratio" may not be the best term to use here. The high contrast with the ambient NO_2 background is what the authors are talking about, which is analogous to SNR, but when using the term SNR in this context, the authors have not specified the specific parameters that they consider signal or noise.

AAR: We agree; the text is revised as follows:

Line no: 44

"Nitrogen dioxide (NO_2) is a useful co-emitting tracer for CO_2 in fossil-fuel combustion due to its shorter lifetime, which can result in more spatially distinct emission plumes against the ambient background concentration (Reuter et al., 2019; Kuhlmann et al., 2026)."

C5: Line 49: It is worth expanding to "cross sectional flux and gaussian plume models" since Nassar et al. (2021) and (2022) along with Lin et al. (2023) all used the Gaussian plume model approach rather than the cross-sectional flux approach.

AAR: We agree; the text is revised as follows:

C6: Line 96: Is there actual "chemistry" such as chemical reactions, or just "tracer transport"? I believe it is actually the latter.

AAR: Atmospheric CO_2 mixing ratio is simulated as a passive tracer without chemical reactions, but the implementation requires chemistry subroutines in WRF-Chem to ensure adequate passive tracer transport. As a result, we follow the formal model reference as 'WRF-Chem with the GHG module' to indicate this clearly.

C7: Table 1: Surprisingly absent from the table is the Sasan Ultra Mega Power Plant (https://www.gem.wiki/Sasan_Ultra_Mega_Power_Project) near Vindhyachal and Rihand. Furthermore, is there any objective system behind the power plant listed order? Perhaps re-order from largest installed capacity to smallest or north to south latitude.

AAR: Table 1 is now ordered by decreasing installed capacity.

Sasan UMPP (3960 MW) and Singrauli STPS (2000 MW) are part of the Singrauli cluster and are not well separated in the EDGAR emission inventory. Because of this, those power plants are not treated as separate target emission sources. A note is also added to the Table 1 caption, as follows:

"[a]VDL and RHD are part of the Singrauli cluster, which also includes the Sasan Ultra Mega Power Plant (3960 MW) and the Singrauli Super Thermal Power Station (2000 MW); these plants are not treated as separate targets because their EDGAR emissions lie within the VDL emission aggregation (Section 2.4.7 and Fig. 1)."

C8: Figure 1: Nassar et al. (2021) Figure 7 shows the region of the Rihand Reservoir (Govind Vallabh Pant Sagar Reservoir) with multiple power plants: Sasan, Vindhyachal, Singrauli and Rihand. The additional power plants would be relevant secondary sources in emission estimation. One would expect a paper focused on emission estimation from Indian coal power plant to include these other known sources. Singrauli apparently has a planned expansion of 2 additional 800 MW units: https://www.gem.wiki/Singrauli_Super_Thermal_Power_Station

AAR: We have updated the Fig. 1 inset to mark Sasan and Singrauli STPS together with VDL and RHD. Those emissions are part of the cluster aggregation used for VDL (see C7). Though the planned Singrauli expansion does not affect this OSSE, we agree that it can be relevant to emission estimates when using actual CO2M observations. This point is mentioned in the revised manuscript as follows:

Line no: 85

"*As the analysis is carried out within an observing system simulation experiment (OSSE) framework (Section 2.4) using the EDGAR inventory, the actual operating conditions of those TPPs do not influence the results. However, these conditions become relevant when the analysis is extended to real atmospheric CO_2 observations such as CO2M.*"

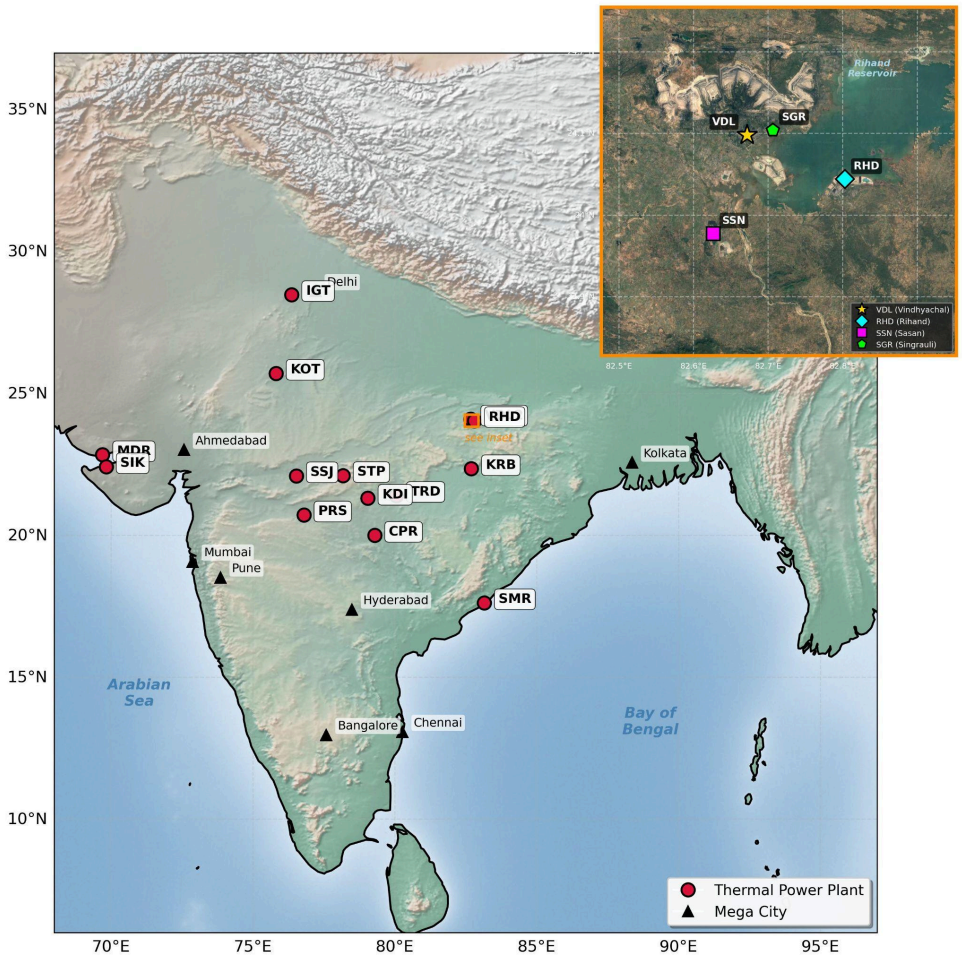


Figure 1. Geographic locations of the thermal power plants used in this study. The VDL TPP location is shown in the inset (Satellite imagery: © Google Earth. All rights reserved.). The background shows topographic relief from Natural Earth (Kelso and Patterson, 2009). Major cities are shown for geographic reference.

C9: Line 111: The fossil fuel scaling factors of Nassar et al. (2013) are commonly referred to as "TIMES (Temporal Improvements for Modelling Emissions by Scaling)". It would be clearest for readers if the scale factors were referred to with this name.

AAR: Thank you, "TIMES" reference added to the text.

Line no: 117

"*Annual EDGAR emissions were temporally disaggregated to hourly resolution by applying the TIMES (Temporal Improvements for Modelling Emissions by Scaling) scaling factors of Nassar et al. (2013) prior to input into WRF-GHG.*"

C10: Figure 2: Panel e) EDGAR seems to show two distinct sources. The source at 24.5°North and 83.0°East is clearly smaller, but the WRF-GHG simulation has no real plume there, although there is a feature that is at about the level of the noise. Would the authors not expect a secondary plume in CO_2 and NO_2 based on the EDGAR emissions?

AAR: Based on the EDGAR emission inventory, a plume can be expected from those two weaker sources. However, as R01 pointed out, the plume signal is sufficiently weak to become indistinguishable from the background variability. As a result, our plume detection algorithm may fail to detect such weaker plumes and, subsequently, the emission source. Note that the NO_2 field is derived from actual TROPOMI retrievals, which are independent of the EDGAR emission inventory. Therefore, the presence of a CO_2 emission source in EDGAR does not necessarily imply that a corresponding NO_2 plume can be present. This is further clarified in the manuscript as follows:

Line no: 223

"*In the case of weak sources, further refinements in the algorithm may be needed in conjunction with the availability of more high-resolution observations. For example, the plume signal of a weak source nearby the target source location (Fig. 2e) is not distinguishable from the background variability (Fig. 2b) and therefore, our plume detection algorithm may fail to detect such plumes.*"

C11: Line 159: There is inadequate discussion of the temporal offset between TROPOMI NO_2 and CO2M CO_2 observations. The authors need to elaborate more on their assumptions and the implications of a temporal offset when using observations from a different satellite compared with fully coincident observations in both space and time.

AAR: Thank you for this comment. We agree. We have now included a paragraph in Section 2.4.2 that explains the temporal offset between TROPOMI NO_2 - CO2M CO_2 observations and its implications in the scope of the present study.

Line no: 181

"TROPOMI and synthetic CO2M overpasses are paired when their acquisition times differ by less than 1.5 h (nominal local overpass times of approximately 13:30 for TROPOMI and 11:30 for CO2M). The synthetic XCO_2 field and the transport wind are taken from the WRF-GHG output nearest to the CO2M overpass time, whereas the NO_2 field corresponds to the TROPOMI overpass time. Within this offset, a plume is advected by roughly 5–30 km for wind speeds of 1–5 m s−1, and its orientation can change if the wind direction varies. The 50 km buffer is chosen to exceed this displacement. Since the NO_2 plume is used only to delimit the search domain, and not to prescribe the shape, width, or flux of the CO_2 plume, the temporal offset affects which pixels are searched but does not enter the flux calculation. A residual risk is that, under rapidly

changing winds, part of the CO_2 plume lies outside the constrained domain or that the domain includes enhancements from other sources. The comparison with retrievals without NO_2 guidance (Sects. 3.1 and 4.1) quantifies the overall effect of the NO_2-derived domain. For CO2M, the simultaneous NO_2 observations of the onboard spectrometer will remove this offset."

C12: Line 218-219: Are both the speed and direction of the representative wind used? How?

AAR: Yes, both are used. The wind direction is used to choose between the 925 hPa and PBLH/2 levels, and overpasses with a wind–centreline angle above 45° are rejected. The flux is then computed with the wind component normal to each cross-section, $v_\perp = |v| \cos \theta$.

We apologize that this 45° filter had not been described in the Methods before. The definition of $v_\perp$ after Eq. 3 has been corrected; both are now included in Sect. 2.4.5.

Line no: 251

"Overpasses for which the angle θ between the selected wind vector and the centreline exceeds 45° are rejected, because the plume orientation is then inconsistent with the local transport direction. Both wind speed and wind direction therefore enter the calculation: the direction is used for the level selection and the alignment check, and the flux through each cross-section is computed with the wind component normal to the cross-section, $v_\perp = |v| \cos \theta$."

C13: Figure 3b – surprising that T tied for lowest number of "Excellent" and had second highest number of "Poor". For Vindhyachal, secondary sources contaminating the plume would give systematically high emissions, while sources contaminating the background would give systematically lower emissions, which the only possibility that the authors seem to mention.

AAR: Thank you for the comment. We have now included an explanation in Sect. 3.2. Scenario T uses the full anthropogenic tracer, so overlapping secondary plumes are attributed to VDL and bias the estimate high. Low T values occur when part of the plume falls outside the swath or the 30 km range, or when the single-level wind is not representative. In the D scenarios, as the referee notes, secondary sources can also raise the estimated background and bias the estimate low. The larger number of Excellent overpasses in D1 therefore partly reflects compensating errors and does not indicate better skill.

Line no: 389

"The Scenario T uses the only anthropogenic tracer; enhancements from secondary sources that overlap with the VDL plume within the cross-sections are attributed to the selected TPP and bias the estimates high. Underestimation in Scenario T occurs when part of the plume core falls outside the swath or beyond the 30 km cross-section range, or when the single-level wind is not

representative of the effective transport wind. In the D scenarios, secondary sources can additionally raise the estimated background and bias the estimate low. The larger number of Excellent overpasses in D1 than in T (Section 3.3) therefore partly results from compensating errors, where the background over-subtraction of M1 offsets the overestimation caused by overlapping plumes, and does not indicate better skill."

C14: Line 313: What were the criteria for an emission estimate to be considered valid/invalid?

AAR: An estimate is valid when:

- A CO_2 plume is detected in the constrained domain.
- The wind–centreline angle is 45° or less.
- At least one cross-section with positive flux remains after the exclusions.

We revised Sect. 2.4.7 to include this criterion, as follows:

Line no: 297

"*An emission estimate is considered valid when (i) a CO_2 plume is detected within the constrained domain, (ii) the wind--centreline angle does not exceed 45° (Section 2.4.5), and (iii) at least one cross-section with a positive flux remains after the distance and outlier exclusions. For the scenario comparisons, only overpasses with a valid estimate in all five scenarios are retained.*"

C15: Line 322: It should be noted that by applying the Nassar et al. (2013) TIMES scaling factors to the model simulation, the emissions observed during daylight/midday will be higher by a small percentage compared with the reported annual emissions disaggregated uniformly.

AAR: Yes, we agree. This is a limitation of satellite-based monitoring of point sources when calculating annual emissions using information from a single overpass. However, in our OSSE setup, this will not affect the results, since our true annual emission is derived solely from the emission at the overpass time.

We revised section 2.4.7 as follows:

Line no: 287

"The true emission (E_{truth}) is the annual-equivalent emission derived from the TIMES-scaled EDGAR emission at the time of the CO2M overpass, summed over a 3 × 3 neighbourhood of grid cells centred on the peak emission pixel at the source location rather than the single source grid cell alone. Because the satellite-derived estimates are converted to annual-equivalent

emissions in the same way, both quantities refer to the same overpass time, and the comparison is not affected by the temporal variability introduced by the TIMES factors. The 3 × 3 aggregation partially accounts for the contribution of the immediately surrounding grid cells (Fig. 2e) to the detected plume. The emissions reported in this study therefore represent the aggregated emissions in the close proximity of each TPP, which we refer to as TPP emissions for simplicity."

C16: Line 370: Explanation says that CPR and KDI sources are embedded in dense industrial regions where neighbouring sources contribute to the detected plume. Do they also contribute to EDGAR (True emissions)? CPR E_T emissions are extremely overestimated.

AAR: Not fully. True emissions include neighbouring sources only if they lie within the 3×3 EDGAR neighbourhood. Sources farther away are transported into the cross-sections but are not part of true emissions, which explains the large overestimation at CPR.

C17: Line 377: Bovensmann et al. (2010) do not use the CSF method. In fact, it is not even mentioned in their paper.

AAR: Removed.

C18: Line 385: The explanation for underestimation seems to be an important finding. Is there a set of figure panels resembling Figure 2 that could be used to illustrate this better for the reader rather than just a text description?

AAR: We agree. We have added Fig. B3 that shows the $XCO_{2,ANT}$ and plume detected in T, D1 and D2 scenarios for a sample scene and how the background estimation and noise field influence the enhancement inside the plume, which influences the emission estimation.

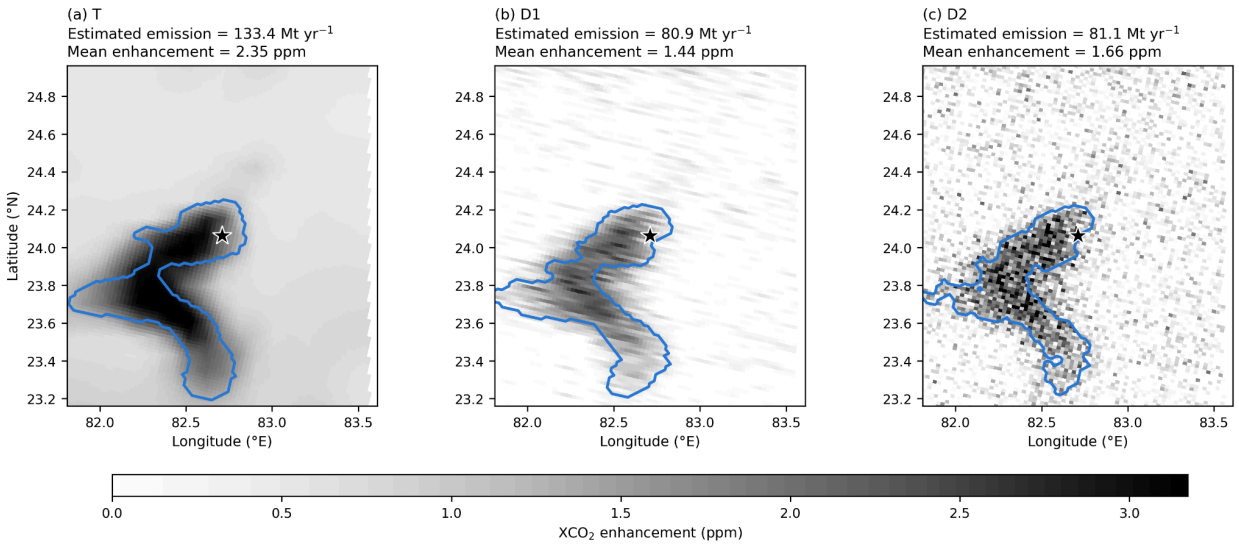


Figure B3. XCO_2 enhancement and detected plume for Vindhyachal (VDL) on a selected day (13 March 2019) in (a) the T scenario (true enhancement) and in (b) the D1 and (c) the D2 scenarios after removal of the polynomial background, with and without Gaussian pre-filtering, respectively. The blue outline marks the detected plume and the star marks the source. The estimated emission and the mean enhancement within the detected plume are given above each panel. Because the polynomial background is fitted to pixels that still contain anthropogenic enhancement (the plume tail beyond the 0.4° exclusion radius and other sources), the background is overestimated, and the enhancement within the plume and the resulting emission are reduced in D1 and D2.

C19: Line 408: In fact, the OCO-2 swath is usually less than 10 km, depending on the position of the sun an orientation of the satellite. Perhaps replace "10 km" with "≤ 10 km".

AAR: Corrected.

C20: Line 422: Missing beginning of sentence or word: "he contrast …"

AAR: Corrected.

C21: Line 429: VDL is not really the most isolated emitter, with Rihand nearby according to figure 1, which does not even account for all nearby sources.

AAR: We agree. Corrected and rephrased.

C22: Line 449: It might not be worth reporting a detection limit of 20 $MtCO_2$/year based on the inconclusive finding of this study in contrast to the Kuhlmann et al. (2021) value of 3/5 $MtCO_2$/year.

AAR: We agree and have revised.

C23: Line 474-476: Neglecting transport error could have a major impact on results.

AAR: Yes. In Sect. 4.5, we now state that the same WRF-GHG run provides both the observations and the wind, so wind errors (a dominant CSF uncertainty) are not represented and the absolute accuracies are optimistic. We also note that T is not fully free of transport-related error: the plume moves with the full 3-D wind field, while the CSF uses a single-level wind.

Section 4.5 is revised as follows:

Line no: 550

"The present OSSE employs synthetic CO2M observations generated from WRF-GHG model output, with retrieval noise added post-hoc at 0.7 ppm. In this OSSE design, we have the following limitations: (i) cloud screening, aerosol interference, and surface albedo effects are not represented; (ii) the spatial correlation structure of real CO2M retrieval noise may differ from the uncorrelated Gaussian model used here; (iii) the model transport error is not considered; and (iv) the WRF-GHG simulations at 10 km resolution are coarser than the 2 km synthetic CO2M pixels.

Regarding (iii), Scenario~T is not entirely free of transport-related error. The plume is transported by the full three-dimensional WRF-GHG wind field, whereas the CSF uses a single-level wind, so the spread in Scenario~T partly reflects the representativeness of the transport wind. With real observations, additional errors in modelled or reanalysis winds would increase the overpass-to-overpass spread and could introduce systematic biases. The absolute accuracies reported here should therefore be regarded as optimistic. Regarding (iv), sampling the model at 2 km reproduces the observation geometry and swath structure, but not variability at scales below the model grid. Near-source plumes are therefore broader and their peak enhancements lower than in reality, which affects the SCI and BEC plume masks, the cross-section widths, and the near-source flux estimates. Whether this leads to an over- or underestimation of the achievable accuracy cannot be determined from the present simulations, which prevents a fully quantitative interpretation of the absolute emission estimates. Both limitations affect all five scenarios in the same way and therefore have less influence on the relative comparison between background estimation methods."

C24: Line 476-478: Surely, the fact that the model resolution at 10-km is coarser than the satellite observations will introduce an error that could prevent a fully quantitative interpretation of the results.

AAR: Yes, we expanded the resolution caveat. At 10 km, near-source plumes are broader and have lower peak values than in reality, which affects the plume masks, the cross-section widths, and the near-field fluxes. The direction of the resulting bias cannot be determined, which prevents a fully quantitative interpretation. The relative comparison of background methods is less affected. Please see the revised text in the AAR for C23.

C25: Line 485: I believe this sentence should refer to "utilizes synthetic CO2M observations", not just synthetic CO_2 observations in general.

AAR: We agree. Corrected

Detailed Response to Referee #2 (RC2)

C1: This study presents a satellite-based framework that combines TROPOMI NO2 observations with simulated CO2M XCO2 retrievals to identify power-plant plumes and quantify CO2 emissions over India. The topic is timely and relevant, particularly for future satellite-based monitoring of large point sources. The manuscript also includes several useful methodological developments for plume identification and emission estimation.

However, several aspects of the methodology and its evaluation need clarification and more thorough testing before the main conclusions can be fully supported. In particular, the added value of using NO2 to guide CO2 plume detection should be demonstrated more explicitly, and the sensitivity of the plume-boundary definition and quality-screening procedure deserves further analysis. I therefore recommend major revision and encourage the authors to address the following points.

AAR: We thank the referee for the comments. We included the comparison with and without NO_2 guidance (Sects. 3.1 and 4.1, Fig. 4, Table A3) and a BEC sensitivity test (Table A3). We also clarified how the quality classes are used and why flagged TPPs fail.

Major comments:

C2: Since one of the main ideas of this study is to use NO2 as a guide for identifying the XCO2 plume, I think it would be useful to show more clearly how much this actually improves the result. A simple comparison between retrievals with and without NO2 guidance would already be very informative. For example, does NO2 mainly help with plume detection, plume-boundary definition, or the final emission estimate? It would also be useful to show the consistency between the NO2 and CO2 plumes on a case-by-case basis. Since the two species have very

different atmospheric lifetimes, their plume shapes and downwind extents may not always match well.

AAR: We thank R02 for this important suggestion. We agree; we have performed the full retrieval with and without NO_2 guidance (Sect. 3.1, Table A3, Sect. 4.1, Fig. 4):

- Detection: guidance increases the number of near-source detections in T and D1, but not in D3.
- Plume boundary: the NO_2 shape is not used; NO_2 only sets the constrained search domain.
- Emission: in T, the results are essentially unchanged (median ratio 1.00). In D1, removing the guidance inflates the median E/E_{true} from 0.81 to 1.28. In D3, there is no systematic effect. The benefit is therefore mainly a stabilization of large-domain background estimates.
- On the different lifetimes: The direction of NO_2 matters most. We are using the NO_2 plume shape only to define the constrained search domain for CO_2 plume detection. Therefore, plume shape mismatch doesn't matter. Then, regarding the downwind extent, we are using a buffer distance of 50 km in all directions on the NO_2 plume to define the constrained domain.

We have updated the manuscript text as follows:

Line no: 350

"Plume detection was performed both with and without the TROPOMI NO_2 retrievals. Because the NO_2 plume is used only to define a constrained search domain for the CO_2 plume, its shape does not determine the shape of the detected CO_2 plume; only its direction and downwind extent matter, as they set where the CO_2 plume is searched for. The number of detected plumes nevertheless differs between the two approaches (Table A3). However, a detected plume does not necessarily correspond to the emission from the target power plant. Even in the noise-free T scenario, 65 of the 83 plumes (78 %) detected only with NO_2 guidance originate more than 20 km from the source, i.e., they are enhancements from other sources or background structures within the search domain. Counting only plumes that originate within 10 km of the source, NO_2 guidance increases the number of detections from 176 to 189 in the T scenario and from 216 to 257 in the D1 scenario, whereas no information gain is found in the D3 scenario (279 without and 262 with NO_2 guidance). In the D scenarios, the background and retrieval noise added to the CO_2 field cause the algorithm to frequently identify noise or background structures as plumes, so the numbers in Table A3 should be interpreted with care. These differences in plume detection are also reflected in the emission estimation process. The results are discussed in Section 4.1."

Line no: 443

"In the estimation phase, the influence of the NO_2 guidance depends on the background estimation rather than on the plume itself. The number of valid estimates is similar with and without NO_2 guidance in all scenarios. In the noise-free T scenario, both retrievals give almost identical emissions when they detect the same plume: for the 97 overpasses with an estimate in both, the median absolute change is 4 % (interquartile range 1–8 %), 80 % of the pairs agree within ± 10 %, only 8 % differ by more than 30 %, and the median E/E_{truth} is unchanged (1.45 and 1.44) (see Figure 4). In the D1 scenario, the polynomial background fitted across the entire CO2M track does not capture the regional enhancement around the source, leading to systematic overestimation. The estimates change by a median of 64 % (interquartile range 24–169 %), 71 % of the pairs differ by more than 30 %, mostly towards higher values (66 of 77), and the median E/E_{truth} increases from 0.81 to 1.28. In the D3 scenario, where the background is estimated locally, there is no systematic shift (median ratio 1.04). However, the individual estimates vary by a median of 21 % (interquartile range 8–72 %). 43 % of the pairs differ by more than 30 %, in both directions (29 higher, 21 lower), mainly because the two retrievals select different plumes in the noisy field. Differences in the detected plumes may partly arise from the constrained search domain imposed by the NO_2 plume. Differences between the TROPOMI and CO2M overpass times and errors in the WRF-GHG simulations used to generate the CO2M pseudo-observations may cause the CO_2 plume to lie partly or entirely outside this domain. In such cases, NO_2-guided detection may identify a different CO_2 enhancement within the search domain and incorrectly attribute it to the target source, whereas unguided detection may identify the actual source plume outside the constrained domain. Using actual CO_2 retrievals acquired at overpass times coincident with the NO_2 observations would reduce these temporal and model-related mismatches, potentially improving plume correspondence and source attribution. The NO_2 guidance therefore mainly prevents a systematic bias when the background is estimated over a large domain, whereas with a local background estimate it causes no systematic change but affects individual estimates considerably in the current OSSE setup."

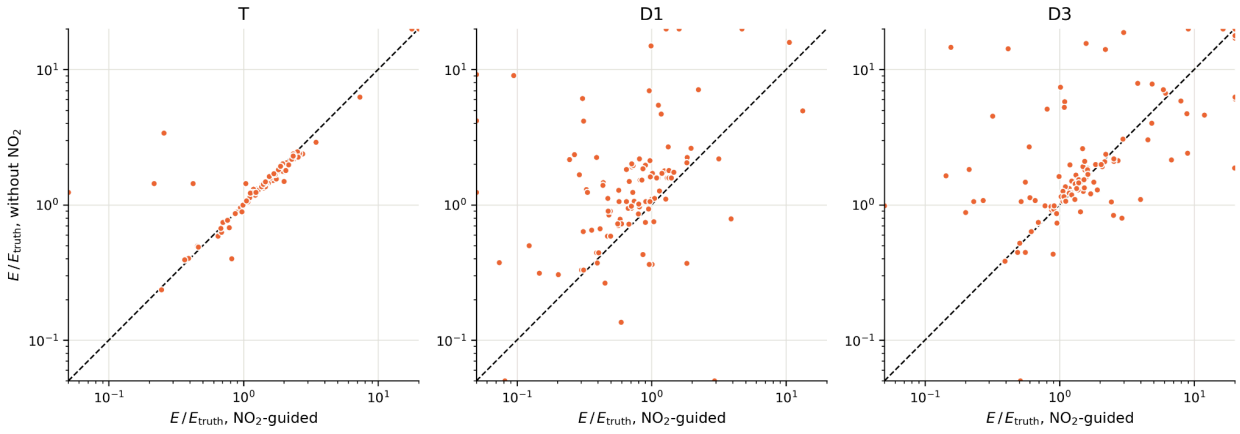


Figure 4. *Retrieved emission relative to the true emission (E/E_{truth}) with TROPOMI NO_2 guidance (horizontal axis) and without NO_2 guidance, i.e., with the whole CO2M track as search*

domain (vertical axis), for the T, D1 and D3 scenarios. Only overpasses with a valid estimate in both retrievals are shown (n = 97, 109 and 117 of 427 overpasses, respectively). The dashed line is the 1:1 line; values outside the range 0.05–20 are plotted at the axis limits.

Conclusions: Line 612:

"*NO_2 guidance improved source-proximal plume detection in some scenarios and mitigated systematic emission overestimation when the background was estimated over a large domain, although its benefits varied across scenarios. Differences in plume attribution partly resulted from the constrained search domain imposed by the NO_2 plume, which may not encompass the simulated CO_2 plume because of atmospheric transport errors in the CO2M pseudo-observations and differences in TROPOMI and CO2M overpass times. Our results emphasize that coordinated CO_2 and NO_2 observations from future satellite missions can improve plume correspondence and source attribution by reducing these temporal mismatches.*"

C3_A: The definition of the final plume boundary in the BEC procedure is still not fully clear to me. The manuscript says that the outermost enhancement contour that remains spatially connected to the SCI core is taken as the final boundary. However, this boundary may depend quite strongly on the selected enhancement levels, contour spacing, and the way spatial connectivity is defined. Please explain more clearly how these choices are made, and what happens if the expanding contour merges with another nearby plume or background enhancement.

AAR: Thank you for this suggestion. We expanded the BEC description in Sect. 2.4.1. It now gives the contour levels (80/60/40/30 % of the core peak and 50 % of the core mean), the 8-neighbour connectivity, and the rule that selects the lowest level overlapping the core. We added two new figures: Figure B1, which shows how sensitive the plume boundary is to the selection of the lowest contour level, and Figure B2, which explains each step of plume detection in more detail.
The text is revised as follows:

Line no: 146

"In BEC, a set of enhancement levels is defined relative to the enhancement statistics of the SCI core, namely 80, 60, 40, and 30 % of the maximum core enhancement and 50 % of the mean core enhancement. The enhancement field is thresholded at each level, starting from the lowest, and spatially connected regions are identified using 8-neighbour connectivity. The connected region at the lowest level that overlaps the SCI core is adopted as the final plume boundary (See Fig. B1). Because the levels scale with the core enhancement, no absolute enhancement threshold is required, and the same procedure can be applied to different sources and species. By doing this, we capture plume edges that fall outside the statistical significance threshold while preserving the statistical anchor provided by the SCI core. If the region at the lowest level merges with a

nearby enhancement from another source or the background (as in Fig. B1c), the merged region is retained in the plume mask; the influence of such merged features on the emission estimate is limited by the distance cut-off and the sequential outlier check applied to the cross-sections (Section 2.4.5).”

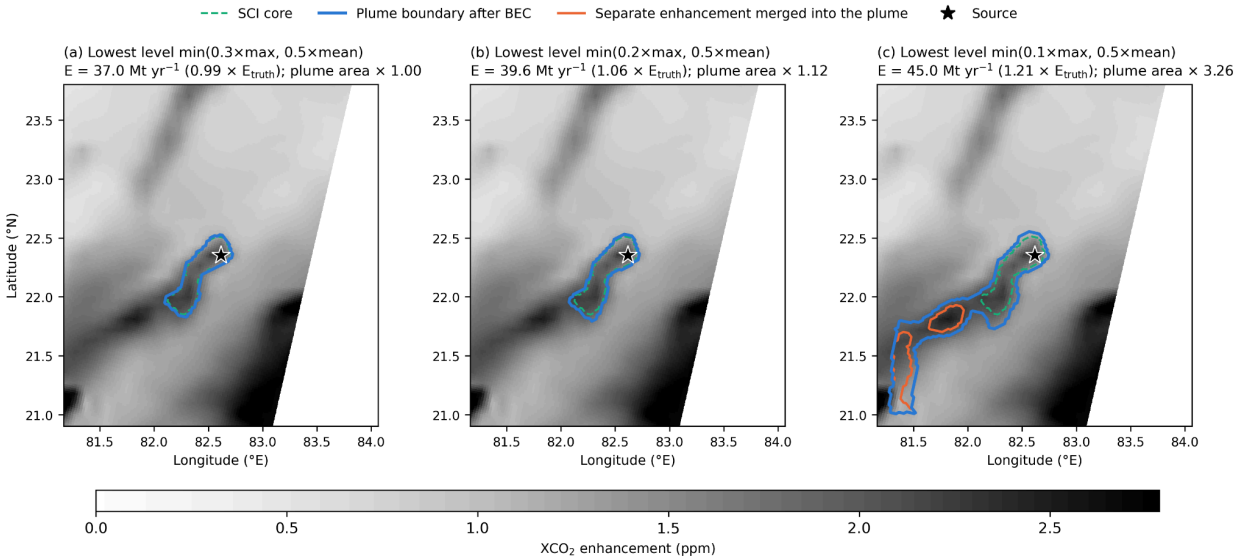


Figure B1. Sensitivity of the final plume boundary to the lowest contour level of the boundary extension by contours (BEC). (a) the lowest contour level used in this study, min(0.3 × max, 0.5 × mean), (b) min(0.2 × max, 0.5 × mean), and (c) min(0.1 × max, 0.5 × mean), where max and mean are the maximum and mean enhancement of the SCI core. At the lowest level, the boundary merges with two separate enhancements (orange). The estimated emission, its ratio to the reference emission, and the plume area relative to (a) are given above each panel.

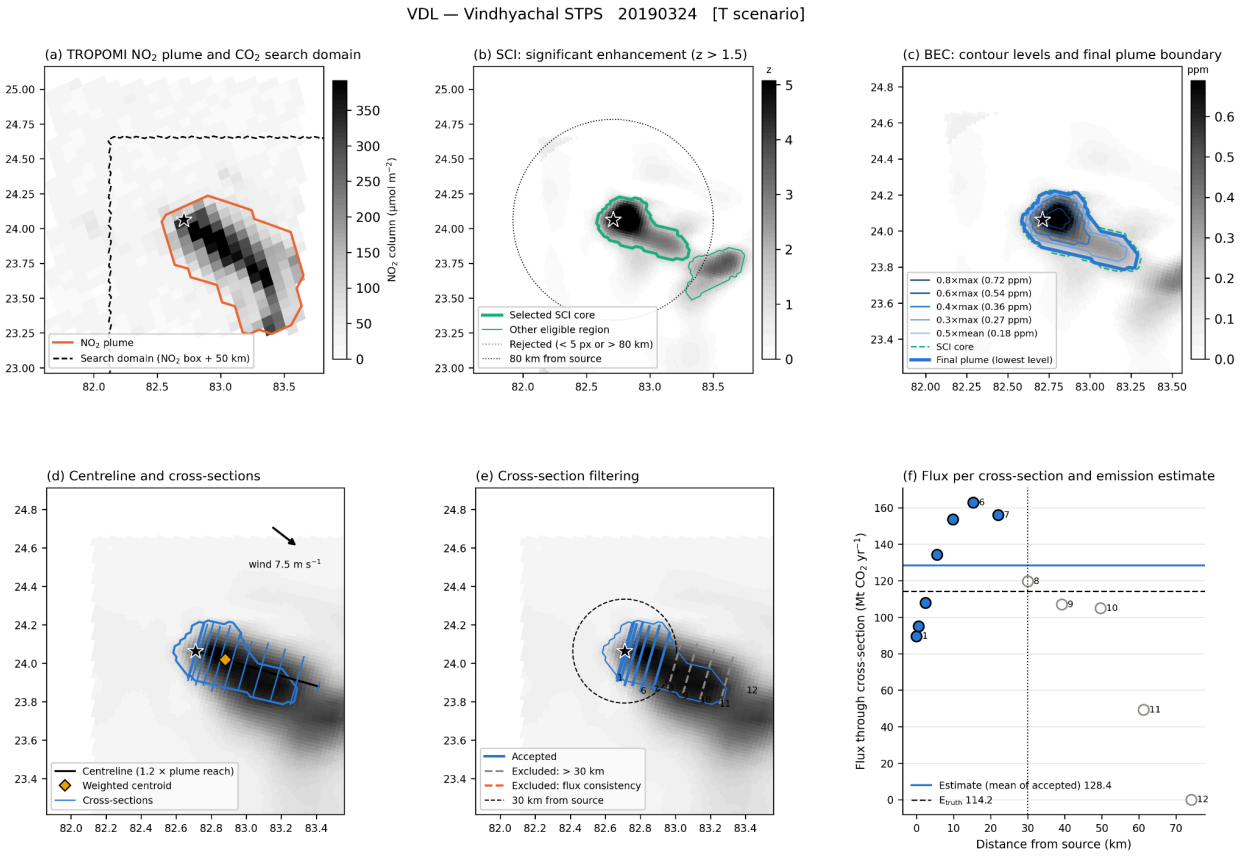


Figure B2. Steps of the plume detection and emission estimation for Vindhyachal (VDL) on a selected day (24 March 2019) T scenario. (a) TROPOMI NO_2 plume (orange) and the resulting CO_2 search domain, i.e., the bounding box of the NO_2 plume extended by 50 km (dashed). (b) Statistical core identification (SCI): significance z of the XCO_2 enhancement; regions with z > 1.5 that contain at least five pixels and lie within 80 km of the source (dotted circle) are eligible, and the selected SCI core is shown in bold. (c) Boundary extension by contours (BEC): regions connected to the SCI core at the contour levels 0.8, 0.6, 0.4 and 0.3 times the maximum and 0.5 times the mean core enhancement; the region at the lowest level forms the final plume boundary (bold blue). (d) Enhancement-weighted centroid, centreline and cross-sections perpendicular to the centreline, with the wind at 925 hPa. (e) Cross-sections used in the flux calculation (sections 1–7) and excluded beyond 30 km from the source (sections 8–12). (f) CO_2 flux through each cross-section as a function of distance from the source.

C3_B: I would also suggest adding a sensitivity test to show how much the final plume area and retrieved emission rate change when these settings are varied.

We have also conducted the sensitivity analysis with different levels assigned for the lowest contour level, which determines the plume boundary. Table A1 shows the results of the analysis. The results show that when the contour level is kept high, the boundary shrinks and the emission

estimates decrease, as expected. The important finding is that lowering the lowest contour level increases the area, but the emission does not increase as expected. When the plume area increases, there is a chance that secondary plumes, if they exist, also get merged with the primary plume (See Fig. B1c). But in our estimation step, we already addressed such situations by filtering out the cross-sections accordingly (see Section 2.4.5, Fig. B2e-f). The manuscript text is updated as follows:

Line no: 155

"*The sensitivity of selection of the lowest contour level towards the plume area and estimated emissions was also tested. The results show that lowering the contour level further enlarges the plume area (by up to a factor of 2) but raises the estimates by only 12–17% (See Table A1) because of the cross-section filtering explained above.*"

Table A1. Details of the sensitivity analysis conducted. Fractional change in plume area and emissions relative to the reference plume are shown.

Lowest contour level (Plume boundary)	*Fractional change in plume area w.r.t. the reference plume* [a]	*Fractional Change in emission estimation* [b]
*0.5 * Max. Enhancement*	*0.39 - 0.48*	*0.66 - 0.73*
*0.4 * Max. Enhancement*	*0.59 - 0.70*	*0.80 - 0.84*
*0.2 * Max. Enhancement, 0.5 * Mean Enhancement*	*1.00 - 1.21*	*1.00 - 1.03*
*0.1 * Max. Enhancement, 0.5 * Mean Enhancement*	*1.44 - 1.97*	*1.12 - 1.17*

[a]*The "reference plume" is derived using the lowest contour level condition used in the manuscript.* [b]*Expressed as a factor of difference relative to the emission estimation using the reference plume.*

C4: For overpasses classified as "excellent", "good", or "poor", is this ranking actually used when calculating the final emission estimate? For example, are better-quality overpasses given larger weights, or are all accepted cases treated equally and then summarized using the mean and one standard deviation? If all accepted cases are treated in the same way, then it is not entirely clear what the practical role of this classification is. It would be helpful to explain this more clearly and, if possible, show whether the different quality classes have different retrieval errors or uncertainties.

AAR: No weighting is applied. The tiers only describe how close the estimates are to the true emission in the OSSE. We have now added this information to the manuscript as:

Line no: 303

"The tiers are used only to describe the distribution of per-overpass estimates; they are not used to weight or select overpasses, and therefore all valid overpasses contribute equally to the reported mean and standard deviation."

C5: In Section 3.4, the manuscript mainly shows results for TPPs with quality_flag = 0. I think it would also be useful to briefly look at the cases that do not pass this quality check. What are the main reasons for failure? Are they mostly related to weak emissions, clouds, unfavorable wind conditions, large background variability, overlapping plumes, or insufficient satellite coverage? This could help readers better understand the practical limits of the current method and under what conditions it is likely to work well.

AAR: We revised statements of flag criteria. Also, we have included Table A2 that shows the primary reason for excluding a TPP from reporting the emission estimate. Also, we revised the manuscript text in sections 2.4.9 and 3.4.

TPP	***Flags***
Mundra (MDR)	*F2*
Tiroda (TRD)	*F2*
Shri Singaji (SSJ)	*F2*
Indira Gandhi (IGT)	*F1, F2*
Paras (PRS)	*F1*
Kota (KOT)	*F2*
Satpura (STP)	*F1, F2*

Table A2. Thermal power plants (TPPs) that did not pass the quality check, with the quality flags assigned to each.

Line no: 342

"We find that emissions from seven active TPPs, IGT, KOT, PRS, STP, MDR, SSJ, and TRD, are either severely underrepresented or omitted in the EDGAR emission inventory. Consequently,

they are positioned for filtering based on one or more of the flags described above (See Table A2) and are excluded from further analysis."

Line no: 416

"In the case of flagged TPPs, as discussed in Section 2.4.9, the flags assigned within our OSSE framework are primarily associated with weak emission signals resulting from the relatively low emission estimates reported in the EDGAR inventory. However, in a real-world scenario, these conditions may vary depending on changes in emission intensity and the influence of other factors, such as cloud cover, aerosols, and background CO_2 concentrations."

Specific comments:

C6: Line 30: "uncertainties in transport modelling over India." Is there any specific reason why transport modelling is particularly challenging over India? For example, does this refer to complex topography, monsoon circulation, boundary-layer dynamics, or strong spatial variability in meteorological conditions?

AAR:

We thank the reviewer for this remark. Uncertainties in transport modelling over India arise from several factors, such as complex topography, monsoon circulation, and boundary-layer dynamics, as well as the scarcity of measurements to evaluate the models. Sukumaran et al. (2024) demonstrated that errors in representing PBL dynamics play a significant role among these factors. In addition, the spatially heterogeneous distribution of surface emissions across India complicates the interpretation of observed variability in atmospheric mixing ratios. We have revised the sentence to state these reasons explicitly and to refer readers to the cited studies for further detail.

Line no: 30

"A*tmospheric inverse modelling using atmospheric monitoring networks offers a promising pathway for independent assessment, but its performance is constrained by sparse measurement networks and uncertainties in transport modelling, particularly in the representation of planetary boundary layer dynamics, which are further compounded by the spatially heterogeneous distribution of surface emissions over India (Thilakan et al., 2024; Sukumaran et al., 2024).*"

C7: Line 35: "anthropogenic enhancements" is a little vague. I suggest using something more specific, such as "anthropogenic emission-induced concentration enhancements" or simply "XCO2 enhancements caused by anthropogenic emissions". It would also be helpful to give a rough quantitative value here, for example, the typical plume enhancement in ppm compared with the background XCO2 level.

AAR: Done. We use the term XCO_2 enhancements caused by anthropogenic emissions to refer to those.

C8: Line 45: Please give the full name of TROPOMI when it first appears.

AAR: Added. Thank you

C9: Line 45: Please define XCO2 when it first appears, for example, as the column-averaged dry-air mole fraction of CO2.
AAR: Added. Thank you

C10: Line 85: What does "IST" stand for? Please define it when first used.
AAR: Corrected. Thank you

C11: Line 90: Please give an approximate lifetime of NO2. Since NO2 is used here as a guide for CO2 plume identification, it would also be useful to briefly mention that its lifetime varies with photochemical and meteorological conditions.
AAR: Added to Sect. 2.2: NO_2 has a lifetime of a few hours in the daytime boundary layer, which varies with photochemistry and meteorology.

Line no: 96

"*The relatively short atmospheric lifetime of NO_2, typically a few hours in the daytime boundary layer and varying with photochemical conditions (solar radiation, OH and NOx concentrations) (Lammel and Graßl, 1995) and meteorology, together with its high contrast against the ambient background, makes it an efficient proxy for detecting CO_2 emission plumes from fossil-fuel combustion (Georgoulias et al., 2020; Liu et al., 2020; Goldberg et al., 2019).*"

C12: Figure 3(a): There seems to be an extra "1" in panel (a).
AAR: Corrected. Thank you.