Anonymous reviewer #2

We thank the reviewer for these helpful insights, we have tried to address the reviewer's comments and questions to the best of our ability and feel this has added clarity to the manuscript. With respect to questions about the relevance of this manuscript to ACP - we feel that the work presented is firmly within the subject area of the journal ('The main subject areas comprise atmospheric modelling, field measurements, remote sensing,…'). Similar works describing both top-down and bottom-up global biomass burning emissions inventories have been previously published in ACP (e.g. Van der Werf et al., 2010; Ichoku & Ellison, 2014) along with other remote-sensing based approaches to estimating biomass burning emissions (e.g. Chang and Song, 2010; Mao et al., 2014; Wu et al., 2018).

The paper presents the retrieval of carbon monoxide emissions over Africa from a 16-year archive of geostationary satellite images, with several auxiliary data derived from other satellite instruments, such as Tropomi. It's a dense technical document, which certainly represents a huge amount of work. The reader of the ACP might be surprised by its contents, because in the portfolio of the EGU journals, it would fit much better with the AMT ("advances in remote sensing") than with the ACP ("studies investigating the Earth's atmosphere and the underlying chemical and physical processes"). That aspect aside, it is full of technical processing details and acronyms that do not make it easy to read, but I still did not find some of the key information I was looking for, leaving me to wonder about the quality of other information about which I know less:

> What Tropomi data was used? NRTI, OFFL? The document states that the 2020 data was used at 7 km resolution as for 2018, but the Tropomi data was at 5.5 km resolution along the track during the second period: how was the degradation of the resolution made? Was data with variable "qa_value" less than 1 correctly discarded? How were the stripes of CO data in the direction of flight treated for the two time periods? Can the selection of Tropomi data bias the results towards certain types of fires (with low aerosols optical thicknesses for example)?

We thank the reviewer for this comment, there are several elements and we have separated their questions for ease of addressing them.

What Tropomi data was used? NRTI, OFFL? The document states that the 2020 data was used at 7 km resolution as for 2018, but the Tropomi data was

at 5.5 km resolution along the track during the second period: how was the degradation of the resolution made?

This work uses Sentinel-5P TROPOMI data from 01/07/2019 to 31/12/2019 plus that from the entirety of 2020. The NRTI_L2_CO product is typically only available on the Copernicus datahub (https://s5phub.copernicus.eu/dhus/#/home) for 24 hours after the current time. Since we used many months of observations we rather used the OFFL product. We have added this clarification to the methodology section and added text detailing the change in the along-track resolution of the S5P product in 2019 and how this was dealt in when calculating Total plume CO.

Lines 117:

"The offline (OFFL) S5P total column carbon monoxide (TCCO) product used in this work can be downloaded from Sentinel-5P Pre-Operations Data Hub (https://scihub.copernicus.eu/) and has a spatial resolution of 7 × 7 km until August 2019 after which the along-track resolution increased to 5.5 km."

Line 178:

"For each fire in the final match-up set we calculated i) the total CO [g] contained in the plume from the S5P TCCO retrievals, and ii) the total FRE [MJ] released by the fire from Meteosat FRP-PIXEL data – integrating FRP from the time of the fires first AF pixel detection on that day to the moment of the S5P overpass. The minimum TCCO value ($mol.m^{-2}$) from each plume buffer was taken as the appropriate CO 'background' value for that plume and subtracted from all plume pixel TCCO values. The resulting 'excess' TCCO pixel values were then converted to units of grammes by multiplying by the molecular weight of CO and by the pixels' area calculated from their geographic corner coordinates (thus accounting for the change in the along-track spatial resolution of S5P data in 2019 and any view-angle dependent pixel area growth). Summing these values across all plume pixels of a fire then yielded the total amount of fire-emitted CO in that plume, which was compared to the total amount of FRE the fire generated over the time that it released that CO. Each match-up fire FRE and total CO pair constitutes one datapoint on the relevant 'fire biome' scatterplot of **Error! Reference source not found.**. FRE and Total CO uncertainties were calculated from the uncertainty values provided within the FRP-PIXEL product and the TCCO product respectively, these are also plotted in the form of error bars on Error! Reference source not found.."

Was data with variable "qa_value" less than 1 correctly discarded?

Thank you for this question. The *qa_value* in the S5P TCCO product gives an indication of whether a S5P pixel passes tests for contamination by mid-level and high-level cloud. Our methodology does not eliminate TCCO pixels with qa_values less than 0.5 (as recommended in the product metadata) because the presence of cloud is already determined and delt with through the use of the two other (far higher spatial resolution) satellite products used in tandem with the S5P product. Firstly, we use the 3 km spatial resolution SEVIRI FRP-PIXEL Quality Product (associated with the FRP product) to inspect and filter out any potential match-up fires that might be contaminated with cloud during the duration of the fire. These 15 min temporal resolution data ensure that no clouds were present over the location of the fire at any point during it's lifetime. Secondly, we use VIIRS RGB imagery (375 m spatial resolution at nadir) acquired within 3.5 minutes of the S5P data to ensure the S5P observations are cloud free. For these reasons we do not need to use the S5P CO qa_value.

How were the stripes of CO data in the direction of flight treated for the two time periods?

We are certainly aware of the Fixed Mask De-striping (FMD) method proposed by Borsdorff et al. (2019) for Sentinel-5P data. Indeed, the FMD is now applied in the Sentinel-5P OFFL operational product from July 2021 onwards, but is not applied to the same product prior to that date. You can create the FMD method yourself and apply it to such data – however we chose not to apply it to the retrospective dataset analysed in the main text. This choice was made based on our understanding of the FMD method, its "validation" in Borsdorff et al. (2019) and the situations under which it was evaluated therein, plus the results of our own tests of the impact that applying the de-striping had on fire plumes.

The FMD "correction" is essentially an adjustment of the TCCO values delivered at each pixel position across TROPOMI swath, based on adding or subtracting a constant value at each location. The adjustment is thus irrespective of the actual TCCO value in each pixel in the swath, and constant at each swath pixel position. In Borsdorff et al. (2019) an evaluation of this de-striping procedure was carried out by comparing daily averaged de-striped S5P TCCO values against a TCCO measurement made by one of the TCCON ground stations. Whilst the TCCON station was a single point observation, the S5P data were averaged over a colocation radius of 50 km and typically the CO values varied rather mildly across this 50 km region.

The temporal and spatial scales of the CO variations seen in these validation locations are clearly very different from the application of the data in our work, where very strong spatial CO gradients exist at the location of the fire plumes. Although allusions to the impact of the (de)striping at smaller scales are made in Borsdorff et al. (2019), no solid analysis is presented showing the impact of the striping, or de-striping on such phenomena. We tested this with our data, however, and observed that in the case of smoke plumes (which are represented in the TROPOMI product by small groupings of pixels with high TCCO concentrations) the striping clearly has less impact on the high TCCO values in the plume than it does on the low TCCO pixels of the plume background. This can be seen in the below plots (Figure 1) that show two different regions containing smoke plumes on 23-07-2022 at 11:01 UTC. The plumes are circled in red, and the image comparisons (top vs bottom images) show that over these relatively small areas, low "background" TCCO pixels are more affected as a percentage of their TCCO by both the striping and the FMD 'de-striping' procedure than are the high "plume" TCCO pixels. It is not at all clear here, nor in the original Borsdorff et al. (2019), that a constant de-striping correction such as the FMD approach is appropriate to apply in an area where TCCO concentrations have such a high spatial gradient between pixels. Furthermore, the FMD method does not actually remove all the stripes, neither here nor in the original examples of Borsdorff et al. (2019). Hence, we have chosen not to apply the FMD procedure until more quantitative evidence of it's appropriateness to such high CO gradient regions is published. For completeness however, we have applied it to the data from which we calculate the CO emission coefficients (EC_{CO}^{b}) for each fire biome – and included these results in a new Appendix along with an explanation. The main manuscript still includes our CO emission coefficients derived with the standard CO product, whilst the Appendix includes the values calculated from the same data but with the FMD procedure applied (it also shows the % difference between these and the ones in the main manuscript – see Table 1 (below)). The FMD procedure has a greater impact on low CO value "background pixels" than on the high CO "plume" pixels - as observed in Figure 1 below – and the maximum smoke emission coefficient change is 10% when FMD is applied compared to when it is not (see Table 1) – and in all cases EC_{CO}^{b} is slightly lower with destriping applied.

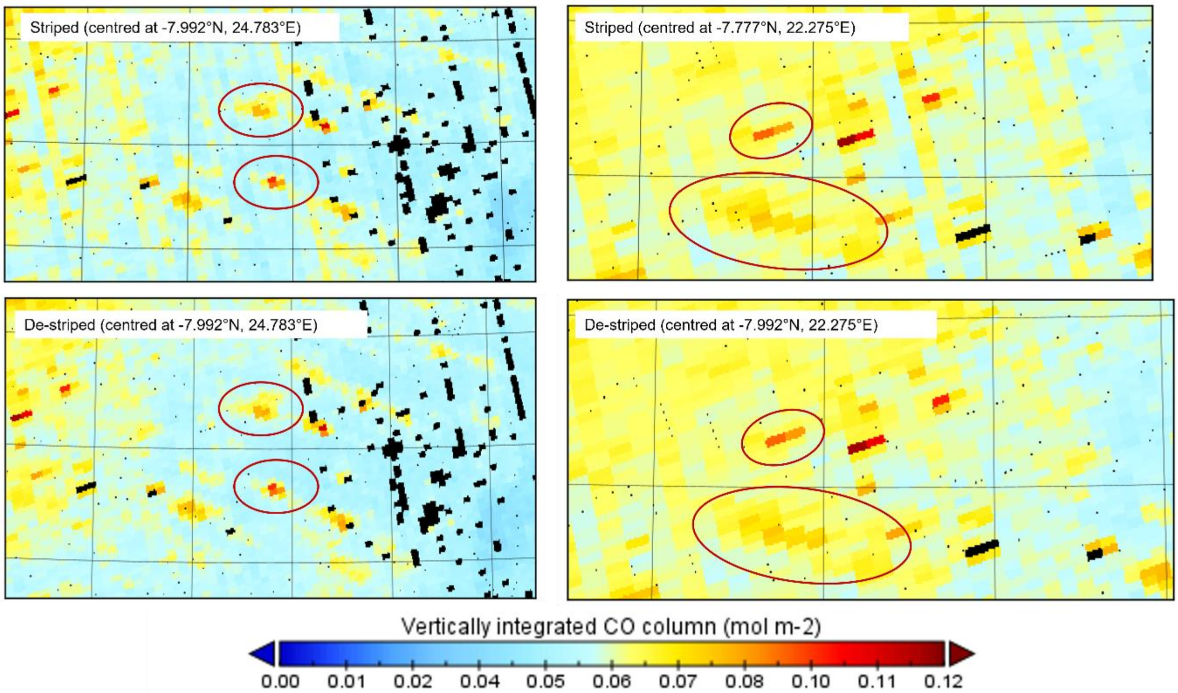

Figure 1: Comparison of standard and Fixed Mask De-striped S5P CO data of two fire plume regions (Region 1 at left and Region 2 at right).

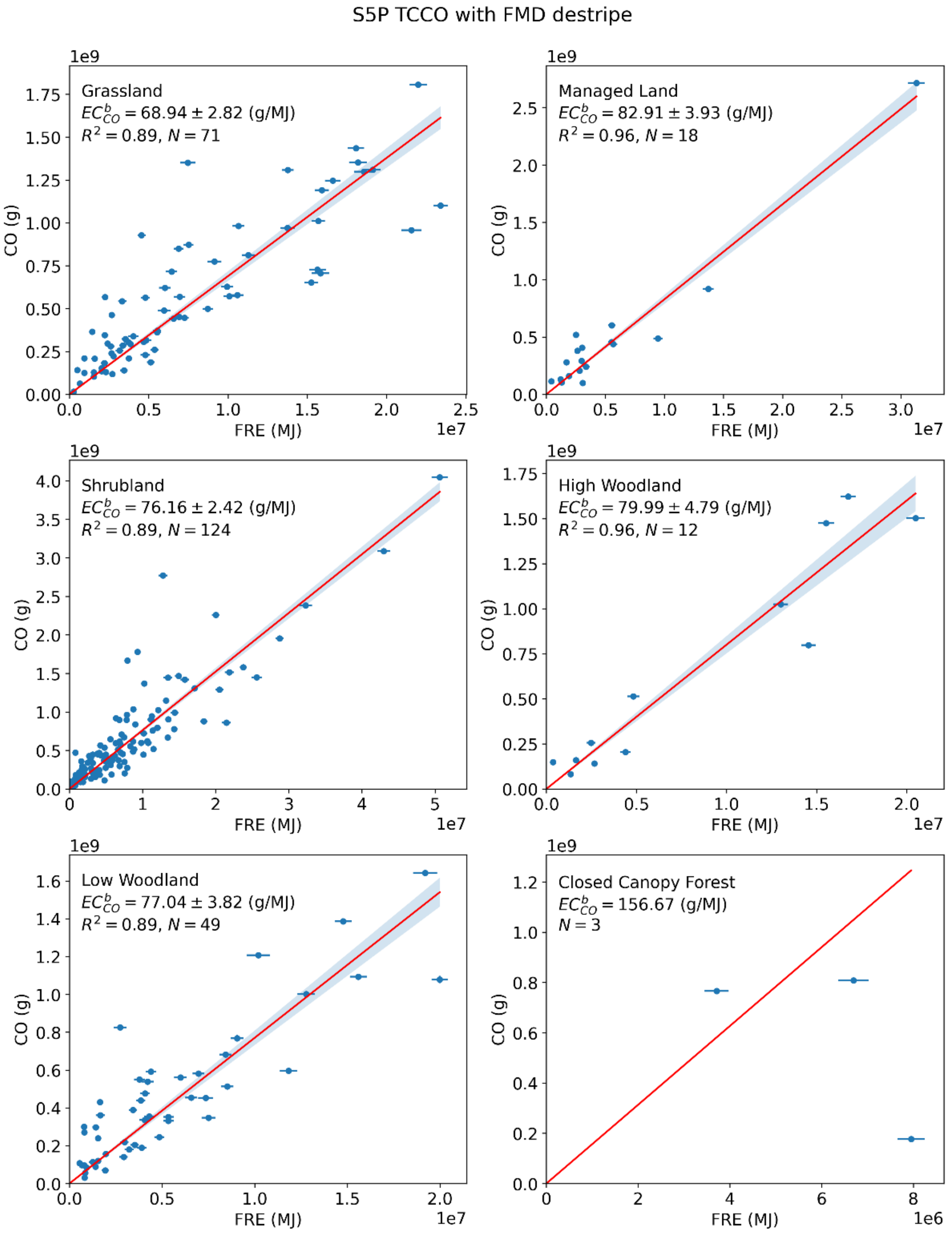

Figure 2: CO Emissions coefficients derived with FMD applied to S5P CO data.

Table 1: Comparison of smoke emission coefficients with and without the FMD de-striping procedure applied.

	S5P without FMD applied	S5P with FMD applied	Percentage difference (%)
Grassland	75.51	68.94	-8.7
Shrubland	81.07	76.16	-6.1
Managed land	88.35	82.91	-6.2
High-woodland savanna	81.85	79.99	-2.3
Low-woodland savanna	85.49	77.04	-9.9

Can the selection of Tropomi data bias the results towards certain types of fires (with low aerosols optical thicknesses for example)?

We are somewhat unclear what the question is aiming at here. We think it is related to the fact the TROPOMI products mask out data with very high aerosols, so therefore our dataset might not include fires that produced such characteristics. We don't expect such biases on the basis of what we see – which includes fires with significant aerosol plumes. What we do have is a bias towards larger and longer-lived fires (No shorter than 30 minutes with 70% of matchups more than 2 hours), as very small or short-duration fires will not generate the sort of data record necessary for the matchups. However, since we find a linear relationship across lower and higher FRP fires in the matchup scatterplots (see Figure 2 in the manuscript) we have no evidence for the introduction of bias.

How is the plume buffer computed (I cannot guess it from Fig 1)? It is easy to bias the "background" estimation high or low with a slight change in the buffer definition, in particular due to the unstable choice of the minimum value.

The fire match-ups and plume polygons were identified manually and the buffer around each plume was selected such that i) it visibly incapsulates the high CO pixels of the plume and ii) gives several pixels' worth of close-to-background values surrounding the plume. Defining a buffer of a set pixel width is not appropriate as this has the potential to result in contamination from neighbouring plumes, which can be close by during periods of greater fire activity. In all previous top-down approaches that derive emissions coefficients from a dataset of fire match-ups, no single or consistent method has been used for defining the background level of the variable of interest (previously this was always AOD). For example, Ichoku & Ellison (2014) and Fisher et al. (2020) define the AOD background as the value of a pixel or group of pixels up-wind of each plume, while Mota & Wooster (2018) use a large area average (500 km^2) of AOD surrounding the plume. In the previous version of FREM (Nguyen & Wooster, 2020) and in the current work, the use of a percentile threshold approach to define the ambient background was explored - this method relies on defining the lowest, say, 5% of pixels within the plume polygon to be the background CO or AOD value while the rest are defined 'plume' pixels. However, this approach was found to be more unstable than a minimum pixel approach because a percentile approach means that an individual pixel has a different weight contribution to the background value depending on the size of the plume. The minimum pixel value was shown to be the more stable of the two for this work.

What is the time scale of the FRP temporal integration? Can some of the CO have been oxidized during that time?

The temporal integration of these fires is in the order of hours, the average integration time of the match-ups was 3.4 hours and 80% of the plumes had integration times below 5 hours. With a typical atmospheric lifetime of 2 months, significant oxidation of the plume CO is highly unlikely in the time between the first AF detection and the time of the S5P overpass.

How are the boundary conditions treated in the model simulations?

Both Initial and Boundary conditions for the meteorology (WRF) are taken from the NCEP FNL global model which has a grid resolution of 0.25° × 0.25° at 26 levels and 6 hour timesteps. Initial and boundary conditions for the chemistry (CMAQ) are taken from the WACCM global chemistry model having 0.9° × 1.25° spatial resolution, 88 levels, and 3 hour timesteps. These details are mentioned briefly in Section 4.1 and further model details are listed in full in Appendix B.

Apart from the uncertainty of the slope (and not the "error" of the slope as wrongly put in the legend, in fact without any definition of the uncertainty variable used), there is no notion of uncertainty in the product shown here.

We thank the reviewer for this valid and insightful comment. Both the FRP and CO satellite observations have associated uncertainties, and the uncertainties in total plume FRE and CO were calculated from these data but were previously not plotted on Figure 2 of the manuscript. These uncertainty measures have now been included in the form of error bars on Figure 2. These uncertainties were not included in the original manuscript as they represent only part of the uncertainty associated with the total fire FRE and CO i.e. the uncertainty inherent in the mathematics to calculate these variables and the radiometric measurements of the satellite. There are contributors to the uncertainty in the match-up FRE and CO that are not included and not quantifiable in the same manner. For example, the plotted uncertainty doesn't include uncertainty in FRE associated with how much of the fire emitted energy is intercepted by tree cover. Weighting the derivation of the EC_{CO}^{b} values with data whose weights depend on incomplete and poorly constrained uncertainties was deemed inappropriate, and thus ODR regression (which does just that) was rejected in favour of OLS regression. A discussion has been added on this point (see below).

Lines added regarding OLS selection. Line 198: following

"Matchup data for each 'fire biome' are shown in Error! Reference source not found. and were used to derive the set of biome-dependent CO smoke emission coefficients [EC_{CO}^{b}] listed in **Error! Reference source not found.**. Zero-intercept ordinary least squares (OLS) regression was used for this, rather than the orthogonal distance regression (ODR) used by Nguyen and Wooster (2020) during derivation of TPM smoke emission coefficients [EC_{TPM}^{b}]. OLS was used in this work for two main reasons. Firstly, although the ODR method considers the uncertainty in each of the variables, these uncertainties are themselves rather poorly constrained, with the known uncertainties only representing part of the total uncertainty sources. There are contributions to the uncertainty of FRP that are not quantifiable, for example due to variations in the amount of interception of a fire's FRP signal by any overlying tree canopy. We therefore deemed use of a regression method in which the slope is strongly driven by datapoint uncertainty to be unsuitable for use. Secondly, weighting based on uncertainty often resulted in undue weight being given to high value points (e.g. match-ups with high FRE and high plume-species amounts) due to them typically having lower relative uncertainties (see Wooster et al., 2015). Due to their typically being very few high value datapoints in each 'fire biome' due to the heavy tailed nature of fire size distribution (Freeborn et al., 2009), these few large fires were potentially being too strongly weighted in the resulting calculation of EC_{x}^{b}. For these reasons, we opted to use OLS regression, and to ensure a consistent methodology for emission coefficient derivation we also applied the same approach to the Nguyen and Wooster (2020) dataset to re-derive their EC_{TPM}^{b}values using OLS regression (see **Error! Reference source not found.**). The updated EC_{TPM}^{b} for closed canopy forest, managed land, grassland, shrubland, low-woodland savanna and high-woodland savanna are 26.07 $g.MJ^{-1}$, 12.23 $g.MJ^{-1}$, 9.39 $g.MJ^{-1}$, 9.88 $g.MJ^{-1}$, 10.65 $g.MJ^{-1}$, and 14.18 $g.MJ^{-1}$ respectively. On average these new values are 14% lower than those reported in Nguyen and Wooster (2020) derived via ODR, and are the ones referred to and used hereafter. The WRF-CMAQ model-based approach to evaluating our final CO emissions rates and totals in Section **Error! Reference source not found.** was also used to carry out an analogous evaluation of the TPM emissions generated from the updated EC_{TPM}^{b} values of **Error! Reference source not found.** (see **Error! Reference source not found.**)."

Just by looking at table 1, I understand that the 5 key numbers derived from the Tropomi data are very close to each other: are their differences statistically significant? –

This is a very valuable comment and we thank the reviewer for this input. We have taken pairs of fire biomes and assessed whether their derived EC_{CO}^{b} values are statistically significantly different. We have also carried out the same analysis for the updated EC_{TPM}^{b} values of Appendix A. These results have now been added in two tables and a discussion included in the main text.

Line 255:

"Tests were carried out to determine the statistical similarity of the EC_{CO}^{b} values shown in **Error! Reference source not found.** for each fire biome. Results are shown in **Error! Reference source not found.**, and indicate that only one pair of biomes (grassland and managed land) have CO emissions coefficients (EC_{CO}^{b}) that are statistically distinct at the 95% confidence interval. Two additional fire biome pairs are statistically distinct at the 85% confidence interval (grassland and low-woodland savanna; managed land and shrubland). This analysis indicates that based on the current match-up dataset, there is relatively little inter-biome difference in the relationship between the amount of thermal radiant energy (i.e. FRE) emitted by a fire and the amount of CO emitted by the same fire. However, smoke emission factor databases (e.g. Akagi et al., 2011; Andreae, 2019) indicate that EF_{CO} values do differ significantly between certain biomes. The updated FREM TPM emissions coefficients (EC_{TPM}^{b}) of Appendix A were similarly tested for statistically significant differences, and the emission coefficients generated from this (much larger) match-up dataset (primarily due to the higher spatial resolution of the MCD19A1 AOD product allowing more plumes to be identified) were found overall to be more statistically significantly different than those of CO. Of the fifteen biome pairs only four did not have EC_{TPM}^{b} that were statistically significantly different at a 95% confidence threshold. This reduces to 2 pairs at the 85% confidence limit. It may be the smaller sample size in the case of the CO data is not allowing the biome-driven differences to manifest themselves in a significant enough way - and limitations to producing a larger sample size in this study have previously been discussed along with the future research work needed to address this key issue."

An error budget for the complete processing chain is surprisingly lacking to support, for example, the discussion of p. 12-13. As of now, this discussion looks like mere speculation.

One of the largest challenges in the assessment of biomass burning inventories of any kind (top-down or bottom-up) is the lack of uncertainties – and this includes for example the GFAS and GFED fire emissions inventories that are by far the most widely used of any worldwide. There is also at present no feasible way to directly compare emissions estimates with an absolute ground truth. As a result, the methods used to evaluate the validity of a fire emission inventory are typically; sense-check comparisons to other fire emissions inventories; large spatio-temporal aggregation of emissions that are then compared to satellite observations of gas or aerosols for specific cases; validation of the input data, i.e. FRP or Burned Area; or an in-direct evaluation through atmospheric modelling and comparison of the output fields to atmospheric observations (as is carried out in Section 4 of this work). The discussion to which the reviewer refers simply serves the purpose of highlighting what previous studies have shown about the current and previous versions of GFED and how our results compare with those. Perhaps our suggestions referring to how the results of this work support the conclusions of previous work may be overstated, so we have re-worded these points. The bulk of the discussion, however, simply makes reference to other studies, or states the results of this work.

In addition, two minor aspects should be addressed:

- The title contains a typo ("Sentinal") and a not-so-common acronym (FRP)

The title has been changed

- The term "measure" used in many places about remote sensing is not appropriate (see the BIPM metrology definitions).

Remote sensing does include a measurement – of light of certain wavelengths. Therefore, we feel it is appropriate to use the term – and also just for readability we have included it. Where possible we say "estimated" or "derived" since for example CO column amounts are derived from Sentinel-5P observations, rather than CO being directly measured.