Response to comments from Referee 2

Black: Referee's comments; Blue: Authors' answers

We thank referee #2 for the review and for providing useful feedback.

Review of Sha et al.: Intercomparison of low and high resolution infrared spectrometers for ground-based solar remote sensing measurements of total column concentrations of CO_2, CH4 and CO

This paper details the results from an intensive campaign wherein 4 portable spectrometers (Bruker EM27/SUN, Bruker IRcube, Bruker Vertex40, RAL LHR) were hosted near the Sodankylä TCCON station from March-September 2017. AirCore profiles were also flown throughout the campaign. The campaign resulted in some interesting data collected that helped reduce uncertainties in the Vertex40, pointed out where the LHR needs improvement, and identified a problem unique to the Sodankylä TCCON station that was resolved.

This paper may be suitable for publication after the following concerns are satisfactorily addressed.

Major Comments

Referee:
1. The non-linearity problem at Sodankylä is unique to Sodanklyä, and this is not made clear in the paper. In general, TCCON stations have limited the light incident on the detectors by a combination of reducing the input field stop (on the aperture wheel) and placing an aperture stop after the input CaF2 window. From what I can glean from Section 5.8 (specifically, "The TCCONmod data set is a better representation of the true atmospheric signal. As TCCON is our primary data reference for the intercomparison study for this campaign, the non-linearity correction has been applied to the TCCON data."), this non-linearity correction has been applied to all comparisons/figures/tables throughout the paper (except where the non-linearity is discussed directly). If I have interpreted this correctly, I believe it is the correct approach, since it is more representative of the data a TCCON station would produce. That said, it should be stated up front (i.e., in section 2.2.1) that this correction has been applied, with a clear statement that this does not generally affect all TCCON data; just Sodankylä's. The details included in section 5.8 (and 5.9) ought to be relegated to Appendix A.
Authors' response:
Although the referee is correct in pointing out that non-linearity is not a problem that affects all TCCON stations, it is as well unjustified claiming that it is a special problem affecting only the single TCCON site at Sodankylä. The Sodankylä TCCON station is a well-maintained and carefully operated site performing measurements since 2009. The input field stop on the aperture wheel was set to 1 mm and the aperture stop after the CaF_2 window was set to 32 mm following the standard TCCON recommendations. However, it was only during this campaign, while comparing with the EM27/SUN spectrometer, that we discovered the issue with the non-linearity and its

associated influences on the trace gas results. This is a very important finding of this campaign. The detection of non-linearity should be incorporated in the TCCON processing chain (as it is done for COCCON), flagging spectra with non-linearity issues before the trace gas analysis is attempted. As this has not been done in the TCCON data quality management, it may well be possible that other sites might also be affected by non-linearity and remain unnoticed during a longer period of time.
In our paper, the non-linearity corrected TCCON data are labelled as "TCCONmod" and the standard TCCON data are labelled as "TCCON". The standard TCCON data is also what is made publicly available via the TCCON data repository. The TCCONmod data is a product of this campaign and not yet submitted to the TCCON database. Following the suggestion of both referees, we have moved the section of the discussion with the non-linearity corrected TCCON data to the beginning of the results section and the discussions with the uncorrected data to the appendix.

Referee:
2. Your comparisons of Xair between the various instruments is misleading because it is not an apples-to-apples comparison. It is not only instrument effects that impact Xair: it is also spectroscopy. PROFFAST and GGG2014 do not include the same spectroscopy, line shapes, etc., and thus would be expected that they have different Xair values. It would be helpful to have a short (1-paragraph) discussion about the differences between the two retrieval software algorithms, as relevant to this issue. It is unclear to me why you would choose to use PROFFAST to analyse the EM27/SUN data and not also use the EGI wrapper for GGG2014 to maintain consistency with the other retrievals. I suggest you add the EGI retrievals of the EM27/SUNs to your analysis. (https://bitbucket.org/em27gi/egi/wiki/Home)
Authors' response:
The PROFFAST software has been used for evaluating the performance of the EM27/SUN, because this is the official code to be used for the COCCON analysis. From this perspective, we demonstrate an end-to-end evaluation of instrument hardware and data processing, as targeted by the FRM4GHG project. The PROFFAST code has been developed on behalf of ESA in order to provide a source-open and freely available code (without any licensing restrictions) as required by the growing COCCON user community. Note that the use of an independent code does not imply generating "oranges" instead of "apples" as both processors intend to provide an optimal reconstruction of the true atmospheric state. Instead, it allows us to uncover the discrepancies introduced by the various design considerations of the pre-processing, the retrieval codes and generated outputs. We agree that a future alignment of line lists between GFIT and PROFFAST is desirable for revealing those changes introduced by more subtle aspects of the setups (e.g. treatment of background continuum). It is therefore planned to upgrade PROFFAST to the line lists used by the upcoming GFIT version in the framework of a follow-up project. We also have added a paragraph describing some details of the PROFFAST setup.
See lines 173 – 188: "PROFFAST is a code for retrieving trace gas amounts from low-resolution solar absorption spectra. It has been developed on behalf of ESA, in order to provide a source-open and freely available code (without any licensing restrictions) as required by the growing COCCON user community, e.g. for TROPOMI validation work. It is a least-squares fitting algorithm, which adjusts the trace gas amounts by scaling atmospheric a priori profiles. The

retrievals are performed on spectra generated with the included PREPROCESS tool. This tool produces spectra out of the measured DC-coupled EM27/SUN interferograms. It includes a DC correction of the interferogram, a dedicated phase correction scheme for double-sided interferograms and several quality control tests (e.g. testing for the presence of out-of-band artefacts). The lookup table for cross-sections used by PROFFAST is created on the basis of HITRAN spectroscopic line lists: For H2O, CH4, N2O, HITRAN 2008 line lists are used (in case of H2O including some minor empirical adjustments), for CO2 and CO HITRAN 2012 line lists are used. PROFFAST uses the solar line list compiled by Geoff Toon, JPL, for GGG2014. In contrast to the TCCON GGG2014 processing, the empirical airmass-independent and airmass-dependent post-calibrations are applied species-wise including molecular oxygen. Thereby, the Xair equivalent provided by PROFFAST is on average normalised to unity, while it remains an un-calibrated intermediate result in GGG2014, which calibrates only the Xgas results. The PROFFAST approach of calibrating Xair is transparent for users, as the calibration factors can be directly related to deviations of the spectroscopic band intensities, and gives the user a more sensitive diagnostic tool at hand, as airmass-dependent artefacts in the reported quantity are also reduced."

Referee:
3. It is unclear to me why you include the time period before the hardware upgrades of the Vertex and IRcube in your subsequent analysis. It makes more sense to me to show the significant improvement in their data after fixing the hardware in the time series plot, and then do not show the pre-improvement data in subsequent analyses, focusing only on the "good" data.
Authors' response:
The campaign began with an initial blind intercomparison phase where the instruments were operated with the optimised setting best known to their PIs to get good SNR comparable to the TCCON. However, it was found that the settings were not optimal for Vertex70 and the improvements were significant after the instrument modification. The change of the optical fibre also affected the IRcube results. Therefore, we find it relevant to report these data. The other instruments did not undergo any modification and therefore the data for the whole period is considered as "good" data. In the paper, we present two sets of results for each intercomparison section where the results from the full year are relevant for the instruments, which did not undergo any modification, and the results for the shorter period, which are focusing on the comparison of the Vertex70 instrument results relative to the other instruments for the same period.

Minor comments

Referee:
1. I found the authors' motivation for the need for the COCCON (or another complement to TCCON) misleading. I believe it is true that the TCCON could be usefully supplemented by LR portable spectrometers, but the atmospheric and surface conditions you list are generally already covered by TCCON stations. For example:

a. “A denser distribution of ground-based solar absorption measurements is needed to cover various atmospheric conditions (humid, dry, polluted, presence of aerosol), various surface conditions (high and low albedo) and a larger latitudinal distribution.”
i. Humid –> Darwin
ii. Dry –> Armstrong, Eureka, Sodankylä, Pasadena
iii. Polluted –> Pasadena, Tsukuba, Wollongong (sometimes)
iv. Aerosol –> Pasadena, Tsukuba, sometimes Ascension
v. High albedo –> Armstrong
vi. Low albedo –> Park Falls, etc.

b. Where I do agree, is that if we want to cover geographic gaps in the locations of these stations, we need more stations, and the low-resolution instruments may be well suited to that.
Authors’ response:
Taking the referee’s comments into account we have added further details in the paper (see lines 4 – 9) to clarify our statement:
“The number of stations in the network (currently about 25) is limited and has a very uneven geographical coverage: the stations in the Northern hemisphere are distributed mostly in North America, Europe, Japan and only 20% of the stations are located in the Southern hemisphere leaving gaps in the global coverage. A denser distribution of ground-based solar absorption measurements is needed to improve the representativeness of the measurement data for various atmospheric conditions (humid, dry, polluted, presence of aerosol), various surface conditions such as high albedo (> 0.4) and very low albedo and a larger latitudinal distribution.”

Referee:
2. It is unclear to me whether you use identical surface pressure values in your retrievals for each instrument. If they are different, this would also cause biases in Xair. For a fair comparison, they should be identical and calibrated to a meteorological standard. Please discuss and resolve if necessary.
Authors’ response:
Yes, all spectrometers used the identical set of ground-pressure data collected at the Sodankylä site. We have added this information to the paper (see line 363): “The spectrometers used an identical set of ground-pressure data collected at the Sodankylä site for the retrieval.”

Referee:
3. It is unclear to me why the HR125LR has its own section. Why not include it in the comparisons with the other LR instruments wrt TCCON?
Authors’ response:
The HR125LR data set provides interesting results when compared to the reference data set. It is a unique data set as both HR125LR (low-resolution) and the reference high-resolution measurements were performed using the same instrument. In order to highlight these features we have provided a separate section for the discussion of the results.

Referee:

4. The water vapour dependence section is not conclusive, since Sodankylä is generally dry (XH2O <4500 ppm). This is therefore not an exhaustive test of XH2O dependence. I suggest moving this section to an appendix, noting in the main text that little XH2O dependence was found over the (relatively dry) conditions at Sodankylä.

Authors' response:

Sodankylä is not the most humid TCCON site. The maximum XH2O measured by the TCCON is < 6000 ppm during the summer period (see Figure 1 below). In comparison, the TCCON site at Darwin, which is a relatively humid site, show a maximum measured XH2O of < 10000 ppm during the summer period. The year 2017 was relatively dry where the range of XH2O measured at the Sodankylä site was between 500 and 4500 ppm. Following the suggestion of the referee, we have moved the detailed discussion to the appendix F and introduced a small paragraph as section 5.6 in the main text focusing on the main results.

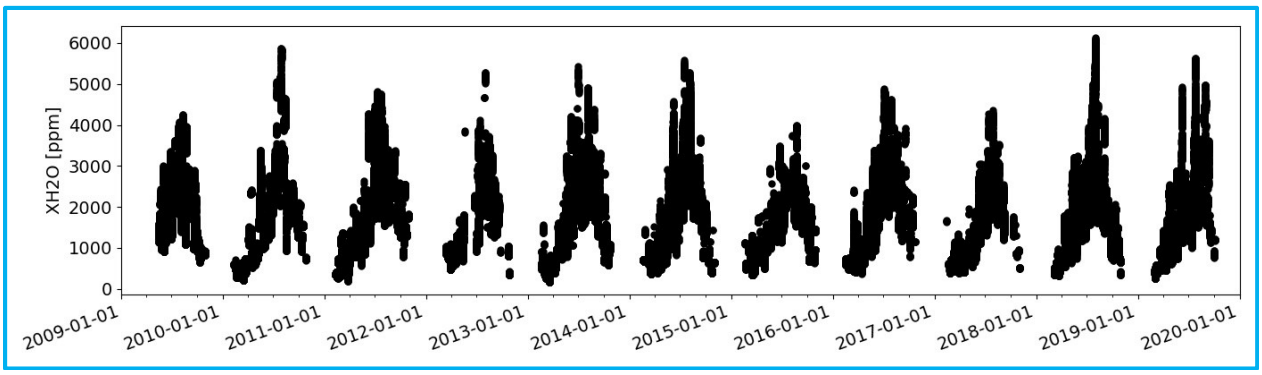

Figure 1: Timeseries of the XH2O measured at the Sodankylä TCCON site.

Technical comments

Referee:

1. P1L1: … the baseline *ground-based* network of instruments …

Authors' response:

Done

Referee:

2. P1L5: Northern America –> North America and again in L72.

Authors' response:

Done

Referee:

3. P2L25-26: This seems to imply that the non-linearity is a problem throughout TCCON, which it is not. Please revise.

Authors' response:

We have added further explanation to emphasise that the non-linearity results are for the Sodankylä campaign. See lines 21 – 25: "The reference measurements performed with the Bruker IFS 125HR were found to be affected by non-linearity of the Indium Gallium Arsenide (InGaAs)

detector. Therefore, a non-linearity correction of the 125HR data was performed for the whole campaign period and compared with the test instruments and AirCore. The non-linearity corrected data (TCCONmod data set) show a better match with the test instruments and AirCore data as compared to the non-corrected reference data."

Referee:
4. P2L41: increasing in ** recent years (no "the")
Authors' response:
Done

Referee:
5. P2L43: You may want to mention the (important) role of VOCs in the production of CO.
Authors' response:
Done

Referee:
6. P2L48: positive radiative forcing*, therefore it is* considered as an indirect …
Authors' response:
Done

Referee:
7. P3 first paragraph: Generally unnecessary paragraph. It's unclear what you mean by "To ensure equal dependency on the input spectral data, …"
Authors' response:
This paragraph provides the rationale for the ground-based total column measurements and its usefulness for satellite validation. Following the suggestion of the referee, we have added further explanations in the paragraph to make our statements clearer for the reader. See lines 63 – 65: "To ensure equal dependency on the measurement parameters, the best validation method for the satellite data is to use the total column amounts of the trace gases calculated from the solar absorption measurements performed from the surface and the satellite in the same spectral region."

Referee:
8. P3L74: Again, there are TCCON stations that span all those conditions. Rephrase.
Authors' response:
We have modified the sentence, see lines 76 – 79: Furthermore, for the complete validation of the satellite data set, a denser distribution of ground-based solar absorption measurements is needed to cover geographical gaps and to improve the representativeness of the measurement data for various surface and atmospheric conditions (e.g., high and very low surface albedo, pollution, aerosol presence, humid, dry).

Referee:

9. P3L80. “However, there has been little characterization, intercomparison and harmonization of these new instruments in comparison to the standard instrument used in TCCON.” There is some literature on just this topic. Please cite:

a. Hedelius, J. K., C. Viatte, D. Wunch, C. M. Roehl, G. C. Toon, J. Chen, T. Jones, S. C. Wofsy, J. E. Franklin, H. Parker, M. K. Dubey, and P. O. Wennberg (2016), Assessment of errors and biases in retrievals of XCO2, XCH4, XCO, and XN2O from a 0.5 cm-1 resolution solar-viewing spectrometer, Atmos. Meas. Tech., 9(8), 3527–3546, doi:10.5194/amt-9-3527-2016.
b. Hedelius, J. K., H. Parker, D. Wunch, C. M. Roehl, C. Viatte, S. Newman, G. C. Toon, J. R. Podolske, P. W. Hillyard, L. T. Iraci, M. K. Dubey, and P. O. Wennberg (2017), Intercomparability of XCO2 and XCH4 from the United States TCCON sites, Atmos. Meas. Tech., 10(4), 1481–1493, doi:10.5194/amt-10-1481-2017.
Authors’ response:
Done

Referee:
10. While I understand that Sodankylä was chosen for (important) practical considerations, it is a challenging place for a ground-based campaign for a number of reasons, and these challenges ought to be clearly stated near the beginning of the paper with your thoughts on how those challenges may manifest:
a. High latitude means higher SZA
b. Lack of full seasonal cycle
c. Proximity to the polar vortex means increased likelihood of poor a priori profiles as GGG2014 does not handle vortex air
d. Dry atmosphere does not provide full range of XH2O seen in other locations
Authors’ response:
a. We have added an explanation in the text. See lines 110 – 113: “Due to the location of the site at a high latitude, measurements are possible for solar zenith angle (SZA) range between > 43° and < 90°. The coverage of high SZAs is important to check the dependence of the airmass on the retrieval results. The airmass dependent correction factor applied to the remote sensing data is relevant for measurements at higher SZA.”
b. Yes, the Sodankylä TCCON site lacks measurements during Nov – Jan. However, it has the benefit of having high variability of signal (e.g. for CO_2) during the year.
c. The same a priori was used for all instruments. The low-resolution instruments are aiming to complement TCCON type measurements. It is therefore very interesting to compare them also in conditions where TCCON stations exists.
In addition, validation of the XCH4 product from the TROPOMI instrument on board the Sentinel-5 Precursor (S5P) with respect to the TCCON station at Sodankylä shows a high bias during the spring period as compared to the rest of the year (see Figure 2 below). It is therefore very interesting to identify the cause of the bias. Measurements performed with the low-resolution ground-based instruments during this campaign will help us to understand some of the causes of the bias.

d. The year 2017 was a dry year in general. Typically, XH2O values for Sodankylä cover higher range and reach up to 6000 ppm. However, we agree that comparing the low-resolution instruments for conditions with very high humidity would be useful in a future campaign.
As mentioned in the paper, after carefully analysing the available site options and the requirements to host simultaneously all instruments and the availability of AirCore facilities, we have selected the Sodankylä TCCON site for our campaign.

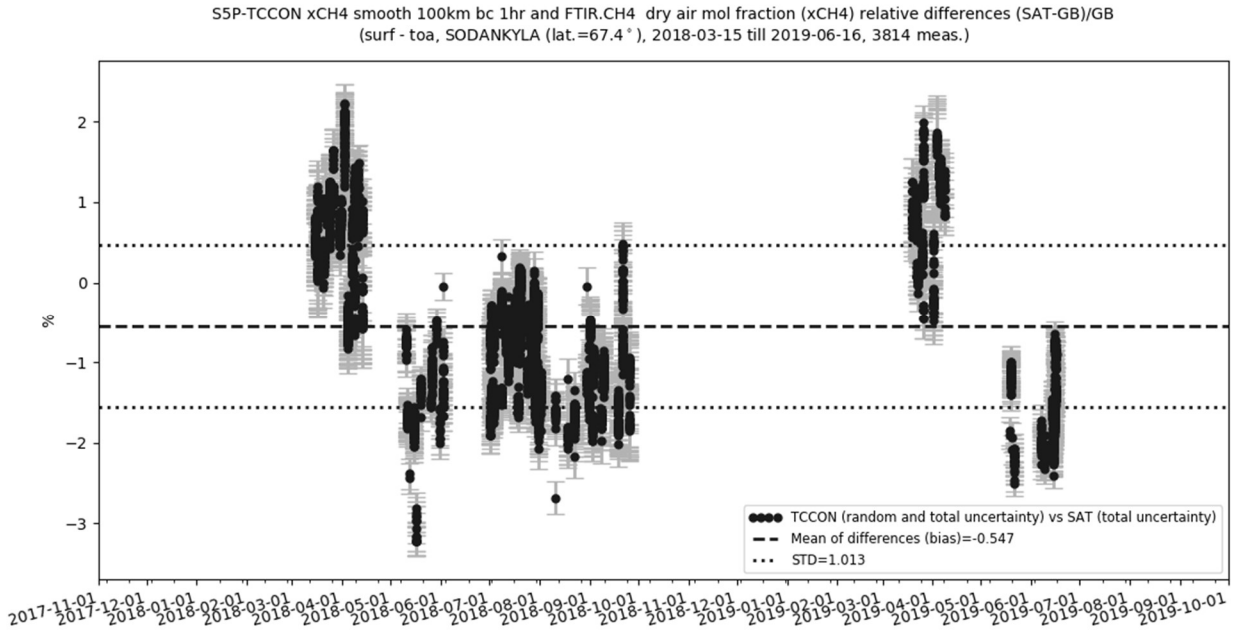

Figure 2: Relative difference of the XCH4 from the (S5P-TCCON)/TCCON data sets for the Sodankylä site. The coincidence criteria for the validation are given in the presentation by Sha et al. 2019.

Referee:
11. P3L85-85: Awkward sentence. Please clarify.
Authors' response:
Indeed, it was a long complex sentence. Therefore, we replaced it with two sentences to convey our message. See lines 88 – 92: "For this reason in 2017, the European Space Agency (ESA) initiated an intercomparison campaign within the project Fiducial Reference Measurements for Ground-Based Infrared Greenhouse Gas observation (FRM4GHG). The campaign was performed in Sodankylä (Finland) with the aim to assess the performance of different spectrometric instruments for remote sensing of atmospheric trace gases and to quantify their performances regarding precise measurements of column-averaged dry-air volume mole fractions of CO_2, CH4 and CO."

Referee:
12. P4L88: Please cite the Karion paper when AirCore is introduced.
Authors' response:

Done

Referee:
13. P5L124: What does “The number of usable detector positions differs for the three instruments.” mean in practical terms for this work?
Authors’ response:
We have elaborated this point by providing further details on the number and type of detectors for each instrument in the paper. This is mentioned in Table1. See lines 131 – 135: “The number of usable detector positions differs for the three instruments. The EM27/SUN can accommodate two room temperature (RT) Indium Gallium Arsenide (InGaAs) detectors covering different frequency ranges. Also the Vertex70 can accommodate two detectors, one InGaAs and a second channel with either a liquid nitrogen (LN2) cooled Indium Antimonide (InSb) or an RT InGaAs detector. The IRcube can only accommodate one InGaAs detector and has no room for a second detector.”

Referee:
14. P4L136: Please cite: Keppel-Aleks, G., G. C. Toon, P. O. Wennberg, and N. M. Deutscher (2007), Reducing the impact of source brightness fluctuations on spectra obtained by Fourier-transform spectrometry., Appl. Opt., 46(21), 4774–4779, doi:10.1364/AO.46.004774.
Authors’ response:
Done

Referee:
15. P4L138: Please state that this is the GGG2014 software that you are using.
Authors’ response:
Done

Referee:
16. P4L141: Are you recording just on the InGaAs detector alone? Are ghost corrections performed?
Authors’ response:
Yes, the double-sided DC coupled interferograms at 0.5 cm-1 are recorded using the InGaAs detector. The laser sampling error (LSE) caused by any asymmetry is minimized by collecting data employing the interpolated sampling option provided by Bruker. Gisi, 2014 showed that no ghost were found when the option of interpolation was enabled with the M16 electronics. In our measured spectra, we did not find any LSE ghost.

Referee:
17. P7L187: build –> built
Authors’ response:
Done

Referee:
18. P8L232: atmospheric ** and local oscillator *beams* are mixed

Authors' response:
Done

Referee:
19. P8L244: 13.4 mbar for ** ambient pressure*s between the surface and* 232 mbar, and 3.9 mbar for ** ambient pressure*s* lower than 232 mbar.
Authors' response:
Done

Referee:
20. P9L260: Why are you listing this CO value in ppb?
Authors' response:
We reported it now in ppm.

Referee:
21. P9L275: The sentence beginning with "The continuous operation" is awkward. Please revise.
Authors' response:
Done. See lines 302 – 303: "The measurements preformed helped to observe diurnal variation of the target gases."

Referee:
22. P10L293: All teams performed *a* full functionality test …
Authors' response:
Done

Referee:
23. P10: Change "highest OPD" to "maximum OPD".
Authors' response:
Done

Referee:
24. P11L312: documentation (no 's')
Authors' response:
Done

Referee:
25. P11L322: Is the surface pressure identical between instruments?
Authors' response:
Yes, all spectrometers used the identical set of ground-pressure data collected at the Sodankylä site.

Referee:
26. P11L328: What scaled ratio are you referring to? You refer to scaling several times throughout the paper and it is not clear what you are referring to. Please clarify.

Authors' response:
We have modified the text giving further explanation on this and added the definition of Xair following TCCON publication.
TCCON uses airmass dependent and airmass independent calibration factors to scale the Xgas values to the WMO standards. The results of the low-resolution instruments analysed with GFIT were scaled in the same way as TCCON. The scaling factor used for the EM27/SUN results are discussed in Frey et al. 2015.

Referee:
27. P11L335: These two lines are repetitive with information stated earlier in the text.
Authors' response:
We have removed both lines as suggested by the referee and added this information in Section 3.1.

Referee:
28. P12L350: State your reasoning for the SZA cutoff earlier.
Authors' response:
We think that the reasoning for the filtering of SZA cutoff is better suited here. Therefore, we have modified the text in the first paragraph of section 5.1 such that we do not need to mention the filtering criterion there. See lines 348 – 349: "However, these measurements performed were recorded with SZA > 75°."

Referee:
29. In general, I do not think that uncommon phrases, and/or phrases that are used only a few times in a manuscript should be given acronyms. I find these acronyms detract from the readability of the paper. For example, I would appreciate it if you would type out the acronyms for:
a. LO
b. FO
c. NA
d. ME (though admittedly common in our field, but used only a few times)
e. PE (again, common in our field)
Authors' response:
Done

Referee:
30. P13: consider reporting XCH4 differences in ppb instead of ppm.
Authors' response:
The TCCON XCH4 are reported in ppm. As TCCON is our reference, we also stick to reporting XCH4 values and differences in ppm.

Referee:
31. In general, in these Xgas-specific subsections: I do not think the seasonal cycle needs to be described in words.

Authors' response:
We prefer to keep the description of the seasonal cycle as it emphasises the variability of the signal for the Sodankylä site. Also for the comparison of the LHR XCO_2 values, we could observe that it was able to capture the annual summer drawdown. Having the description of the seasonal cycle helps to point to this kind of aspects.

Referee:
32. P14L430: "Any difference relative to the ideal case is an indicator for the instrument and retrieval code performance." And the spectroscopy, which can be distinct from the code itself. Please add.
Authors' response:
Done

Referee:
33. P14L440: Again, the EM27/SUN data are processed using *an entirely different retrieval code with different spectroscopy* so it is not surprising (or an indication of instrument performance) that the EM27/SUN Xair shows the smallest airmass dependence. Please clarify.
Authors' response:
Done, please see our reply to point no. 2 of the major comments.

Referee:
34. P15L451: "However, no such instrument specific calibration factors were applied for the other instruments …". Why not? You know the biases wrt TCCON now.
Authors' response:
The EM27/SUN spectrometer participating in this campaign is part of the COCCON network. As discussed in Frey et al., 2015, a species dependent calibration factor is defined for each EM27/SUN spectrometer. The calibration factor has been checked at the Karlsruhe TCCON site before the instrument was shipped for the campaign. The EM27/SUN retrieval results we scaled using these calibration factors. However, the other instruments were deployed for the first time in the configuration discussed in the instrument description section in the paper. Therefore, no such calibration measurements were performed for the other low-resolution measurements before this campaign. As a result, we want to report the absolute bias seen by the individual instrument. We need to investigate further if this bias remains constant over longer period. This is currently under study as part of the extension of the campaign.

Referee:
35. P15L457: Why not just truncate the HR interferograms instead of recording special HR125LR interferograms?
Authors' response:
The recording of low-resolution double-sided interferograms (HR125LR) has added benefits as compared to truncating the high-resolution single-sided interferograms.

1. The acquisition time of the HR125LR interferograms is much shorter as compared to that of HR interferograms. Valid low-resolution measurements can be acquired under weather

conditions that are too poor for high-resolution measurements; therefore, there is the possibility of adding more observations with the HR125LR type observations.

2. HR125LR are double-sided interferograms that make the phase correction easier as compared to the HR interferograms that are single sided and truncating it might lead to artifacts. The centre burst of the interferogram is near the ramp-up section of the forward scan.
3. The signal to noise of the truncated HR interferograms will not be comparable to the LR interferograms since most of the observation time is omitted.

Referee:
36. P16L499: Again, without the LR AKs to compare with, we cannot assess the impact of errors in the prior on the total column comparisons.
Authors' response:
The LR and HR AKs are shown in the Figure 6 of Hedelius et al., 2016, this reference is already stated by the referee in point no. 9 of the technical comments.

Referee:
37. P17S5.4.4: Is there a figure this section refers to?
Authors' response:
Yes, this section referred to Figure 5. A description of this has been given in the introduction part of section 5.4 in the original version of the paper.

Referee:
38. P18L553: This paragraph is a non sequitur wrt the previous paragraph regarding XCO.
Authors' response:
Done

Referee:
39. P19L576: "From the plots it can be seen that the SZA dependency … is related to the spectral resolution and the AK of the instruments." While I agree that it's probably true, it cannot be seen from the plots, since you have not shown the LR AKs.
Authors' response:
Please see our reply to point no. 36 of the technical comments.

Referee:
40. P20L606: Any idea why this is occurring? This is a potentially interesting result.
Authors' response:
This is potentially due to the difference in the retrieval methods between the GFIT and PROFFIT. The data from other low-resolution instrument were analysed using GFIT and therefore show similar dependency and lower difference when compared to TCCON.

Referee:
41. P20L617: Again, these have different retrieval code and different spectroscopy. Please make that clear.

Authors' response:
Done, see lines: 930 – 933: "The EM27/SUN and the HR125LR results retrieved with PROFFAST do not show SZA dependence for species where an airmass correction factor, which was previously determined, was applied except for carbon monoxide where no correction was applied. The other instruments show a SZA dependence to some degree. In order to minimise the effect of the SZA, measurements with an SZA<75° should be used for the instruments."

Referee:
42. Section 5.7.1: Suggest moving to an appendix for the reasons stated above.
Authors' response:
Done

Referee:
43. Section 5.8: Suggest putting brief sentence earlier in manuscript, and moving this section to Appendix A for the reasons stated above. Also, "These higher values could come from the spectral double passing of the signal within the interferometer." This sounds interesting but requires far more explanation. Where is the double-passing coming from? How can it be removed?
Authors' response:
We have moved the discussion with the intercomparison results using the uncorrected data to the appendix. This is in line with what has been suggested by the referee. We have added explanation about double-passing in the paper, see lines 389 – 391: "These higher values can be explained by the presence of unintended double passing of the infrared beam in the interferometer that occurs if some radiation is reflected back from the detector system."
If present, then this is a feature of the detector. However, this is not relevant as the signal is in the out-of-band spectral region and not affecting the retrieval of trace gases.

Referee:
44. P22L687: "the annex 1" –> Appendix A.
Authors' response:
Done

Referee:
45. P25L775-780: This discussion is confusing. Please clarify.
Authors' response:
We have added further explanation, see lines 479 – 486: "For example the AK values for CO_2 for lower altitudes are >1 for measurements performed at higher SZA, which means that the retrieval will overcompensate any over- or underestimation of the a priori: If the a priori is underestimating the lower partial column values in comparison to the true atmospheric state, then these will be overestimated by the retrieval in the total column amount; and vice versa if the a priori overestimates the lower partial columns then the retrieval will underestimate their contribution in the total column amount. Similar reasoning is applicable to the case where the AK<1 for lower SZA measurements typically at local noon. From Fig. 2 we can see that the TCCON a priori is underestimating during the summer months and therefore the SZA dependence in the

bias (TCCONmod - TCCONmodAC) in Fig. 4 can be explained from the shape of the AK and it is higher for the 28 August case as compared to the 15 May measurements.

Referee:
46. P26L817: This is not surprising, because the GGG2014 TCCON CO prior is a climatology. It will generally not capture pollution events.
Authors' response:
We have added this information in the paper. See lines 521 – 523: "The AirCore profile measured on 28 August captured a large signal in the troposphere but it is not seen in the TCCON a priori. The TCCON CO prior is a representation of the climatology and so it will generally not capture pollution events."

Referee:
47. Section 5.9.2: Move to Appendix.
Authors' response:
Following the suggestion of the first referee, we have moved this section to the start of the results section and the discussions with the uncorrected data to the appendix.

Referee:
48. P29L891: The airmass dependence of the retrievals is an effect of the software *and spectroscopy*.
Authors' response:
Done

Referee:
49. P29L892: What is this "airmass correction factor" you're speaking of? Please make this language clearer throughout the paper, and define your terms.
Authors' response:
Done, please see our reply to point no. 2 of the major comments.

Referee:
50. P29L904: showed ! provided
Authors' response:
Done

Referee:
51. Figures. In general, there are far too many figures, the font size on the figures is far too small, the point sizes are too small, the point styles are indistinguishable, and there are too many points on the figures. This will be even more problematic after typesetting when the sizes of the figures are shrunk to fit into the AMT two-column format. Consider relegating some figures to appendices, using shading instead of individual points, removing whitespace between multi-panel figures, simplifying the content, etc.
Authors' response:

Following the referee's suggestion, we have moved several figures to the appendix. We now have only 11 figures (from before 21) in the main part of the paper. The font size for several figures was increased.

Referee:
52. Figure 6. Does the "PF" at the end of the HR125LRPF indicate "PROFFAST"? If so, please state clearly in the caption.
Authors' response:
Done

Referee:
53. Figure 17. When plotting vertical profiles in the context of a total column retrieval, it is more natural to plot the vertical axis in linear pressure. This is because it better approximates the mass-weighting that the total column represents. Please change these plots to display them in linear pressure. Also, you show column averaging kernels for TCCON up to 85 SZA, but do not use them above 75 SZA. Why plot the other two SZAs? Furthermore, you do not show the HR125LR, EM27/SUN, Vertex, IRcube AKs, which is important to understand how the differences between the a priori profile and the true profile would affect the total column. Please add the AKs of the LR instruments (if they are basically the same, you can just add one representative set).
Authors' response:
Done, we have replaced the vertical profiles plots plotted against altitude and added the plots plotted against pressure. For the rest of the comments please see our reply to point no. 36 of the technical comments.

Referee:
54. Figure 18. What exactly is being plotted? 10 minute averages or all individual points? If averages, please make this clear in the caption, and describe the number of points being averaged in each averaging period for each instrument.
Authors' response:
These are 5-minute averages. We have added further explanation in the text, see lines 462 – 465: "In order to make the intercomparison, data from each instrument were sorted and all data within a time interval of a 5 min sequence were averaged and associated to the respective start time of the bin. The timestamp of the reference data set (e.g., TCCONmod) was matched with the same timestamp as the other instruments to find the coincident data pairs, which were used for the difference and the correlation calculation."
We have also added further explanation in the text regarding the number of points being averaged, see lines 488 – 493: "Within the period of five minutes, it is possible to average five measurements for the EM27/SUN data set whereas a maximum of only two measurements is possible for the TCCONmod data set. The Vertex70 measurements on 15 May were performed before the instrument modifications. As a result, a high bias relative to the TCCONmod was seen. This bias is not present for the measurements performed after the instrument modification on 28 August. The scatter in the IRcube and Vertex70 is comparable to the TCCONmod due to the averaging of the similar number of measurements within the five minutes time interval.

Referee:
55. Figure 21. If you are comparing data to a reference (in this case, AirCore is the reference), I believe it is customary to put the reference on the x-axis. You have plotted TCCON on the x-axis here.
Authors' response:
We have removed this plot to reduce the number of figures in the paper.

Referee:
56. Tables: In general, some tables could be combined and descriptive text should be in the caption, not embedded in the table (e.g., Table 4).
Authors' response:
We have removed the descriptive text embedded in the tables 4,5,6,7 and 9 and included it in the caption of the table.

Reference:

1) Hedelius, J. K., Viatte, C., Wunch, D., Roehl, C. M., Toon, G. C., Chen, J., Jones, T., Wofsy, S. C., Franklin, J. E., Parker, H., Dubey, M. K., and Wennberg, P. O.: Assessment of errors and biases in retrievals of X_{CO2}, X_{CH4}, X_{CO}, and X_{N2O} from a 0.5 cm^{-1} resolution solar-viewing spectrometer, Atmos. Meas. Tech., 9, 3527–3546, https://doi.org/10.5194/amt-9-3527-2016, 2016.

2) Gisi, M.: EM27/SUN, Oral presentation at: Annual Joint NDACC-IRWG & TCCON Meeting, Bad Sulza, Germany, May 12–14, available at: http://www.acom.ucar.edu/irwg/IRWG_2014_presentations/Wednesday_PM/Gisi_Bruker_EN27.pdf (last access: 1 February 2020), 2014.

3) Frey, M., Hase, F., Blumenstock, T., Groß, J., Kiel, M., Mengistu Tsidu, G., Schäfer, K., Sha, M. K., and Orphal, J.: Calibration and instrumental line shape characterization of a set of portable FTIR spectrometers for detecting greenhouse gas emissions, Atmospheric Measurement Techniques, 8, 3047–3057, https://doi.org/10.5194/amt-8-3047-2015, https://www.atmos-meas-tech.net/8/3047/2015/, 2015.

4) Sha, M. K. et al., Sentinel-5P methane and carbon monoxide product validation using global TCCON and NDACC-IRWG data (TCCON4S5P and MPC results), Oral presentation at the Copernicus Sentinel-5 Precursor Validation Team Workshop, 11-14 Nov. 2019, ESA/ESRIN, Frascati (Rome) Italy, available at: https://nikal.eventsair.com/QuickEventWebsitePortal/sentinel-5-precursor-workshop-2019/sentinel-5p/ExtraContent/ContentPage?page=5