Authors' Responses to Reviewer #1 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 #1

This work is a useful description of the improvements introduced in the latest version 8 of CO data derived from measurements recorded with the MOPITT instrument onboard Terra. The empirical foundation of the detected drifts and implemented countermeasures is the comparison of MOPITT data with airborne in situ observations of CO collected regularly by NOAA and during several campaigns. The paper is well written, in some parts perhaps slightly too concise (see minor comment at the end of this review) and the statistical analysis is carried out in a careful manner. I recommend publication after minor revisions. I have only two reservations concerning the presented work worth mentioning:

(1) It would be a valuable independent confirmation of the significant reduction of drifts claimed for version 8 to incorporate the rich dataset of CO provided by ground-based FTIR networks TCCON and NDACC, as these observations also cover the complete lifetime of the satellite. The TCCON data would be useful for checking the spurious trends of CO total column in version 7 and version 8, the NDACC observations even offer a moderate vertical resolution. I agree that one will assume that the in-situ dataset is the primary reference (as also used for calibrating ground-based remote sensing observations), but the remote sensing network observations cover additional locations and times. Moreover, by the inherent character of a spectroscopic solar absorption measurement performed by a distributed network, a common drift of the network as a whole is highly unlikely, so these datasets seem a useful asset especially in the context of trend evaluation.

Response: We agree that comparisons between MOPITT and TCCON/NDACC could be useful for validation of trends, but we consider comparisons against in-situ profiles more definitive. Significant differences in column averaging kernels for MOPITT and TCCON/NDACC retrievals complicate the comparisons considerably and inevitably raise questions related to the possible influence of long-term changes in CO concentrations on apparent biases (and drift) between MOPITT and TCCON/NDACC. The use of in-situ profiles for validation avoids this issue entirely. Ongoing MOPITT validation work includes comparisons with TCCON (Hedelius et al., in prep) and NDACC (Buchholz et al., AMT, 2017 – to be updated for V8) as well as other long-term satellite CO datasets (e.g., IASI).

(2) For the non-experts (including myself) with just a schematic understanding of the working principle underlying the MOPITT instrument, the idea of the proposed radiance-based correction remains somewhat obscure. The explanation provided “potential bias sources including errors in instrumental specifications, forward model error. . ., errors in assumed spectroscopic data, and geophysical errors” is focusing mainly on aspects of the data analysis and the spectral simulations required in this context (as does the provided reference Deeter et al., 2014). I understand that model errors will evoke retrieval errors, however, as a steady trend is observed this seems to be an instrument issue? Could you make a clearer statement in this respect?

Response: We refer to Drummond et al., 2010 for more details on MOPITT measurement concepts. Possible sources of the bias drift are briefly discussed in the second paragraph of the Conclusion. While an instrumental source is quite possible, it is also conceivable that the quality of the MERRA-2 meteorological data (temperature and water vapor profiles) used in the retrieval algorithm has varied over the MOPITT mission (due to the assimilation of an ever-evolving set of in-situ and satellite datasets). We feel it would be premature to make further statements on this topic.

Might it be even possible to provide further speculations concerning what is actually drifting here, detectors, sources, window transmissions of gas cells, . . .?

Response: Collaborations with the instrument developers will be essential to identify possible physical sources to the bias drift, if in fact the bias source is instrumental. Preliminary discussions with the Canadian MOPITT team (who designed the instrument) have taken place, but no conclusions have yet been reached. We do not feel that it would be appropriate to speculate further on this issue in this manuscript.

Although the improvements in the version 8 data are obvious, such kind of intervention by incorporating an explicit time-dependent correction is to some degree problematic. The previous publication on version 7 data was entitled "a climate-scale satellite record for carbon monoxide". Unfortunately, as version 8 introduces an ad-hoc time-dependent correction, the confession shines through that MOPITT is not capable of measuring CO trends with higher confidence than 0.5 to 1% / a on a decadal scale, as an ad-hoc correction is needed for adjusting the results (so the trend on decadal scales is actually determined by the aircraft data, not MOPITT). If some sort of instrumental degradation could be identified and quantified independently from atmospheric observations using routine onboard calibration measurements (I do not know the procedures – observations of internal blackbody sources or a solar diffuser?) and finally be made responsible using a quantitative error-propagating scheme built on an instrument model, this point of criticism could be removed. It might be appropriate to include a deeper discussion of these issues.

Response: Our chosen strategy to rely on the stability of the NOAA aircraft in-situ data to determine the time-dependent radiance bias correction factors does imply that the fidelity of CO trends in the radiance bias-corrected MOPITT data (i.e., the V8 retrieval product) are ultimately limited by the stability of the NOAA in-situ measurements. However, the NOAA measurements are widely accepted as the standard for long-term CO analyses, and the stability of the NOAA flask-sample measurements are certainly much better than 0.5 to 1% per year. The NOAA dataset is therefore quite suitable for climate analyses. On the other hand, we acknowledge (e.g., in Section 3.2) that bias drift in the V8 product for regions not represented by the NOAA aircraft network (primarily covering North America) could be substantially different. This seems unlikely though, particularly if the source of the bias is instrumental. We are currently investigating methods for quantifying MOPITT bias drift globally, using field campaign measurements as well as both ground-based and satellite-based remote sensing datasets.

Minor comment: Facing this multitude of validation datasets is a bit confusing for the reader. Although all activities and campaigns are cited appropriately, it would be useful to provide some kind of tabulated or graphical overview in order to summarize the temporal (and perhaps spatial?) coverage of each dataset. This would also incorporate useful information on the relation of the subset of AC, KORUS-AQ and Atom with the datasets used for the development of the bias minimization.

Response: A table has been added (introduced in Section 3) in the revised manuscript summarizing these aspects of the different in-situ datasets.

Finally, in the several tables reporting bias values, SD, r, (and drifts in case of table 2) . . . it would be useful to also specify how many aircraft profiles were used for calculating these statistical parameters.

Response: The number of aircraft profiles for each in-situ dataset is included in the new table.