Authors' Responses to Reviewer #3 of 'Radiance-based Retrieval Bias Mitigation for the MOPITT Instrument: The Version 8 Product' by M. Deeter et al.

Original reviewer's comments in blue. Authors' responses in black.

Replies to Comments of Reviewer #3

This paper describes version8 of the MOPITT CO retrievals and of its associated data record. Despite being short and technical, the paper deals with an increasingly important aspect of satellite remote sensing, which is the consistency and homogeneity in the time series at level 2. By targeting CO, a so-called indirect climate variable, and relying on the longest time-series available from satellites (almost two decades with MOPITT), this paper is a very useful reference. To achieve this consistency in the CO time series for version 8, an empirical and "dynamical" radiance correction scheme is proposed. The methodology is well described and the results give convincing evidence of its benefit in V8 over previous versions. New validation results using recent aircraft measurement campaigns (from ACRIDICON-CHUVA, KORUS-AQ and Atom) are given as well, and these provide an additional support for the proposed radiance bias minimization scheme. The paper is extremely clear and very well written. For these reasons, I recommend publication of the paper in AMT. I list below a few comments or questions that the authors may want to consider to possibly improve or clarify some aspects of the paper.

The authors appreciate the reviewer's useful comments.

- Page2, discussion around Figure 1: the drifts compared to the measurement flasks are mostly positive below 500 hPA and negative above. In the total columns this results in almost no drift. This could be indicated from the start (around line 25).

Response: The following sentence has been added in the fourth paragraph of the Introduction: 'However, opposing drift in the upper and lower troposphere appears to mostly cancel with respect to the retrieved total column [Deeter et al, 2013].'

- Page 2, line 30: To explain the latitudinal-dependent bias in the V7-TIR only CO, the water vapor field is suspected to be a cause. From Figure 3 it seems indeed obvious that water plays a role and this is one of the reason for the bias correction » suggestion to remove "perhaps" on line 30?

Response: The word 'perhaps' is used since either of two quite different effects (errors in the radiative transfer modeling of water vapor or, alternatively, the accuracy of water vapor profiles in the MERRA-2 reanalysis) could conceivably explain the results.

- Would a figure showing the (global? hemispheric?) averaged retrieved profile from v7 and V8 with the associated standard deviation not be insightful? Is this possible and does this add something?

Response: We think this is an interesting idea, but have not yet pursued it. Ideally, such an analysis should involve V7/V8 comparisons at different stages of the MOPITT mission (to expose the effects of bias drift minimization) and for multiple regions and seasons (to expose the effects of water vapor corrections). Differences between V7 and V8 might also vary with surface type (land or ocean) and thermal contrast (day vs night). Such an analysis would significantly add to the length of the paper.

- Page 5 top; it would be useful to see how to new way to calculate the averaging kernels in v8 affects the averaging kernel calculations. Also, is there an impact on the validation between V7 and V8 that could potentially be related to the AVKs calculations?

Response: The following sentence has been added to the last paragraph of Section 2.3: 'While the new method for calculating the total column averaging kernel is more rigorous than the previously used method, resulting differences in total column validation statistics (correlation coefficient, bias, and standard deviation) were found to be insignificant.

- Page 5, line 10. Remove parentheses for that sentence?

Response: Agreed.

- Page 5, equation 5. Is x_true different than x_cmp from eq. (3)? I understand that x_cmp is more general but for this paper is the distinction useful?

Response: These two terms are really not interchangeable. We begin using x_true starting with Eq. 5 since validation involves comparisons with some type of 'truth' as the reference. In cases where MOPITT data are used to evaluate models, MOPITT may actually be the reference, and therefore it would be inappropriate to use the term x_true for the model.

- Page 6, last sentence before section 3.1; I would suggest to move this sentence earlier in section 2.3

Response: Agreed.

- Figure 5 is similar to Figure 1 for the new version. When discussing Figure 5 on page 7 it would be good to refer the reader to Figure 1 for a useful comparison

Response: Agreed.

- Figures 5 and 6 are convincing on what the dynamic bias-correction brings. Nice results; well Done.

Response: Thanks!

- Figure 8 compares the new CO profiles from V8 to more recent reference measurements from aircrafts. As said before, this is useful and again convincing. It would, however, be nice to show (or at least comment) how V7T performs compared to these reference dataset (in comparison to V8T).

Response: We feel that the V7/V8 comparisons using the NOAA and HIPPO datasets are sufficient to establish the improved performance of the V8 product, which is the main subject of the manuscript.

- Page 8. The last sentence before section 3.3 refers to biases at high northern latitude but looking at Figure 8 one has the feeling that there are almost no measurements in the latitude band 60-90 N. Is that the case and could you please then indicate how significant this bias is?

Response: We agree that the sparseness of in-situ profiles at high northern latitudes during ATom decreases the statistical significance of the corresponding validation results - the large error bars for the latitude band 60N-90N in Fig. 8 may well be the result of a small number of 'outliers'. We have therefore revised the text in Section 3.2 to read: 'Biases outside of this range are most evident between

60 N and 90 N. However, this could be related to the sparseness of profiles in this region and the influence of a small number of outliers.'

- Comparing Figure 12 and Fig8 one sees that the biases in v8 are larger for the TIRNIR retrievals than for the TIR only. Was this feature also there for v7? Is there a known reason for this?

Response: Both systematic errors and retrieval noise tend to be larger in the TIR-NIR product due to the strategy to amplify the weight assigned to the NIR measurements. Text explaining this point, along with an additional reference, has been added to Section 3.4.

- Check if the reference Worden et al., 2013 is cited in the text.

Response: This reference should have been cited at the end of the fourth paragraph of the Introduction. This has been corrected.

- Caption of table 2: correlation coefficients are given between retrieved quantities and corresponding a priori quantities. Is it a priori or reference?

Response: That statement in the caption is meant to clarify that correlations are calculated for the quantities (x_rtv - x_a) vs. (x_sim - x_a), rather than x_rtv vs x_sim. Subtracting the a priori from x_rtv and x_sim before making the correlation calculation eliminates 'false correlations' that are just related to variability in the a priori. A sentence has been added at the end of the first paragraph in Section 3.2 to clarify this point: 'Correlations due simply to the variability of the a priori are avoided by basing correlation coefficient calculations on (x_rtv - x_a) rather than x_rtv.' The issue is no longer discussed in the caption.

- Would it not be possible to make a single table out if Tables 4-6? This would be nice and also helpful to easily compare the results from each validation set.

Response: Agreed.