Reviewer #2 comments:

The authors have clarified the points that I raised in a convincing way, except for the following ones:

- The authors explain that they do not need the qa_value filter because they have other means to detect clouds in the scenes. Their use of ancillary data is commendable, but they cannot ignore the flags that come with the retrievals. For example, if the retrieval assumed there is a cloud, its averaging kernel will likely peak higher in the atmosphere and will less represent a total column. Authors should remove all data with qa_value less than 1 to avoid misinterpretation.

We are happy to hear that Reviewer #2 is convinced by our replies, and appreciate the expansion on the point originally raised. As suggested by the reviewer on the remaining point, we have now used the qa_value in the Sentinel-5P CO product to threshold the data used. The reviewer suggests removing all data with qa_value < 1.0 ; though we believe this rather conservative threshold is both unnecessary and impractical – based on the detail we provide below. We have rather removed all data with qa_value < 0.5 from all match-up plumes in our dataset, based on the recommendations and information in the Sentinel-5P CO User Manual and Sentinel-5P CO Product Readme document on how to use this qa_value.

Detail of threshold selection: A discussion of the sensitivity of the vertical averaging kernels used in the TROPOMI Total column CO retrieval can be found in Borsdorff et al. (2014). The Figure below is taken from that work and shows the difference in the mean global averaging kernel on a single day (10^{th} November 2017) of observations with differing qa_values . We can see that at the typical altitudes at which the large smoke plumes from the types of fires used in our matchup process are observed (typically 800 m to 4000 m altitude), the weighting of the averaging kernel [a_{col}] applied to pixels containing mid-level cloud (i.e. pixels with qa_values > 0.5 but < 1.0; black line) differs by no more than 0.1 from the weighting of the averaging kernel applied to clear-sky pixels (yellow line). The weighting applied to pixels with qa_value = 1.0 at these altitudes is also well within the standard deviation (horizontal bars) of the that applied to pixels with qa_values > 0.5. In fact, out of all altitudes, the sensitivity to total column CO is closest between pixels with qa_values > 0.5 & < 1.0, and pixels with qa_value = 1.0 at the altitudes which the bulk of a fires smoke plume is likely to be observed at after several hours of burning (approximately 1500-2000 m). Indeed, Borsdorff et al. (2014) state that in remote regions, such as the locations of the plumes in our matchup dataset, the error introduced by the choice of averaging kernel is comparable between clear-sky observations and observations containing mid-level cloud (qa_value > 0.5 but < 1).

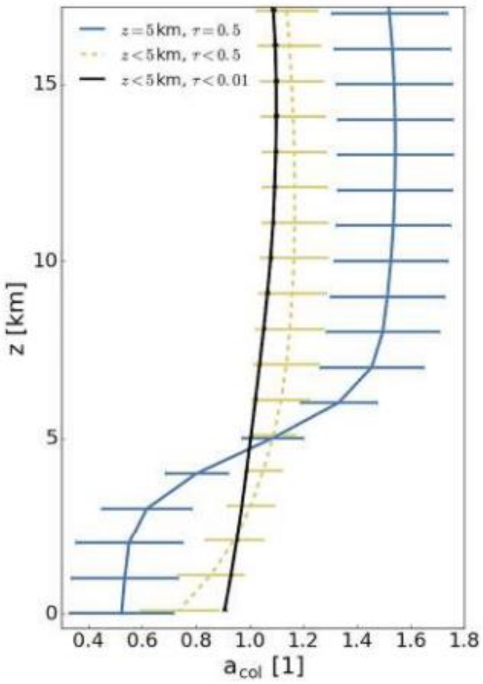
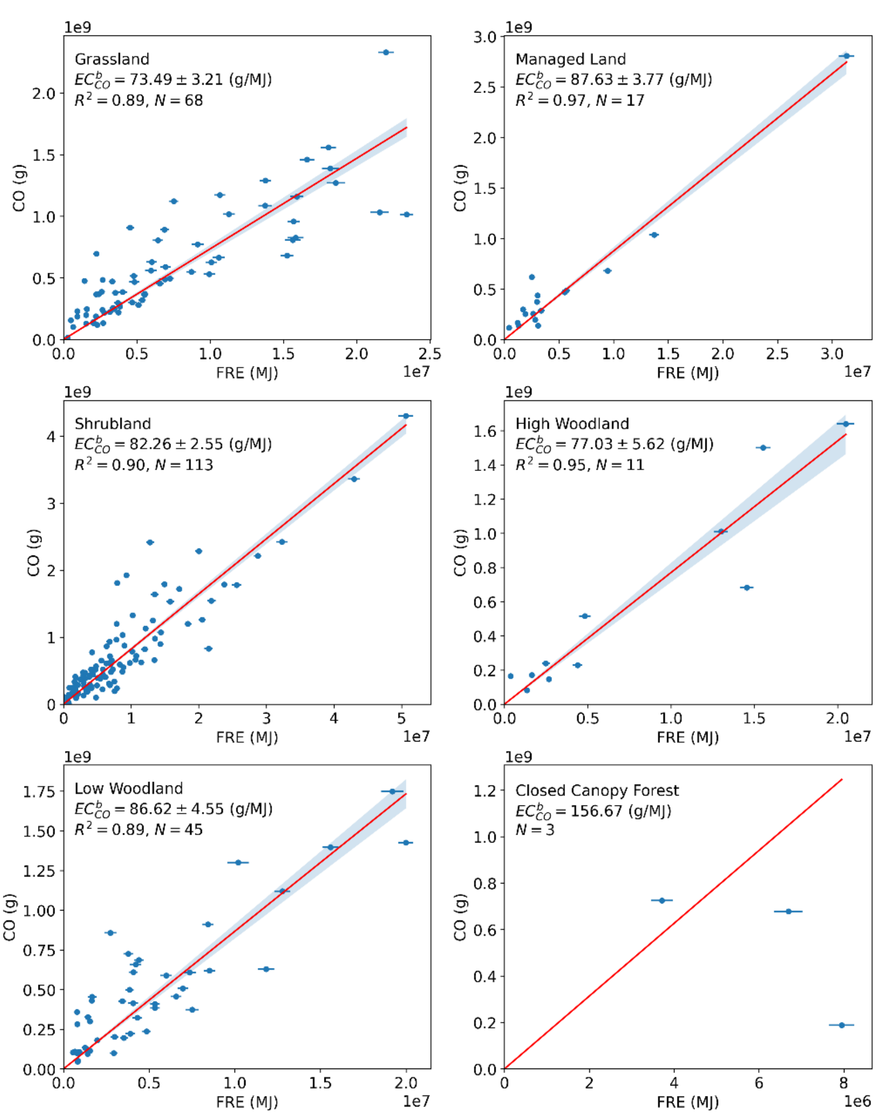


Figure 1. TROPOMI CO total column averaging kernels for 10 November 2017. The global average is shown for three different categories of cloudiness strict cloud clearing (black), clear-sky equivalent (yellow), and high optical thick clouds (blue). The standard deviation is indicated as error bars.

It should also be noted that the qa_values provided in the S5P CO product have not been validated in conditions anywhere near comparable to the application in which we have used the product. S5P CO data have been evaluated using the TCCON and NDACC ground-based solar FTIR monitoring networks, with this conducted by averaging the S5P CO product over 50km^2 areas around each ground measurement site and comparing the total column CO values. None of these sites were located in Africa, and the validation was based on large area averaging of ambient-type total column CO data from clear-sky conditions (Borsdorff et al., 2014). Our matchup dataset features total column CO observations over far smaller areas with extremely elevated CO and strong spatial CO gradients. The application of the qa_values applied in Borsdorff et al. (2014) is therefore unlikely to be fully representative of our application, and we hence treat the qa_values assigned in the S5P data with caution. Therefore, we set a qa_value threshold of 0.5, with this slightly more lenient threshold being especially appropriate as we have also used the extremely cloud-sensitive LSA SAF Meteosat cloud mask and the VIIRS imagery (which has a 750 m spatial resolution) of the same time as the S5P overpass to already remove any cloudy matchups from our candidate matchup plume dataset. We therefore confirm via the qa_value thresholding and the image analysis that no matchups that are cloudy exist in our matchup plume dataset. We trust this new use of the qa_value and our confirmation of cloud free data now satisfies the reviewer.

Having applied the qa_value threshold as detailed above, the resultant emissions coefficients (EC_{CO}^{b}) are shown below in Figure 2. These EC_{CO}^{b}values are summarised in Table 1 along with the EC_{CO}^{b} values derived when the previously discussed Fixed Mask De-striping (FMD) is applied, and the respective percentage difference between these and the original EC_{CO}^{b} values presented in the original manuscript. The qa_values > 0.5 thresholding results in an average EC_{CO}^{b} difference of only

2.4% across all five biomes (maximum difference is 5.9%) demonstrating that the impact of the mid-level cloud on the retrieval of CO in our study is rather minimal and well within the uncertainty bounds of the EC^{b}_{CO} values. As discussed above, it is unclear from the validation carried out by Borsdorff et al. (2014) on the S5P CO product whether the qa_values defined for each pixel are really appropriate for use in the extreme conditions in which we utilise the S5P CO data. Combined with the apparent minimal impact that the removal of qa_values < 0.5 has on our calculated EC^{b}_{CO} values, we have included the coefficients with the qa_value threshold applied in the Appendix - for the benefit of those that may prefer to use these values and interested to understand. We have also added a condensed explanation of the above in the main manuscript text and in the Appendix.



	S5P without FMD applied	*S5P with FMD applied*	*Percentage difference (%)*	*S5P excluding qa_value < 0.5*	*Percentage difference (%)*
Grassland	75.51	68.94	-8.7	73.49	2.7
Shrubland	81.07	76.16	-6.1	82.26	-1.5

Managed land	88.35	82.91	-6.2	87.63	0.8
High-woodland savanna	81.85	79.99	-2.3	77.03	5.9
Low-woodland savanna	85.49	77.04	-9.9	86.62	-1.3

Reference

Borsdorff, T., Aan De Brugh, J., Hu, H., Hasekamp, O., Sussmann, R., Rettinger, M., Hase, F., Gross, J., Schneider, M., Garcia, O., Stremme, W., Grutter, M., Feist, Di. G., Arnold, S. G., de Mazière, M., Kumar Sha, M., Pollard, D. F., Kiel, M., Roehl, C., … Landgraf, J. (2018). Mapping carbon monoxide pollution from space down to city scales with daily global coverage. *Atmospheric Measurement Techniques*, *11*(10), 5507–5518. https://doi.org/10.5194/amt-11-5507-2018

- "It is rather more likely that the sample size of the current work is not sufficiently large to enable statistically distinct $ECCOb$ values to be derived". I do not understand the argument because a small sample size will unlikely bring the values close to each other (ie by accident) as it happens here.

We thank the author for their comment. This statement was certainly unclear and in need of further expansion, which we provide below and in the main text.

Each of the fire-plume matchups in our dataset are classified into one of the 6 distinct 'fire biome' classes, based on the location of the active fire pixels that make up the fire. As such, in most cases a smoke plume is not 100% generated by fire that burns vegetation of a single biome – because the active fire pixels can be from more than one biome. We therefore use a filter to ensure that the majority (>50%) of the active fire pixels responsible for a smoke plume come from a single biome, and only if this condition is met do we include the fire and plume within the matchup dataset for that biome (for its EC_{CO}^{b} derivation). If not we discard it from the matchup. A larger plume dataset would enable us to apply an even more stringent condition (e.g. 70% of active fire pixels from a single biome) whilst still maintaining a reasonable matchup sample size to generate EC_{CO}^{b} values from for each biome, and ultimately more matchups that are even more dominated by a single biome may make the EC_{CO}^{b} values more statistically distinct. In our prior study that used AOD data rather than CO data (Nguyen & Wooster 2020), the 1 km spatial resolution of the AOD product used allowed us to include more smaller plumes and smaller fires to generate the EC_{TPM}^{b} values. And the fact that these plumes were generated by fires covering smaller areas meant that there were more plumes able to be included in the matchup dataset which had a larger majority of the AF pixels coming from a single biome. The resulting EC_{TPM}^{b} values were indeed more statistically distinct from one another. We have re-worded the text in manuscript to explain this point more fully.

- "Remote sensing does include a measurement". Sure, but the processing of this measurement and its combination with auxiliary information and models for remote sensing is not a measurement. I can understand that the word is kept occasionally for simplicity, but it should not be used in the title.

We have amended the manuscript title and scaled down its use throughout the manuscript

- There are a few "Sentinal" left in the text.

These have now been amended