Black: referee's comments red: authors' answers
First of all, we want to thank the two referees for the detailed analysis of our paper.
For the details, please look into the paper with keeping track of changes.

Anonymous Referee #2

Zhou et al. provide a comprehensive intercomparison between two remote sensing Fourier transform infrared (FTIR) spectrometer measurement techniques: sensing carbon monoxide (CO) in the near infrared (NIR) and mid-infrared (MIR) spectrum. They compare column average CO volume mixing ratios (XCO) at 6 sites that are a part of both the TCCON and NDACC networks. The study makes a direct comparison, as well as analyzes the impact of the uncertainty budget on the difference found between the two methods. Overall, this study presents a valuable comparison that allows for future synergy when using CO measurements from both the TCCON and NDACC networks. The subject matter of measurement intercomparison is well within the scope of AMT. The paper meticulously describes the methodology, and draws on previous work in the field. Writing style is clear and descriptive. The study includes rigorous analysis and sound mathematical procedures. Subsequent to addressing and/or responding to the minor comments and clarifications below, I recommend that the manuscript be accepted for publication in AMT.

Specific Comments:
Comment 1: AirCore comparison Using the NDACC a priori profile in section 4.3 to extend the AirCore profile presents a circular argument. It is not necessarily surprising that the AirCore total column was higher than the TCCON total column, because the scaled NDACC a priori profile was used to "fill in the blanks". (a) First of all, it is unclear where the "scaled NDACC a priori" (P14, L5) originates, because it is stated that Sodankylä is not an operational NDACC site (P13, L8). Please clarify how the scaled a priori is calculated - i.e. are there non-operational data available? If so, are they up to the same standard as the rest of the NDACC network? (b) I suggest to use the TCCON a priori above the top of the AirCore measurements, perhaps scaled by the AirCore: TCCON profile difference that exists between 15 to 20 km. Figure 5 suggests that the shape of the TCCON a priori profile above 10 km might be better than the NDACC at Lauder and Wollongong. While the profile above 20 km may not contribute a large amount to the total column, it is important to remove potential bias in the comparison. Alternatively, the authors could calculate how much the "extended profile" contributes to the total column and explain why it doesn't convolute the results.
(a)"scaled NDACC a priori profile at Sodankyla" means "the scaled WACCM profile at profile". Although Sodankyla is not a standard NDACC site yet, we ask our NOAA colleague to give the WACCM output at this site. This information is added in the revised version.

(b) The TCCON a priori profile in the stratosphere is mainly generated based on the MkIV balloon profiles in 2005 measured in the 30-40˚N latitude range. Figure 3 shows that the TCCON a priori CO VMR in the stratosphere is almost same at these sites. It is not realistic, because the CO is variable in the stratosphere and mesosphere (Garcia et al., 2014). Figure 1a shows that the CO VMR measurement by ACE-FTS v3.5 data (mean profile in the time period of 2007-2017) at Sodankyla and Orleans together with the TCCON and NDACC a priori profile. For some more information about the ACE-FTS CO data please refer to Clerbaux et al., (2008). Above 25-30 km, the WACCM model profile is close to the ACE-FTS measurements, while there is large bias between the TCCON a priori profile and the ACE-FTS measurements. Using the TCCON CO a priori profile to extend the AirCore measurement will bring a systematic error.

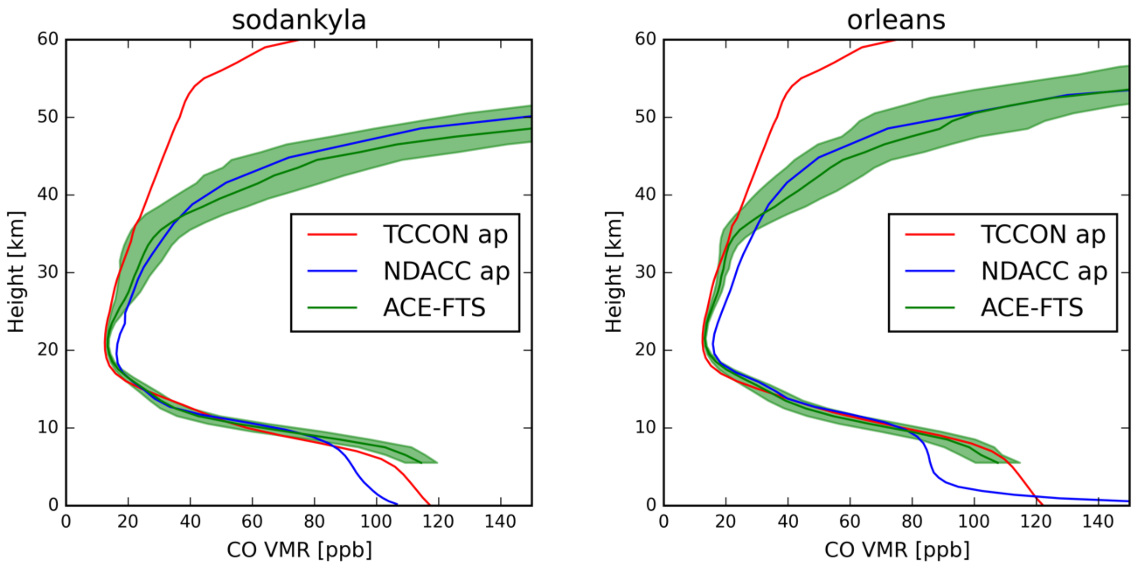

Figure 1a. the TCCON and NDACC a priori CO profile at Sodankyla and Orleans, together with the ACE-FTS observed CO profiles. The TCCON a priori profile is the mean in 2013. The NDACC a priori profile is from WACCM v6 model. The ACE-FTS profile is the mean of all measurements located within ±10° latitude band of the FTIR site during 2007-2017, together with the mean retrieval error (green shadow).

The scaled mean ACE-FTS profile is also applied to extend the AirCore measurements, we found that the AirCore smoothed XCO is 6.1±1.6% larger than the TCCON XCO data at Orleans, and is 8.0±3.2% larger than the TCCON XCO data at Sodankyla. The results using the ACE-FTS measurements are within the uncertainty of the result using the NDACC a priori profile. In the revised version, the ACE-FTS profile is applied to extend the AirCore measurement above.

Comment 2: "Low" Mean Bias for the Southern Hemisphere It is stated that the Southern Hemisphere has a low mean bias of 0.3% (P6, L6) and that the values "agree well" (P18, L2). I am a little nervous about how good the 0.3% seems, because it is the result of averaging positive and negative numbers. In particular, the Lauder station uses different a priori protocol than Wollongong and St Denis (P8, L25-28). If only the two latter stations are included, the Southern Hemisphere bias is 1.5%, which may be more representative. I would appreciate a small discussion of this in the manuscript. Perhaps it would also be useful to discuss the absolute relative error.
Thanks for the suggestion. Instead of the mean at these three sites, we rewrite it as "A direct comparison shows that the NDACC X$_{CO}$ measurements are about 5.5\% larger than the TCCON data at Ny-\r{A}lesund, Bremen, and Iza\~{n}a (Northern Hemisphere), and the absolute bias between NDACC and TCCON data are within 2\% at St Denis, Wollongong and Lauder (Southern Hemisphere)".

Concerning the difference a priori profile at Lauder, since the smoothing error is relatively small (systematic/random 0.1/0.2 %; see Table 5) of the NDACC retrievals, the choice of the NDACC a priori profile has limited impact on the retrieved total column. As discussed in the Section 4.1 and 4.2, we think the a priori profile of the TCCON retrieval plays a key role in this systematic bias at Lauder.

Comment 3: Double-check main text values with figures and tables for consistency (a) Figure 1 - Bremen shows a mean difference of 6.518 +/- 7.044 ppb. This seems inconsistent with Table 4 Bremen 6.4 +/- 4.3%. (b) P16, L21-22 - the values in the data range do not match Table 6. For example, I think 2.1 to 6.1% should be 2.1 to 8.1%.

(a)Thanks for pointing out the mistake, we have double checked all numbers, and made the correction for Bremen site.
(b)The P16, L21-22 - the values in the data range do match Table 6. We mention that "Table 6 shows that the changing of the model X_{CO} data after smoothing with TCCON data ranges from 2.1% (Bremen) to 6.1% (Lauder)"

Comment 4: Relevance of calculations on Page 10 It was unclear to me why the systematic uncertainty was kept here (Eq. 12) while on P8, L12 it was eliminated. Also, how is the choice of optimal common a priori made to ignore the first component (P10, L10)?
On P8, it is assumed that the systematic uncertainty of the public TCCON data is eliminated by the post-correction (the scaling factor). On Page 10, we write that "We keep the systematic uncertainty here, in case the correction of the TCCON data does not get rid of the systematic uncertainty completely."

We mentioned here that "If the optimal common a priori profile is close to the true status… ". To explain how is the uncertainty from the first term by using the scaled NDACC a priori priofile as the common a priori profile, the smoothing error is estimated in Section 4.2.

The comment on P12, L2-3 suggests that the first component should not be ignored. There is a missing link/clarification needed between the theory presented on this page and exactly how it relates to this study. Also, would the main text on page 10 work better inside section 4.1?
Thanks for the suggestion, we arrange this part to section 4.1.

Minor Comments:
P1, L10-11: Reword to clarify what is meant by this sentence. For example, does this convey what is meant: "The TCCON smoothing error is significant because it is higher than the reported uncertainty".
Reworded. "For TCCON data, the smoothing error is significant because it is higher than the reported uncertainty, particularly at Southern Hemisphere sites."

P2, L21: Mention briefly why the measurements are assimilated into models.
Done

P3, L29 L30: Describe acronym ATM and JPL.
Done

P3, L32: Add the equation definition after the "CO total column".
Done

P4, Table 1: I suggest site reference papers rather than "Research group".
We prefer to leave the research group, since no direct NDACC reference are found for these sites. TCCON reference are already listed in the paper.

P4, L1: What is meant my "indirectly validated" against AirCore measurements?
"indirectly validated" means the TCCON product are validated by the aircraft or AirCore measurements, and the aircraft or AirCore measurements are calibrated/validated by other in situ measurements. Therefore, we think it is appropriate to say "indirectly validated" here.

P4, L11: Please clarify why 3.0% and not 3.5% is used as the upper limitation on random uncertainty.
We use the 3.5% in the revised verison.

P4, L12-14: I am confused by the last sentence in section 2.1. How does it relate to WMO scaling? Was public data not used in this study?
We used the public TCCON data in this study. Since the TCCON data have been corrected according to the aircraft measurements (Wunch et al., 2015), we assume the systematic uncertainty of the TCCON data is removed at this stage. However, by comparing the public TCCON data with AirCore measurements, there is an about 7% underestimation in the public data. So, we infer that the scaling factor of the TCCON CO data is wrong, and the systematic uncertainty is still there in the public TCCON data.

P5: I suggest to swap the order of Tables 2 and 3, and move sentence line 8-9 to the end of the paragraph.
Done

P5, L12: Unclear purpose for the last sentence of Section 2.2. Are the authors suggesting these smoothing errors be included in reported NDACC data?
No, we do not suggest to include the smoothing error in the reported NDACC data, but we could like to have a smoothing error estimation here for the NDACC CO data.

P5, Table 3 caption: Mention that the "Total" uncertainties are calculated by adding the sub-types in quadrature. Also, why can "-" be ignored? Are they simply too small to worry about?
Done. Yes, they are just too small. We write" '-' means that the uncertainty is less than 0.1 and then can be ignored"

P5, Fig1: Is the seasonality in the Lauder difference plot related to that NDACC station using a different a priori than the other NDACC stations?
It might be one reason. Since this paper mainly focus on the mean bias between TCCON and NDACC XCO data, we try to avoid to discuss the seasonality in the bias. But I think this is a very interested topic, and would like to continue to figure it out in the near future.

P7, Table 4 caption: I suggest to work on reducing the repetitive wording.
Done

P7, L4-L9: Is the conclusion from this section that NDACC adequately accounts for SZA dependence in the retrieval algorithm?
It is hard to make such conclusion, because we did not check the NDACC XCO retrievals with SZA directly. We could prefer to say that " the differences between public TCCON and NDACC XCO data resulting from SZA are very small in both hemispheres, compared to the large scatter "

P11, L3: Describe why the scaled NDACC a priori were used as common a priori, rather than the NDACC a priori.
Done

P11, Figure 5: Please add uncertainty/standard deviation to the HIPPO data points.
Done

P11, L13-14: Unclear why 7% (links to specific comment #4).
according to Eq. 12, the correction procedures and using O2 to calculate the dry air brings 7% in the TCCON XCO data. In the revised version, we mention that "(scaling TCCON data by +7\% according to Eq. 12; see Table 4)"

P12, L8: The 2.0% assumption on diagonal systematic bias is lower than any value presented in Table 2 - clarify why this value was chosen.
Clarify in the revised version: "Since the scaling factor of the NDACC a priori profile is based on the NDACC retrieved total column, and the systematic uncertainty of NDACC XCO data at Iza\~{n}a, St Denis, Wollongong and Lauder are about 2.0\% (see Table 3), it is assumed that the systematic bias for the diagonal values are 2.0\% For Bremen and Ny-\r{A}lesund, the systematic uncertainty might be underestimated."

P12, L25-26: Discuss why the random smoothing error is larger here than in Wunch et al. (2015).
Wunch et al., (2015) calculated the smoothing error by the following method: "A priori profiles are modified for X_{CO_2} by shifting the profiles down by 1 km". We think they underestimate the variability in atmospheric CO.
More information is added in the revised version.

P14, L2: Briefly mention how diffusion inside the AirCore coil can impact the measurement, and how it is minimized.
Done

P14, L6-7: How are the 3.0% and 6.0% assumed uncertainties chosen?
3% is set based on the uncertainty of AirCore and surface in situ measurements.
6% is set based on the systematic uncertainty of the retrieved partial column between 25 km and 100 km NDACC retrieval, and we assume that the accuracy of the scaled NDACC a priori profile is mainly determined by the retrieved NDACC retrieval. Now, we use the ACE-FTS measurement to extend the AirCore measurements, and the uncertainty is set as 25% according to the ACE-FTS measurement uncertainty.

P15, L13: What is "a truncation of T511"? Does it relate to the model altitude level?
No, it is not related to the altitude level but related to the horizontal resolution, please find detail information in this book Page 223 < The Emergence of Numerical Weather Prediction: Richardson's Dream>

P16, L4: Is the CAMS model interpolated in both time and space?
No, we use the bilinear interpolation in space (latitude and longitude) but no change in time. More information is added in the revised paper.

P16, L10: Mention that fewer satellite observations improve the CAMS model at higher latitudes due to measurement difficulties, which may cause the poorer performance at Ny-Ålesund.
Done

P16, L13: Mention that high locally impacted values are not expected to be captured by the model due to dilution: both temporally (6 hr compared to minutes) and spatially (40 km square compared to site location).
Done

P17, Figure 9: Consider moving this to a supplement because details are in Table 6.
We prefer to keep it here, as some text are described based on this Figure.

Technical Corrections: The paper is well written and there are only minor technical corrections.

P1, L5: A direct comparison shows the NDACC XCO measurements...

P1, L14: To determine the source of the bias, regular...
P2, L8: ...), and biomass burning (...
P2, L8: There are also small quantities of CO...
P2, L11: biomass burning
P2, L13: ...thus affects the...
P3, L1: Despite the similar...
P3, L14: ...carried out in Section 3.
P3, L15: ...investigated in relation to...
P4, L13: ...in public TCCON...
P4, L19: The reference spectroscopy database is...
P5, L15: At Northern Hemisphere stations...
P5, L17: At Southern Hemisphere stations...
P6, L1: ...dominated by biomass burning...
P8, L1: ...investigate the causes of the difference between...
P8, L27: The a priori profile...
P9, L7: ...is relatively clean, coming mainly...
P9, L9: At Bremen, the CO VMR in...
P9, L10: ... free troposphere, because there are strong local anthropogenic emissions (European Commission, 2013).
P9, L16: ...total columns correctly capture a...
P9, Fig. 3 caption: ...a priori profiles change every day,...
P11, L9: ...is the most reasonable...
P13, L5: ...measurements have been performed...
P13, L9: ...a long coiled tube...
P13, L10: ...tube is transferred to a gas analyser...
P15, L15: The model output...
P16, L10: ...model at this site.
P16, L19: According to the AirCore measurements in Sect. 4.3, the bias...
P18, L6: ...data products are more consistent (5.6...
Corrected

Reference:

Clerbaux, C., George, M., Turquety, S., Walker, K. A., Barret, B., Bernath, P., Boone, C., Borsdorff, T., Cammas, J. P., Catoire, V., Coffey, M., Coheur, P.-F., Deeter, M., De Mazière, M., Drummond, J., Duchatelet, P., Dupuy, E., de Zafra, R., Eddounia, F., Edwards, D. P., Emmons, L., Funke, B., Gille, J., Griffith, D. W. T., Hannigan, J., Hase, F., Höpfner, M., Jones, N., Kagawa, A., Kasai, Y., Kramer, I., Le Flochmoën, E., Livesey, N. J., López-Puertas, M., Luo, M., Mahieu, E., Murtagh, D., Nédélec, P., Pazmino, A., Pumphrey, H., Ricaud, P., Rinsland, C. P., Robert, C., Schneider, M., Senten, C., Stiller, G., Strandberg, A., Strong, K., Sussmann, R., Thouret, V., Urban, J., and Wiacek, A.: CO measurements from the ACE-FTS satellite instrument: data analysis and validation using ground-based, airborne and spaceborne observations, Atmos. Chem. Phys., 8, 2569-2594, https://doi.org/10.5194/acp-8-2569-2008, 2008.

Garcia, R. R., M. López-Puertas, B. Funke, D. R. Marsh, D. E. Kinnison, A. K. Smith, and F. González-Galindo (2014), On the distribution of CO2 and CO in the mesosphere and lower thermosphere, J. Geophys. Res. Atmos., 119, 5700–5718, doi:10.1002/2013JD021208.