Thank you for your comments. Please find our responses below. Page and line numbers refer to the manuscript as submitted for the discussion phase.

Interactive comment on "1.5 years of TROPOMI CO measurements: Comparisons to MOPITT and ATom" by Sara Martínez-Alonso et al.
Anonymous Referee #2
Received and published: 6 April 2020
This manuscript compares TROPOMI carbon monoxide retrievals to data from the MOPITT satellite and in situ airborne profiles (ATom-4). The manuscript is well written and falls into the scope of AMT. I recommend publication after the following comments have been addressed.

General Comments
My main concern is that the significantly different vertical sensitivities of the instruments and the different apriori profiles used in the algorithms are not taken into account in the comparison of the TROPOMI and MOPITT data. It is alleged that the corresponding comparison methodology is not applicable to profile scaling retrievals. However, I do not agree with this view as TROPOMI's averaging kernels (AKs) take into account that it is a profile scaling retrieval. The AK value of the i-th layer quantifies as usual the sensitivity of the total column to a change of CO in the i-th layer. It is also not a question of constraining the results with the apriori or not. If the AKs are not a direct output of the retrieval, you can simply compute them for every kind of algorithm by confronting the retrieval with simulated measurements and doing the following for each layer i: 1) change the abundance in the i-th layer, 2) perform the retrieval, 3) compare the retrieved column to the "true" column. For a meaningful comparison, at least the individual apriori profiles of both retrievals should be replaced by a common prior by using the AKs (see e.g. Section 4 of Rodgers and Connor (2003) or Appendix A of Wunch et al. (2011)). The common prior can be the TROPOMI prior, the MOPITT prior, or a different third prior. Please improve the comparison method by taking these aspects into account or give a justification why the consideration of the AKs is negligible in this analysis and prove by example that the figures of merit like the global bias between the two data sets do not critically depend on whether the individual apriori profiles are replaced by a common prior.

Thank you for this comment. New text and two tables with results from an additional analysis have been included in the manuscript to 1) better justify the direct comparisons without transformation and 2) investigate the effect on biases of the differences between MOPITT *a priori* CO profiles and TROPOMI reference CO profiles. New Section 3.1 discusses in more detail the differences between the MOPITT and TROPOMI CO retrieval algorithms, as well as the challenges these differences impose when comparing the two datasets. New Section 3.3.1 discusses the main sources of error in satellite CO retrievals; it also discusses sources of error when comparing satellite datasets, e.g., differences in *a priori* information used by each dataset and differences in vertical sensitivity (represented by the averaging kernels, or AKs) between instruments.
Determining whether or not observed differences in retrievals from these two instruments are consistent with differences in their *a priori*, AKs, and instrument noise would require knowledge of the true atmosphere during observation; this information is often unavailable, here included. Our main goal in comparing MOPITT and TROPOMI total CO column retrievals is to quantify differences between the two retrieval products available to users, rather than quantify the actual bias of either product. This goal is addressed by direct "end to end" comparisons of the two untransformed products in various

regions of interest, after colocation of the MOPITT and TROPOMI retrievals. These comparisons quantify the MOPITT/TROPOMI difference statistics due to all effects: AK differences, *a priori* differences, and instrument noise.

Additionally, we now investigate the effects of differences between the *a priori*/reference information used by MOPITT and TROPOMI in their retrievals; we do so by applying a null-space adjustment (based on the MOPITT *a priori*) to TROPOMI. We present results from this additional analysis in Sections 4.1.4 and 4.2.3 and show that differences in *a priori*/reference CO profiles affect MOPITT/TROPOMI relative biases by 1-2 percentage points, well below TROPOMI's required 15% accuracy.

Specific Comments

Page 1, Lines 4-5: TANSO-FTS-2 on the GOSAT-2 satellite (launched in 2018) is also deriving CO from solar reflected radiances in the 2.3 μm spectral region. Thank you for bringing this point to our attention; similarly, SCIAMACHY should also have been included in the list of satellite instruments that derived CO from solar reflected radiances. We have reworded the sentence as follows: "MOPITT and TROPOMI are two of only a few satellite instruments to ever derive CO from solar reflected radiances." We have also added an introduction to SCIAMACHY and TANSO-FTS-2 later on in the manuscript, please see below.

Page 1, Lines 16-17: see general comments. Please note that, for focus, we have moved the MOPITT/TROPOMI above and below cloud comparison to the Supplement Materials. Because of this, the Abstract now does not refer to this particular type of comparison. Please see response to General Comments for more details.

Page 3, Lines 49-50: TANSO-FTS-2 on the GOSAT-2 satellite is also deriving CO from solar reflected radiances in the 2.3 μm spectral region. To address this point we have reworded the sentence in lines 49-50 and introduced both SCIAMACHY and TANSO-FTS-2 as follows: "TROPOMI was, until recently, the only other operative satellite instrument retrieving CO from NIR measurements. (ENVISAT SCIAMACHY (2002-2012; Bovensmann et al., 1999) and GOSAT-2 TANSO-FTS-2 (since 2019; NIES, 2019) are two other instances.) Thus, understanding how MOPITT and TROPOMI retrievals compare to each other is important."

Page 7, Lines 157-160: see general comments. The content of these lines regarding MOPITT and TROPOMI algorithm differences has been expanded and clarified in new Section 3.1. Please see our response to General Comments for additional details.

Page 8, Line 195: The units in Eq. (1) do not match.
For added clarity and to show explicitly that units do match, we now provide the units of the constant $2.12*10^{13}$, which are molec. cm^{-2} hPa^{-1} ppb^{-1}. (Please note that Eq. 1 is now Eq. 3.)

Page 9, Lines 223-225: Why not use the actual TROPOMI averaging kernels here instead of a binary step function? Please note that, for focus, the analysis described in Section 3.2.2 has been moved to Supplement Materials. The purpose of the method described in Section 3.2.2. is to calculate the worst-case scenario errors (the maximum errors) that could be introduced by the use of modeled CO in TROPOMI retrievals over water. In a way, the method can be understood better by thinking of a step function, since it is assumed that TROPOMI sensitivity to CO above cloud top would be 1 (i.e., no modeled CO involved), while below cloud top would be 0 (i.e., only modeled CO involved). As stated

in the manuscript, this method would be most accurate in case of optically thick clouds. To explain the motivation for this section better and thus clarify this point, we have reworded the text as follows (page 9, lines 223-227): "The goal of this analysis was to calculate the maximum error caused by the use of reference CO profiles in TROPOMI retrievals over water. To this effect, we assumed that TROPOMI retrievals are only sensitive to CO above cloud top, while CO below cloud top is fully approximated by TROPOMI's scaled reference profiles. This scenario would be most accurate in case of optically thick clouds. To quantify this error, we compared TROPOMI retrievals over bodies of water (total columns and their above cloud partial column components) to their colocated MOPITT TIR counterparts."

Page 10, Lines 242-243: Why not use the actual TROPOMI averaging kernels? Please see response to previous comment. To explain the motivation for this section better and thus clarify this point, we have reworded the text as follows (Supplement Materials page 2, lines 43-44): "The goal of this analysis was to calculate the maximum error caused by the use of reference CO profiles in TROPOMI retrievals over water. To this effect, [...]"

Page 13, Lines 342-343: The negative bias could simply be a consequence of the different sensitivities and apriori profiles used in the estimation of the true atmospheric state for the two individual instruments. Thus, consideration of the averaging kernels is important. Please note for example the change in sign for the biases in Figure 6 due to the AKs. Thank you for this comment. New Table 4 shows that land MOPITT/TROPOMI bias values after accounting for *a priori* differences are very similar to (and retain the same negative sign as) the original bias values shown in Table 1 and discussed in lines 342-343. As discussed in the response to General Comments, the effect of *a priori*/reference CO profile differences on relative biases is very small, only 1-2 percentage points.

Page 14, Lines 378-380: How do these error estimates change when considering the averaging kernels and apriori profiles? Please note that, for focus, we have moved the MOPITT/TROPOMI above and below cloud comparison to the Supplement Materials. Because of this, the Discussion section now does not refer to this particular type of comparison. The manuscript does, however, still include a comparison of MOPITT/TROPOMI total CO column values over ocean plus its MOPITT/null-space adjusted TROPOMI comparison counterpart. New Table 5 shows that ocean MOPITT/TROPOMI total column bias values after accounting for *a priori* differences are very similar to the original bias values (shown in Table 3) and retain the same sign.

Page 14, Lines 384-386: In addition to the global accuracy and precision, it would also be interesting to quantify the regional relative accuracy quantifying region-to-region biases, e.g. the standard deviation of the individual biases for the regions of Figures 2-5. We assume that "region-to-region" (here and in other comments below) means "for each ROI". Please note that the bias (accuracy) and standard deviation of bias (precision) had been quantified for each ROI, both in percentage and in CO column values; results are summarized in Table 1, described in the Results section, and revisited in the Discussion section. (We assume the comment refers to Fig. 3-5.) The precision requirement of 10 % is not satisfied. Thank you for pointing this out. Calculated precision versus required precision had been discussed elsewhere in the manuscript (e.g., page 13 lines 341-342 and 366-367). However, the wording in page 14 lines 384-386 is indeed insufficient. For clarity, the sentence has been reworded to: "Our results show that the accuracy of TROPOMI retrievals with respect to MOPITT and ATom far exceeds Sentinel-5P mission requirements (Veefkind et al., 2012; Landgraf et al., 2016). The precision values calculated for some of the ROIs analyzed surpass the target value by a few percent."

Page 14, Line 390: To validate TROPOMI retrievals over land, ground-based measurements from the Total Carbon Column Observing Network (TCCON), which are calibrated using aircraft profiles, can also be used. In contrast to aircraft data this would also allow the validation of seasonal variability at a fixed location. We agree, thanks for bringing this point up. To address this comment we have reworded/added the following text in the manuscript: "To that end, *in situ* data from other airborne measurement programs are required. Ground-based measurements (e.g., NDACC, TCCON) could also be used; this would allow the validation of seasonal variability at fixed locations."

Page 14, Line 392-393: I would call a comparison to other satellite data sets verification instead of validation. Thank you for this comment. We have reworded this sentence as follows: "The MOPITT dataset represents the longest global CO record available (2000-present); because of extensive validation efforts with respect to *in situ* measurements and comparisons with other satellite datasets, it is well characterized." Has the seasonal variability of MOPITT also been validated, e.g. by using TCCON or NDACC ground-based measurements? MOPITT has been compared in the past to ground measurements, as discussed in page 3 line 52. (In that line, for simplicity and focus, neither the names of these ground networks nor the names of the individual satellite datasets used in previous MOPITT validations are provided.) We have added elsewhere (page 13, line 363) the following text describing these previous efforts: "The seasonal variability of MOPITT has been validated in the past using ground-based measurements. In their comparison to NDACC data (Network for the Detection of Atmospheric Composition Change; De Maziere et al., 2018), Buchholz et al. (2017) found no significant seasonally dependent bias for MOPITT products. Hedelius et al. (2019) compared MOPITT to the TCCON dataset, reporting no persistent seasonal trend globally and some seasonal variability for individual sites." Relevant references have been added to the manuscript.

Page 14, Line 395-396: TANSO-FTS-2 on the GOSAT-2 satellite is also deriving CO from solar reflected radiances in the 2.3 µm spectral region. Thanks for pointing that out. We have reworded the sentence as follows: "Furthermore, TROPOMI and MOPITT were, until TANSO-FTS-2 became operational in 2019, the only working satellite instruments retrieving CO from NIR solar-reflected radiances."

Page 14, Line 399: Please replace "do not fully account for" by "do not account for". Done.

Page 15, Line 408: What about transport of CO from major sources in coastal regions to the ocean? Thanks for this observation. We have address the comment as follows: "Since there are no major CO sources over water, CO values closer to the surface (and, therefore, most likely to be below cloud top) tend to be spatially homogeneous and stable through time. Thus, they are well characterized by the reference profiles. (Caution should be exercise in case of sporadic CO sources near open water, e.g., fires near a coastline, which could in some cases result in plumes transported off the coast and below cloud top. Larger errors could occur in such retrievals over water, if sources were not well represented in the TM5 model.) "

Page 15, Line 411-412: or validation with ground-based measurements (TCCON or NDACC). Reworded to "These errors require further characterization with colocated *in situ* data and ground measurements over land."

Figures 3-5: Please add regional mean bias and standard deviation of the differences to all individual subplots. Figures 3-5 currently show daily mean of regional bias for each ROI. Please note that regional mean bias values for each ROI and each MOPITT product (TIR, NIR, and TIR+NIR) are shown in Fig.

13. Adding this information to Fig. 3-5 would be redundant and would make these figures harder to read. Regional standard deviations for each ROI and each MOPITT product are also shown in Fig. 13; adding this information to Fig. 3-5 would make these figures harder to read.

Figures 13-14: Please add the standard deviation of the individual regional biases as a measure of the region-to-region bias to the plots (for TIR, NIR, and TIR+NIR). Please note that Fig. 13 shows results for colocated retrievals while Fig. 14 shows results for non colocated retrievals. The solid black lines in Fig. 13 represent, for each ROI, the standard deviation derived from individual biases between each pair of colocated observations. We have added a few words early in the manuscript (Methods section, page 7, line 168) to clarify this point: "We quantified, among others, daily bias (i.e., accuracy) and standard deviation (i.e., precision; calculated from individual biases between each pair of colocated observations) between TROPOMI and each of the three MOPITT products (TIR, NIR, and TIR+NIR)." In contrast, the dashed lines (not "solid lines"; caption has been corrected accordingly, here and in Supplement Materials) in Fig. 14 represent, for each ROI, ±1 standard deviation of mean daily relative biases (i.e., inter-daily bias variability). In this case a standard deviation cannot be calculated from individual biases between each pair of colocated observations, since no colocation was performed.

Table 1: Please add region-to-region biases in case "all ROIs". Please note that Table 1 already contains, for each ROI, the biases as well as the standard deviation of the individual biases, both in percentage and in CO column values.

Technical Corrections
Page 3, Line 50: Please rephrase "is key." Reworded to "is important".
Page 8, Line 195: Please replace $\sum_{i=1}^{i=n} ...$ by $\sum_{i=1}^{n} ...$ in Eq.(1). Thank you. The equation (now Eq. 3) has been slightly simplified and now it does not include a summation symbol.
Page 13, Line 340: Please add the unit % for the relative biases. Done.

References
Rodgers, C. D. and Connor, B. J.: Intercomparison of remote sounding instruments, J. Geophys. Res., 108, D3, 4116, https://doi.org/10.1029/2002JD002299, 2003.
Wunch, D., Wennberg, P. O., Toon, G. C., Connor, B. J., Fisher, B., Osterman, G. B., Frankenberg, C., Mandrake, L., O'Dell, C., Ahonen, P., Biraud, S. C., Castano, R., Cressie, N., Crisp, D., Deutscher, N. M., Eldering, A., Fisher, M. L., Griffith, D. W. T., Gunson, M., Heikkinen, P., Keppel-Aleks, G., Kyro, E., Lindenmaier, R., Macatangay, R., Mendonca, J., Messerschmidt, J., Miller, C. E., Morino, I., Notholt, J., Oyafuso, F. A., Rettinger, M., Robinson, J., Roehl, C. M., Salawitch, R. J., Sherlock, V., Strong, K., Sussmann, R., Tanaka, T., Thompson, D. R., Uchino, O., Warneke, T., and Wofsy, S. C.: A method for evaluating bias in global measurements of CO2 total columns from space, Atmos. Chem. Phys., 11, 12317-12337, https://doi.org/10.5194/acp-11-12317-2011, 2011.