Reviewer1

The authors would like to thank Prof. Krol (R1) for his time and his interesting discussions on our draft. Following the reviewers' comments, we have expanded the methodological assessment presented in the Supplementary information. Several of these additions directly address comments made by R1, including sensitivity analyses of NO_2 retrieval uncertainties in fire scenes, background selection, the performance of the synergy plume-masking approach and the possible effect of NO_x plume ageing on the MDR.

Please, find our point-by-point responses to major comments below. Afterwards, we have also addressed the minor comments that R1 made on the manuscript's PDF and in the supplementary PDF. Herein, all R1 comments are shown in black and our responses in blue. Please note that, when referring to line numbers in the revised manuscript or Supplement, we use those from the clean versions (without tracked changes) as, depending on the settings, Word may change the line numbering when tracked changes are displayed.

This paper addresses the 2025 fires at the Iberian Peninsula using TROPOMI CO and NO_2 data.

The main development presented in the paper is the "synergy" method to quantify CO and NO_2 emissions from individual fires. These top-down estimates are compared to bottom-up estimates from GFAS, GFED, and FINN.

A main finding is that the CO emissions from the top-down method are much higher (3x) than the bottom-up estimates. Concerning the NOx/CO mole density ratio (MDR: NOx emissions relative to CO emissions: a metric for combustion efficiency), the paper finds variability in time and space, which is sometimes consistent with bottom-up estimates.

Overall, the paper presents a new "satellite data only" methodology that holds some promise to independently quantify wildfire emissions in complex fire scenes.

The paper is well written and clearly presents the results. I have a number of small comments, which I add as annotated manuscript and supplement. I have some major remarks that could improve the quality of the manuscript further. These points are listed below.

1. Satellite data quality

TROPOMI columns are derived from fingerprints in the absorbed spectrum (UV/vis for NO_2, Short-wave IR for CO). Specifically, for NO_2, the sensitivity to near-surface NO_2 depends on the amount of scattering. Specifically, in the quoted Wang et al. (2026) paper it is written: " However, due to the lack of explicit fire-related a priori information in current satellite NO_2 retrieval algorithms, the resulting data products exhibit large uncertainties under fire conditions.", and "Retrieved tropospheric NO_2 VCDs increase by up to 100 % at locations greatly impacted by fires and by about 80 % in surrounding areas. ". I am a bit surpised that this source is error is not discussed at all. These errors will both affect the plume mask and the MDR. For CO TROPOMI, errors related to fire scenes are likely smaller, because of the vertically flat averaging kernel and limited effects of aerosol on the CO retrieval.

We thank R1 for pointing this out. Indeed, the effect of scattering, particularly on the TROPOMI NO_2 fields, should be addressed. This is now done in Sect. S1.2 of the new

Supplementary information. In the revision, we discuss previous findings and conduct a sensitivity study how errors propagate into our results. Overall, our tests indicate that moderate perturbations of the NO_2 VCDs have little systematic effect on the retrieved CO emission rates, although the response of individual cases becomes more variable as the NO_2 perturbation increases. The MDRs are more directly sensitive to NO_2 retrieval uncertainties, since NO_2 enters explicitly in their calculation. We have also added references to these tests in the revised manuscript (L392-400 and L559-560).

2. Plume masking algorithm

The main issue in plume masking is to determine the CO and NOx added to the background. Thus, a background is proposed (section 4.1.1). However, in the case of overlapping plumes, the background might be influenced by upwind plumes (which is OK, because you want to analyse the amount that is added by the new plume). Although this becomes clear from the paper (e.g. figure 5) and the supplement, it would be worthwhile to explain this better in the manuscript. Moreover, in case of an upwind plume, the size of the background (0.8 degree by 0.6 degree) will encapsulate likely only part of the upwind plume and hence lead to uncertainty. Since the "synergy" methods is central to the paper, it would be helpful to present a more elaborate error analysis for background selection (now figure 3 presents a plume-free background, while figure 5 presents a more complicated case). Concerning figure 5 and potential biases, it remains a bit unclear how the "total" result from the "standard threshold" method differs from the newly developed "synergy" method. Mentioning these numbers would be valuable information.

In the Supplement additional information is provided. However, the "standard", "threshold", and "synergy" methods are not very clearly explained. It would be good to start with the "standard threshold" method (e.g. used for NO2) and then explain how the synergy method applies the NO2 mask to CO satellite data, i.e. define the "standard threshold" method as standard.

We agree, in the original text (manuscript and Supplementary) we used "standard method" and "threshold method" while in fact we were referring to the same plume masking method. The new manuscript and Supplementary text have now been corrected so only two terms are used for the plume masking methods: (1) standard threshold method and (2) synergy method.

Regarding sensitivity tests performed on the background selection, as R1 mentioned, in Sect. S1.1.3 of the original supplement draft, we compared CO emission estimates inferred after applying the reference upwind background method, against estimates calculated using a fixed regional background. To assess the background selection even further as suggested by anonymous referee 2, Sect. S1.4 of the new Supplement also includes an additional sensitivity test of the retrieved CO emission rates based on the size chosen for the background region. This is now added in L380-384 of the new manuscript: "The sensitivity of our results to background selection was also investigated (Sect. S1.4). Changing the size of the upwind background region produced little systematic effect on the retrieved CO emission rates, whereas the use of a fixed pre-fire regional background resulted in systematically higher estimates (~39%). The latter reflects the elevated regional CO background during the multi-fire episode and supports the use of a scene-specific upwind background for the conditions investigated herein"

3. Plume chemistry

CO is relatively long lived, and emissions are estimated using the IME method. For NOx the application of the IME method is questionable. Apart from the NO (emitted) to NO2 conversion, also the lifetime of NOx (~ hours) might bias the IME method. In line 536 "plume chemistry" is explicitly mentioned, so it is important to quantify potential errors due to (lack of) chemistry.

We clarify that, in our study, the IME method is applied exclusively to CO. Thus, there are not reported NO_2 or NOx emission rates derived from IME. Nevertheless, we agree that NOx chemistry could potentially affect the interpretation of the NO_2-derived NOx/CO MDR. A sensitivity test on the MDR is now included in Sect. S3 of the new Supplement. The results suggest that the finite NOx lifetime does not introduce a systematic bias in the MDR values reported in our work. In the new manuscript we have now added (L554-559): "Similarly, restricting the MDR calculation to the pixels closer to the hot spot (i.e, fresh plume) gives a slope close to unity relative to the full-plume calculation, indicating no significant systematic effect of plume ageing at the spatial scales considered here."

Minor comments on the manuscript's PDF:

- L69 and L72 "van der Werf et al., 2025" has been corrected.
- In L74 "atmospheric chemistry" has been also added as uncertainty factor for observation-based approaches.
- L84: I think the CO emissions are always higher (hence the unit mol/mol for MDR), but indeed the flaming fires emit relatively more NOx compared to smouldering fires. So the relative is important: maybe good to give emission factors as an example is this stage.

  We agree and have revised the sentence to emphasize that flaming combustion emits relatively more NOx with respect to CO. We note that the flaming–smoldering balance depends not only on fuel type but also on the combustion phase and temporal evolution of the fire. Since the discussion at that point of the manuscript refers mainly to changes during the lifetime of an individual fire, rather than to differences among fuel-specific emission factors, we have clarified the wording without introducing specific emission-factor values here (L86-87 of the new manuscript).

- In L89 "independently" has been corrected.
- In L97 "387 897 ha" has been changed by "0.388 Mha". Similarly done in L98 and L99.
- L99: The sum of these numbers is higher than the 387 897 quoted above which is confusing.

  The first number shown (0.388 Mha) in the original manuscript corresponds to the area burned studied in this work, while the numbers of 0.334 and 0.219Mha correspond to the total area burned within August 2025 in Spain and Portugal, respectively. The sentence has been modified (L99-101 of the new manuscript).

- L103, "Lands" has been corrected.
- Fig. 1: Confusing: 15 areas with different points? Total 46?

  We have now clarified this in the new manuscript by adding L102-105.

- Fig. 2: Points area outside the gray period?

  The gray line in the figure corresponds with the start and end date reported by EFFIS supervised MODIS and Sentinel-2. Note that, as detailed in their documentation (https://effis-gwis-cms.s3-eu-west-1.amazonaws.com/apps/effis.viewer/userguide.pdf, page 23): "The dates reported as

Start date and Last update may not correspond to the date of ignition and extinction of the fire. ". Thus, the grey line in Fig. 2 should be taken as an approximation. For instance, we found that for that specific wildfire (ID 8) the NRT data reports a period expanding from August 10th until 26th , while the supervised data reports the period shown in Fig. 2 (14th August). Although the NRT data contains the fire points, for consistency we show only the supervised data. Note however that this case was the only one we found with this considerable disagreement between NRT and supervised data.

In L114 of the new manuscript we have added: "It should be noted, however, that these dates should be only regarded as indicative of start and end dates of active fires (EFFIS, 2020).". Thus, we have also included the EFFIS (2020) reference to the bibliography. We have also included "approximate" in the caption of Fig. 2.

- Fig. 2: I count 44 "dots" (not 46?).

  The reviewer is right. The reason is that for two different hotspots (ID #10 and #12), there were two overpasses for August 14th. This possibility is now clarified in Fig. 2 caption.

- L155: not correct representation of units? This seems not the standard to me

  All units have been changed from e.g. "g/day" to "g day^{-1}" format throughout the manuscript and supplement, including figures.

- L155: Below you discuss uncertainties, but not for GFAS? Why?

  Note that GFAS does not seem to provide any uncertainties assessment. This is clarified in the new manuscript in L172.

- L161: I guess units is then g/day per gridbox. At least same unit as GFAS (incl m2)

  "per grid cell" has been added (also in Table 2).

- L169: 1x1 km2? Corrected, also in Table 2.
- L192: Is this the case of #8 in figure 2? Yes (now clarified in L209).
- Table 2. Units and temporal resolution.

  Units are now clarified. Since units and temporal resolutions are provided in the table, all units have already that temporal resolution (i.e., per day or per hour).

- Fig. 3: The "background" works well for region 3, but will fail for #7 (&8) because plume #3 is located in the background…?

  We understand RC1. However, that is not the case. By using the upwind method for the background selection, we subtract possible emissions from fires located upwind of target fire. In the example, when studying the emission from wildfire #7, emissions from wildfire #3 are part of the background region of fire #7 and therefore subtracted from the CO fields from #7. This is now better clarified in L232-235 of the new manuscript.

- L260: Include then in the table above. PBLH now included in Table 2.
- L269: Per grid box? For clarity, "per grid cell" and "per km^2" have now been added for GFED and FINN units respectively.
- L273: What do you mean by densities? And "retrieve"? Calculate?

  It has now been corrected "GFAS provides emissions per area".

- L279: Yes, so the only difference is that units are 'per grid cell', while GFAS gives per m2, but to not make this to big!"

We considered useful to distinguish equations for different inventories to give a full comprehension of our methodology.

- L326: look at the recommendations how to present units. Corrected
- Fig. 5c: Here it seems that part of plume in panel (b) is in the background

  As mentioned above, that is indeed to goal of our chosen upwind method for background subtraction of each fire (i.e., to account for possible contributions from fires upwind).

- Caption of Fig. 6: Maybe good to mention that these numbers refer to the 'synergy' method"

  In L315 of the original manuscript, we mentioned that all the CO estimations reported in the main manuscript were computed using the synergy method. We have now clarified this better in L342 of the new manuscript.

- L381: two times the (double as high is incorrect I think). Corrected.
- L394: maybe some reference to Table 3 would help the reader to see some of the unlying causes.

  Thank you for the suggestion. We have added L450 int the new manuscript.

- L400: "Although maybe not the right word here? Next to the moderate correlation, there is also a bias? Corrected
- L409: GFFEPS acronym. Now "Global Forest Fire Emissions Prediction System" has been added as definition to GFFPS acronym (L471). Similarly, to assess RC1 in L424 of the original manuscript, in L492-493 of the new manuscript we have added Quick Fire Emission Database (QFED) and Global Biomass Burning Emission Product (GBBEP).
- L413: Indeed the 13:30 overpass time of TROPOMI came directly to my mind! This is maybe the "most active period" for fires, while inventories report "diurnal averages" (actually, do they? please add information in the relevant sections).

  Indeed, the temporal resolution of the inventories used in our work are daily means (as indicated in Table 2 of the original manuscript). That is the reason behind our comment (L414 of the original manuscript) "Although their specific explanation referred to GFFEPS, the broader issue is also relevant to this work, since TROPOMI estimates are based on conditions close to the overpass, whereas the inventories used herein provide daily emissions or temporally distributed estimates that must be related to the overpass time for direct comparison." Please, refer also to our response to the major comment 2 (temporal sampling) of anonymous referee 2, where we clarify that the TROPOMI estimates should be interpreted as overpass-centred, short-timescale emission estimates, whereas the inventories provide daily emission estimates. See also L477-484 of the revised manuscript.

- Title of Fig 11. The figure title has been changed.
- L491-492: The maximum in the TROPOMI derived NOx/CO is however only 20: so what is the value of this statement?

  We agree that the original wording could be misleading. We did not intend to suggest that the maximum inventory MDR is too low relative to TROPOMI. We meant that the inventories MDR seem to show an upper limit. We have now deleted "Fig. 12 also shows an upper plateau in the inventory ratios, even for FINN and GFED. Several fires reach similar maximum inventory NOx/CO values, whereas the TROPOMI-derived ratios vary

more continuously. This behaviour likely contributes to the moderate correlations and suggests that the upper range of biome- or fire-type-specific emission ratios used in the inventories may be too constrained or not fully representative of the observed combustion variability.“. Instead, we have added L699-701 on the new manuscript.

- Titles in Fig. 13: POI? POI (point of interest) has been changed to “hotspot ID” (Table 1).
- L526: From the CO emissions (in panels (a) and (b)) it seems that magnitudes differ specifically during peak events

  We agree that the largest differences in absolute CO emission rates seem to occur mainly during the peak emissions, as already discussed in the original manuscript in the preceding paragraph (L510-515) for both Jarilla and Arganil. We did not intend to attribute the lower TROPOMI-derived MDR specifically to enhanced smoldering combustion (i.e, relatively higher CO than NOx), but rather to mention this as one possible interpretation. We have now clarified this in the new manuscript L565-567.

- L527: Well, this sounds like a long stretch! If you claim this (certainly interesting), you should try to obtain information on the fire-fighting activity in this period. Now it sounds like an ad-hoc argument.
  We agree that the original wording was speculative. Our intention was to point out that the evolution of large wildfires is not controlled only by fuel characteristics and meteorological conditions, but can also be affected by firefighting intervention, a factor that is not explicitly represented in the emission inventories considered here. However, we agree that linking firefighting activity specifically to the lower MDR observed around the TROPOMI overpass would require event-resolved information on the timing and intensity of those operations, which is not available in the present study. We have therefore removed this specific phrase and added L605-610 on the new draft: “Firefighting intervention may also influence the temporal evolution and combustion conditions of individual fires. Such activities are not explicitly represented in the emission inventories used in this work, although their effects may be indirectly reflected in the satellite fire observations on which those inventories rely. In contrast, the atmospheric columns observed by TROPOMI at the overpass time integrate the actual state of the plume and therefore may also reflect any preceding influence of firefighting activities on the fire emissions. However, the effect of firefighting activities cannot be evaluated with the information available.”
- L536 (plume chemistry): this I have not seen being analysed…what would be the effect on plume chemistry on NOx/CO MDR?.

  Regarding plume chemistry, the effect of the lifetime of NOx on MDRs has now been assessed. Please refer to our responses to your major comment 3 and to Sect. S3 of the new supplement material. See also L545-560 of the revised draft.

- L552: with reference to figure 13 it is fair to remark that specifically the inventories seem to miss peak events

  Thank you for the suggestion. L634 is now added the new manuscript.

- L583: For NO2 there is an issue with the averaging kernel. There might be a bias because the AK used is not fire-aware (i.e. has a too low NO2 abundance in the prior profile). Thus, in reality, NO2 columns might be higher.

  We agree, each gas may have independent retrieval issues. The sentence has now been modified in L668 to “…but also to analyse complex emission scenes by taking advantage of their different chemical lifetimes and complementary spatial information.”. Please refer also to our response to your major comment 1.

Minor comments on the Supplementary's PDF:

- Sect. 1.1.1: Here uncertainty in TROPOMI should maybe be larger (aerosol, NO2 AMF, ..)

  Now investigated in the new supplementary information (Sect. S1.2)

- Sect. 1.1.1: PBL to PBLH. Corrected
- Sect. 1.1.3: "We compute the average VCDs for the whole region..."?? In the new version of the Supplementary (Sect. S1.4), we have now described better that sensitivity test.
- Fig. S2: Here, some more context is needed. The "orange" points are quite a bit higher than the "mean" green dashed line. So, logically, the enhancements are larger. But the "orange" points are often also higher than the "blue" mean values on that day. Does this mean that plumes are indeed "stacked" and that the background gets high due to one fire and then another fire adds CO and NOx to it?

  We clarify that the orange crosses in Fig. S2 of the original Supplementary text (Fig. S3 in the revised one) are background CO VCD values, not plume enhancements. As mentioned above, that Section S1.4 of the new supplement has been modified to explain better our results.

- Sect. 1.1.4: "we used the nearest wind field from ERA5 to the hotspot". wind field nearest to the hot spot. Corrected
- Sect. 1.2.1: the threshold (what is exactly the threshold method??. I tend to loose track here: I read (1) threshold method, the standard method, and the synergy method. In the end to differences are not 100% clear after reading this.

  Please, refer to our response to your major comment 2.

- Sect. 1.2.2: "make easy to isolate" to "facilitate easy isolation of the plume". Corrected.
- Sect. 1.2.2: "lower CO emission rate values (10%)". Is this due to a lower background? Or due to slightly different plume areas?

  As shown in Fig. S7 of the original Supplementary material, the synergy method generally selects fewer plume pixels than the standard threshold method because the synergy plume mask is constrained by the shorter-lived NO_2 field. Now clarified in L190-192 of the revised Supplementary information.

- Sect. 1.2.2: "is expected closer" Sharper gradients are expected for NO2 close to the hotspots. Corrected, as well as the other corrections suggested and typos found by R1 in that section.
- Sect. 1.2.2: "NO2 density" It would be instructive to also show the NO2 plume (i.e. the red in panel (b)?) Included (now it is in Fig. S9 of the revised Supplementary information).

Reviewer2

The authors would like to thank anonymous referee #2 (R2) for her/his time. After both referees' inputs and comments, we believe that our draft has substantially improved. In response to R2, we have added sensitivity analyses on background selection, ERA5 wind representation, spatial resolution and the NO_2-to-NOx conversion, and expanded the discussion of temporal sampling and plume masking. Please, find our point-by-point responses below. R2 comments are shown in black and our responses in blue. Also, note that when referring to line numbers in the revised manuscript or Supplement, we use those from the clean versions (without tracked changes) as, depending on the settings, Word may change the line numbering when tracked changes are displayed.

The manuscript quantifies CO emission rates and NOx/CO mole density ratios (MDRs) for 46 individual TROPOMI measurement points spanning 15 major wildfire events during the August 2025 Iberian fire season, and compares these top-down estimates with three fire emission inventories: GFAS, GFED, and FINN. The authors developed a synergy plume-masking approach that uses TROPOMI NO_2 plumes to spatially constrain the CO plume mask in complex, multi-fire scenes. The results show that inventories underestimate CO emissions by approximately a factor of three and that TROPOMI-derived NOx/CO ratios are systematically lower than inventory-based ratios. The approach and findings are of interest; however, several methodological choices require clearer justification or more detailed description before the reported "factor of three" difference and MDR results can be fully assessed. I provide detailed comments below.

Major comments

1. TROPOMI observes plumes at a much finer spatial resolution than GFAS and GFED. Although the manuscript sums inventory emissions within a fixed area around each fire, it does not discuss whether differences in spatial resolution contribute to the scatter in the results. Coarse inventories can mix emissions from nearby fires within the same grid cell, whereas the TROPOMI plume-masking method is designed to separate individual plumes. This may partly explain why TROPOMI–inventory correlations are weaker than inventory–inventory correlations. A sensitivity test, or at least a discussion of this issue, would strengthen the analysis

As detailed in Table 2 of the initial manuscript, each inventory has a different spatial resolution (GFED 0.25º x 0.25º, GFAS 0.1º x 0.1º and FINN 1km x 1km), from coarser to finer resolution compared to TROPOMI (5.5 x 7 km^2). However, the TROPOMI-derived factors with regard to all the inventories were found to be approximately 3 (2.9 for FINN, 3.4 for GFAS and 3.3 for GFED). Note that the correlations found (Fig. 9 of the initial manuscript) were ~0.48 for FINN (with finest spatial resolution) and GFED (with coarsest resolution), very similar to the correlation e.g. GFED-GFAS (Fig. 7a). These results suggest that although, as suggested by R2, the spatial resolution may indeed be a source of the rather low TROPOMI-inventory correlations found in our work, it seems unlikely that is the reason for the systematic x3 factor found. This is now included in the new manuscript (L462-469) and in Sect. S4 of the new supplement.

2. Related to the previous comment, the manuscript does not clearly explain that the quantities being compared are fundamentally different. TROPOMI provides an instantaneous emission estimate at the satellite overpass time (~13:30), whereas GFAS, GFED, and FINN provide daily average emissions derived from bottom-up

methods. Part of the reported factor-of-three difference between TROPOMI and the inventories may therefore be due to differences in temporal aggregation. A short explanation of these differences and their potential impact on the comparison is recommended.

Indeed, this is a limitation of comparisons involving observations from polar-orbiting instruments such TROPOMI. This issue was acknowledged in Sect. 5.1.3 of the original manuscript, where we noted that the conclusions of Griffin et al., (2024) are also applicable to our study, i.e. that part of the differences between inventory and TROPOMI derived emission estimates may relate to differences in temporal sampling. Moreover, as a caveat, in the original draft we stated (L414) that "TROPOMI estimates are based on conditions close to the overpass, whereas the inventories used herein provide daily emissions or temporally distributed estimates that must be related to the overpass time for direct comparison". We agree with the reviewer that this limitation and its potential implications for the comparison should be made more explicit. This is now done in the new manuscript L478-485: "The inventories used in our work only provide daily mean estimates. However, since the integrated plume mass observed by TROPOMI (Eq. 2) contains emissions released during transport prior to the overpass over a characteristic timescale of approximately L/u. Thus, our TROPOMI-derived emissions should be interpreted as overpass-centred, short-timescale emission estimates. If fire activity around the TROPOMI overpass (~13:30) is not representative of the corresponding daily inventory estimate, this temporal mismatch may contribute to both the scatter and part of the systematic TROPOMI–inventory difference. However, the similar x3 offset found independently for GFAS, GFED and FINN suggests that differences in the temporal sampling are unlikely to be the only explanation for the discrepancy shown in Fig. 9."

3. The choice of the 0.8° × 0.6° upwind background region is not fully justified. It is unclear why this size was selected, whether it was applied to all cases regardless of plume conditions, and whether sensitivity tests were performed. In complex fire scenes, the background region could contain smoke from nearby fires, potentially biasing the results. A sensitivity analysis and discussion of this uncertainty would be valuable. The authors should also explain why they chose the upwind method instead of the alternative approaches cited in the manuscript.

As mentioned in L211 of the original manuscript, "the background region is taken as a 0.8° x 0.6° square shape, similar to the work of Griffin et al., (2021, 2024).". However, the background choice is a very interesting comment also put forward by R1. Please, refer to our responses to R1 for detailed background selection sensitivity tests (Sect. S1.4 of the revised Supplementary information). In brief, our studies indicate that changing the size of the upwind background region produced little systematic effect, whereas using a fixed pre-fire regional background resulted in systematically higher CO emission estimates. This is now mentioned in the revised manuscript in L380-384.

As for the selection of the background method, we chose the upwind background method because we specifically want to know the composition of the air parcel before it reaches the target fire. Also, because this method will incorporate in the background subtraction possible contributions of upwind fires. This is now clarified in L232-235 of the new manuscript.

4. The use of ERA5 winds at 0.25° resolution may introduce uncertainty because these winds may not accurately represent plume transport, especially in mountainous terrain. Since wind speed is already identified as the largest source of uncertainty,

the manuscript should discuss wind representativeness and plume injection height more explicitly.

In our study, plume transport is represented by ERA5 winds averaged from the surface to the planetary boundary layer height (PBLH). This choice provides an effective transport wind representative of the boundary layer, where most of the observed smoke is expected to be transported. However, we agree with the reviewer that the effective transport level may vary between fires and also that the complex terrain might be even more complicated and hardly resolved in 0.25º resolution. Indeed, depending on e.g. the fire intensity or atmospheric stability, the smoke might be concentrated in the lower or upper part of the PBL or extends above it. This effect is now investigated in Sect. S1.5 (Fig. S6) of the new Supplementary information. In general, results show only a limited systematic effect of the selected vertical averaging layer on the retrieved CO emissions (~10%), although extending the averaging layer above the PBLH increases wind variability and the uncertainty of the CO emission estimates for some individual cases. These results are also reflected now in the revised manuscript (L375-379).

5. The choice of qa > 0.5 for TROPOMI data is not fully justified. Although the manuscript explains that this threshold increases the number of observations near fires, it does not explain why 0.5 was selected instead of the more commonly used qa > 0.75 threshold.

In our work (L146 of the original draft) “The quality assurance flag (“qa value”) used herein is set to higher than 0.5 in order to increase the number of observations close to fire hotspots, while still removing low quality data (Adams et al., 2019; Griffin et al., 2021; Wan et al., 2023; Griffin et al., 2024).”. Also, in Sect. 1.1.2 of the original supplementary information, we tested our results with qa > 0.5 to those obtained if setting qa > 0. If, on the other hand, we apply the suggested standard qa > 0.75, the number of analysable fire cases reduces from 46 to only 6. We therefore used qa > 0.5 to retain sufficient observations near the fire plumes while excluding low-quality retrievals. Note also that our estimates do not correlate with qa value used. We have clarified this point in the revised manuscript L160-164. Also, we now provide as supplementary asset (L687 of the revised draft), a data table providing information of retrieval details of each of the fires investigated, including the qa values.

6. The plume-masking method is the main contribution of the paper, but the description in the main text remains largely qualitative. Important details, such as threshold values and the procedure used to match NO_2 plume pixels with CO pixels, are only described in the Supplement. A more complete description in the main text is recommended.

Sect. 4.1.2 in the main text and Sect. S1.2 in the supplementary have been rewritten in order to better introduce both plume masking methods (i.e., standard threshold and synergy). We still keep the more detailed formulation and assessment of the threshold method in the Supplementary information in order to maintain the clarity and readability of the main text.

Note also that, in the newly provided supplementary data table referred to above, we also provide additional retrieval information such as plume length, threshold value (n), PBLH, wind speeds, ERA5 coordinates, polygons defining each background region, mean NO_2 and CO in the background and the inferred CO emission rates.

Minor comments

1. The study uses a fixed conversion factor of 1.3–1.5 to estimate NOx from NO_2. Since this factor is important for interpreting the MDR results, a sensitivity analysis using a wider range of values would help demonstrate the robustness of the conclusions.

This is now assessed in Sect. S3 of the revised Supplementary Information by expanding the NO_2-to-NOx conversion factor from 1.3–1.5 to 1–2. This wider range increases the spread of the TROPOMI-derived MDRs, but does not alter the main conclusion that they remain generally lower than the inventory-derived values. We added also a reference to this regard in L555-567 of the revised manuscript.

2. Table 1 provides only a single latitude and longitude for each fire. For reproducibility, it would be helpful to provide the actual spatial extent used in the analysis, such as bounding boxes or fire polygons.

The afore mentioned new Supplementary data provided contains that sort of information.

3. The Abstract refers to NOx/CO MDR, although TROPOMI measures NO_2 rather than NOx directly. For clarity, consider using a term such as "NO_2-derived NOx/CO MDR."

We thank the reviewer for pointing this out. Indeed, Sect. 4.3 states that TROPOMI measures NO2 rather than NOx directly. The TROPOMI NO2/CO ratio is converted to an approximate NOx/CO MDR using the literature-based NO2-to-NOx conversion factor. We therefore prefer to retain the term "NOx/CO MDR" for the derived quantity, including in the title, as this is also the quantity compared with the inventories. To avoid any possible ambiguity, we have revised the Abstract in the new manuscript to explicitly state that the satellite-derived NOx/CO MDR is estimated from the observed TROPOMI NO2/CO ratio (L23-24). This is also referenced in L89.