We would like to thank Reviewer 2 for their second review. In response to the feedback from both reviewers, we have implemented significant revisions to the manuscript. Our answers are highlighted in red font, while the reviewers' comments are in black.

Here I share an anonymized and sometimes paraphrased summary of Reviewer 2's comments to the editor on the revised manuscript, which I think would be relevant in another revision round. I would be interested in author responses to these comments if a revision is pursued.

Reviewer 2 expressed overall concern that the revised manuscript failed to demonstrate the manuscript is ready for publication.

There are concerns that language in the abstract had not been revised to avoid repetitions and typos, and for sentence integrity, as requested in the first round of reviews, e.g.:

"using inputs such fuel type",

"wildfire emissions have been demonstrated to be estimated directly".

The reviewer strongly suggests more effort should be made to improve the writing and readability of the manuscript throughout.

We have made the following changes the abstract to address these concerns:

"These emissions can be estimated by a bottom-up approach that relies on fuel consumed and standardized emission factors. Emissions are also commonly derived with a top-down approach, using satellite observed fire radiative power (FRP) as proxy for fuel consumption. Wildfire emissions can also be estimated directly from satellite trace gas observations, including carbon monoxide (CO)."

Reviewer 2 also expressed that many vague statements are included instead of quantitative statements, e.g.:

From the abstract: "The CO emission totals derived from satellite data align reasonably well with those from bottom-up emission inventories for various global regions. However, notable discrepancies are evident in specific regions, such as Southern Hemisphere South America, Southern Hemisphere Africa, and Southeast Asia". According to this, the comparison is poor (but we don't know how poor) for the regions with the most fires and, thus, the most CO emissions. The reviewer therefore has concerns regarding how well the results from this study represent actual emissions.

Specifically: "...emissions [] align reasonably well with ...". "Reasonably well" is not informative; difference values, either absolute or relative, should be provided. Same for "notable discrepancies".

In the few instances where the manuscript was modified to correct this issue, the reviewer expressed concern that the new wording is often still vague (e.g., "Many of the fitted emissions" becomes "a substantial portion of the retrieved emissions", "are very close" becomes "closely matches"). The reviewer desired more effort to use quantitative statements rather than vague, subjective language.

We included the changes as suggested.

Line 2: "global CO wildfire emissions have, on the whole, decreased". Line 30: "Given the increase in fire intensity and number of fires". Readers may find these statements contradict each other but no attempt is made to address the contradiction.

The sentence about increased intensity and number of fires was talking about fires in North America. To avoid confusion we removed the sentence in question.

The reviewer shared several specific comments on using TROPOMI retrievals with qa_value <1 without applying the averaging kernels (AK) follow:

From the responses: “We also found a recent publication by Rowe et al., 2022 (https://pubs.acs.org/doi/full/10.1021/acsearthspacechem.2c00048) that investigates the uncertainty of the TROPOMI CO observations in smoke, we added the reference to our manuscript, showing approximately 10% higher CO than the aircraft measurements. This uncertainty is consisted with the uncertainty assumed in our error budget analysis (Table 1).” Rowe et al. (2022) found that TROPOMI data was higher than aircraft data by 36%. Only after correcting the TROPOMI data to account for transport and applying the AK did the TROPOMI positive bias go down to 10%.

The statement “The CO averaging kernel from the TROPOMI observations predominantly registers values close to 1 within the boundary layer” as used in the response and in the revised manuscript is incorrect. Plots in the TROPOMI L2 user manual for CO shows that the AK values within the boundary layer depart from 1 when clouds are present; several peer reviewed articles include similar plots illustrating this concept.

“the presence of clouds diminishes the sensitivity of the averaging kernel beneath them” This language could use refining. The reviewer notes that AK themselves have no sensitivity, they indicate/represent/describe sensitivity. The sentence “attempting to correct the averaging kernel in regions near fires would introduce additional uncertainty to the analysis” further suggests that the AK concept is not well understood, or the authors need to clarify what they mean. The reviewer argues that AK themselves are not to be corrected; by applying the AK to the TROPOMI retrievals it is the retrievals that are corrected. Rowe et al. (2022) findings show that correcting the TROPOMI by (among other things) applying the AK reduces the TROPOMI positive bias by 26% (please see above); it does not introduce additional uncertainty.

“Consequently, including observations with a quality flag of 1 would result in the exclusion of a substantial number of data points, primarily due to the influence of smoke.” Can the authors confirm their language here? The authors say that “including data” will result in the “exclusion of a substantial number of data points”. Is the wording incorrect?

“Thus, the averaging kernel is not considered in this study, but is taken into account in the overall uncertainty (uncertainty of VCDs) of the emission estimate (see Table 1). Rowe et al. (2022) investigated TROPOMI CO in thick fire plumes and found agreement within 10 % with the aircraft observations, which has been used to estimate the overall uncertainty of the emissions.” The reviewer expresses concern that this statement is incorrect, since the Rowe et al. 10% agreement is based on applying the Aks.

Overall, the reviewer expresses concern that the revisions failed to demonstrate that not applying TROPOMI averaging kernels to retrievals with qa_value < 1 is the best (or at least acceptable) approach. The reviewer expressed their opinion that applying the AK in those cases is advised in the TROPOMI CO documentation and is standard practice in studies described in peer reviewed articles.
Could the authors please summarize a response/rebuttal to the reviewer's comments on TROPOMI AK usage here? Is there a misunderstanding in how the data has been assessed and processed that can be resolved?

We examined the averaging kernel and the effect of it more closely. Rowe et al (2022) used aircraft measured profiles to examine the impact of AVKs. Other than these we do not have profiles inside the fire plumes that can be applied to correct the AVK. We were looking at different averaging kernels and different profiles, see figure below, depending on the model profile that is applied to the averaging kernel the VCDs change: The stringer the enhancement near the surface the higher the VCDs are, for a plume with an enhancement of 0.4ppm near the surface the impact on the VCDs are small (0 if the the averaging kernel is 0.7 near the surface and 0.03 mol/m2 if the averaging kernel goes to 0.2 near the surface), however, if the enhancement in the applied profile is larger (shown is 1.4 ppm) the VCDs are approximately 0.1-0.14 mol/m2 higher than the original. The profiles are taken from a GEM-MACH model run (10x10 km) with fire emissions (GFFEPS), same fire as in Figure 5 (a) which has a great overlap with the TROPOMI observed plume. Note that this GEM-MACH model output of fire plume profiles is not available globally at present.

Rowe et al. applied the averaging kernel and examined the effect using aircraft observations and aircraft observed profiles (which are not available for all fire globally). In his paper he notes that the averaging kernel improves the agreement but to a lesser effect than other (such as FLEXPART): " When the additional corrections are applied (FLEXPART, and to a lesser degree also AVK),". Table 3 give more insight into the specific effect of the comparison between TROPOMI VCDs and aircraft observations showing that the difference without considering averaging kernels (but including other correction such as viewing geometry) leads to 13-16% difference between the aircraft and the TROPOMI CO VCDs. Based on this we increased the uncertainty due to the TROPOMI VCD to 17% which made the total uncertainty

44% instead of 42%.

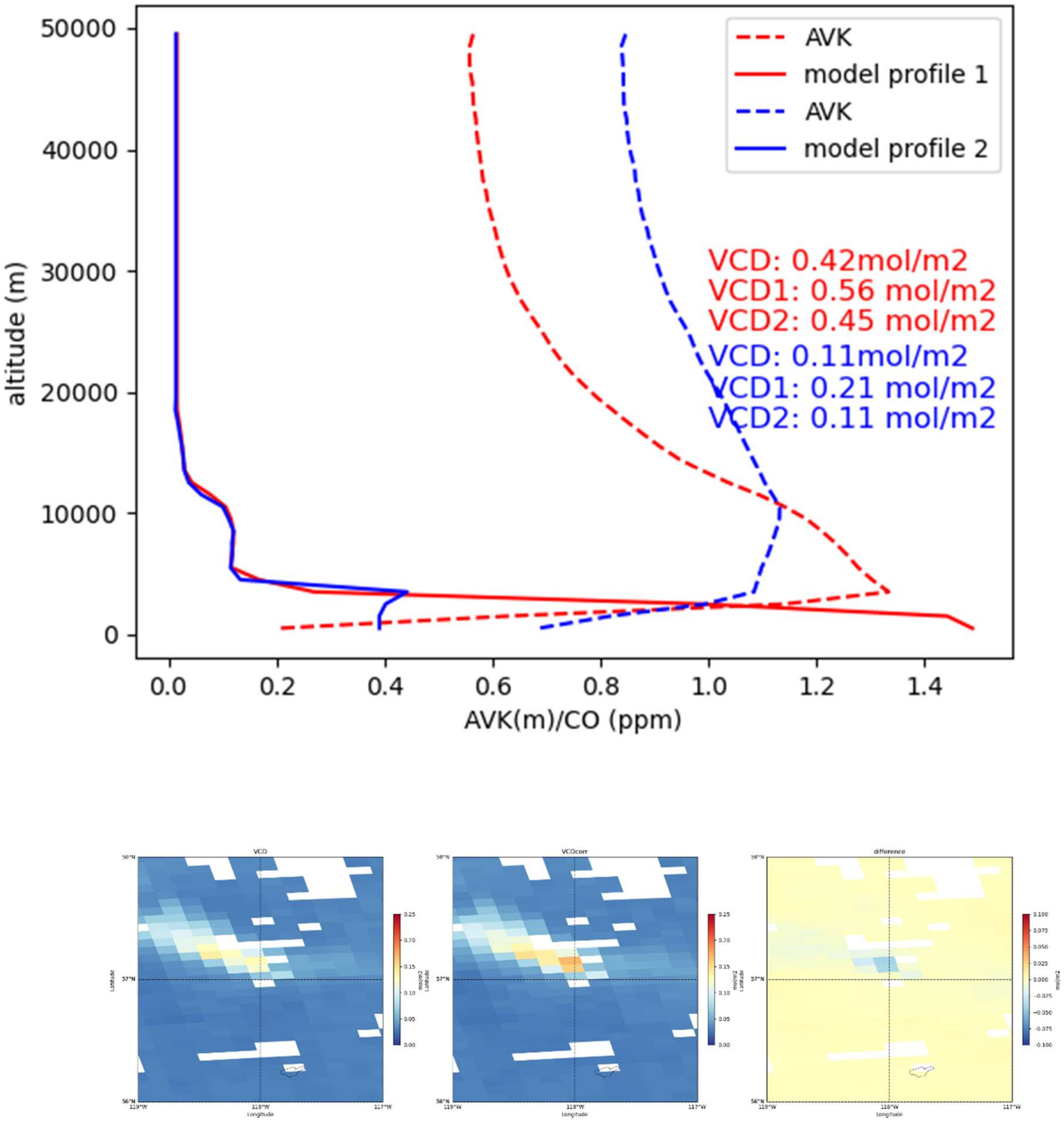

Figure B1. Effect of applying the TROPOMI averaging kernel for a fire that overlapped well with the GEM-MACH model (using GFFEPS fire emissions). The averaging kernel corrected columns are estimated using the corrected averaging kernel AVKcorr = int(AVKz · Nz dz)/int(Nz dz) that is then applied to the CO colummns: VCDcorr = VCD/AVKcorr, where Nz is the GEM-MACH CO profile. The original CO columns are shown on the left panel, the corrected ones in the middle panel, and the difference (VCD-VCDcoor) on the right panel. The emissions were estimated with the original and corrected VCDs and show an increase of 17 % when the corrected VCDs are used. The example shown is for the same fires as in Fig. 5 top panel (May 21, 2019 at 57◦N and 118◦).

Section 2.1:

“Consequently, only including observations with a quality flag of 1 would result in the exclusion of a substantial number of data points, primarily due to the influence of smoke (we found most pixels inside smoke plumes have a quality flag value of 0.7). The CO averaging kernel from the TROPOMI observations predominantly registers values close to 1 within the boundary layer for cloud-free conditions, specifically around 0.95 with a narrow range of variability (approximately ±0.05) (Schneising et al., 2020). Nevertheless, the presence of clouds diminishes the sensitivity of the averaging kernel beneath them. It is important to note that smoke is primarily composed of fine particles, 0.25 um or smaller. (~1 µm). At a wavelength of 2.3 µm these small particles scatter minimal light. Looking at the TROPOMI averaging kernel, we found that in case of fires the sensitivity close to the surface is typically lower than 1. Rowe et al. (2022) investigated TROPOMI CO in thick fire plumes and found agreement within 13-16 % (Table 3) without considering the averaging kernel which has been used to estimate the overall uncertainty of the emissions. The effect of the averaging kernel depends on 1) the shape of the averaging kernel and 2) on the CO profile, looking at different profiles and averaging kernels, we found the largest effect is for an averaging kernel that is close to 0 at the surface and for a strong enhancement of the CO profile, the magnitude of this enhancement determines the magnitude of the averaging kernel corrected columns. We found that the effect that applying the averaging kernel inside the smoke plume always increases the CO columns. Other than (Rowe et al., 2022) (who investigated fires during FIREX-AQ) we do not have profiles (globally) that can be applied for an averaging kernel correction. Testing the effect on the emissions we used GEM-MACH model profiles and applied the averaging kernel correction (see Appendix Fig. B1) which showed a 17 % increase of the emissions, and thus we attribute 17 % uncertainty due to the uncertainties of TROPOMI CO columns (the uncertainties are further discussed in Sect. 2.3).”

“Geostationary satellite sensors, such as TEMPO (covering North America), Geostationary Environment Monitoring Spectrometer (GEMS), or Sentinel-4 (covering Europe and Africa) will help to validate the diurnal pattern of emissions.” The type of emissions should be clarified. Please note that neither TEMPO nor GEMS measure CO and CO is not among the species to be measured by Sentinel-4.
We have included the following to the sentence:
“…will help to validate the diurnal pattern of emissions (for example for NO2 and HCHO, note that none of the current geostationary satellites have the ability to measure CO).”

“This discrepancy is likely due to an underestimation from GFFEPS rather than an overestimation of TROPOMI emissions [...]” Rowe et al. (2022) showed that TROPOMI retrievals overestimate CO from fires by 36% (10% corrections, including AK correction). That could arguably result in an overestimation of TROPOMI emissions.
See averaging kernel discussion above -> applying the AK leads to higher total columns and higher emissions.

The reviewer also noted some specific comments about how emission coefficients (EC) are derived from TROPOMI and MODIS:

- "Using emission coefficients from 2019,2020, and 2019-2021 combined did not impact the total emissions (see Fig. C1), only for 2021 the total emissions reduced by approximately 20 %" Fig. C1 does not support this statement: SHAF (South. Hem. Africa), depending on what year is used as the reference (either 2019, 2020, or 2021), annual emissions vary by > 20 Mt, a 25% departure from average values. NHAF (North. Hem. Africa) vary by 20 Mt, or 40%. Global emissions vary by ~ 100 Mt, or 33%. The reviewer expressed concern therefore that contrary to what is stated in the revised manuscript, the magnitude of the emission coefficients will vary depending on what year is selected to calculate them.
Also included other biomes in addition to GLC2000 , the effect is xx. The uncertainty is expected to be 40% based on the variation on the emission coefficients thus all the values previously stated are with the estimated uncertainties.

- Focusing on the emissions coefficient (EC) calculated for biome #1: Tree cover, broadleaved, evergreen. This biome is 1) one of the most extensive by area and 2) covers regions with the most CO emissions from fires, e.g., Amazon, Equatorial Africa, Indonesia, etc. Depending on the year considered, the EC calculated from TROPOMI and MODIS for biome #1 goes from 150 (2019) to 101 (2020) to 59 (2021), according to Tables C1, C2, and C3. That is an absolute change in EC=91 or a relative change in EC=88%. This shows that TROPOMI-derived EC values are highly dependent on what year is used to calculate them. The opposite is argued in the manuscript over the years, we conducted an extensive analysis of the entire time series and calculated CO emissions spanning from 2003 to 2021")
The latter comment may again be a question of using more qualitative language when quantitative language would be preferred. What do the authors mean by "relatively stable over the years?". I think the reviewer is wondering what metric is used to assess stability, given there is a reasonably large range in interannual EC.

Based on this concern we included the following discussion:

"To assess the uncertainty of the total annual emissions of our estimates (TROPOMI-FRE), we also used emission coefficients derived from fires of individual years (2019 to 2021, using the GLC2000 classification). Using emission coefficients from 2019,2020, and 2019-2021 combined did not impact the total emissions for individual regions or globally (see Fig. E1). 2021 seems to be the anomaly for which the total global emissions reduced by approximately 20-25 % , due to overall lower EC_CO (for biomes 1-3, see Table E3 which affected the SHAF region the most and almost halved the emissions in 2021 compared to using the ECs from other years). This shows that the uncertainty of our approach is at least 25 % for the global emissions but for individual regions an uncertainty of 50 % should be assumed. The impact of using ECs from individual years is greater than the impact of using different definitions of biomes. It should be noted that the uncertainty discussed here is only due to the ECs, the uncertainty due to GFAS FRE (not provided) does not go into this estimate thus the total uncertainty of the TROPOMI/FRE top-down emissions is expected to be higher. For most wildfire emissions inventories an uncertainty of a factor of two is assumed (Pan et al., 2020; Wiedinmyer et al., 2023)."

We would like to thank Reviewer 3 for their review. In response to the feedback from both reviewers, we have implemented significant revisions to the manuscript.
Our answers are highlighted in red font, while the reviewers' comments are in black.

Review of manuscript "Towards an improved understanding of wildfire CO emissions: a satellite remote-sensing perspective" by Griffin et al.

Revised version.

This manuscript describes a methodology to evaluate CO emission coefficients (EC) that may be used to estimate CO biomass burning emissions from observations of the fire radiative energy (FRE). This method relies on an inversion of CO emissions from TROPOMI satellite observations of total columns CO for a large set of fires detected in different biomes.

A first section describes the methodology for CO emissions inversion as well as an analysis of the corresponding uncertainty. The derived EC are then presented. Finally, the method is applied to the 2003-2021 time period based on the GFAS FRE database.

This study brings a valuable contribution to the efforts on improving the estimates of biomass burning emissions, for which the uncertainty is still estimated to be about a factor of 2. The approach chosen is original because it targets ECs and not just final emissions, so that it may be used in other emission inventories and time periods not covered by CO observations. I think this work is worth publishing in ACP, and that the authors did a significant effort to improve the manuscript and add new material to respond to the reviewers' comments. However, I also think that major revision is necessary before publication, which I'm confident the authors will be able to address.

Major comments.

I think the title should be revised to be closer to the added contribution of the manuscript, i.e. information on CO emission coefficient derived from TROPOMI observations. It is not clear if emissions are improved, which does not prevent the manuscript from making an interesting contribution.

The title has been changed to reflect this suggestion: "Forest Fire CO Emissions: Exploring Insights through TROPOMI-derived Emissions and Emission Coefficients"

Emissions inversion:

Section 2.3, uncertainty evaluation using synthetic data:

The authors were able to construct a database of 208 fires between May and September 2019, and only 105 remain after the filtering. It's not clear to me how these fires are representative, especially for regions in which the fire season is not during boreal spring and summer and for large fires (both very low wind and high wind may be favorable to large fires, e.g. during heatwaves).

Several criteria are listed for successful retrieval (l. 230—240 & FRP > 1000MW): are all inversions successful if these criteria are met?

All of section 2.3 is based on synthetic columns to test the accuracy of the direct emission estimate. This is done to determine the uncertainty of the method. The model is available at a resolution of 10x10km only in North America that is needed to be comparable to the satellite resolution. These are all synthetic and do not go into any estimates in the later sections. The sole purpose is to estimate the uncertainty of the method. For the emissions we apply the direct emission estimate and then apply the filter. Specific number and more details have been included in the manuscript in Sect. 2.4:

“For a typical year using the described MODIS clustering, we are left with approximately 13-18 thousand fire clusters globally for which we attempt an emission estimate. For about 3-4 thousand fires the estimate fails entirely. And further 9-12 thousand fire emissions are filtered due to poor quality, leaving 4-6 thousand successful fire emissions globally per year.”

Section 2.5, evaluation top-down vs bottom up:

In this section, the authors state that about 5000 fires are analysed (l.336); but approx. 4000 fires according to the legend of Fig.4… What is the exact number?

How many fires were detected during that time period, in total, and how many above the threshold of 1000MW (which is already very high)? What fraction of fire detection may be analysed using this method?

1000MW (1GW) is the threshold using the sum of all hotspots of the fire cluster and it’s not a very large number; to make this clearer we included the term “fire cluster” in the manuscript. As can be seen in Fig.6 fires with FRP<1GW (=1000MW) are not really relevant for the estimate of the emission coefficient (scale is from 1 to 100GW). I would also like to point out that the minimum number of hotspots is set to 5. Reducing the minimum FRP threshold or number of hotspots will lead to more failed fires as those are likely the TROPOMI detection limit, or false positives such as flaring from industry). We included the following in the manuscript to make this clearer:

Sect. 2.4:

…”1 GW (note, this threshold is not for individual hotspots but for the fire cluster) and a confidence of at least 75 % (for individual hotspots).

“**EC estimates:**

This estimate relies on a landuse map to attribute specific fires to different biomes. For this purpose, they use the GLC2000 database. How is this quite old land use map representative

of vegetation in 2019-2021? I think this adds serious doubts to the results presented since it has been demonstrated that vegetation attribution is also a very large contribution to uncertainties in emission estimates (e.g. Turquety et al., GMD, 2020). There are many more recent land use classifications that may be used (e.g. from MODIS). : *The reviewer is correct in stating there are other, more recent land use classifications than GLC2000. The decision to use GLC2000 was to remain consistent with that used by GFEPPS . We needed a global land use of sufficient resolution that was easy to employ (for proof of concept) and GLC2000 was well suited. It provided single map global coverage at a 1-km resolution. We considered the reviewers concerns and added the emission coefficients using MODIS land classification (0.05 degree resolution), as well as GFED land classification (0.25 degree resolution, and only 6 types of vegetation).*

In terms of timeliness, this was seen as less critical as vegetation changes relatively slowly and most land use changes, whether they were a result of disturbance (fires, deforestation) or urbanization, would result in landscapes less fire prone and would be reflected by a reduced number of hotspots.

We modified section 5 (Figure 8, and text) to include the estimated emissions using the emission factors from different biome classifications (GLC2000, MODIS, GFED).

“Section 4:” To differentiate biomass burning emissions in different biomes we use the GLC2000 (European Commission, 2003) as used for the development of GFFEPS, additionally we also tested the classification from MODIS (MCD12C1) and the GFED 425 partitioning.

[...] The results using the MODIS MCD12C1 land classification instead are shown in Table 3. For water, snow/ice, urban, and sparsely vegetated no EC has been derived as there were unsurprisingly no fires found in the direct emissions from TROPOMI on these land types. As expected the EC for broadleaf evergreen forest are the highest (as for the GLC2000 definition) with 129g/MJ and the lowest are found for closed shrublands (23 g/MJ).

GFED relies on only 6 different types of vegetation (AGRI: Agricultural waste burning; BORF: Boreal forest fires; DEFO: Tropical deforestation and degradation; PEAT: Peat fires; SAVA:Savanna, grassland, and shrubland fires; and TEMF: Temperate forest fires) and the results of the emission coefficients using the TROPOMI direct emissions estimates (and MODIS FRP) are shown in Table 4. This gives the opportunity to be able to compare with the emission factors from GFED and GFAS directly (shown in Table F1 although the conversion factor to convert the ECs to Efs can change by a factor of four depending on the study used (Wooster et al., 2005; Kaiser et al., 2012), limiting a meaningful comparison). From our analysis: PEAT the highest Ecs (183 g/MJ), and the lowest are found for agricultural waste burning.”

Section 5:” The total wildfire related CO emissions using TROPOMI-FRE are approximately 290 Mt in 2019. Using different land classifications we get 308 Mt using the MCD12C1 classification and 265 Mt for GFED classification, these help to determine an overall uncertainty for the methodology and the total emissions of the TROPOMI-MODIS top-down emissions. Using the coarser classification of GFED leads to lower emissions (~ 15Mt) in the SHSA and SHAF region compared to using the ECs for GLC2000 or MCD12C1 classification.”

I think that table 2 is the most important result of the manuscript. Since the same approach is used in other work (e.g. GFAS), it would be necessary to compare results to previously published values. I understand that the authors mention that a coarser classification of biomes degrades the correlation coefficients, but a comparison is necessary in order to understand the potential added value of this work in estimating uncertainties on EC used in the literature. If I understand correctly, GFAS uses a conversion coefficient to estimate dry matter consumed from the FRE; and then derives emissions using tabulated emission factors. An equivalent emissions coefficient for CO could be estimated (or at least an order of magnitude).

We included emission coefficients using the GFED biome classification (Table 4), additionally we converted these to emission factors to compare to GFED and GFAS emission factors directly (in appendix Table D1). Note that the conversion factor to estimate emission factors varies significantly for different studies, by as much as a factor of 4. We added the following to the manuscript:

"Section 4: The results using the MODIS MCD12C1 land classification instead are shown in Table 3. For water, snow/ice, urban, and sparsely vegetated no EC has been derived as there were unsurprisingly no fires found in the direct emissions from TROPOMI on these land types. As expected the EC for broadleaf evergreen forest are the highest (as for the GLC2000 definition) with 129g/MJ and the lowest are found for closed shrublands (23 g/MJ). 460 GFED relies on only 6 different types of vegetation (AGRI: Agricultural waste burning; BORF: Boreal forest fires; DEFO: Tropical deforestation and degradation; PEAT: Peat fires; SAVA:Savanna, grassland, and shrubland fires; and TEMF: Temperate forest fires) and the results of the emission coefficients using the TROPOMI direct emissions estimates (and MODIS FRP) are shown in Table 4. This gives the opportunity to be able to compare with the emission factors from GFED and GFAS directly (shown in Table F1 although the conversion factor to convert the Ecs to Efs can change by a factor of four depending on the study used (Wooster et al., 2005; Kaiser et al., 2012), limiting a meaningful comparison). From our analysis: PEAT the highest Ecs (183 g/MJ), and the lowest are found for agricultural waste burning."

Section 5:" The total wildfire related CO emissions using TROPOMI-FRE are approximately 290 Mt in 2019. Using different land classifications we get 308 Mt using the MCD12C1 classification and 265 Mt for GFED classification, these help to determine an overall uncertainty for the methodology and the total emissions of the TROPOMI-MODIS top-down emissions. Using the more coarse classification of GFED leads to lower emissions (~ 15Mt) in the SHSA and SHAF region compared to using the Ecs for GLC2000 or MCD12C1 classification."

Uncertainty analysis:

I appreciated finding an analysis of the method's uncertainties and of the uncertainties on the final emissions, which is a very difficult exercise.

It is evaluated through:

Section 2.3: academic case study with synthetic data allow an estimate of uncertainty on CO emission's inversion to 42%.

Section 3: detailed comparison of inversions with bottom-up inventory GFFEPS

Section 5: intercomparison of annual emissions for 2003-2021 using 5 other emissions inventories.

Throughout the paper, the authors estimate an uncertainty to 40% or 42% (a consistent number would probably be better).

However, I think it is strongly underestimated.
This estimate assumes that the only uncertainty in the calculation of the EC values is the uncertainty on the inversion but I don't think that this is fully demonstrated. For example, the authors mention that an inversion is not possible for large wildfires because of multiple plumes. These wildfires are likely to emit a very large mass of CO. May this filter (and others) induce a bias? Does the fraction of fires with successful inversion depend on the biome?
We made sure the number is consistent. We added the following to Sect. 5:
"To assess the uncertainty of the total annual emissions of our estimates (TROPOMI-FRE), we also used emission coefficients derived from fires of individual years (2019 to 2021, using the GLC2000 classification). Using emission coefficients from 2019,2020, and 2019-2021 combined did not impact the total emissions for individual regions or globally (see Fig. E1). 2021 seems to be the anomaly for which the total global emissions reduced by approximately 20-25 % , due to overall lower EC_CO (for biomes 1-3, see Table E3 which affected the SHAF region the most and almost halved the emissions in 2021 compared to using the Ecs from other years). This shows that the uncertainty of our approach is at least 25 % for the global emissions but for individual regions an uncertainty of 50 % should be assumed. The impact of using Ecs from individual years is greater than the impact of using different definitions of biomes. It should be noted that the uncertainty discussed here is only due to the Ecs, the uncertainty due to GFAS FRE (not provided) does not go into this estimate thus the total uncertainty of the TROPOMI/FRE top-down emissions is expected to be higher. For most wildfire emissions inventories an uncertainty of a factor of two is assumed (Pan et al., 2020; Wiedinmyer et al., 2023)."

Yes, the success-rate of the direct emission estimate does depend on the biome. For example, the large fires during the black summer in Australia would largely affect biome type 1. However when the EC are estimated (sect. 4), only successful fires with total FRP of those specific fires are considered. No inversion is used for the annual total emissions. Fires that were not successfully estimated from TROPOMI, but detected by MODIS are part of GFAS FRE (here, no minimum FRP threshold is applied) and are still considered in the total emissions (sect. 5) through the use of the GFAS FRP and applying the emission coefficients that are derived in section 4. We are not applying an inversion technique in this paper to quantify total fire emissions, rather we are using 1) direct emission estimates (flux method) based on winds and observed total column CO, and 2) top-down estimates by applying TROPOMI/MODIS-derived EC to total FRP. Throughout the text we refer to "direct" estimates and "top-down" estimates to distinguish the two.
We included changes in the abstract, Section 3, Section and Conclusions to make this clearer.

The authors discuss the difficulty of such exercise due to overpass times and plume transport (p. 14). Would the very classic approach of comparing plumes simulated using a CTM may be more adapted in this case? It allows to compare the resulting enhancements regardless of under-constrained parameters (like diurnal variations in this case).
We have used synthetic total column CO from a CTM in sect. 2.3 to obtain better understanding of the uncertainties associated with the direct emission estimate and to explore how best to filter the emission estimates. Figure 2 shows how the estimate (using the model CO columns) compares with the input emissions to the CTM. Which allows to evaluate the emission estimate method directly.

We also compare some plumes, from the model with TROPOMI (total columns), see Fig.5, however this kind of comparison might be difficult to do globally (we have no global run GEM-MACH version available at 10km resolution), and often the accuracy of the model meteorology plays an important role in how similar the plumes in reality vs model look like.
The diurnal variability is included in the model emissions, but TROPOMI alone cannot provide the diurnal variability.
The authors mention that there are no uncertainty estimates for other inventories which is not true. Many publications use these inventories to simulate BB plumes using chemistry-transport models, that are compared to atmospheric observations. Some studies are included in the publications describing the inventories. It is commonly assumed that uncertainties on BB emissions are a least a factor of 2.
Different recent papers present an intercomparison of BB inventories and discuss uncertainties, e.g. :

Wiedinmyer, C., Kimura, Y., McDonald-Buller, E. C., Emmons, L. K., Buchholz, R. R., Tang, W., Seto, K., Joseph, M. B., Barsanti, K. C., Carlton, A. G., and Yokelson, R.: The Fire Inventory from NCAR version 2.5: an updated global fire emissions model for climate and chemistry applications, Geosci. Model Dev., 16, 3873–3891, https://doi.org/10.5194/gmd-16-3873-2023, 2023. Pan, X., Ichoku, C., Chin, M., Bian, H., Darmenov, A., Colarco, P., Ellison, L., Kucsera, T., da Silva, A., Wang, J., Oda, T., and Cui, G.: Six global biomass burning emission datasets: intercomparison and application in one global aerosol model, Atmos. Chem. Phys., 20, 969–994, https://doi.org/10.5194/acp-20-969-2020, 2020.
We included the suggested publications and added the following sentence in Section 5:
"For most wildfire emissions inventories an uncertainty of a factor of two is assumed (Pan et al., 2020; Wiedinmyer et al., 2023)."

Lastly, many conclusions are vague (e.g. l.338, 355). Although they are mostly supported by figures and tables I would have appreciated a few summary numbers.
We have made revision to the conclusions as suggested:
"The comparison of TROPOMI-FRP derived top-down emissions with other inventories reveals disparities and highlights the substantial uncertainties associated with fire emission estimates. Notably, GFFEPS generally exhibits the closest agreement with TROPOMI-FRE (e.g. total emissions for 2019: 337 vs 290Mt of CO for GFFEPS versus TROPOMI-FRE, respectively) , although some exceptions are evident in regions such as SHSA, NHSA, SEAS, and AUST. Possible reasons for the discrepancy are the omission of small fires, the absence of detected hotspots, and an underestimation of FRP, ultimately contributing to an overall underestimation of total emissions. FINNv2.5 shows the largest discrepancies with TROPOMI-FRE where the FINNv2.5 global emissions (579 Mt) are almost twice of TROPOMI-FRE, in the SEAS region the FINNv2.5 emissions are higher by a factor of five: 20Mt versus 100 Mt). Examining the trends over the past two decades (corresponding to the MODIS lifetime), it appears that global CO wildfire emissions have, on the whole, decreased, the trend (between -8.7 to -5.1 Mt/yr) is significant for all inventories, except GFED (this is driven by the high increase for BOAS of almost 20Mt/yr). The trend of wildfire emissions is highly region-specific, with the highest reductions occurring in SHSA, SHAF, NHAF, and CEAS. Conversely, wildfire emissions in TENA are on the rise (by 0-4 Mt/yr). For all other regions, the variability within the past two decades has been too substantial to determine a statistically significant trend."

Global inventory and intercomparisons:
Recent intercomparison exercises should be mentioned and discussed (see previous comments). The authors find the highest agreement in trends with GFAS which may not be surprising since both approaches rely on the same FRE database. These are not independent estimates. Since the vegetation is assumed constant, the trends obtained reflects the trends in FRE.
To better understand the differences obtained for some regions in total emissions (e.g. BONA, SHAF, EQAS), it would be important to compare the values of EC used in GFAS with the values used in this study (see previous comment).
We included this suggestion, see previous comments.

Minor comments.

Throughout the manuscript: check remaining typos, extra spaces in front of points, etc. I think it would be more accurate to talk about "emission inventory" rather than "emission budget"?
We have changed emission "budget" to emission "inventory" or simply "total emissions" throughout the manuscript as suggested.
Abstract:
l. 4: remove 'More recently' since first emissions inversion were performed before the approaches based on FRP…
Removed as suggested.
Main conclusions should provide some key numbers, as well as uncertainties.
Changes to the conclusions were made see above comment. The following is in the conclusion section:
"The sensitivity tests show that the methods uncertainty is approximately 34% (and 44% total uncertainty including the uncertainty of VCDs and winds)."
"The TROPOMI-FRE top-down emissions (based on the uncertainties of the ECs) are at least 50% regional or 25% globally. Note that the total uncertainty is likely larger as the uncertainty of GFAS FRE is not accounted for."
Introduction:
l. 54, 75: laboratory measurements and field experiments.
Unfortunately, I'm not sure what the refers to.
Section 2.5 should introduce all emission inventories used in the manuscript.
We changed the name of the section to "Emission inventories" and included the following paragraph:
"We compare our retrieved CO emissions to several existing biomass burning CO inventories, namely GFAS, GFED, FINN v1.5 and v2.5, as well as GFFEPS. GFED (van der Werf et al., 2017) and FINN are both based on bottom-up approach. Here, we use GFEDv4.1 which has a 0.25◦ resolution, developed by NASA, in Sect. 5 we use the annual total emissions for different geographical regions. The FINN inventory (Wiedinmyer et al., 2006, 2011) by NCAR is not on a regular grid but based on the location of the MODIS or VIIRS detected hot spots, these are then summed to obtain annual totals in Sect.5. The GFAS fire emission inventory (Kaiser et al., 2012) by ECMWF utilizes a top-down approach based on MODIS FRP and is on a 0.1◦ regular grid, here we use v1.2. Additionally, we also compare to a new global biomass burning algorithm, GFFEPS, developed by Environment and Climate Change Canada. "
What is the horizontal resolution of GFFEPS?

The emissions are directly based on the hotspot measurements, they are not gridded. We added the following to the manuscript:

"The resulting emissions are not gridded, but distributed to the location of the detected fire-hotspots. "

Section 3: 2nd sentence should be revised.

We changed the sentence to the following:

"Stockwell et al. (2022) have shown good agreement between the TROPOMI-derived to aircraft-derived CO fire emissions as part of the FIREX-AQ campaign (Warneke et al., 2023). The sensitivity tests (Sect. 2.3) using synthetic total columns also suggest that emissions can be reliably estimated using the flux method within 42 % uncertainty.