Response to Reviewers (Second Round)

Manuscript: Global variability in the detectability of power plant NO_2 plumes from space

Authors: Ruizhe Huang, Sherrie Wang

Journal: Atmospheric Measurement Techniques (AMT)

Dear Editor and Reviewers,

We thank both reviewers for their continued careful reading. We are grateful that Reviewer 2 considers Sections 1 to 4 to be of excellent quality, and we have worked to bring the discussion in Section 5 to the same level. Below we reproduce each comment in full, followed by our response and the corresponding change to the manuscript.

Esri attribution

Reviewer comment. Regarding Figures 1, 10, H1, H2, H3, H4 and H5: since Esri requires the use of "Powered by Esri" in any case, please include this statement in the caption.

Response. We have added the "Powered by Esri" statement to the captions of Figures 1, 10 and H1–H5.

Response to Reviewer 1

1. Figure 5

Reviewer comment. Fig. 5: please add lat/lon coordinates to x and y labels; please add arrows for wind direction to the panels; please provide information on the respective measurement dates in the figure caption.

Response. We have revised Figure 5 accordingly. The x and y axes now carry longitude and latitude coordinates, each panel includes an arrow showing the wind direction, and the caption now reports the acquisition date and time of both examples.

Response to Reviewer 2 (A. Barr)

1. Implications for future satellite missions

Reviewer comment. There is a missed opportunity to connect this paper to forthcoming satellite missions, of which plume detection of NO2 will be a main focus, which is really a pity. GOSAT-GW (1 km), CO2M (4km), TANGO (300 m) and CO2Image (50 m) will provide a new suite of sensors with higher spatial resolution, increasing the capability of plume detection. I strongly suggest to the authors that they, at the very least, include a paragraph in the discussion on addressing the question of spatial resolution, and in doing so highlight the importance of their work to the science community.

The real question is: to what extent do subpixel variations in the surface albedo and elevation affect plume detectability? Such questions are of high value at this time, reflected in the literature (Weimer et al. 2026 for CO2M https://egusphere.copernicus.org/preprints/2026/egusphere-2026-1458/). While it is beyond the scope of this paper to provide an analysis on this issue, it should certainly be the focus of a follow-up paper. A comparable study with a sensor of higher spatial resolution would go a long way in solidifying the results of Figure 9. It may well be that the same top 5 features also govern plume detectability in these new missions (higher emissions would still lead to more detectability; higher albedo leads to more signal thus higher detectability). On the other hand, surface albedo and elevation are also highly correlated with retrieval biases which might lead to the scenario that these features are less important if they are better characterised at small scales. Thus how this work generalises to finer scales is of utmost importance. Then we could get rid of the 'would' and 'should' in the author response #6.

Furthermore, a sensor with better spatial resolution implies that the field of view is also smaller. This could further help reduce interference from neighbouring power plants/cities, which would allow more power plants to be characterised and, more importantly, a higher percentage of global emissions.

Based on their results can the authors give a statement in the discussion about what would most improve plume detectability? Better forecasting models, better satellite instruments (SNR, spatial/spectral resolution), more accurate emissions reporting?

Note: Please do not argue that any of the features are instrument-agnostic. This ignores the interdependece of information. Detectability, by definition, depends on the amount of plumes detected by a sensor. Thus, detectability is governed by how the features listed in Figure 9 impact satellite data. The authors themselves discuss this point several times (L443, L487, L511, section 5.1, L584).

Response. We thank the reviewer for this in-depth comment. It indeed strengthens the paper to connect our findings to forthcoming NO_2 missions. We added the following new section to the Discussion (section 5.4) to expand on this:

"Several new and forthcoming satellite missions are moving point-source plume sensing toward spatial resolutions finer than TROPOMI's 3.5–7 km pixels. GOSAT-GW/TANSO-3 targets 1–3 km footprints in Focus Mode, CO2M will provide collocated CO_2, CH_4, and NO_2 through CO2I/NO2I at ~4 km^2, and TANGO-Nitro targets facility-scale NO_2 imaging at ~300 m. How do the TROPOMI-scale relationships quantified here translate to sensors with smaller footprints? We discuss implications for plume detectability and feature importance in turn.

Spatial resolution affects detectability through competing mechanisms. First, a smaller instantaneous field of view reduces mixing between the plume and background, so a point-source enhancement may be less diluted within a pixel. Second, finer pixels can reduce interference from neighboring power plants and cities. This is particularly relevant for our study because interference filtering excludes a large fraction of plants from the second-stage detectability analysis. Higher-resolution sensors could therefore allow more power plants, and a larger fraction of global emissions, to be characterized. Third, for a fixed instrument design, smaller pixels collect fewer photons, which can increase retrieval noise and background variability. Thus finer spatial resolution should improve plume detection when TROPOMI is limited by plume dilution or source interference, but this gain depends on maintaining sufficient SNR and retrieval quality.

How feature importance changes at finer spatial resolution is harder to predict. The important features identified in Fig. 10 should be interpreted as properties of the TROPOMI retrieval, TROPOMI pixel size, the detection algorithm, and the sampled observation conditions, rather than as feature rankings that transfer unchanged across sensors. Surface albedo and surface altitude are especially important examples. In our TROPOMI-based model, these variables may influence detectability both by affecting retrieval and by acting as proxies for systematic bias associated with unresolved surface heterogeneity. Recent work on CO2M-like retrievals shows that correlated subpixel variations in surface reflectance and altitude can produce retrieval biases, and that improved subpixel surface characterization can reduce those biases (Weimer et al., 2026). At finer resolution, some of the apparent importance of albedo or altitude could therefore decrease if it currently reflects unresolved subpixel bias. Repeating our analysis for higher-resolution sensors would show which predictors remain dominant across instruments and which are specific to TROPOMI-scale retrieval and resolution. We leave this comparison to future work."

We also appreciate the suggestion to offer a statement on what would most improve detectability. We first want to clarify that the three categories listed in the reviewer's comment (better forecasting models, better satellite instruments, more accurate emissions reporting) play different roles. Improvements in satellite instrumentation (e.g., spatial resolution, SNR, spectral resolution, and retrieval algorithms) have the potential to directly improve plume detectability (and we discuss spatial resolution above). In contrast, improved meteorological forecasts and more accurate emissions inventories do not change the NO_2 enhancement in a satellite observation and do not affect detectability. Rather, they improve our ability to predict, model, and attribute detectability.

Regarding instrument design: our study provides only limited guidance because nearly all instrument characteristics are fixed for TROPOMI. The primary instrument-dependent quantity that varies across observations is the viewing geometry, which also changes the effective pixel size across the swath. Our analysis shows that detectability is substantially higher near the center of the swath, where pixels are also smaller, than near its edges. Other important design parameters—including SNR, spectral resolution, and retrieval algorithm performance—do not vary within the TROPOMI dataset and therefore cannot be isolated by our framework. Quantifying those effects would require repeating our framework using observations or realistic simulations from future sensors.

2. Sensor properties versus prior information (Table 1, L162)

Reviewer comment. L162: Important to distinguish sensor properties from information obtained by the sensor. Table 1 can mislead: some parameters (e.g. surface pressure) are prior information from other databases. I believe surface albedo at 440 nm comes from OMI climatology. Add a clarifying note at the start of §2.4.

Response. We have added a clarifying sentence at the end of the introductory paragraph of §2.4, immediately before the table is presented, to make clear that the "Source" column reflects where the variable is retrieved in our pipeline rather than its original provenance. The added sentence reads:

"Several variables labeled as 'TROPOMI' in Table~\ref{tab:combined_features} are not direct sensor measurements but auxiliary fields packaged into the Level-2 NO$_2$ product from external datasets (e.g., surface albedo and surface pressure)."

3. Appendix E moved into the main text

Reviewer comment. Appendix E is never referenced in the main text. The result in Fig E1 is very important and increases the reliability of the results. This needs to be discussed in the main text, appropriately in the emission-uncertainty discussion. Can you move Appendix E to §5.3?

Response. We agree the cross-pipeline check is a substantive robustness result and have promoted it into the main text. We placed it in §4.2 rather than §5.3 because the comparison is between two empirical detectability estimates from independent pipelines (driven by detection labels and overpass counts); emissions do not enter the comparison, so it is a robustness statement about the detectability values reported in §4.2, not about emission uncertainty. Appendix E has been removed.

4. Appendix A never referenced

Reviewer comment. Appendix A is never referenced in the main text. Please include at least one reference so the reader knows what these refer to.

Response. We have added a cross-reference to Appendix A from §3.3, attached to the very statistic the appendix table breaks down by year (the U.S. modeling-subset retention of 189,713 / 666,222 = 28.5%). The revised sentence reads:

"...the final modeling subset retains 189,713 of 666,222 U.S. observations (71.5% excluded across 2019–2024; see Appendix A for the year-by-year breakdown) and 161,118 of 875,686 global observations (81.6% excluded)."

5. Figure 3 colour

Reviewer comment. Figure 3: What is orange referring to? Should this say green?

Response. We thank the reviewer for catching this. It should read “green,” and we have corrected the caption accordingly.

6. L53: additional references

Reviewer comment. L53: Please add the references. I would mention here Kurchaba et al. 2024 who studied NO2 plume detectability from ships from TROPOMI.

Response. We have added references at this location and explicitly distinguished the prior power-plant and ship-plume literature. The Kurchaba et al. (2024) study is now cited as the ship-plume analog to our power-plant work, since it characterizes single-ship plume detectability with TROPOMI across four maritime regions over a multi-year period.

7. L214: detectability and overpass count

Reviewer comment. L214: Please include that detectability accounts for the number of satellite overpasses, as explained in §4.2.

Response. Updated the Detectability definition so it spells out the overpass normalization. It now reads as "the per-plant ratio of detections to the number of valid TROPOMI overpasses, which normalizes out differences in overpass count across plants."

8. L344: title of Section 3.3

Reviewer comment. L344: Can the title of §3.3 be changed to 'Training Sample Filtering'?

Response. §3.3 renamed to "Training Sample Filtering"

9. L434: meaning of P90

Reviewer comment. L434: Can it be clarified what P90 refers to? Is this a metric of detection frequency, or are these detectability values? I think the second.

Response. Thanks for catching this. In that paragraph P_{90} was used for both distributions — first detection frequency, then detectability — and the lack of units on the first value made it easy to read both as detectability.

Fixes:

At first use (§2.1), we now state explicitly that P_X denotes the Xth percentile of whichever per-plant distribution is under discussion (observation counts, detection counts, or detectability).

In the §4.2 paragraph, we added units (detections/yr) to the detection-frequency values so the count and the subsequent % values are unambiguous.

10. L454: detection frequency versus detectability

Reviewer comment. L454: How does detection frequency correlate with detectability? Could this be quantified as a correlation coefficient, quoted for both US and global datasets?

Response. We thank the reviewer for this suggestion. Detection frequency (per-plant count of detections) and detectability (detections per valid overpass) are strongly correlated across plants: Pearson $r = 0.96$ (U.S.) and $r = 0.89$ (global). This is expected, since detection frequency is the product of detectability and the number of valid overpasses. We have added this to Sect. 4.1:

"Detection frequency is strongly correlated with empirical detectability across plants (Pearson $r = 0.96$ for the U.S. and $r = 0.89$ for the global dataset), as expected since detection frequency is the product of detectability and the number of valid overpasses."

11. Section 5.3: worked example of emission-estimate error

Reviewer comment. §5.3: Give an example of how the emission estimate affects model performance. E.g. a plant that over-reports its emissions would be expected to have high detectability, but TROPOMI detection is harder than expected because real emissions are lower. Bearing on detectability-prediction accuracy, given NOx importance (L490).

Response. We have added a concrete example to Sect. 5.3 illustrating how inventory error in the reported emissions propagates into model performance, and reorganized the surrounding text so the example follows directly from the "noise in the reported inventories" factor:

" To make the first factor concrete: because NO_x emission rate is the model's most important predictor (Sect. 3.4), a plant that over-reports its emissions is presented to the model with a high emission value, leading it to predict high detectability; yet TROPOMI detects the plume less often than expected because the true emissions are lower, so the prediction overshoots. A plant that under-reports produces the opposite error.

We also clarify the downstream consequences: such errors mean the reported F1 and AUC values should be read as lower bounds, and the SHAP ranking of NO_x emission rate is a conservative lower bound that cleaner, co-temporal inventories would only reinforce.

12. L618: characterisation scale

Reviewer comment. L618: Please state that these results are characterised at 5 km.

Response. We have added an explicit qualifier to the headline-result sentence in the Conclusions so the spatial scale of the F1/AUC numbers is made clear.

"In this work, we systematically mapped power plant NO$_2$ plume detectability at U.S. and global scales using TROPOMI observations (nadir pixel size 3.5–7,km), and then demonstrated that detectability can be predicted by a suite of meteorological, environmental, sensor, and power-plant variables, with our trained models achieving F1 score >0.66 and AUC >0.8 across U.S. and global datasets at TROPOMI's ~ 5,km native resolution."

13. L635: material in the conclusions

Reviewer comment. L635: As in the first iteration, material should not be presented in the conclusion that does not appear in the main body text.

Response. Thank you for raising this. The paragraph is explicitly framed as future-work outlook ("Several additional variables ... could be considered in future work"), not as a summary of results, which is a standard convention for the closing section.

On the specific items: aerosol optical depth is already in the main body (Sect. 2.4, and the paragraph cross-references it); vertical wind shear and boundary layer mixing are listed only as candidate future-work variables; CAMS and GOSAT-GW appear only as forward-looking pointers to a possible data source and a future sensor, both as cited works rather than new content. We therefore prefer to keep the future-work framing, but are happy to relocate any specific item (e.g., the GOSAT-GW remark) if the reviewer would like.