Original reviewers' comments in blue. Authors' responses in black.

Authors' Replies to Comments of Reviewer #1

In this paper the MOPITT v6 multi-spectral product is validated with aircarft profiles flown over the Amazon basin. This is an important subject and the results indicate a significant negative bias in retrieved lower-tropospheric CO concentrations. This is important for the use of MOPITT data in an inverse modelling framework, e.g. to infer biomass burning emissions and interannual variability therein.

De paper is well written and consice, and I have only a few major remarks, which I list below.

We appreciate the Reviewer's comments and have revised the manuscript as described in detail below.

1. Role of the prior profile

As the authors clearly outline, the MOPITT product comes at 10 vertical layers, and depends on the averaging kernel and the prior profile (a multi-year model-based climatology). First, I was confused by the fact that no mention is made that profiles are used as log(VMR). I think this should be added. In the comparison with the aircraft profiles, the authors mention the role of the prior, especially in the TIR-only product. In fact they mention that extra noise is added by including NIR observations. A valid question to be asked is: “What is now exactly the added value of MOPITT in observing CO biomass burning plumes?”. I think this question can at least partly be answered by presenting the skill of the prior data to explain the aircraft data. How would figure 4/5 look for the prior climatology? The prior data are already presented in figure 7, so it would be relatively easy to construct a comparison between aircraft CO and the prior. Since the role of the prior is important, it would be good to give a bit more explanation about how the prior has been evaluated and its time resolution (monthly?). Finally, at specific locations (e.g. line 131) it would be good to mention the value of the prior.

In the revised manuscript, the paragraphs immediately preceding and following Eq. 1 now clearly state that retrievals are performed in terms of log(VMR). With respect to the effects of the a priori on the validation results, we have added a new subsection to Section 3.2 in which the a priori influence is explicitly removed from the validation analysis. While this reduces the correlation coefficients, the correlations are still robust. New figures are attached below.

2. Comparison of the interannual varibility to bottom-up variability

The paper starts with a brief explanation of “bottom-up” emission inventories, like GFED and FINN. Also, the conclusion section starts with these datasets. In that respect, the paper could provide slightly more context by presenting a comparison with these datasets. One option would be to compare the (scaled) seasonal cycles and interannual varabilty in the MOPITT data (figures 7 and 8).

It is not clear what would be learned by direct comparisons between MOPITT products for CO concentrations and bottom-up emissions inventories. Bottom-up inventories are highly uncertain, and a direct comparison of MOPITT CO products with these inventories would not account for the effects of dynamics and chemistry on atmospheric concentrations. Inverse modeling is the most appropriate methodology for estimating CO sources using MOPITT data. Such an effort is underway for quantifying CO emissions from the Amazon Basin, but is not described in this manuscript.

Minor points:

Line 26: “To account for the chemistry and dynamics that affect trace gas concentrations after their production, inverse modeling methods are exploited”. I think the main issue of inverse modelling is to relate emissions to observations. This relation indeed involves transport and chemistry. I suggest rewording this.

This text has been reworded to "Inverse modeling methods combine atmospheric observations with chemical transport models to relate emissions to atmospheric concentrations."

Line 85: Here it would be good to provide more information about the prior (time resolution, climatology, which model?)

Details about the model-based prior have been added to Section 2.2.

Line 126: This is confusing. Why not present the second overpass after the first overpass?

The text and figures have been revised to discuss the two overpasses sequentially.

Line 131: include values of the prior (e.g. refer to figure 7).

Surface-level a priori values have been added to the text.

Line 154: Maybe some explanation what causes this would be in place?

The origin of this effect (i.e., the dependence of the weighting functions on CO concentrations) is described earlier in the same paragraph. Readers are directed to the Deeter et al., 2015 reference for further discussion.

Line 210: Is this the same climatology as the prior? Yearly, monthly?

Yes, this is now clarified.

Line 257: The correlation coefficients for the three retrieval levels decrease with increasing altitude, from 0.94 at the surface to 0.82 at 600 hPa. In the plot I see values of 0.98 and 0.86. Please check.

We thank the reviewer for finding this error; it has been corrected in the revised manuscript.

Line 260: Here you could add that the dynamic range (max-min, see figure 7) largely explains the high r-values. However, the noise in MOPITT data lowers the r-values I guess. Would a time series provide additional information?

The influence of the a priori on the correlation coefficients is addressed in detail in the revised manuscript. While eliminating the a priori influence reduces the correlation coefficients, the correlations are still robust.

Line 278: “Therefore, the larger standard deviation associated with the TIR-NIR product does not by itself necessarily indicate lower retrieval quality”. Indeed, but here I wonder about the performance of the “prior”. I have the impression that the NIR-data induce quite some noise in the MOPITT product.

It is true that the inclusion of NIR radiances in the retrieval increases random retrieval error, however, it also decreases smoothing error (i.e., the inclusion of NIR radiances produces more optimal averaging kernels). Random errors can be reduced by data averaging, whereas averaging has no benefit for smoothing error.

Line 314: I missed in the description where these T-profiles used in the retrieval come from.

Temperature and water vapor profiles in the MOPITT Version 6 product are extracted from the MERRA reanalysis. This is now stated in Section 3.3.

Line 361: “analyzing Level 3 monthly-mean products is much more efficient than for Level 2”. If this would introduce errors in the analysis, I still would prefer the level-2 analysis. Efficiency is not a good argument.

For analyzing monthly-means, there are no new types of errors introduced by the use of Level 3 data instead of Level 2 data.

Figure 4/5 caption: Are the statitics based on the log-log fit, or on the lineair concentration scale?

Validation statistics are based on log(VMR), to be consistent with the retrieval state vector. This is now stated in the caption to Fig. 4. Biases in log(VMR) correspond to relative errors (e.g., percent) rather than absolute errors (e.g., ppbv).

Authors' Replies to Comments of Reviewer #2

This paper has 2 main foci: (i) an evaluation of MOPITT TIR and TIR-NIR CO retrievals over the Amazon Basin; and (ii) an analysis of seasonal and interannual variations of CO over the Amazon Basin since 2002. Both of these issues are of scientific relevance. However, the manuscript is lacking in some important aspects in terms of both these issues. My major concerns are:

1) It is unclear as to the extent to which the generally good agreement and correlations shown in Figures 4 and 5 arise because of the influence of the prior when aircraft data are smoothed with the MOPITT averaging kernel. Typically, in comparing insitu data (or model output) to satellite retrievals, comparison points are carefully selected to excluded instances when the prior has a significant influence on the posterior. In fact, the MOPITTT users’ guides themselves recommend the use of an Observation Quality Index (OQI), with thresholds depending on the application, to filter out retrievals with large "observation-dependent" noise values. It is not clear that any such quality control was done in the analysis presented here, and it is therefore unclear whether the statistics presented with regards to bias and standard deviation are scientifically robust.

We intentionally avoided data filtering because of the lack of an objectively-defined filtering threshold. With respect to the effects of the a priori on the validation results, we have added a new subsection to Section 3.2 in which the a priori influence is explicitly removed from the validation analysis. While this reduces the correlation coefficients, the correlations are still robust. New figures are attached below.

2) The analysis on the seasonal and interannual variation is quite superficial. The seasonal analysis is presented in terms how the seasonal cycle of the retrievals differs from that of the prior. While this is true, a more important issue is whether the V6 retrievals provide a qualitatively or quantitatively different picture of the seasonal cycle compared to previous studies based on MOPITT data or other independent data. Similarly, while the interannual variation is presented in terms of interannual variations in monthly mean anomalies, there is no information presented as to how these variations compare to independent estimates of interannual variations of biomass-burning emissions of CO in the region. For example, does the year-to-year percentage variation in the MOPITT TIR-NIR surface retrieval during the dry season correspond to percentage variations in bottom-up biomass burning CO emission estimates in this region? What is the cause of the strong anomaly in Nov-Dec in 2015? A more robust analysis is needed to get a sense of the added value provided by MOPITT V6 retrievals in this region.

While they are not the main focus of our work, we believe the analyses of the Amazonian CO seasonal cycle and interannual variability described in Section 4 reveal new information about CO variability over the Amazon Basin and therefore add value to the manuscript. The seasonal cycle analysis compares MOPITT data against the mean annual cycle derived from a multi-year run of the CAM-chem model using GFED2 emissions; this climatology serves as the MOPITT prior. The analysis of CO interannual variability demonstrates that the MOPITT record clearly exposes CO anomalies during drought years superimposed on a long-term trend of decreasing CO concentrations.

It is not clear what would be learned by direct comparisons between MOPITT products for CO concentrations and bottom-up CO emissions inventories, since (1) such inventories are characterized by large uncertainties and (2) atmospheric concentrations of CO generally depend on meteorological and chemical processes as well as CO sources. Discrepancies in such a comparison would result from a variety of effects having nothing to do with the quality of the MOPITT data. Inverse modeling is the most accepted methodology for estimating CO sources using satellite retrievals. Such an effort is in fact underway for quantifying CO emissions from the Amazon Basin using MOPITT data.

Other comments:

I assume that the points shown in Figures 4 and 5 represent individual aircraft flights. No rationale is provided for the specific choice of the 200 km and 24 hour windows used to select MOPITT retrievals for comparison with these aircraft flights. How sensitive are the comparisons to alternate choices of these windows?

The collocation criteria used for this work are based on previous experience. The selection of collocation criteria involves a tradeoff between the number of analyzed retrievals, which affects the statistical robustness, and errors due to CO spatial and temporal variability. This is now stated in the manuscript.

Based on analysis of the retrieval error as a function of AOD, the authors rule out the possibility that retrieval bias is related to high aerosol loading. Can the authors speculate on other possible causes of the retrieval bias?

Our best guess for this effect is stated at the end of Section 3: "The greater standard deviations observed for the AOD >= 0.5 subsets might be explained by stronger CO geographical and temporal variability during the Amazonian dry season, however this is only a hypothesis."

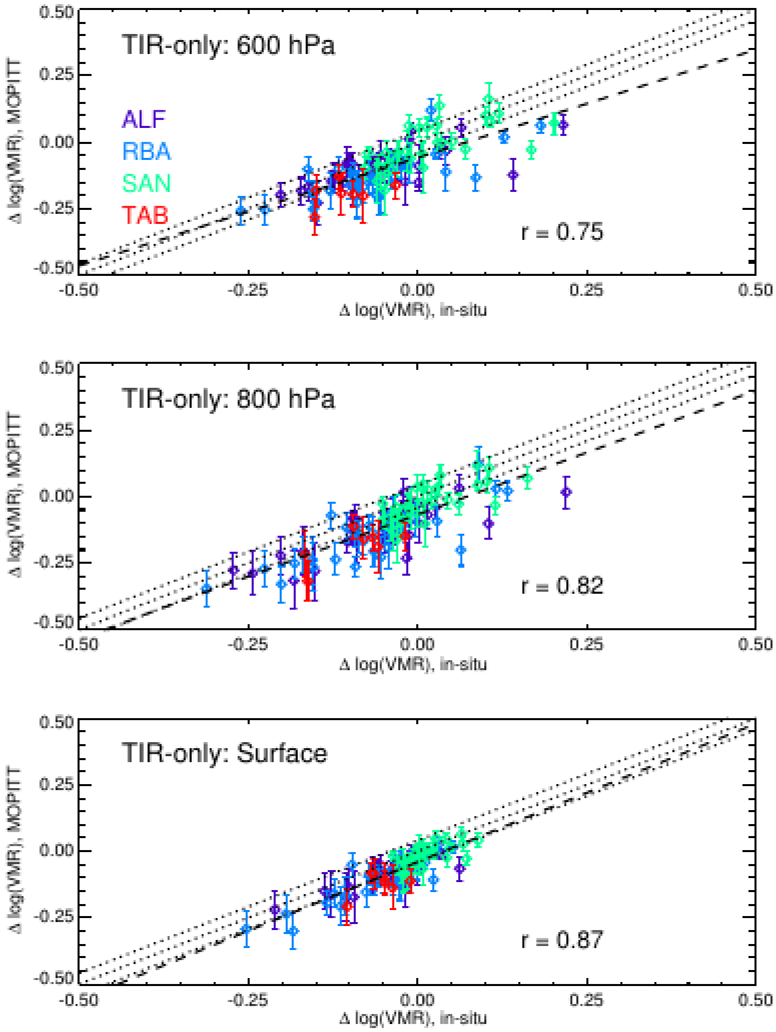

Figure 6. MOPITT Version 6 TIR-only validation results (as presented in Fig. 4), with a priori influence removed from both the MOPITT retrieved values and simulated retrievals based on the in-situ data. As discussed in Section 3.2.3, plotted values correspond to the difference of retrieved (or simulated) log(VMR) values and the a priori log(VMR) value.

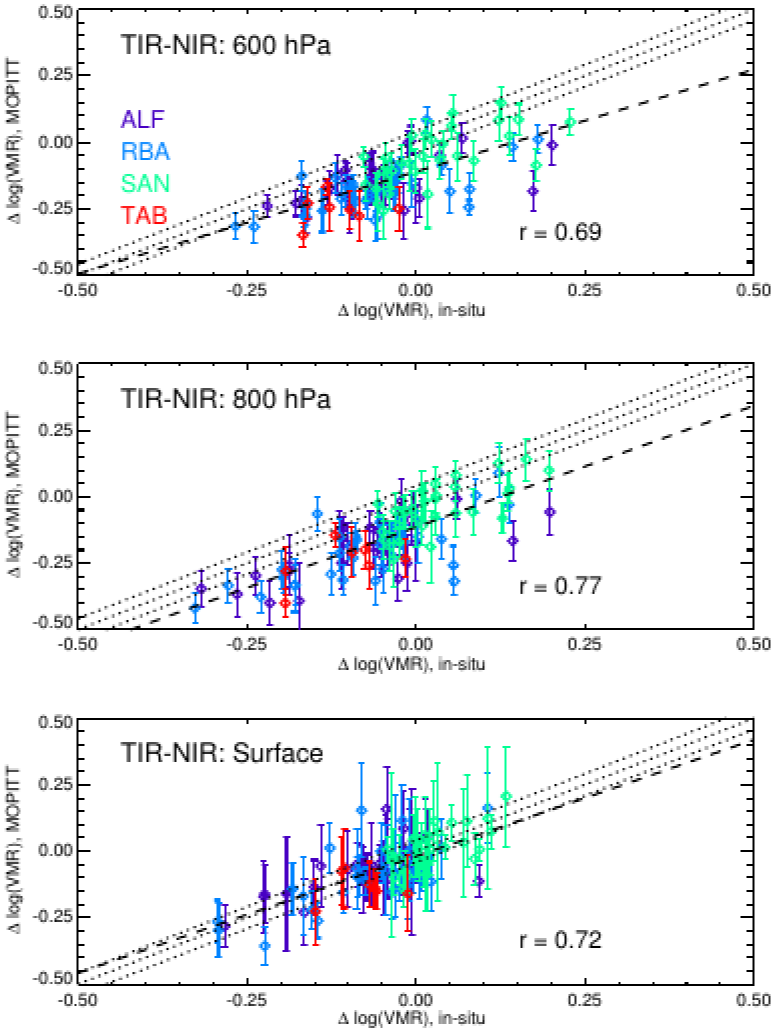

Figure 7. MOPITT Version 6 TIR-NIR validation results (as presented in Fig. 5), with a priori influence removed from both the MOPITT retrieved values and simulated retrievals based on the in-situ data. As discussed in Section 3.2.3, plotted values correspond to the difference of retrieved (or simulated) log(VMR) values and the a priori log(VMR) value.