

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 NO₂-to-NO_x 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 NO_x/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 NO₂ 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 NO_x/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 0.25° x 0.25°, GFAS 0.1° x 0.1° and FINN 1km x 1km), from coarser to finer resolution compared to TROPOMI (5.5 x 7 km²). 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 as 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 ($\sim 13:30$) 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 $\times 3$ 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 $0.8^\circ \times 0.6^\circ$ 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 $0.8^\circ \times 0.6^\circ$ 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 0.25° 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 NO₂ 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 NO₂ 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 NO_x from NO₂. 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 NO₂-to-NO_x 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 NO_x/CO MDR, although TROPOMI measures NO₂ rather than NO_x directly. For clarity, consider using a term such as “NO₂-derived NO_x/CO MDR.”

We thank the reviewer for pointing this out. Indeed, Sect. 4.3 states that TROPOMI measures NO₂ rather than NO_x directly. The TROPOMI NO₂/CO ratio is converted to an approximate NO_x/CO MDR using the literature-based NO₂-to-NO_x conversion factor. We therefore prefer to retain the term “NO_x/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 NO_x/CO MDR is estimated from the observed TROPOMI NO₂/CO ratio (L23-24). This is also referenced in L89.