Anonymous Referee #3

General comments:
The aim of this paper is to evaluate two versions of MOPITT CO (V7 and V8) by comparison with aircraft observations from diverse campaigns all over the globe. Each version has two sub versions (V7-8T, V7-8N, V7-8J for thermal, NIR and TIR+NIR, respectively). Urban and non urban areas are the focus of the evaluation. This is a paper that complete the list of publications of the evaluation of the different versions of MOPITT CO. Lots of statistics are provided and the MOPITT users community could find some interest in order to interpret MOPITT data over urban areas.
Response: Thank you for your time and effort in reviewing our manuscript.

However, I found the comparison sometimes difficult to follow because of the large number of campaigns, the number of aircraft profiles by campaign, number of aircraft profiles over urban regions, the number of MOPITT CO profiles in different circles,.. Table 1 helps but if possible it would be nice to simplify in the text. Moreover, the title does not reflect totally the subject of the paper: the validation of the MOPITT CO retrievals is also over non urban regions. I suggest to change the title in that way.
Response: Please see the responses to the comments # G2 and S1 of the reviewer 1.

1) Moreover, the distinction of urban and non urban regions for the comparison of MOPITT CO with aircraft observations could mislead the reader. What is important in this study, is it the carbone monoxide emitted from the urban region or just the urban region with surface parameters different from non urban regions? Such surface parameters that are used in the retrievals of MOPITT CO (surface temperature, emissivity). At 600 hPa, some comparisons are done but this is above the boundary layer. There is a great chance that the CO measured by both MOPITT and the aircraft is transported from other regions that are not representative of urban regions. The author should clarify this point.
Response: We thank the reviewer for bringing this question up. The urban regions often have different surface parameters (e.g., surface temperature and emissivity), and usually but not always have higher CO concentrations than non-urban regions. However, the surface parameters are unlikely to impact the ultimate quality of MOPITT retrieval products (Pan et al., 1998; Ho et al., 2005). The goal of this study is to understand if MOPITT retrievals are able to represent conditions over urban regions given sampling, and cloud cover. In addition, the relatively large spatial and temporal variability of CO concentrations over urban regions makes the validation even more complex. Because of the complexity of urban regions and their connection with non-urban regions nearby, we also provide analysis at high CO concentrations regardless of landcover type. As the reviewer pointed out, the comparisons are done for the 600-hPa layer (usually in the free troposphere). It is possible that CO concentrations at this layer are transported from other regions that are not representative of urban regions. Even so, MOPITT retrievals at the 600-hPa layer are still impacted by the CO concentrations at other layers including the surface layer (equation 1). Therefore, the comparisons at 600 hPa is necessary. We have added the discussions above to the section 2.2. See the manuscript for details.

2) Also, it would be nice to have a clear recommandation on which MOPITT CO version to use. For example, after reading the table 2 of the paper, I found difficult to conclude on which version to use for urban or a non urban study as well. The statistics are often very similar and I was wondering what is the added value of V8 vs V7 and how significant the values are? It would be nice the authors discuss this point and conclude with clear recommendation in the conclusions on the use of the different versions of MOPITT CO.

Response: We thank the reviewer for pointing this out. The main goal of this study is not to compare MOPITT V8 and V7 products, but rather to validate the performance of MOPITT products over urban regions versus non-urban regions. The finding is that in general, MOPITT agrees reasonably well with the in-situ profiles over both urban and non-urban regions. As the reviewer pointed out, the statistics are often very similar, therefore we do not have recommendation for which version to use in terms of urban versus non-urban regions.

The MOPITT TIR-only, and TIR-NIR products both have their own advantages and disadvantages. MOPITT TIR-NIR products usually have higher DFSs and have enhanced the sensitivity to near-surface CO but may have larger retrieval noise compared to the TIR-only products (Deeter et al., 2011, 2013; Worden et al., 2010). The MOPITT V8 uses a new parameterized radiance bias correction method to minimize retrieval biases, therefore in general the MOPITT V8 performs better than V7 and is recommended (Deeter et al., 2019). A detailed description of MOPITT V8 products and their comparisons to MOPITT V7 products can be found in Deeter et al. (2019). We added the discussion below to the section 5 of the manuscript:

> *"The statistics are often very similar between different versions and products over urban and non-urban regions, and in general, MOPITT agrees reasonably well with the in-situ profiles in both cases. There is not, therefore, any reason to recommend the continued use of MOPITT versions earlier than V8 based on urban or non-urban region considerations. In general, MOPITT V8 is recommended (Deeter et al., 2019) as it uses a new parameterized radiance bias correction method to minimize retrieval biases, and has updated spectroscopic data for water vapor and nitrogen."*

3) The Section 4.4 (Sensitivity to the signal-to-noise ratio (SNR) filters) is unclear to me. What are the conclusions we can draw from this section? Is level 3 useless? I didn't catch the point of this section. Maybe the authors could clarify on how to use Level 3 data over urban and non urban regions in the light of the use of such SNR filter.

Response: Please see the response to the comments # S37 of the reviewer 1.

Specific Comments:

Abstract:
In the paper, V7 and V8 of MOPITT CO are evaluated whereas only V8 is mentioned in the abstract.

Response: MOPITT V7 products is only used as a reference in the sub-Section 3.1 and is not the focus of this study. To avoid the confusion, we changed the sentence "*We focus on evaluating the recently released Version 8, as well as the Version 7, of the MOPITT TIR, NIR, and multispectral*

TIR-NIR products." in the Section 2.1 (MOPITT retrievals and products) to "*We focus on validating the recently released Version 8 of the MOPITT TIR, NIR, and multispectral TIR-NIR products. We also include comparisons with the MOPITT Version 7 TIR, NIR, and multispectral TIR-NIR products in the Section 3.1 for reference.*"

Section 3.3

L 300-301: This means MOPITT CO concentrations are highly variable in circles where true concentrations are high. In this condition, what are the retrieval errors for these MOPITT pixels?
Response: We thank the reviewer for the question. We have conducted the calculation of the retrieval uncertainties, and added the statement below to section 3.3:

"*At higher 50% CO concentrations, the averaged retrieval uncertainties for the 600-hPa, 800-hPa, and surface layers, are 28%, 28%, and 29%, respectively. This is smaller than the averaged retrieval uncertainties at lower 50% CO concentrations (28%, 29%, and 30% for the 600-hPa, 800-hPa, and surface layers, respectively). We therefore conclude that the larger apparent biases at high CO concentrations are related to greater CO variability and representativeness error of the in-situ profile within the co-location radius used for analyzing the MOPITT data, rather than indicating larger retrieval uncertainties. Theoretically, MOPITT retrievals perform better with higher CO concentrations. The larger biases at high CO concentrations in Figure 7 implies that the relatively greater CO variability may overcome the impact of high CO concentrations. Addressing representativeness error and spatial variability in the comparisons between satellite and in-situ profiles is challenging, and will be discussed further in Section 5.*"

L 334: please correct the sentence
Response: The sentence is changed to "*At the 600-hPa layer, the agreements between the values of MOPITT and in-situ profiles are affected more by P_{interp} compared to the those at the surface layer and the 800-hPa layer for comparisons with all the campaigns.*"

L 360: please correct the sentence
Response: We changed the sentence to "*We note that the usage of the largest radius (200 km) in this paper does not appear to degrade the overall results, even though representativeness errors generated from CO spatial and/or temporal variability are expected to increase. However, the use of the smallest radius (25 km) degrades the overall results by reducing the number of included MOPITT retrievals.*"

L 369: The sentence 'we note..' is unclear to me. Please clarify if necessary.
Response: We changed this sentence to
"*We note that when comparing to the ARIAs campaign, using 1h as the allowed maximum time difference decreases the biases at the surface layer, the 800-hPa layer, and the 600-hPa layer, compared to the cases using longer allowed maximum time difference (i.e., 3h, 6h, and 12h). This implies that the temporal variability is relatively large in the region.*"