Referee # 2

We thank the referee for his/her helpful comments. Our response to referee #2 is below. Items in bold and italics are the referee comments.

While they establish a bias and variance between the NR and independent observations, their fundamental threshold is that the difference appears "reasonable".
What they have not done is to relate the errors between the NR and independent observations to the interpretation of the performance of the AR. So, if the accuracy and precision of the NR is twice as bad, what should one infer about the performance of the AR?

We will rewrite this part of the paper avoiding vague statements and quantifying the differences between the NR and the GS and the CR and the GS. We will also compare these differences to typical errors in the MOCAGE data assimilation system derived from published papers (e.g., work by El Amraoui et al., 2014).

I would argue that a more important implementation of their guiding principle is to assimilate real observations, e.g., MOPITT and/or SCIAMACHY, into their system and compare the analysis fields to independent observations. Then, they could do an OSSE for the same observing system and assess the statistical difference between the AR and NR sampled at independent observations versus AR(real) against independent observations. That would provide a better sense of what the OSSE limitations actually are.

We think that the suggestion of the referee is valuable, and suitable for a study in its own right. However, it would only provide direct information on the error of the OSSE for MOPITT and/or SCIAMACHY CO and not directly for S5P. The S5P instrument has different characteristics to MOPITT and SCIAMACHY. Instead, and to keep the study tractable, we focus on the evaluation of the uncertainty in the NR, and compare it to published studies using the MOCAGE DA system. This in line with standard practice in AQ OSSEs as discussed in Timmermans et al (2015).

As it stands, I'm still suspicious of the overall performance. Furthermore, we don't know how well S5p will perform given other sensors, e.g., MOPITT, CrIS, are already taking CO data with comparable performance.

In this AQ OSSE, we follow standard practice by comparing the performance of S5P against a free model run. If the proposed satellite data are to have added value, they must perform better than a model. This is the first step one must take to evaluate the added value of a proposed satellite instrument. We will mention this in the revised text.

Otherwise, the overall work is reasonable and the authors have performed some nice statistical analysis of the results. Of course, in principle, this OSSE should have been performed *before* S5p was funded to assess its potential. But, practice is still catching up with theory. I've attached comments of the manuscript in the accompanying pdf.

We thank the referee for this comment. We address below the other points from the referee.

4-1
Need to include CrIS.

We will add information on CrIS

4-2
What about planned?

We will check if there are plans for GEO missions to measure CO.

4-3
Why is methane defined twice?

We identify that the formula for methane is CH_4.

4-4
How is this point relevant if you're not discussing geostationary options?

We think it helps the reader to contrast the characteristics of GEO and LEO satellite platforms with respect to atmospheric composition.

5-1
Relative to what, MOPITT? Not clear how since they both have NIR channels.

We mean that the S5P with its SWIR band will do better than our model in the PBL. Furthermore, compared to TIR instruments such as IASI, we expect S5P to do better in the PBL (Veefkind et al., 2012). We will mention this in the revised text.

5-2
Poorly formed sentence. And not quite true. The resolution of an inverse emission estimate is controlled by the data and transport/diffusion. Not clear which is the limiting factor without analysis.

We will reword the sentence and take account of the referee's comment.

5-3
True, but how is relevant?

We think it is helpful to remind the reader where the PBL is located.

5-4
So, TROPOMI needs to be assessed relative to the existing satellites.

See response to the general comments from this referee.

6-1
So, if one knew CO concentrations or emissions better, what societal or scientific benefit would have been achieved? Improved forecasts? Better attribution?

We will provide here an example of the benefits of improved knowledge of the CO distribution.

7-1
Shouldn't the control run include assimilation of the existing observing system, e.g., MOPITT, AIRS, CrIS, rather than a free run? This would be true for your 2003 test case when MOPITT and AIRS was available.

See the response to the general comments of this referee.

7-2
I shouldn't have to look up the figure. How does it relate to the specific OSSE elements listed?

We will provide this figure and relate it to the OSSE elements listed (this will be Fig. 1 in the revised paper). We will renumber the other figures.

9-1
Did I miss something? Where is the CR described?

We will refer to section 2.4, which describes the CR.

9-2
There is also some high frequency component that is missed. What is that frequency? is that the night time values? Needs to be discussed.

We will discuss this high frequency component.

9-3
That statement needs to be limited to the ultimate performance of the OSSE. The assimilation can't saying anything better than 10-20% in accuracy.

We will rephrase this sentence following the referee's suggestion.

10-1
At what accuracy?

We will quote Veefkind et al (2012) on this point: 15% (accuracy) and 10% (precision).

10-2
How low?

We will quote Veefkind et al on this point (the value is 2%)

10-3
More accurately compared to what?

We will rephrase the sentence.

Original sentence reads:

"The use of S-5P CO total column measurements with inverse modelling techniques will also help quantify more accurately biomass burning emissions and map their spatial distribution."

A simple solution would be: remove "more accurately"

The statement refers to the current observing system, consisting of, e.g., IASI, MOPITT, OMI, GOME-2, including measurements of the species CO and NO_2. S5P will provide global coverage, enhanced sensitivity for CO at the surface (compared to, e.g., IASI) and 3.5 to 7 km high spatial resolution observations.

11-1
But OMI is quite a bit larger in footprint than TROPOMI. MODIS would be a better choice. Describe how the cloud fraction is related to the 7km footprint. How different are cloud ODS between the UV and the NIR? It seems like you are assuming they are the same.

The cloud fractions were derived at the resolution of the ECMWF 0.25 x 0.25 degree grid. This is ca. 30 x 30 km^2 at the equator and decreases as a function of latitude. The ground pixel of OMI UV-2 and VIS channels is 13 x 24 km^2 at nadir increasing to 13 x 128 km^2 at edges of the swath. We consider that the ECMWF grid cells and OMI pixels are of comparable size for the purpose of comparing the cloud fraction distributions (ca. 0.5 million pixels or cells in each distribution). We model clouds with a simple Lambertian reflectors and ignore any wavelength dependency of cloud fraction.

We will include this information in the revised version.

12.1
Provides

We think provide is appropriate

12-2
Is it released now?

Yes, it is available from the ESA Sentinel-5P TROPOMI document library:

https://sentinel.esa.int/web/sentinel/user-guides/sentinel-5p-tropomi/document-library.

We will provide this information in the revised paper.

12.3
sea

We will correct this.

13.1
Cloudy

We understand clouded and cloudy are both appropriate here. If insisted upon, we will change this.

13-2
It looks like you are assuming that the retrievals will work under partially cloudy scenes. MOPITT NIR only works under clear sky. Provide a reference on NIR CO retrievals under partially cloudy conditions and justify why a weighted approach works. Also I haven't heard any discussion of aerosols. These will be important for emissions like biomass burning and industry.

TROPOMI NIR CO retrievals in partially cloudy conditions are discussed in Landgraf et al.: “Carbon monoxide total column retrievals from TROPOMI shortwave infrared measurements”, Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2016-114. They are also discussed in Vidot et al.:”Carbon monoxide from shortwave infrared reflectance measurements: A new retrieval approach for clear-sky and partially cloudy atmospheres”, Remote Sens. Environ., doi:10.1016/j.rse.2011.09.032. We will add these references to the paper. We did not include aerosol effects in our study.

13-3
This is the first time MOCAGE is mentioned even though it is implicitly referenced in the CR run. Needs to be mentioned earlier.

At the start of section 2, when we first mention the CR, we will introduce MOCAGE and provide appropriate references.

16-1
That is not exactly true. The existence of burning and the burnt area can be obtained from optical measurements. That is very important a priori information. Is that being ignored?

The visible information on burnt area and burning does not provide knowledge on CO concentrations. We will add this information in the text.

17-1
In the introduction, the authors argued for the value of TROPOMI to resolve emissions. Here the focus has shifted to concentration estimation. What is the scientific rationale? What are the limitations of this OSSE, then, to make statements about resolving sources?

We will reword the text in the introduction to state that we focus on CO concentrations. We will clarify the scientific rationale of the paper and indicate the limitations of the OSSE.

21-1
The correction in the free trop points to the role of boundary conditions in the assimilation, which would be an important consideration for a GEO. How much change is occurring at the boundary of the nested grid?

First, note that the OSSE concerns S5-P, which is a LEO. Nevertheless, we focus on the surface level and we assume that the effect of the free troposphere on the boundary layer is secondary. Due to the revisit time of S5-P, we expect the impact at the boundaries to be small. Second, for efficiency reasons and storage limitations, we set up our DA system to only store the data over the regional domain. This means that without rerunning the OSSE, we cannot quantify the response to the reviewer's comment. If requested by the reviewer, we could rerun the OSSE for a short period and address the reviewer's question.

21-2
That is not obvious. The biggest weakness is using 3D-var and having poor prior statistics. Given that significant fires generally last longer than a day, a proper inversion system would pick those up. That does not diminish the value of GEO sounders.

We will clarify that we are talking about concentrations and not emission inversions.

21-3
Please elaborate as to whether it is the diurnal sampling or the effective sampling density that is more important for a geo.

The comment from the referee is not clear to us. We think that diurnal sampling (high temporal resolution) will be the determining factor, as the relatively coarse model resolution would compromise the high spatial sampling.

21-4
That's too qualitative. Could please provide some simple metrics, e.g., means, to quantify these statements?

We will quantify this difference in the revised paper.

23-1
Why?

Owing to the relatively small variability of CO over remote land regions, the S5-P data can provide a larger benefit compared to regions where the variability is relatively high. We will make this point in the revised version.

24.124
Unbiased

We will correct this.

24-2

Why is this called a systematic error? It looks like you are merely removing the bias term in the error, which is simply a statement that the assimilation is not an unbiased estimator.

When calculating the RMSE we remove the bias between the AR and the NR and between the CR and the NR. We then make the common equivalence between systematic error and bias. We will clarify this point in the revised version.

25-1
However, both the CR and AR miss the high frequency min/max. Why?

The AR and the CR capture the variability but not the values of the peaks. However, the LEO only samples at most twice a day over Paris and may not capture the peaks. We indicate the S5-P revisit time by the plus signs at the top of the panel and when you zoom in one sees that the peaks do not coincide with the time of the S5 P measurements. Thus, S5P cannot capture the value at the peak. Another factor could be that the emission inventory used in the AR has lower values than the one used in the NR. We will clarify this in the revised version.

25-2
This is a weakness of the OSSE design, not the measurements.

Maybe we have mis-understood the referee's comment, but our view is that because we do not know the fires a priori, we cannot include them in the CR and the AR. In our view, this result shows the benefits from the measurements regarding the identification and quantification of fire emissions.

25-3
Please explain why the variability is high in Paris but not in E. Europe.

The variability is higher over Paris than over E. Europe, because there are higher emissions over Paris than over E. Europe (as shown in Fig. 7 – old paper submission). We will clarify this in the revised version.

25-4
This seems like a discovery for the authors post-facto. I recommend using this version of the OSSE only rather than devoting a whole section to it. It's what should have been done originally.

We agree with the reviewer. However, we think it is relevant to present the results in this way because it shows the limitations of using standard operational criteria as we did in the first experiment of the OSSE. We will make this point in the revised version.

26-1
The first several paragraphs are repetitious of the introduction. I recommend removing. In fact, the OSSE has shown that S5P will have a similar or bigger impact on the free trop rather than merely the surface.

We will edit the conclusions following the reviewer's comments. The focus of this study is the surface; however, a study of the increments (see figs. 6 and 8 in the old paper submission) indicates that S5-P has benefits in the free troposphere. We will mention this in the revised version.